San Jose State University
SJSU ScholarWorks

Master's Projects

Master's Theses and Graduate Research

Fall 2015

Energy Efficiency and Quality of Services in Virtualized Cloud Radio Access Network

Khushbu Mohta
San Jose State University

Follow this and additional works at: https://scholarworks.sjsu.edu/etd_projects

Part of the Computer Sciences Commons

Energy Efficiency and Quality of Services in Virtualized Cloud Radio Access Network

A Project

Presented to

The Faculty of the Department of Computer Science

San Jose State University

In Partial Fulfillment

of the Requirements of the Degree

Master of Science

By

Khushbu Mohta

December 2015

ABSTRACT

Energy Efficiency and Quality of Services in Virtualized Cloud Radio Access Network

By Khushbu Mohta

Cloud Radio Access Network (C-RAN) is being widely studied for soft and green fifth generation of Long Term Evolution - Advanced (LTE-A). The recent technology advancement in network virtualization function (NFV) and software defined radio (SDR) has enabled virtualization of Baseband Units (BBU) and sharing of underlying general purpose processing (GPP) infrastructure. Also, new innovations in optical transport network (OTN) such as Dark Fiber provides low latency and high bandwidth channels that can support C-RAN for more than forty-kilometer radius. All these advancements make C-RAN feasible and practical. Several virtualization strategies and architectures are proposed for C-RAN and it has been established that C-RAN offers higher energy efficiency and better resource utilization than the current decentralized radio access network (D-RAN). This project studies proposed resource utilization strategy and device a method to calculate power utilization. Then proposes and analyzes a new resource management and virtual BBU placement strategy for C-RAN based on demand prediction and inter-BBU communication load. The new approach is compared with existing state of art strategies with same input scenarios and load. The trade-offs between energy efficiency and quality of services is discussed. The project concludes with comparison between different strategies based on complexity of the system, performance in terms of service availability and optimization efficiency in different scenarios.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

CHAPTERS

List of Figures

List of Tables

1 INTRODUCTION

Cloud Radio Access Network (C-RAN) is feasible due to recent progress of network and virtualization technologies. Software defined radio (SDR) and network function virtualization (NFV) has enabled physical resource sharing between virtualized base band processing stations. The general purpose processors can support software defined baseband units (BBU). Different virtual network protocol like OpenFlow[1] enables separation of control and data plane in network devices. Also, advancement in the optical transport network provides high bandwidth and low latency channels for communication between decoupled remote radio heads and centralized and virtualized baseband units [2]. These advancements paved path for Cloud RAN. Resource utilization and energy efficiency has been extensively studied for cloud data centers (CDC). Different virtual machine (VM) clustering and packing algorithms are proposed for CDC's energy efficiency. These algorithms can be modified to meet BBU pool's requirement for clustering and placement of virtual base stations (VBS) on the serving general purpose processing (GPP) servers. The facility of C-RAN is same as CDC, but there are major differences in the bandwidth and latency requirements, number of clients and acceptable jitters [2]. It is important to consider these tighter constraints of C-RAN over CDC and chose appropriate resource management for acceptable quality of services. Centralization of RAN offers numerous advantages which include high resource utilization and better mobility and radio interference management. It is also Traditional research questions address the increased resource utilization of baseband unit (BBU) pool and energy consumption of Remote Radio Heads (RHs). The different CDC's VM clustering and packing algorithms [5, 6] and some proposed high resource utilization architectures for C-RAN [3, 4] are studied and compared. A new load-prediction based algorithm for resource allocation is proposed. The C-RAN can benefit from VBS clustering and packing algorithms to maximize resource utilization and minimize energy consumption by reducing number of active under-utilized physical resources.

The proposed technique is front-haul location-aware virtual base station consolidation and placement algorithm. The virtualized BBUs requires resources and consumes energy

9

on demand. Efficient VBS clustering and packing based on infrastructure similarity between BBU pool and CDC and estimate the energy consumption of the BBU pool are deduced. The proposed approach tries to minimize handover distance within BBU pool to optimize VBS clustering and placement. To evaluate the algorithms, a simulator for Cloud RAN is implemented to produce energy consumption and quality of services metrics.

This project targets the challenge of deducing new technique for Virtual BBU clustering in the pool to prevent under and over utilization for underlying physical resources. This project proposes two algorithms:

- *VBS clustering based on location of associated RH, such that reduces the handover distance between VBS in the BBU pool*
- *VBS cluster packing algorithms that places the clusters of VBS on Hosts in most optimized fashion that keeps inter-host communication minimum and minimizes overall active number of hosts*

In section 2 the related works for Cloud RAN is discussed. The different CDC's VM clustering and packing algorithms and some proposed high resource utilization architectures for C-RAN are highlighted. The section 3 explains the proposed algorithms and its comparison with other algorithms. In section 4, the energy model for C-RAN and Quality of services metrics used in this project are explained. Section 5 contains implementation details. The section 6 discusses the results of simulation which is followed by the section 7 which concludes the project with possible future works.

2 RELATED WORKS

2.1 Cloud Radio Access Network

Recent mobility predictions suggest that mobile operators will need to accommodate twenty-five percent more Long Term Evolution (LTE) subscriptions in the next six years and an almost similar increase in data traffic [7]. Cloud RAN or Centralized RAN is basically centralization of baseband units (BBUs) into a pool of BBU resources. Over the years, network technologies have evolved and so did the base stations. Fig. 1 shows the C-RAN architecture [2] for Mobile network. The major advantage of C-RAN is scalability and elasticity. Centralization offers higher resource utilization, network utilization and power efficiency. Apart from that, as described by Checko et al [2], it can achieve high throughput and less delays by easier implementation of Cooperative Multiplexing (CoMP) [8], Fractional Frequency Reuse (FFR) [9] and dynamic control of transmit power to manage interference. As it is software defined, it will be easier to upgrade and maintain. The only bottleneck was front-haul network capacity and latency, which can be overcome by recent advances in network technologies such as Dark Fiber optical cables.
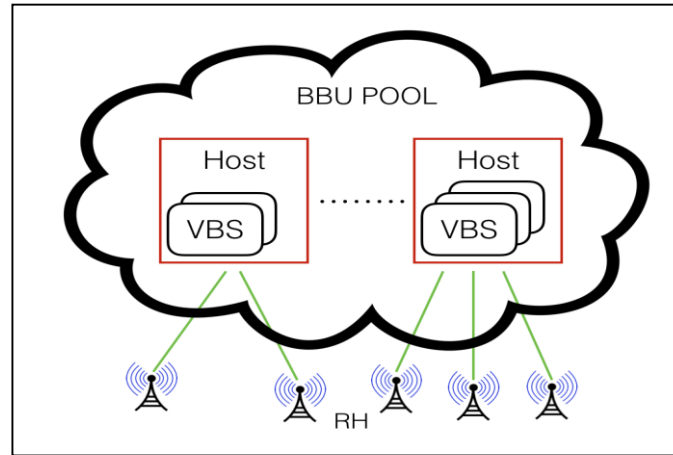

Fig. 1. Cloud Radio Access Network Architecture.

The Capital Expenditure (CAPEX) is very high for setting up new cell sites to meet the increasing demand. The operational cost (OPEX) is also high, most of which is power

consumption. In a typical cell site, there are several components that need power. Antennas require transmission power, base band unit require compute power and site require lighting and cooling power. To reduce these cost, centralization is studied as a next generation solution. Centralization directly reduces the lighting and cooling power. Recent studies has shown that the current network technology has reduced latency and increased bandwidth making it possible to centralize the BBU processing for up-to a forty kilometer radius [2]. C-RAN also allows to activate/deactivate or control transmission power on antennas, as it is less complex to have a global status of all the BS being virtual. Network Function Virtualization can optimize compute power by optimizing resource sharing. There are several papers [10,11,12] that talk about compute power required for baseband processing. Boyapati et al [10] lists all the baseband functions in uplink and downlink and their compute requirement. They analyze different green architectures, algorithm-architecture mappings, energy management strategies that can improve the energy efficiency of the baseband sub-system. Bhaumik et al. [11] provides a compute load estimate for VBS as a linear function of Physical Resource Blocks (PRBs) in use and modulation and coding scheme (MCS) in use.

2.2   Energy Efficiency approach in Cloud Data Centers

There are many VM clustering and placement algorithms proposed for cloud data centers to achieve compute power optimization. The most related algorithms are based on greedy approaches similar to bin-packing. A slight modification in VM placement strategy can improve energy efficiency of cloud data center. Reguri et al [5] proposes one such optimization algorithm for cloud data center that clusters VMs based on inter-VM data traffic. A logical clustering of virtual instances can save inter-host transmission or migration power within data centers. After clustering, simple allocation algorithm like bin-packing, first-fit and best-fit can perform better than it would without clustering. Liu et al[6] presented a traffic aware VM packing in cloud data center using an approximate graph cutting algorithm to solve bin-packing. This approach also minimizes the inter-host traffic reducing network load and increasing throughput. Each VM is considered a node in the graph and the inter-VM communications is translated as weights on the graph edges

connecting VMs. Their proposed algorithm use this input for determining clusters of VM that are then bin-packed on hosts. These techniques improve energy efficiency of data centers by reducing number of active host while maintaining quality of services. Similar approaches can be applied to a BBU Pool Center for efficient physical resource utilizations, when the BBUs are virtualized and are hosted on a general purpose hardware.

2.3   Energy Efficiency in Cloud RAN

Pompili et al [3] proposed one such logical framework to implement elastic resource utilization where the VBS clusters serving areas with negative correlation in bandwidth demand share resources. When the bandwidth demand of one cluster increases it can request the other clusters to release unused resources as its demand is predicted or known to be decreasing. The authors assume that the correlation exists and is already known, which can be used to place VBS-clusters together to improve resource utilization efficiency. This approach will fail if at some point of time, the negative correlation cease to exist between VBS-clusters. Also, they do not consider VBS migrations from one host to another to minimize pool power. They do achieve better resource utilization than traditional decentralized RAN. And the quality of services is also measured by measuring the blocks due to reactive resizing of cluster. The energy is saved by not allocating compute resources to the towers that does not require it. And putting the Radio Heads on stand-by/less power modes.

Zou et al [4] proposes a resource allocation technique for densely populated area. During low demand time the proposed resource allocation mechanism selectively deactivates radio heads (RHs) when neighboring active RHs can serve the current demand. To identify neighboring RHs. They divide entire area of consideration into virtual square grid where edge of square is same as the coverage radius of RHs. Their two step mapping technique is used for selective resource allocation. This project proposes an algorithm to find neighbor RHs in constant time complexity.

The literature so far compare energy efficiency of C-RAN techniques with the current decentralized RAN. In this project, the proposed approach is compared for the energy

efficiency and quality of services with other C-RAN energy optimization techniques. The proposed approach consider three main areas of problem in CRAN:

- Demand fluctuation in the areas due to mobile users
- Handling handovers – UE Context in Virtual BBUs
- Predicting User data traffic variation to optimize resource allocation – modification on Pompili et al [3]

3 ENERGY EFFICIENT CLUSTERING AND PACKING STRATEGIES

Different VBS clustering and packing algorithms [3, 5, 6] were analyzed. This project discusses a location based clustering algorithms, where RH are clustered together based on their area of service. RH servicing in around same neighborhood are clustered into a VBS cluster that can be packed later on hosts, using a packing algorithm.

3.1 VBS Clustering Algorithm

For purpose of simplicity, each RH is associated with a fixed size VBS, size of VBS here indicates compute, memory and network requirement for baseband processing.

3.1.1 Location based clustering

This clustering strategy puts RHs in a close vicinity to each other in a cluster. The entire area is divided in a grid as done by Zou et al. [4] to recognize neighbors, will be referred as GetNeighborRHs algorithm from hereon. The entire area is divided into square grids with edge size equal to clustering distance (The maximum allowed distance between two RHs in the cluster) and then to search for neighbors for an RH the algorithm only needs to get the enclosing nine squares and RHs in those squares. Fig 2 shows the grid, RH enclosed and its neighborhood. We use GetNeighborRHs to cluster RHs that are within C distance of each other. C is clustering distance – an input to clustering algorithm.
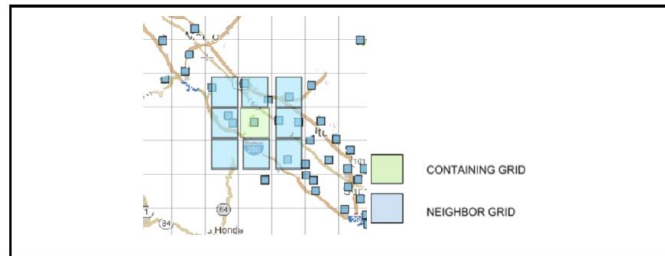

Fig. 2. RH inside a grid and its neighbors

Assuming that **GetNeighborRH** and **Merge** takes linear/log linear time, the complexity of this algorithm is $O(N^2)$, where N is total number of RHs in the area of interest for C-RAN. Other comparable and well known clustering techniques like Affinity

Propagation Clustering also shows similar complexity. This is a reasonable complexity as the number of radio heads for C-RAN is a relatively small.

Location Based Clustering Algorithm (Newly Proposed)

Cluster VBS associated with RH within C Distance

Input: C, AllRHList

Output: ClusteredRHList

For each RH in AllRHList:

   If not already in a ClusteredRHList:

     GetNeighborRH within clustering distance

     For each candidate in clusterCandidate

       If candidate has cluster **And** RH fits in it

        If RH not in cluster

         Add RH to candidate's cluster

       ElseIf RH's cluster not same as clusterCandidate

        Merge clusters

   Else

     Add candidate to RHCluster

  If RH is not in cluster

    Add RH to New RHCluster

  Add RH to ClusteredRHList

3.2 VBS Cluster Packing Algorithms

This project compares previously known and proposed greedy VM packing algorithms with the newly proposed algorithm.

3.2.1 Bin Packings (BP)

This is VM packing implementation of algorithm proposed in [6]. Where the packing and clustering is based on the network traffic between VMs, hence it is a VM traffic

aware technique. The complexity of packing M clusters on N Host for bin Packing is $O(M^2)$

3.2.2 First Fit (FF)

This is well known first-fit greedy solution for np-complete bin-packing problems, where VBS and host are both sorted and maximum size VM are fit into first available host, where all hosts are sorted by host utilization. The complexity of First Fit is $O(NM)$

3.2.3 Traffic Prediction

This is user traffic prediction based, as shown in Fig 3. A VBS cluster increases and decreases in size as discussed by Pompili et al in [3]. For fair comparison, prediction-based migration logic is implemented. Using both proactive and reactive approach for resizing and packing the clusters. The complexity of this approach is $O(M2)$
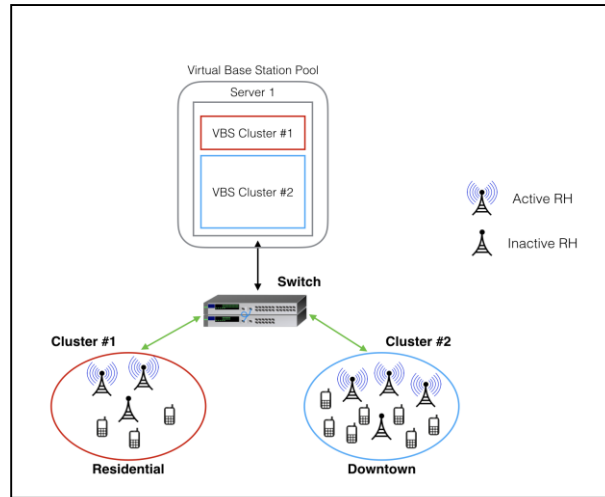

Fig. 3. Traffic Prediction based packing

3.2.4 *Location Aware*

This is user traffic prediction based, as shown in Fig 3. A VBS cluster increases and decreases in size as discussed by Pompili et al in [3]. For fair comparison, prediction-This is newly proposed packing where VBS–clusters are further clustered based on location.

The main aim is to improve quality of services. UE context is migrated between VBS during Handovers and this is viewed as traffic between VBS. If the VBS are in same cluster, this reduces the handover latency. And two VBS on different hosts have higher handover latency.

Location Based Packing Algorithm (Newly Proposed)

Sort clusters by compute utilization ascending in VBBUCLusterList
Sort host by compute utilization descending is HostList
    for each cluster in VBBUClusterList:
        Sort NeighborClustersHosts by Utilization
        for each host in NeighborClustersHosts :
            if cluster fits in host :
                Map cluster to host
            Exit for
        if cluster not mapped to any host:
            if original_host was over utilized:
                doBestFitMapping

    for each cluster in unmapped_cluster_list:
        for each host in NeighborClustersHosts :
            if cluster in host :
                Map cluster to host
                Break For-loop
        if cluster not mapped to any host
            add cluster to unmapped_cluster_list
    for each cluster in unmapped_cluster_list:
        for each host in HostList:
            if cluster fits on host :
                map cluster to host
            if cluster is not mapped:
                activate newhost
                map cluster to newhost;
    Sort host by compute utilization ascending is HostList
    for each host in HostList:
        if host empty:
        deactivate

This algorithm is O (M^2), where M is number of VBS clusters.

3.2.5 Mobilty aware Location aware Packing

We further extend the Location-aware Packing to consider user mobility and predict handover

4 ENERGY MODEL AND QUALITY OF SERVICES METRICS

4.1   Energy Model

In LTE, BBU processing and Radio Frequency contribute to approximately fifty-seven percent of total energy for a cell site. Assuming, that the number of active PRBs with just one MCS is proportional to number of User Equipment (UE) connected and active during a processing cycle. It is deduced that overall compute energy required for baseband processing can be approximated as being proportional to number of UEs active on the corresponding RH. Assuming that this is linear, energy for compute per VBS is given as,

$$E_{computeVBS} = E_{BaseVBS} + E(\int Pc(u(t,u))dtdu \qquad (1)$$

Here,

$E_{compute}$ is compute energy required by a VBS.

$E_{BaseVBS}$ is base power required by baseband signal processing

Pc is compute power as a function of compute utilization

$u(t,u)dtdu$ is energy of a VBS as a utilization function of time t and number of active UE u

VBS may migrate to other active Host to achieve higher energy optimization and the migration power uses same calculation as done by Reguri et al [5] for migration energy Emigration and host activation energy $E_{activation}$. Thus, total energy can be given as

$$Energy_{total} = \Sigma n \ E_{computeVBS} + \Sigma m \ E_{migration} + \Sigma k \ E_{activation}$$

20

where,

n is number of VBS,

m is number of migrations

k is number of host being activated.

4.2   Quality of services

Quality of services is measured in terms of the following metrics.

4.2.1   Percentage UE blocked in each packing

During VBS migration or packing processes, some percentage of connected active UE's may be blocked/denied service. For example, if some of the RHs are deactivated and a sudden surge of active UEs is observed such that the active RHs cannot handle the load, resulting in denial of service that is measured as blocked UEs.

4.2.2   Percentage Handovers between hosts and within hosts

The techniques in comparison are trying to pack VBS such that the resource utilization is maximized and this project also considers handover distance in the pool as evaluation criteria. During handovers the UE context needs to be shared between VBSs and if those VBS are in the same cluster then its more efficient than it being on same host, which is more efficient than between two VBS on different Hosts.

4.2.3   Average allocated bandwidth/requested bandwidth

This is another quality of services measurement. The average allocated bandwidth to requested bandwidth ratio helps in understanding the tradeoff between energy efficiency and quality of services. As, the tower are deactivated when the load is low as the load increases the towers are activated again which takes time and has a small duration when quality of services may be poor.
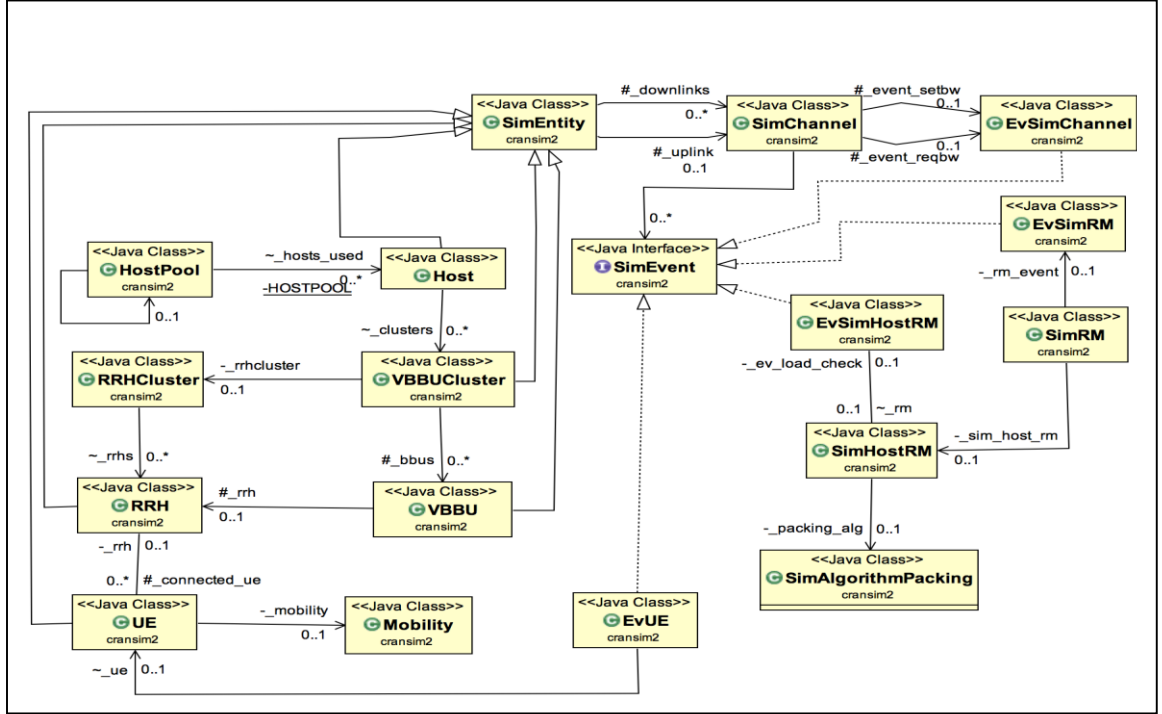
5 SIMULATION IMPLEMENTATION
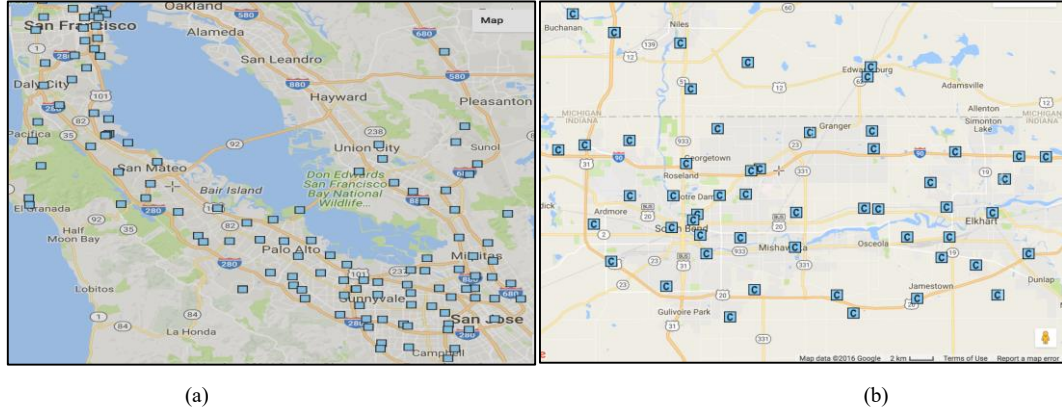

Fig. 4. Simulation class diagram

(a)

(b)

Fig. 5. FCC registered towers as obtained from [13] (a) San Francisco Bay Area, (b) South Bend, Indiana and surrounding sub-urb

For the simulation and results analysis, the input and output parameters were identified and documented. Table I contains the input for the simulation. Also statistical data that is

needed to evaluate the performance of the proposed algorithm was identified as mentioned is Table II. These logs fall into three groups - hosts statistics, users statistics, and algorithm statistics. This data was used to compare the behavior of the clusters and the impact on performance when different algorithms were applied. The simulation records events and status over timestamp where multiple events may occur simultaneously (e.g. mobiles accessing the same RRH). The simulation has a mobility model similar to LTE-Sim: positioning and detection using time slices, and trajectory calculation.

For traffic prediction algorithm, regression analysis was considered to create usage curves based on history and use them to estimate and predict near-future loads on the system. We used normal regression for predicting traffic load.

5.1   Scenario Generation

For implementation, to simulate data that is as close to actual facts as verifiably possible. To generate tower locations and user movement, actual cities are chosen. For example, San Francisco Bay Area, California and South Bend, Indiana as shown in Fig 5 (a) and (b) respectively. The Figure represent cell towers as blue squares. These maps are retrieved from "mapmuse"[13], a website that lists a myriad of things on map. In this case, it is Federal Communication Commission (FCC)[14] registered towers. FCC has information about cell towers registration and cell-site leasing records that include information about different cell tower specifications, owners, location, lease term, etc. mapmuse.com uses this information and puts towers on the map as shown in Fig 5. Approximately a 40 km X 40 km area that is populated with enough towers and freeways and is easily retrievable for simulation purpose is selected. WebPlot Digitizer [14] application is used to generate data-points from the retrieved map images. The process of generating data points is simple. One can upload the image to the app, calibrate the image into (x, y) points, the app asks to select two random points on each X-axis and Y-axis and value for those points. Once calibrated, each point on the image can be retrieved as (x, y) data point by marking it with a pen tool in Manual data generation mode. For this project, the RH locations in the Map were marked and downloaded the generated data points as

comma separated value, which becomes RH location for input. There is another line tool, which is used to get freeways and get data points on freeways to use for user mobility. The mobility events were generated on these freeways in a such a way that indicates a general pattern of people moving from residential to work places with some degree of randomness.

The input is RH locations, UEs with bandwidth request events and mobility event.

5.2   Assumptions

To simplify the simulation while maintaining integrity of comparison between different algorithms, following assumptions were made.

- The bandwidth requirement of all active UEs is uniform.

- The compute power for baseband processing is proportional to the number of active UEs, for simplicity of calculations it is assumed to be linear. The propotionality coefficient depends on the Modulation and coding scheme is use [10].

- All RHs are assumed uniform and after clustering one of the RHs in the cluster can cover the entire area. All RHs has a circular coverage area with overlaps.

- Each RH has a dedicated VBS when active and has a minimum compute and memory requirement.

- A VBS has some compute and memory allocated for each active UE.

- All the General Purpose Processors that act as hosts for VBSs are uniform.

5.3   Simulation Input Load

The main input to the simulation is as below. Each simulation logs comparable data for each packing algorithm in observation.

TABLE I. INPUT LOAD

Input	Value	
Scenario	Scene 1 Bay Area	Scene 2 South Bend
#RH	90	52
#UE	~13000	~10000
#Host	30	30
Packing Algorithm	Bin Packing, First Fit, Traffic Prediction, Location Aware	Bin Packing, First Fit, Traffic Prediction, Location Aware
#Highways	3	3
#MobileUE per Highway	2000	2000
# Stationary UE per Highway	Randomize between (10,100)	Randomize between (10,100)

5.4 Output and Power Calculations

The simulation logs data as mentioned in Table II, which is used for performance analysis:

TABLE II. OUTPUT LOGS

Log	Description
Active Host over Time	Logs with Timestamp when Active Host Count changes
Migrations	Logs number of migration during resource management cycle

Log	Description
	with time stamp
UE bandwidth demand	Logs UE bandwidth requested and allocated with timestamp.
Compute utilization	Logs compute utilization at regular intervals with timestamps
Handover data	Logs handovers between RH, between clusters, between hosts
#UE blocked	Logs UE blocked due to handovers and/or unavailability

These logs are used to calculate and compare E_{total} and QoS metrics.

6 SIMULATION RESULTS

The following observations were made on the average of two scenarios simulated.

6.1 Migrations and Active Hosts

In Fig. 6. the First fit (FF) and Bin-Packing seems to do best with least number of migrations. This is more proactive approach, the number of average active host Fig. 7. is higher than the other three, Traffic Aware (TA), Location Aware (LA) and Location aware with Mobility (LAM). LAM performs highest migrations of all algorithm but it also has least average active hosts across all algorithms. The high number of migration is a result of the reactive approach of the algorithms, the hosts are activated and deactivated as the demand fluctuates. A further investigation we calculate the area under the active host over time graph Fig. 8 and compare the areas in Fig. 9, this shows the active hosts time in days. This helps in calculating the Active host power utilization as well.
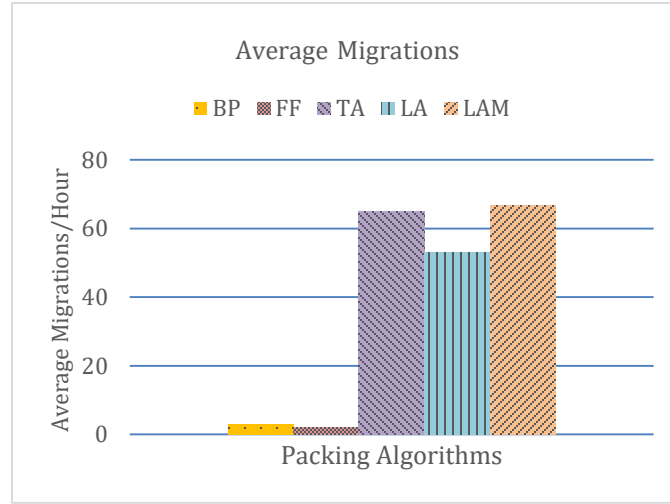

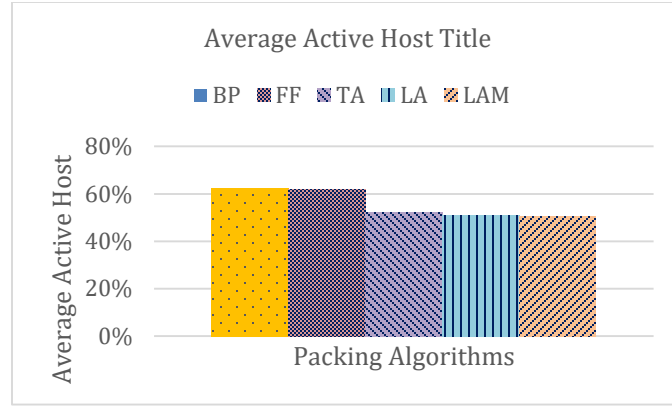Fig. 6. Average Migration comparison between the packing Algorithms

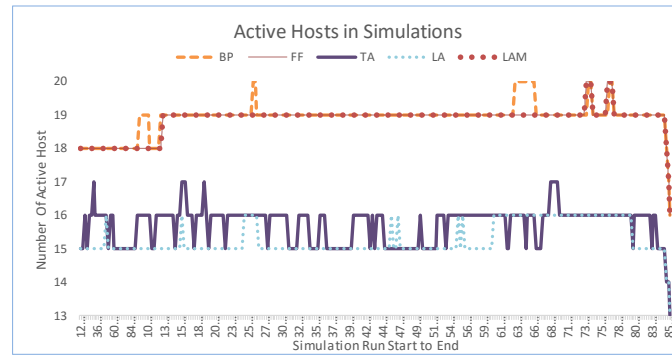Fig. 7. Percent Average Active Host out of total available hosts
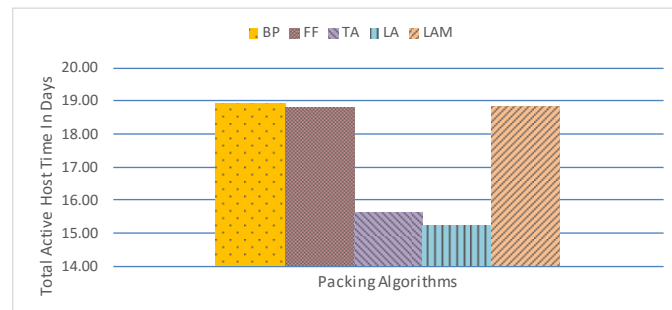

Fig. 8. Active Hosts during Simulation

Fig. 9. Area Under Active Hosts for all Packing

6.2   Handovers between Hosts, With-in Hosts, With-in clusters

The newly proposed algorithm is to minimize the handovers across different hosts while optimizing the resource utilization. Fig 10 shows the handover as observed in different

algorithms. The LAM is best in terms of handovers as it has least percentage of handovers between hosts and highest within hosts as well as between clusters. Although traffic aware showed lesser migrations and active hosts it has higher handovers overall.
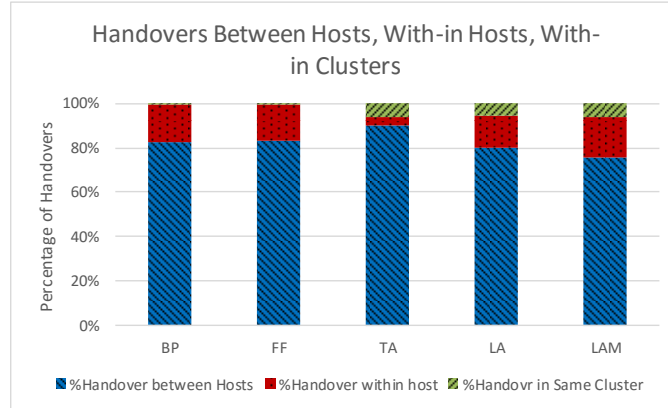

Fig. 10.　　　　Handovers Between Hostsm Wih-in Host and With-in Cluster

6.3　Energy and Power Comparison

As the UE load is same for all the algorithms the Compute power is same for each one of them. The Active Host Power is lowest for LA and TA. The BP, FF and LAM performed similarly in terms of power utilization Fig 11.
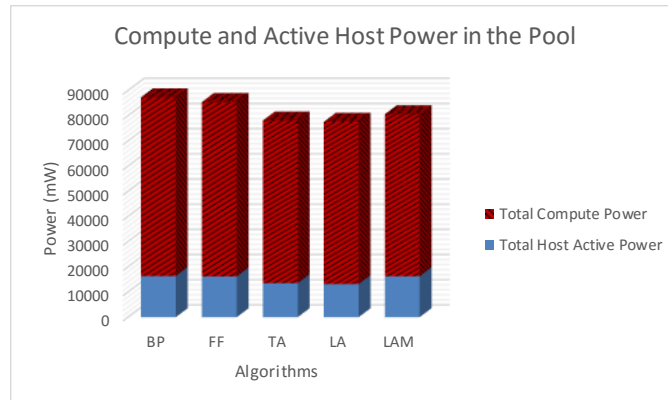

Fig. 11.　　　　Compute and Active Host Power

Fig. 12 shows that the migrations are heavy in TA than LA There are negligible migrations in BP and FF due to less resizing and higher power utilization as well for the same reason.
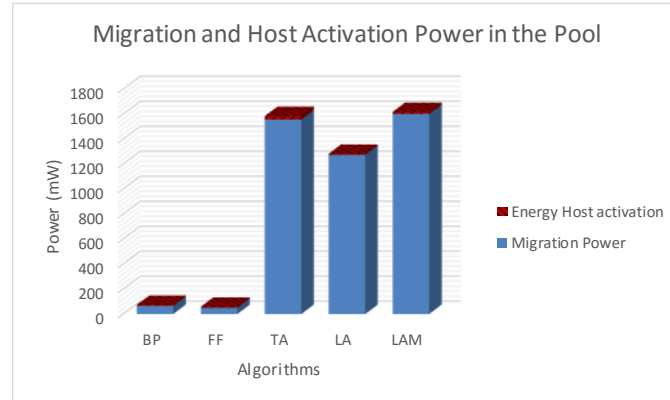

Fig. 12.   Migration and Host Activation Power

6.3.1   Quality of services UE Blocked and Bandwidth Allocation

The Traffic Aware has highest QoS in terms of the total bandwidth allocated to the UEs by the BBU pool over the total requested bandwidth in the simulation. The total requested is same but the total allocated changes due to the sharing of resources on the BBU side. It can be clearly seen that the reactive approach of the TA, LA and LAM performs better than the less reactive and more proactive approach in BP and FF.

Also noticeable in Fig. 13 that Traffic-aware performs best out of location aware in terms of meeting user demands
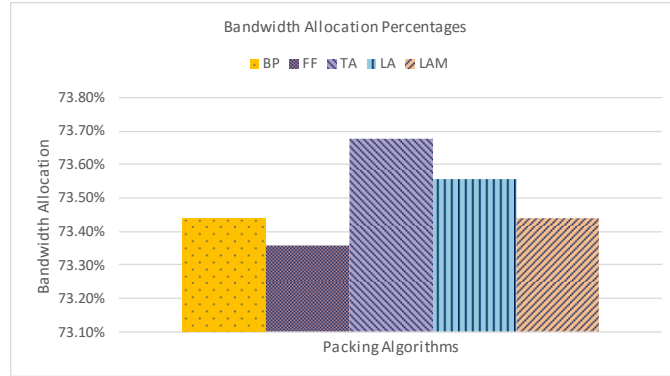
Fig. 13.　　　　Actual bandwidth allocation percentage.

The average UE blocked is very close for all five algorithms as most of the blocks are because of UEs being out of range of RHs by simulation design. Aprroximately 1.6% UEs are blocked either as out of range or during migrations. Fig 14 shows UE blocks
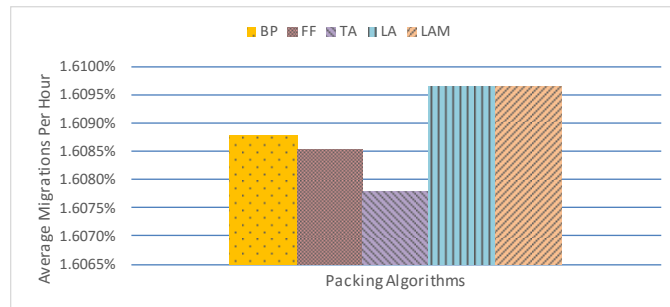

Fig. 14.　　　　Percent UE blocked out of total UE requested data during Simulation

7 CONCLUSION AND FUTURE WORK

The frequent scaling of resources although increases Migration and resource on-off, it is worth it as the total compute power goes down. The Traffic aware shows most number of migrations but better QoS. Location aware and Location aware with Mobility shows better VBU placement but slightly lacks in QoS than traffic aware. But all three perform better than Bin Packing and First Fit packing in QoS and in having less active host and high resource utilization. The FF and BP saves energy by having less migration but at the cost of QoS on the peak time. The reactive approach guarantees better QoS with slight increase in power consumption over proactive approach.

This project assumes user traffic to be uniform as well as uniform resources in BBU pool. Although, it is sufficient for the evaluation of different techniques, it is not realistic. Also, the BBU processing has different functions that require different power utilization, for simplicity of the calculation and sanity of the evaluation some approximation is done. These approaches can be further evaluated by having non-uniform user requests and observe the scale and prediction performance.

8 REFERENCES

[1] https://en.wikipedia.org/wiki/OpenFlow

[2] 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 Communication Survey & Tutorials*, vol.17, no.1, 1st Quarter 2015, pp. 405-426.

[3] D. Pompili, A. Hajisami and T. X. Tran, "Elastic resource utilization framework for high capacity and energy efficiency in cloud RAN," *IEEE Communications Magazine,* Vol.54, Issue 1, Jan. 2016, pp. 26-32.

[4] S. Zou, Y. Zhang and Y. Tang, "Resource Allocation Mechanism Based on Two-Step Mapping for Saving Energy in Wireless Network Virtualization," *IEEE 9th International Conference on Anti-counterfeiting, Security, and Identification*, 2015, pp. 150-154.

[5] V. R. Reguri, S. Kogatam and M. Moh, "Energy Efficient Traffic-Aware Virtual Machine Migration in Green Cloud Data Centers," San Jose State University, Department of Computer Science, 2015.

[6] J. Liu, J. Guo, and D. Ma, "Traffic Aware Virtual Machine Packing in Cloud," presented in 2nd IEEE International Conference on High Performance and Smart Computing (IEEE HPSC 2016), New York, Apr 2016.

[7] "C-RAN the road towards green ran," China Mobile Research Institute, Beijing, China, Oct. 2011, Tech. Rep.

[8] D. Lee, "Coordinated Multipoint Transmission and Reception In LTE-Advanced: Deployment Scenarios and Operational Challenges", *IEEE Commun. Mag.*, vol. 50, no. 2, pp. 148-55, 2012.

[9] H. E. E. O. M. Elfadil, M. A. I. Ali and M. Abas, "Fractional frequency reuse in LTE networks," 2015 2nd World Symposium on Web Applications and Networking (WSWAN), Sousse, 2015, pp. 1-6. doi: 10.1109/WSWAN.2015.7210297.

[10] H. K. Boyapati, R. V. Raja Kumar, S. Chakrabarti, "Energy Consumption Assessment in LTE Baseband of eNodeB and Guidelines for Green Baseband Subsystems", 2010 Annual IEEE India Conference (INDICON), 2010.

[11] Sourjya Bhaumik et al, "CloudIQ: A Framework for Processing Base Stations in a Data Center", presented in ACM MobiCom 2012, Istanbul, Turkey,2012.

[12] Navid Nikaein, "Processing Radio Access Network Functions in the Cloud: Critical Issues and Modeling," Proceedings of the 6th International Workshop on Mobile Cloud Computing and Services, September 11-11, 2015, Paris, France

[13] FCC Tower map - http://find.mapmuse.com/interest/cell-towers

[14] http://wireless.fcc.gov/uls/index.htm?job=transaction&page=weekly

[15] WebPlot Digitizer - App to get data points from images: http://arohatgi.info/WebPlotDigitizer/app/

[16] A. Verdis, G. Stea and G. Nardini, "SimuLTE – A Modular System-level Simulator for LTE/LTE-A Networks based on OMNeT++", SIMULTECH 2014.

[17] http://www.simultech.org

[18] G.Piro, L. A. Grieco, G. Boggia, F. Capozzi and P. Camarda, "Simulating LTE Cellular Systems: an Open source Framework"

[19] "Power Management in Intel® Architecture", White Paper Intel® Architecture Servers

[20] L.Minas, B.Ellison, "Energy Efficiency For Information Technology", Intel Press, 2009

[21]   M. Blackburn, "Five Ways to Reduce Data Center Server Power Consumption",
       2008

[22]   A. Binstock, "Measuring HDD Power Usage with Newer Tech Universal
       Adapter", Green Computing, 2007