On the Energy Minimization of Heterogeneous Cloud Radio Access Networks

Tshiamo Sigwele(✉), Atm Shafiul Alam, Prashant Pillai, and Yim Fun Hu

Faculty of Engineering and Informatics, University of Bradford, Bradford, UK
{t.sigwele,a.s.alam5,p.pillai,y.f.hu}@bradford.ac.uk

Abstract. Next-generation 5G networks is the future of information networks and it will experience a tremendous growth in traffic. To meet such traffic demands, there is a necessity to increase the network capacity, which requires the deployment of ultra dense heterogeneous base stations (BSs). Nevertheless, BSs are very expensive and consume a significant amount of energy. Meanwhile, cloud radio access networks (C-RAN) has been proposed as an energy-efficient architecture that leverages the cloud computing technology where baseband processing is performed in the cloud. In addition, the BS sleeping is considered as a promising solution to conserving the network energy. This paper integrates the cloud technology and the BS sleeping approach. It also proposes an energy-efficient scheme for reducing energy consumption by switching off remote radio heads (RRHs) and idle BBUs using a greedy and first fit decreasing (FFD) bin packing algorithms, respectively. The number of RRHs and BBUs are minimized by matching the right amount of baseband computing load with traffic load. Simulation results demonstrate that the proposed scheme achieves an enhanced energy performance compared to the existing distributed long term evolution advanced (LTE-A) system.

Keywords: Base station sleep · Bin packing · Cloud computing · C-RAN · Energy efficiency · HetNets · Virtualization

1 Introduction

Everyday, the number of connected devices are growing into billions and today's mobile operators are facing a serious challenge. For example, according to Huawei Technologies, in the year 2020, 100 billions of devices will be connected [1]. This will cause an increase in traffic from smart phones like iphone, android and other high-end devices like the iPad, kindle and gaming consoles spawning a raft of data intensive applications, Internet of Things (IoT) and machine-to-machine connections. As a result, fifth-generation (5G) networks have targeted to increase capacity by 1000 times, data rates by 100 times and millisecond-level delay [2]. To fulfil the capacity demands, more base stations (BSs) with a mixture of macro and small cells forming a heterogeneous network (HetNet) have to be deployed by operators, which results to a significant amount of energy consumption.

I. Otung et al. (Eds.): WiSATS 2016, LNICST 186, pp. 223–234, 2017.
DOI: 10.1007/978-3-319-53850-1_22

This contributes to the mobile network's operating expenditure (OPEX) and emits large amounts of CO_2 which causes a greater impact to the environment.

A large amount of power within a base station (BS) is consumed by the power amplifier (PA) and baseband unit (BBU) [3]. The energy consumption of BBU implementation is getting more and more dominant in small cells due to gradual shrinking of cell size and the growing complexity of signal processing [3]. The traditional distributed long term evolution advanced (LTE-A) BSs architecture consumes a significant amount of energy and waste a lot of computing power as the BBU servers are not shared but serve each individual cell [7]. The BSs have been traditionally preconfigured to provide peak capacities to reduce outages. Nevertheless, the mobile traffic varies significantly, irrespective of the either the time of day or traffic profile and is rarely at its peak in practical scenarios [8]. Many energy-efficient schemes for wireless systems have been implemented such as BS sleeping [4–6] where offloading traffic to neighbouring BSs and then completely turning off the BS during low traffic, discontinuous transmission (DTX) where a BS is temporally switched off without offloading and cell zooming. However, current BS processing capacity is only being used for its own coverage rather than being shared within a large geographical area. As a result, during the evening, BSs in residential areas are over-subscribed while BSs in business areas stay under-subscribed. These under-subscribed BSs still consume a significant amount of energy even when they are not necessarily required to be kept active. Therefore, it is imperative to solve this problem and free up the processing capacity and save the corresponding energy.

Meanwhile, cloud radio access networks (C-RAN) have been proposed as a promising solution for minimizing energy within the cellular networks by leveraging cloud computing virtualization technology. With virtualization, baseband workload is consolidated on a minimum number of BBU servers and baseband processing is performed on virtual BBUs (vBBU) and resources are provisioned in accordance to traffic demands. C-RAN comprise of three parts: (i) remote radio head (RRH), which performs lower layer analogue radio frequency functions, (ii) BBU for digital signal processing, and (iii) fronthaul connection between the BBU and RRH. The C-RAN architecture is shown in Fig. 1. Furthermore, more energy savings can be gained from reduced air conditioning cost and reduced equipment room size. This paper integrates C-RAN in HetNets and proposes an energy efficient scheme for reducing energy consumption in C-RAN HetNets by switching off RRHs using a greedy algorithm and also switching off idle BBUs using the first fit decreasing (FFD) bin packing algorithm. The number of RRHs and BBUs are minimized by matching the right amount of baseband computing load with traffic load. The cloud based energy minimization is formulated as a bin packing problem where BS traffic items are to be packed into compute servers, called bins, such that the number of bins used are minimized. The simulations results validates the energy efficiency improvement of the proposed scheme and is compared with the distributed LTE-A system.

This paper is structured as follows: Sect. 2 discusses the related works while the system model and problem formulation is described in Sect. 3. The proposed

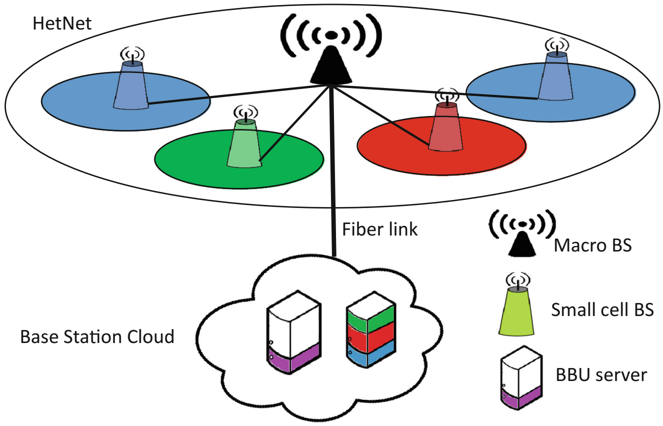

Fig. 1. An illustration of a C-RAN architecture.

scheme with computational resource models and greedy BS switching algorithm are formally are also described in Sect. 3. Section 4 provides the simulation results and discussion, while Sect. 5 provides some concluding comments.

2 Related Works

There are a plethora of solutions towards energy-efficient BSs ranging from energy-efficient hardware design, BS sleeping, to the optimal deployment of BSs [4,9]. This paper will concentrate on BS sleep which is a promising solution for minimizing energy consumption in both the radio side and cloud side of C-RAN HetNet. Authors in [10] proposed a BBU-RRH switching scheme for C-RAN that dynamically allocates BBUs to RRHs based on the imbalance of subscribers in business and/or residential areas. Even though the scheme in [10] reduces the number of BBUs required, the model performs poorly during high-traffic periods and thus still consumes a lot of energy because more BBUs are allocated to meet traffic demands. Authors in [11] developed a BBU pool test-bed using virtualization technology on general purpose processors (GPPs). The BBUs are dynamically provisioned according to traffic load. However, the paper fails to show how the number of BBUs are reduced while traffic load varies. L. Jingchu et al. [12] presented a mathematical model to quantify the statistical multiplexing gain of pooling virtual BSs. The author use a multi dimensional markov model to evaluate pooling gain considering both compute and radio resources. Nevertheless, the author have not considered energy consumption in the BS-Cloud. In [13], the authors considers the energy-delay trade-offs of a virtual BSs considering the BS sleeping approach in general IT platforms. The paper does not show how the energy savings of the virtual BSs model scales with traffic load. S. Namba et al. [14] proposed a network architecture, called colony-RAN, which has the ability to flexibly change cell layout by changing the connections of BBUs and

RRHs in respect to the traffic demand. However, the proposed method has frequent reselections of RRH to BBU, i.e., ping-pong effects.

Since this paper combines HetNet and C-RAN, research on BS sleep in HetNets will be also studied. The author in [15] introduce energy-efficient sleep mode algorithms for small cell BSs in a bid to reduce cellular network power consumption by switching OFF some BS equipments in idle conditions in accordance to variations in traffic load. However, the author assume that the pico and macro cells consume constant power of 12 W and 2.7 kW respectively irrespective of traffic load. The author in [16] combines the sleep mode feature of picocells and load balancing between the different types of base stations in HetNet, hence improving up to 68% for low traffic load and up to 33% for medium traffic load. However, users are assumed to be uniformly distributed whereas users are non-uniformly distributed in reality.

3 System Model and Problem Formulation

3.1 System Model

The proposed system model is shown in Fig. 2. Consider an LTE-A C-RAN HetNet downlink system consisting of a single macro RRH (MRRH) and overlaid by several small cell RRHs (SRRHs). Assume a set of RRHs $\boldsymbol{\mathcal{R}} = \{RRH_j : j = 1, 2, ..., N\}$ where N is the maximum number of RRHs and RRH_1 is the center MRRH. Define a set of users in the entire network as $\boldsymbol{\mathcal{U}}$. Moreover, assume a set of computing servers in the pool $\boldsymbol{\mathcal{M}} = \{GPP_i : i = 1, 2, ..., M\}$ where M is the number of physical computing servers for processing baseband signals of N cells. The global cloud controller (GCC) is located in the BS cloud and it is where the greedy BS switch off and the FFD algorithms are located. The baseband processing procedure of each RRH is divided into L tasks with a set $\boldsymbol{\mathcal{L}} = \{T_k : k = 1, 2, ..., L\}$ where T_k is the k^{th} baseband task for RRH_j. The computing processing power is measured in *Giga Operations Per Second* (GOPS). Each server has maximum capacity C GOPS. The total computing resources required by RRH_j is denoted ρ_j^{req} GOPS such that:

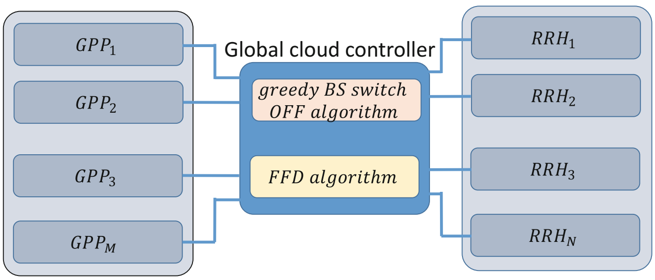

Fig. 2. System Model.

$$\rho_j^{req} = \sum_{k=1}^{L} \rho_{j,k}^{req}; \qquad \rho_{j,k}^{req} \epsilon (0, 1] \tag{1}$$

where $\rho_{j,k}^{req}$ is the computing resource requirement for T_k from RRH_j.

Therefore, the computing resource at server S_i used by RRH_j can be calculated as:

$$\rho_{i,j}^{server} = \sum_{k=1}^{L} \xi_{i,j,k} \rho_{j,k}^{req}; \qquad \xi_{i,j,k} \epsilon \{0, 1\} \tag{2}$$

where $\xi_{i,j,k} = 1$ when T_k from RRH_j is processed by server S_i and $\xi_{i,j,k} = 0$ otherwise. Tasks from RRH_j can be processed by a single server or distributed among different serves such that the constraint below hold:

$$\sum_{i=1}^{M} \sum_{k=1}^{L} \xi_{i,j,k} = L \tag{3}$$

And the BBU server processing is limited by server capacity C as:

$$\sum_{j=1}^{N} \rho_{i,j}^{server} \leq C \tag{4}$$

The energy minimization in the cloud for M BBU servers can be formulated from two components [8]: dynamic and static power consumption. The dynamic energy consumption is dependent on the amount of processing resources on the server while the static part comprises the energy consumption irrespective of traffic load, but other purposes such as coolings, etc. Now, the energy minimization problem can be formulated as:

$$\underset{\xi_{i,j,k}}{min} \; \sum_{i=1}^{M} \left(\delta \sum_{j=1}^{N} \rho_{i,j}^{server} + \varepsilon_i P_{static} \right) \tag{5}$$

$$\varepsilon_i = \begin{cases} 0 & \sum_{j=1}^{N} \sum_{k=1}^{L} \xi_{i,j,k} = 0 \\ 1 & \text{Otherwise} \end{cases} \tag{6}$$

where δ is the power factor in GOPS/watts. ε_i shows the status factor of server S_i whether S_i is ON or OFF. P_{static} denotes the static power that is constant for every BBU server. Constraints are from (3) and (4).

3.2 Computational Resource Model

The baseband tasks from cells need to be quantified, i.e., they need to be mapped into computing processing in GOPS. The computing resource requirement per user per task is calculated based on the energy consumption model in [17]. The model provides energy modelling for different types of BSs such as macro, micro, pico and femto BSs. In this paper, the power equation in [17] for calculating the computing resources required for baseband tasks is adopted. Defining $\boldsymbol{\mathcal{L}}$ as the set of baseband tasks and $\boldsymbol{\mathcal{X}} = \{BW, Ant, Mod, Cod, R\}$ is the list of parameters affecting the scaling of baseband processing tasks, where BW, Ant, Mod, Cod and R are the system bandwidth, number of antennas used by a user, modulation bits, coding rate and number of PRBs respectively. The power equation is written as [17]:

$$P_u = \sum_{i\epsilon\boldsymbol{\mathcal{L}}} P_{i,ref} \prod_{x\epsilon\boldsymbol{\mathcal{X}}} \left(\frac{x_{act}}{x_{ref}}\right)^{s_{i,x}} \tag{7}$$

where P_u and $P_{i,ref}$ are the processing power required by user u and the processing power of reference system in [17]. The variables x_{act} and x_{ref} denotes the actual and reference values of parameters affecting baseband scaling. The variable $s_{i,x}$ denotes the scaling exponents. Users that generate traffic are randomly distributed in the cell area and the generated traffic are mapped into processing resources as per user per task. Even though there are many baseband tasks processed by a BS, this paper considers two baseband tasks for simplicity, i.e., $k = 2$: (i) Frequency-Domain (FD) processing for mapping/demapping and MIMO equalization, and (ii) Forward Error Correction (FEC) denoted by the following equations, respectively, in GOPS:

$$P_u^{FD} = \left(30Ant + 10Ant^2\right)\frac{R}{100} \tag{8}$$

$$P_u^{FEC} = 20\frac{Mod}{6}Cod * Ant * \frac{R}{100} \tag{9}$$

where P_u^{FD} and P_u^{FEC} are FD and FEC processing requirements, respectively, per user u per task k in GOPS. Ant is the number of antennas used per user, Mod is the modulation bits, Cod is the coding rate used and R is the number of PRBs used by u at time t. In the bin packing algorithm, the tasks per cell are packed on servers hence the processing requirements per task per cell for the two tasks is calculated as follows:

$$\begin{cases} \rho_{j,1}^{req} = \sum_{u\epsilon\mathbf{U}} P_{u,t}^{FD}, & \text{when } k = 1 \\ \rho_{j,2}^{req} = \sum_{u\epsilon\mathbf{U}} P_{u,t}^{FEC}, & \text{when } k = 2 \end{cases} \tag{10}$$

where $\mathbf{U}$ is the set of users within a cell.

3.3 FFD Bin Packing Scheme

A classical bin packing problem consists of packing a series of items with different sizes into a minimum number of bins with capacity C. The C-RAN resource allocation problem can be modelled as the bin packing problem where the aim is to pack items, called baseband tasks $\boldsymbol{\mathcal{L}}$, from cell areas $\boldsymbol{\mathcal{R}}$ into a set of servers $\boldsymbol{\mathcal{M}}$ such that the number of servers used are minimized and hence the energy consumption reduction. Since the problem of finding optimal packings is NP-hard, i.e., there is no way of being guaranteed the best solution without checking every possible solutions. Amongst many other solutions, the approximation algorithm is the mostly used because of fast heuristics that generate good but not necessarily optimal packings. The approximation algorithms of FFD is considered.

The FFD algorithm illustrated in Algorithm 1 is adopted, which is a natural way of finding the approximation bin packing. In this algorithm, all bins are initially empty. Sort all item tasks in descending order. Starts with the current number of bins M and item k. Consider all bins $GPP_i : i = 1, ..., M$ and place item task $\rho_{j,k}^{req}$ baseband task in the first bin that has sufficient residual capacity. If there is no such bin, increment i and repeat until all items is assigned.

Algorithm 1. First-Fit Decreasing Algorithm

Input: a set of RRH cells $\boldsymbol{R}$, a set of tasks $\boldsymbol{L}$ within RRH_j, their resource requirements $\rho_{j,k}^{req}$, and GPP list $\boldsymbol{M}$, GPP capacity C_{cap}^{i}
Output: Number of BBUs M
Sort all RRH tasks in decreasing order of $\rho_{j,k}^{req}$.
Launch one GPP of capacity $_{cap}^{i}$.
for *each* $\rho_{j,k}^{req}$ *that arrives* **do**
  if *there is a server where* $\rho_{j,k}^{req}$ *will fit* **then**
    Place $\rho_{j,k}^{req}$ into the left most GPP;
  else
    Launch a new GPP;
    Place $\rho_{j,k}^{req}$ into that GPP
  end
 end
end
Return M

3.4 Radio Side Energy Consumption Model

The power consumption model in [17] is modified to come up with a generalized component based power consumption model of a C-RAN RRH, denoted P_j, which is formulated as:

$$P_j = \begin{cases} N_{TRX} \frac{\frac{\rho_j P_{max}}{\eta_{PA}} + P_{RF}}{(1-\sigma_{DC})(1-\sigma_{MS})}; & \text{if } 0 < \rho_j \leq 1 \\ P_{sleep}; & \text{if } \rho_j = 0 \end{cases} \tag{11}$$

where N_{TRX}, ρ_j and P_{max} denotes the number of transmitter chains, the normalised traffic load of RRH_j and maximum transmission power of RRH_j. The variables P_{RF} and P_{BB} denotes the RF and BBU power consumption, respectively. The variables η_{PA}, σ_{feeder}, σ_{DC}, σ_{MS}, σ_{cool} denotes power amplifier efficiency, RF feeder losses, DC losses, MS losses and cooling losses, respectively.

3.5 Greedy BS Switch OFF Algorithm

The proposed algorithm is called the greedy RRH switch OFF algorithm and it is centralised and runs in the BS cloud inside the GCC where network information is available. In the algorithm, only the small cell RRHs are to be switch off based on a utility function $\boldsymbol{\mathcal{F}}_j(PRB)$, while maintaining quality of service (QoS) i.e., maintaining the minimum datarate, r_{min}.

$$\boldsymbol{\mathcal{F}}_j(PRB) = \frac{\text{number of } RRH_j \text{ PRBs occupied}}{\text{Total number of } RRH_j \text{ PRBs}} \tag{12}$$

It is assumed that there are N_{chn} available channels in every cell for transmission with each having bandwidth $BW = B/N_{chn}$ where B is the cell bandwidth. In this regard, a channel means one PRB which is allocated to each user per scheduling interval. For simplicity it is assumed that different frequency bands are used by adjacent BS so inter channel interference (ICI) has been taken care of. Thus, the minimum data rate of a user u can be formulated as:

$$r_{min} = BW.log_2\left(1 + \frac{\eta_0.P_u}{d^\alpha}\right) \tag{13}$$

where α is the path-loss exponent and $\eta_0 = G_0/N_0$ includes the effect of antenna gain G_0 and thermal noise N_0, and d is the distance from the RRH to the user. P_u is the transmission power per user. The MRRH is always kept on to maintain coverage. The algorithm runs in the BS cloud on the GCC where all information (e.g., traffic load) about other RHHs is present. At constant time intervals of 1 hour, the algorithm is invoked where the utility of all the SRRHs is calculated and the SRRH with the lowest utility first is then tested for switching OFF. The test involves checking if the SRRH traffic can be offloaded to neighbouring SRHHs or to the MRRH while maintaining the minimum datarate. If the traffic can be offloaded, then offloading is perfomed and the SRRH is then switched OFF. If the offloading can not be perfomed due to violation of QoS or due to not enough resources, the SRRH is kept active.

Algorithm 2. Greedy RRH switch off algorithm

Input: RRH traffic load information
Output: Number of switched OFF SRRHs
Define $\boldsymbol{\mathcal{R}}_{off}$ as a set of switched OFF RRHs.
Define $\boldsymbol{\mathcal{R}}_{on}$ as a set of active RRHs
for *each RRH* **do**
  Calculate the RRH utility function $\boldsymbol{\mathcal{F}}(\rho_j)$
end
Sort RRHs by their increasing utility function
Test the lowest utility RRH_j for switching OFF
while $j \neq N$ **do**
  if RRH_j *can be offloaded* **then**
    offload RRH_j traffic
    switch OFF RRH_j
    $\boldsymbol{\mathcal{R}}_{off} = \boldsymbol{\mathcal{R}}_{off} + \{RRH_j\}$
  end
  else
    do not offload RRH_j traffic
    keep RRH_j on
    $\boldsymbol{\mathcal{R}}_{on} = \boldsymbol{\mathcal{R}}_{on} + \{RRH_j\}$
  end
  $j = j + 1$
end

4 Simulation Results and Discussion

4.1 Parameter Settings

To analyse the performance of the proposed scheme, a simulation layout of one MRRH overlaid with 10 small cells is considered. Bandwidth of 10 MHz was considered with up to 50 users randomly generated within the MRRH and up to 5 users within the SRRH. All results using the proposed scheme are compared with the baseline distributed LTE-A system which comprises of distributed BSs with 10 individual BBU processing servers for 10 cells. The users are allocated PRB in a proportional-fair manner. Adaptive modulation and coding (AMC) scheme is used to adapt to the changing channel conditions. As the simulation runs, the values of Ant, Mod, R were captured and mapped into processing requirements and loaded into the bin packing scheme to reduce the number of servers M. For calculating the power consumption, the power factor used is δ = 40 GOPS/watt and P_{static} = 200 GOPS as in [17]. Other simulation parameters are shown in Table 1.

4.2 Results Evaluation

Figure 3 shows the power consumption of both schemes for different traffic loads. The results show that for both schemes, as the traffic load increases, the power

Table 1. Parameters used in the simulations.

Parameter	Value
Bandwidth B	10 MHz
No. of antennas Ant	2
Modulation mod	2, 4, 6 bits
Coding rate Cod	1/3-1
Number of MRRH	1
Number of SRRH	10
Number of users per MRRH	up to 50
Number of users per SRRH	up to 5
MRRH radius	500 m
SRRH radius	40 m
MRRH transmission power	46 dBm
SRRH transmission power	30 dBm
Inter-BS distance	>1 km
BS antenna gain G_0	16 dBi
Noise Power N_0	-141 dBm/Hz
Pathloss Exponent, α	4
power factor, δ	40 GOPS/watt

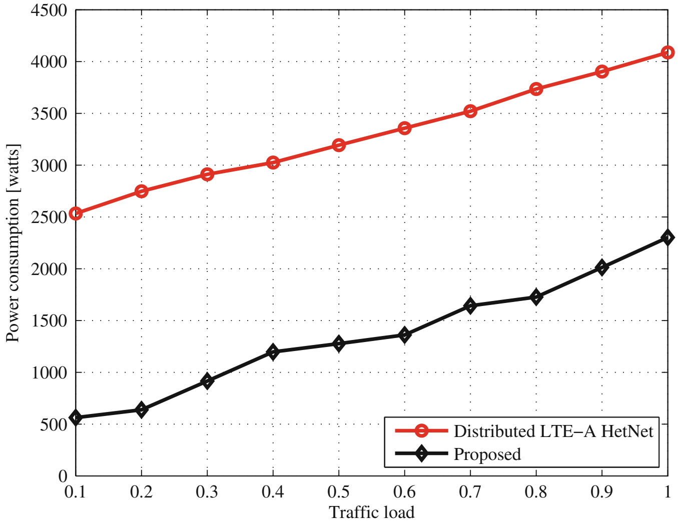

Fig. 3. Power consumption versus traffic load.

consumption increases. The proposed scheme consumes less power compared to the baseline since the proposed scheme combines RRH switch off scheme at the radio side with BBU reduction scheme at the BS cloud which both significantly reduce the overoll system power consumption. The proposed scheme saves up to 44% and 78% of power at during low traffic and peak traffic respectively as compared to the baseline scheme.

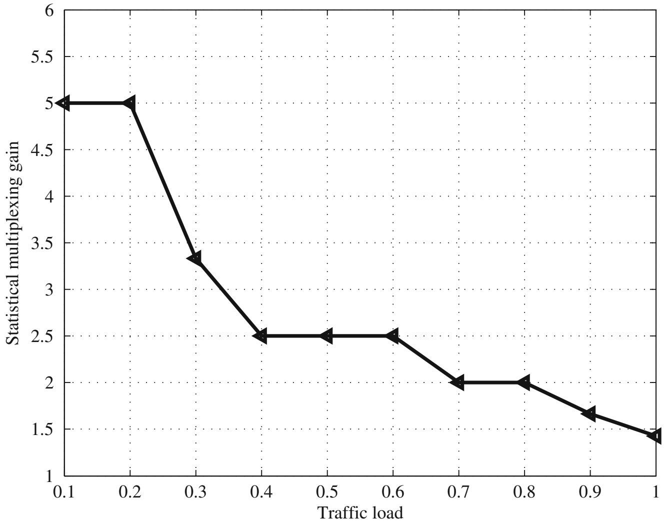

Fig. 4. Statistical multiplexing gain.

Figure 4 illustrate the statistical multiplexing gain for increasing traffic load which is calculated as the ratio of number of BBU servers used by the baseline scheme to those used by the proposed scheme. The graph shows that at low traffic period the multiplexing gain is 5 which means the baseline uses 5 times the number of BBUs compared the proposed scheme. During peak traffic, the multiplexing gain is 1.4 which means the baseline uses 1.4 times the number of BBUs as the proposed scheme due to more traffic being proposed for both schemes.

5 Conclusion

This paper presents an energy efficient scheme for reducing energy consumption in C-RAN by switching off remote radio heads (RRHs) using a greedy algorithm and also switching off idle BBUs using the first fit decreasing (FFD) bin packing algorithm. The number of RRHs and BBUs are minimized by matching the right amount of baseband computing load with traffic load. The proposed scheme saves up to 44% and 78% of power during low traffic and peak traffic periods respectively. The proposed scheme will be extended to include the separation of data and control signalling to further minimize energy consumption.

References

1. Huawei: 5G: a technology vision. In: White Paper of Huawei Tech (2013)
2. Chih-Lin, I., Rowell, C., Han, S.: Toward green and soft: a 5G perspective. IEEE Commun. Mag. **52**(2), 66–73 (2014)
3. Auer, G., Giannini, V., Desset, C., Godor, I.: How much energy is needed to run a wireless network? IEEE Wireless Commun. **18**(5), 40–49 (2011)
4. Yaacoub, E.: A practical approach for base station on/off switching in green LTE-A hetnets. In: IEEE 10th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob). IEEE, pp. 159–164 (2014)
5. Cheng, J.-F., Koorapaty, H., Frenger, P., Larsson, D., and Falahati, S.: Energy efficiency performance of LTE dynamic base station downlink DTX operation. In: 2014 IEEE 79th Vehicular Technology Conference (VTC Spring), pp. 1–5. IEEE (2014)
6. Niu, Z., Wu, Y., Gong, J., Yang, Z.: Cell zooming for cost-efficient green cellular networks. IEEE Commun. Mag. **48**(11), 74–79 (2010)
7. Sigwele, T., Pillai, P., Hu, F.: Call admission control in cloud radio access networks. In: International Conference on Future Internet of Things and Cloud (FiCloud), pp. 31–36. IEEE (2014)
8. Alam, A.S., Dooley, L. S., Poulton, A.S.: Energy efficient relay-assisted cellular network model using base station switching. In: Global Telecommunications Workshop (GLOBECOM 2012) on Multicell Cooperation (MuCo). IEEE, Anaheim, CA, USA (2012)
9. Alam, A.S., Dooley, L.S.: A scalable multimode base station switching model for green cellular networks. In: IEEE Wireless Communication and Networking Conference (WCNC), pp. 1–6 (2015)
10. Namba, S., Warabino, T., Kaneko, S.: BBU-RRH switching schemes for centralized ran. In: 7th International ICST Conference on Communications and Networking in China (CHINACOM), pp. 762–766. IEEE (2012)
11. Kong, Z., Gong, J., Xu, C.-Z., Wang, K., Rao, J.: eBase: a baseband unit cluster testbed to improve energy-efficiency for cloud radio access network. In: IEEE International Conference on Communications (ICC), pp. 4222–4227. IEEE (2013)
12. Liu, J., Zhou, S., Gong, J., Niu, Z.: On the statistical multiplexing gain of virtual base station pools. In: Global Communications Conference (GLOBECOM), pp. 2283–2288. IEEE (2014)
13. Zhao, T., Wu, J., Zhou, S., Niu, Z.: Energy-delay tradeoffs of virtual base stations with a computational-resource-aware energy consumption model. In: IEEE International Conference on Communication Systems (ICCS), pp. 26–30. IEEE (2014)
14. Namba, S., Matsunaka, T., Warabino, T., Kaneko, S., Kishi, Y.: Colony-ran architecture for future cellular network. In: Future Network & Mobile Summit (FutureNetw), pp. 1–8. IEEE (2012)
15. Ashraf, I., Boccardi, F., Ho, L.: Sleep mode techniques for small cell deployments. IEEE Commun. Mag. **49**(8), 72–79 (2011)
16. Abdulkafi, A.A., Kiong, T.S., Koh, J., Chieng, D., Ting, A.: Energy efficiency of heterogeneous cellular networks: a review. J. Appl. Sci. **12**(14), 1418 (2012)
17. Desset, C., Debaillie, B., Giannini, V., Fehske, A., Auer, G., Holtkamp, H., Wajda, W., Sabella, D., Richter, F., Gonzalez, M.J.: Flexible power modeling of LTE base stations. In: Wireless Communications and Networking Conference (WCNC), pp. 2858–2862. IEEE (2012)