A Green Load Balancing Algorithm for Dynamic Spatial-Temporal Traffic Distribution in HetNets

Jichen Jiang(✉), Xi Li, and Hong Ji

Key Laboratory of Universal Wireless Communications, Ministry of Education,
Beijing University of Posts and Telecommunications,
Beijing, People's Republic of China
{jiangjichen,lixi,jihong}@bupt.edu.cn

Abstract. With the increasing users demands, the data traffic in the network reveal different characteristics in both spatial and temporal dimensions, bringing severe load imbalance problem. This may impact resource utilization, users experience and system energy efficiency, and then need further investigation. In this paper, we propose a distributed load-balancing algorithm considering this spatial-temporal variation in a two-tier heterogeneous network. Instead of illuminating the spatial-temporal influence, we make use of this characteristic while designing the algorithm, and accordingly switch ON/OFF small cell base stations (SBSs) for improving the energy efficiency. A load factor described with load variance is derived, based on which the problem is formulated as a non-linear integer programming that seeks to minimize a load function. Then a suboptimal solution is obtained by an effective heuristic algorithm. Simulation results show that our proposed algorithm balances the traffic load better and significantly reduces the total energy consumption, compared with conventional load-balancing scheme.

Keywords: Spatial-temporal variation · Load balance
Small cell ON/OFF

1 Introduction

With the development of mobile communication technology, the data traffic flow of wireless cellular network has grown rapidly in the past decade [1–4]. This growth not only increases the traffic load on wireless cellular network, but also reveals dynamic characteristics of data traffic in both temporal and spatial dimensions. This spatial-temporal variation of data traffic distribution may lead to load imbalance and impact the network performance in energy efficiency, users quality of service (QoS) requirements and resource utilization. So how to solve the load imbalance problem in different scenarios is worth studying.

There have been approaches proposed about traffic load balancing, such as complementary communication technologies, direct transfer, cell range expansion

© ICST Institute for Computer Sciences, Social Informatics and Telecommunications Engineering 2018
B. Li et al. (Eds.): ChinaCom 2017, LNICST 237, pp. 395–404, 2018.
https://doi.org/10.1007/978-3-319-78139-6_40

(CRE), and so on [5–9]. These works have designed algorithms for the load-balancing problem in conventional scenario and achieved good results. However, existing works are mainly aimed at illuminating the influence caused by the variations. In fact, those variations could also be used to cope with the load imbalance problem.

Reference [5] provides a strategy to offload the mobile data traffic from cellular network to WiFi infrastructure for extra capacity and better network performance. This work also proposes a load-based SNR-threshold algorithm that controls the amount of WiFi offloading. Reference [6] analyzes the downlink and uplink transmission performance in two-tier heterogeneous network when micro BS employs CRE and derives an optimal CRE bias and optimal transmit power of SBSs to maximize transmission success probability and energy efficiency. An analytical framework is proposed in [7] which allows to analyze the trade-off between the energy consumption and the traffic offload of cognitive network and can be used for the energy efficient design and operation of cognitive small cell access points. In [8] a new separation architecture called Hyper-Cellular Network (HCN) is proposed, which is supposed to guarantee the coverage. Then two small cell sleeping schemes (random and repulsive schemes) are proposed in HCN. Numerical results show that these two schemes can keep the load of the network balanced and save a lot of energy. In [9] a dynamic pico on/off algorithm is proposed in which the traffic load coefficient is formulated as a function of bandwidth. By controlling this load coefficient under a fixed threshold, the traffic load can be seen as balanced and minimum pico base stations will be switched to active mode.

Although the above works provide many good insights on balancing traffic load and improving energy efficiency, these methods are not appropriate for some typical spatial-temporal variation scenario. For example, in a shopping mall area, there will be not much people early in the morning and these people would more likely to be in the supermarket underground. As time goes by, there will be more and more users arriving at the mall and by noon most of them choose to go upstairs having lunch. Due to this variation, the traffic load would be unbalanced on these two dimensions. Frequent handover from one SBS to another could cause a decline in throughput and result in user experience decreasing.

In this paper, we propose an algorithm of load balancing based on SBSs sleep mode in two-tier heterogeneous networks (HetNets), in which SBSs can decide automatically whether to switch on or off, using the spatial-temporal variation characteristics of data traffic distribution. This self-organized algorithm is able to adjust automatically as the traffic load changes. We use a load factor to measure the traffic load level which is formulated as the variance of traffic load. The goal of the algorithm is to minimize this load factor. We model this problem into a non-linear integer programming problem in which an optimal solution can not be obtained in linear time. So we reformulate the problem into a max-min optimization problem and find a method to solve it through genetic algorithm. Simulation results show that the proposed algorithm can balance the network traffic load efficiently and save the total energy consumption, compared with conventional load-balancing approach.

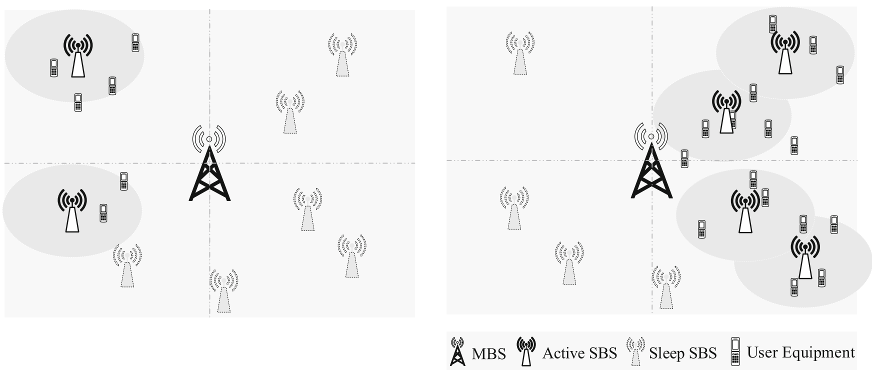
The rest of this paper is organized as follows. In Sect. 2, the system model is defined and the studied scenario is described. Section 3 presents the problem formulation and describes the proposed load-balancing algorithm, followed by simulation results and conclusion in Sects. 4 and 5, respectively.

2 System Model and Scenario

2.1 Network Scenario

We consider a two-tier HetNet, where several SBSs are randomly distributed in a macro base station (MBS) covered area, where users distributed in different location at different time. We illustrate the scenario as in Fig. 1. The upper left area is numbered as region 1, and remaining areas are clockwise marked as region 2, region 3 and region 4. At time t_1, user equipments are mostly distributed in region 1 and region 4, so only the SBSs in these two regions are switched on while others are off, as Fig. 1(a) shows. And as time goes by, at time t_2, some user equipments leave region 1 and region 4 and appear in region 2 and region 3, leading to the fact that SBSs in region 1 and region 4 are switched off meanwhile SBSs in region 2 and region 3 are switched on.

Without loss of generality, we assume that the arrival of users follows a Poisson Point Process (PPP), whose arrival rate at time t is λ. Orthogonal bandwidth is used by different layers to avoid the severe inter-layer interference, especially for protecting the signaling coverage [8]. System bandwidth is $\mathcal{W}$ and the bandwidth of each subcarrier is B. In this paper, we only consider the scenario that the SBSs covered with one MBS. We denote the sets of SBSs and users by $\mathcal{J} = \{1, 2, \cdots, J\}$ and $\mathcal{I} = \{1, 2, \cdots, I\}$, respectively. The transmission power of SBS_j, $j \in \mathcal{J}$, is P_j, and h_{ij} is used to denote the channel gain between SBS_j and $user_i$. In order to realize the load balancing of the whole network, each SBS



(a) User equipments distribution at time t_1 (b) User equipments distribution at time t_2

Fig. 1. Illustration of the studied network at different time.

is set to have two operation modes, one is the active mode, and another is the sleep mode. To express the problem better, an indicator $\xi_{i,j}^{w}$ is used to describe the connection state between stations, users and bands. If $\xi_{i,j}^{w} = 1$, the $user_i$ is associated with the SBS_j and deployed on $band_w$; otherwise, $\xi_{i.j}^{w} = 0$.

2.2 Traffic Model

In this paper, we assume that a user equipment can only be served by one active SBS at one subcarrier frequency. Therefore, the received signal to interference and noise ratio (SINR) of $user_i$ from SBS_j is given by

$$SINR_{i,j}^{w} = \frac{\xi_{i,j}^{w} p_j |h_{ij}|^2}{\sigma^2 + \sum\limits_{k \neq j} \zeta_k^w p_k |h_{ik}|^2} \tag{1}$$

where σ^2 is the thermal noise energy, and $|h_{ij}|^2$ is the channel gain from SBS_j to $user_i$, $j \in \mathcal{J}$, $i \in \mathcal{I}$. If $\zeta_k^w = 1$, it indicates that SBS_j is active and use $band_w$ to transmit signals. Further, according to Shannon Theory, the data rate of $user_i$ is given by

$$r_{i,j}^{w} = B \log_2(1 + SINR_{i,j}^{w}) \tag{2}$$

We use ρ_j to denote the ON/OFF state of SBSs, which is defined as follows:

$$\rho_j = \begin{cases} 1 & \text{SBS}_j \text{ is active} \\ 0 & \text{otherwise} \end{cases}$$

so we can obtain the number of active SBSs in the network from above. Accordingly, the average rate of whole network can be denoted as

$$\bar{r} = \frac{\sum\limits_{i,j,w} r_{i,j}^{w}}{\sum\limits_{j} \rho_j} \tag{3}$$

where $\sum\limits_{i,j,w} r_{i,j}^{w}$ enlightens the sum of data rate brought by the users that associated with SBSs in the network.

3 Problem Formulation and Algorithm Design

3.1 Problem Formulation

Based on the network model, we consider a mechanism to balance the system load by controlling the SBSs' operation mode to active mode or sleep mode, which aims at keeping the load of whole network balanced around an average level while ensuring the QoS of users. In a period of time, the number of users may change and could distribute in different part of the area, so the SBSs of different location may have different operation mode. By switching some SBSs of the network ON/OFF, the load of each SBS can be stabilized near the mean value, ensuring the load of the network to be balanced. In other words, if we keep the difference between the load of each SBS and the average load of the whole network as low as possible, then the network can be seen as a load-balanced system. The load deviation of SBS_j can be written in the following expression:

$$d_j = \sum_w \sum_i \xi_{i,j}^w r_{i,j}^w - \bar{r} \tag{4}$$

As to balance the load of the network, the load deviation of each SBS has to be as low as possible, that is, the load variance of the network has to be minimum. In this paper, we derive a load factor to describe the load variance, which is defined as:

$$\mathcal{L} = \sum_j d_j^2 \tag{5}$$

For the purpose of minimizing the load variance of the network, the optimal load-balancing problem can be formulated as the following:

$$\begin{aligned} & \min \mathcal{L} \\ s.t. \quad & \sum_w \sum_i \xi_{i,j}^w \le 1, i \in \mathcal{I}, j \in \mathcal{J}, w \in \mathcal{W} \\ & P_j \le P_{max}, i \in \mathcal{I} \\ & \xi_{i,j}^w \in \{0,1\}, i \in \mathcal{I}, j \in \mathcal{J}, w \in \mathcal{W} \end{aligned} \tag{6}$$

where P_{max} is the maximum transmit power of SBS.

The goal of this work is to propose a load-balancing algorithm in which the traffic load of each SBS is balanced near a average value by switching the SBS to active or sleep. The optimization formulation is to minimize the traffic load variance as in Eq. (6). Note that (6) is a non-linear integer optimization programming problem which is not suitable to be solved by linear integer programming method. Fortunately, we notice that the formulation can be seen as a max-min optimization problem [10], then the problem can be re-formulated as:

$$\begin{aligned} & \max \sum z \\ s.t. \quad & z \le d_j, i \in \mathcal{I}, j \in \mathcal{J}, w \in \mathcal{W} \\ & \sum_w \sum_i \xi_{i,j}^w \le 1, i \in \mathcal{I}, j \in \mathcal{J}, w \in \mathcal{W} \\ & P_j \le P_{max}, i \in \mathcal{I} \\ & \xi_{i,j}^w \in \{0,1\}, i \in \mathcal{I}, j \in \mathcal{J}, w \in \mathcal{W} \end{aligned} \tag{7}$$

3.2 Load-Balancing Algorithm Design

Problem (7) is still difficult to get an optimal solution in linear time scale, thus, we propose a load-balancing algorithm through switching SBSs ON/OFF based on genetic algorithm (GA).

When there are users arriving at the area, the data traffic is also brought to the network, resulting in the traffic load change of SBSs. As this change happened, SBSs will check their operation modes and decide to change or not according to Eq. (6). If the total traffic load is increasing, then several SBSs should be switched active to share some load. On the contrary, when users depart from the area, the data traffic is also reducing, causing some SBSs have to be switched off. Therefore, different operation mode of SBSs will directly affect the traffic load distribution of the network, so our goal is to choose the appropriate $\xi^w_{i,j}$ for $user_i$ and SBS_j keeping the load function minimum.

Algorithm 1. Load-Balancing Algorithm through Switching SBSs ON/OFF via GA

Initialization:
  a)Users arrive and randomly distributed in the network
  b)Set the maximum number of iterations M, population size Z and encoding length l, p_c, p_m and p_s
  c)Initialize the population according to $\mathcal{Z}$, l, set minV $= V_1$
for $m = 1, 2, \cdots, M$ **do**
  for each individual z in the population,$z = 1, 2, \cdots, Z$ **do**
    Calculate the objective function: the load variance V for each individual
  end for
  if $p_1 < p_m$ **then**
    Calculate the mutation point
    if $mpoint = 1$ **then**
      change the value to 0
    else
      change the value to 1
    end if
  end if
  if $p_2 < p_c$ **then**
    Calculate the crossover point and exchange the value of $individual_z$ with $individual_{z+1}$
  end if
  Calculate $\sum V$ and $\frac{V_z}{\sum V}$
  if $\frac{V_z}{\sum V} < p_s$ **then**
    Select this individual as a member of new population
  end if
  Update the population
  if $Var_z < minV$ **then**
    minV $= Var_z$
  end if
end for
Output the minV, best individual, and obtain the load-balancing policy according to the result

Because the problem is NP-hard which is difficult to obtain an optimal solution, so we derive a suboptimal solution through GA. GA updates a population of solutions via genetic operators such as crossover, mutation and selection to achieve offsprings with better quality until some convergence criteria are met [11]. At each generation, a GA is capable of producing and maintaining a set of feasible solutions, maintaining a population of candidate solutions, and evaluating the quality of each candidate solution according to the objective fitness function. We view each combination of SBS operation modes as an individual of the population, that is, a combination of all the SBSs' binary-state mode value. The original population is set randomly. Equation (7) is the objective fitness function. Mutation probability p_m and crossover probability p_c is used to decide whether to perform mutation and crossover procedures for this individual or not. After the traffic load variance of each individual is calculated, the lower value will be selected, and the corresponding individual will be seen as a good individual and insert to the new population. After several times of iteration, the individuals in the population will tend to be consistent and the solution to our problem can be obtained. This load-balancing algorithm can be summarized as Algorithm 1.

4 Simulation Results

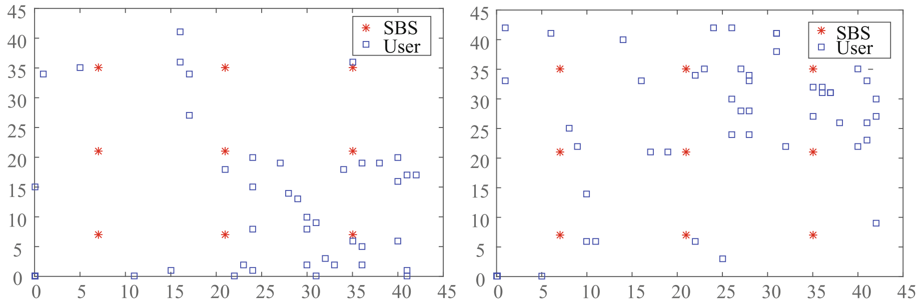
In this section, simulation results are present to show the performance of our proposed algorithm. We consider a scenario that several SBSs are covered by one MBS where those SBSs are deployed uniformly. The parameter used for the simulations are summarized in Table 1. Additionally, we compare the following algorithms for performance evaluating in the this simulation:

– Proposed algorithm: Control the operation mode of each SBS and calculate the traffic load of each SBS. Find the best operation modes combination and obtain the minimum variance of the load in the network.
– Conventional approach: SBSs don't have the capability to switch between sleep-active modes and don't apply the load-balancing scheme either, referred to hereinafter as "SBSs all-on".
– Classical Load-Balancing (LB) algorithm: The SBS with over-loaded traffic will select one potential target SBS with under-loaded traffic to offload, the

Table 1. Parameter

Parameter	Value
Carrier frequency	2 GHz
System bandwidth	10 MHz
Inter site distance of SBSs	40 m
Thermal noise	−174 dBm/Hz
Path loss model (d in km)	$140.7 + 37.6 \log_{10} d$
Maximum transmit power	30 dBm
Users distribution	3/4 in the hot spot

whole process of which is motivated by the settings of self-adjustment of specific system parameters in LTE, referred to hereinafter as "conventional algorithm".



(a) $T = t_1$ (b) $T = t_2$

Fig. 2. Distribution of SBSs and users.

Figure 2 illustrates the scenario. The simulation starts with 20 users, randomly distributed in the area and as time goes by, more and more users arrive in and some of them will centrally distributed in one hot spot which is not a fixed location and will change as time varies. So not only the number of users in the network will change, but also the users' spatial distribution. We assume that each 10 min a maximum of 10 users will join the network and there will be 3/4 of total users are distributed in one hot spot area while others are uniformly distributed in the remaining area within the coverage of MBS.

Figure 3 shows the traffic load variance in the aforementioned scenario about the above three algorithm as time varies from 10 min to 80 min, which indicates the balancing level of the network. To meet the QoS requirement of users, the data throughput offered by SBSs has to be up to 80% of the total data request, and the rest will be transferred to MBS. We can see from the result that although the number and distribution of users in the network is changing over the time, our proposed algorithm manages to stabilize the variance of traffic load and maintain at a lower value by switching SBSs to active or sleep modes, compared with the classical LB algorithm and "SBSs all-on" algorithm.

In Fig. 4, we show the total energy consumption of the network. In this simulation, we only consider the transmission power of the SBSs, for example, if SBS_j is in active mode, then the energy consumption is P_j, otherwise, the energy consumption is 0. As in the conventional LB algorithm, SBSs are always in active mode in which the total energy consumption stays the same while in our proposed algorithm, SBSs' operation modes are different from each other since the number and distribution of users vary from time to time. So when some SBSs are switched to sleep mode, they actually save a large portion of energy, relative to the classical LB algorithm.

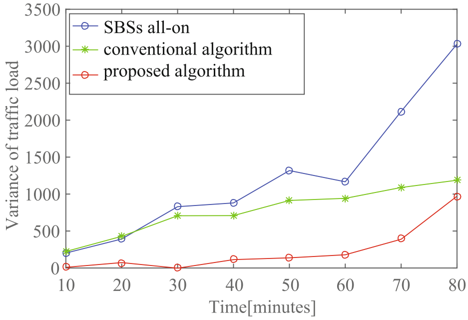


Fig. 3. Variance of traffic load per SBS as time varies.

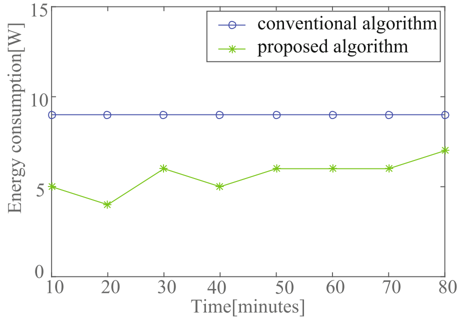
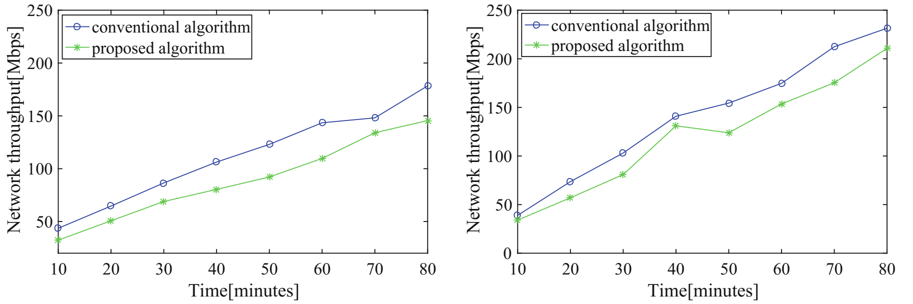


Fig. 4. Energy consumption of the network with respect to time.



(a) 10 users join the network/10 minutes (b) 20 users join the network/10 minutes

Fig. 5. Network throughput with respect to time.

Figure 5 shows the data traffic throughput of the network using our proposed algorithm and classical LB algorithm with different user's arrival rate. We can see from the result that data throughput increases as time flows, for the reason that the number of users increases with time. The value is not much different between those two algorithms, even the classical LB algorithm is a litter larger. The main reason is that our proposed algorithm switches off some unnecessary SBSs while the classical LB algorithm is not, thus a few users have to associate with the MBS which can lower down the throughput slightly (only the data throughput of SBSs is measured in our simulation).

5 Conclusion

In a two-tier HetNet where the users distribution and data traffic are changing from time and geographical location, it may result in traffic load imbalance problem. We propose an energy-efficient load-balancing algorithm in which SBSs can automatically switch between active or sleep mode. In order to describe the load difference among each SBSs, a load factor is derived and modeled as a load variance. Then, to improve the energy efficiency of the system, switching SBSs ON/OFF is considered. Since the problem is NP-hard and an optimal solution is difficult to obtain, we reformulate the problem to a max-min optimization problem and solve it with the proposed algorithm via GA. Simulation results show that the proposed algorithm is energy efficient and can obviously balance the traffic load in the network, compared with classical LB algorithm.

Acknowledgement. This paper is jointly sponsored by the National Natural Science Foundation of China (Grant No. 61671088) and the National Science and Technology Major Project of the Ministry of Science and Technology of China (Grant No. 2016 ZX03001017).

References

1. Chen, L., Yu, F.R., Ji, H., Rong, B., Li, X., Leung, V.C.M.: Green full-duplex self-backhaul and energy harvesting small cell networks with massive mimo. IEEE J. Sel. Areas Commun. **34**(12), 3709–3724 (2016)
2. Mehmeti, F., Spyropoulos, T.: Performance analysis of mobile data offloading in heterogeneous networks. IEEE Trans. Mob. Comput. **16**(2), 482–497 (2017)
3. CISCO: Cisco visual networking index: global mobile data traffic forecast update 2016–2021, February 2017. http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.html
4. Zhang, H., Chen, S., Li, X., Ji, H., Du, X.: Interference management for heterogeneous networks with spectral efficiency improvement. IEEE Wirel. Commun. **22**(2), 101–107 (2015)
5. Hagos, D.H., Kapitza, R.: Study on performance-centric offload strategies for LTE networks. In: 6th Joint IFIP Wireless and Mobile Networking Conference (WMNC), pp. 1–10, April 2013
6. Yang, K., Wang, P., Hong, X., Zhang, X.: Joint downlink and uplink network performance analysis with CRE in heterogeneous wireless network. In: 2015 IEEE 26th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), pp. 1659–1663, August 2015
7. Wildemeersch, M., Quek, T.Q.S., Slump, C.H., Rabbachin, A.: Cognitive small cell networks: energy efficiency and trade-offs. IEEE Trans. Commun. **61**(9), 4016–4029 (2013)
8. Zhang, S., Gong, J., Zhou, S., Niu, Z.: How many small cells can be turned off via vertical offloading under a separation architecture? IEEE Trans. Wireless Commun. **14**(10), 5440–5453 (2015)
9. Jin, Z., Pan, Z., Liu, N., Li, W., Wu, J., Deng, T.: Dynamic pico switch on/off algorithm for energy saving in heterogeneous networks. In: 2015 IEEE 81st Vehicular Technology Conference (VTC Spring), pp. 1–5, May 2015
10. Vlachos, C., Friderikos, V.: Optimal device-to-device cell association and load balancing. In: 2015 IEEE International Conference on Communications (ICC), pp. 5441–5447, June 2015
11. Keskinturk, T., Yildirim, M.B.: A genetic algorithm metaheuristic for bakery distribution vehicle routing problem with load balancing. In: 2011 International Symposium on Innovations in Intelligent Systems and Applications, pp. 287–291, June 2011