Small Cell Operation Mode Control for Energy Saving in Heterogeneous Mobile Networks

Hui Xiao and Mythri Hunukumbure
Future Network Research Division,
Fujitsu Labs of Europe Ltd, UK
Email: Hui.Xiao@uk.fujitsu.com, Mythri.Hunukumbure@uk.fujitsu.com

***Abstract*—Small cells have been increasingly deployed over the coverage provided by macro base stations (BSs) to satisfy the capacity demand at high traffic areas, adding to the costs in energy consumption for running the network. This paper provides a solution for dynamically switching the operation modes of small cells between idle and active according to the traffic variation of network exploiting the distinct characteristics of power consumption models of various BSs to achieve energy saving (ES) gains without compromising the quality of service (QoS) to network users. The validity and effectiveness of the proposed solution have been evaluated through extensive simulation results based on a sophisticated network simulator built in-house, which features realistic scenario setting, spatially and temporally varying traffic emulation, reasonable network planning and a variety of practical network performance metrics.**

Keywords—Small cell; Energy saving; Operation mode switching; Cell deactivation

I. INTRODUCTION

As the demand for mobile data services increases, mobile operators face critical challenges of enhancing network capacity while reducing the costs of operating and maintaining the networks. The peak traffic demands show huge variations in spatial and temporal domains, with much smaller granularity than the coverage foot prints of macro cells. Smaller cells using micro, pico, femto BSs have been increasingly deployed over the coverage provided by macro BSs to provide additional capacity and hence meet this variable traffic demand, leading to heterogeneous cellular networks [1]. However the smaller cells are less efficiently operated during off-peak hours, resulting in unnecessary power consumption and high operational costs. This problem needs a coordinated small cell activation/deactivation solution, which considers the spatial and temporal variations in the traffic demand.

The heterogeneous network (HetNet) can also be an inter-Radio Access Technology (RAT) deployment model, particularly for LTE and legacy 3G RATs. Initially LTE deployments are likely to be with smaller cells covering capacity hotspots, while the wider area coverage will be provided by the existing 3G macro cells [2]. This model also provides ample opportunities for ES, by switching OFF LTE small cells during off-peak demand periods. In this paper the terms “switching OFF” and “switching to sleep mode” are used interchangeably.

The switching smaller cells ON/OFF for energy saving is a fairly established idea in Green Wireless (GW) research. The LTE standardization body, 3GPP have identified several approaches in this regard in [3]. The Green Touch research consortium [4] is also actively looking at related ES aspects. For example, a Self-Organized Networks (SON) based ON/OFF solution is discussed in [5]. In a broad survey of current GW enabling technologies [6], the cell switch ON/OFF is noted as a prominent network management strategy.

The novelty of our proposal includes:

- Adaptive operation mode control based on the spatially and temporally varying traffic distribution trends;
- Flexibility to handle various power models derived from different kinds of BS equipment specifications;
- Capability to deal with a range of radio technologies that are deployed to provide capacity enhancements for high traffic hotspots.

In addition, we demonstrate that it is more energy efficient for the small cells in the active mode to capture additional load from the underlay macro cells. In summary, we provide criterion to switch OFF small cells and also conditions to justify transfer of active users from macro cells to small cells which continue to be ON.

The remainder of this paper is organized as follows. In section II, the system models i.e. the power consumptions models for different BSs, the traffic models and the network performance metrics are presented. The controlling of dynamic operational modes, in terms of traffic distribution and individual cell power consumptions, is presented in section III. Section IV contains the simulation descriptions and related results. Finally in section V, the conclusions from this work are discussed.

II. SYSTEM MODEL

A small cell is a low-power low cost BS primarily designed to improve the radio coverage and boost capacity in residential, enterprise or outdoor hotspot areas. In this paper the terminology “small cell” is used to represent a range of cells deployed with BSs that transmit power much lower than typical macro BSs, including femto cells, pico cells and micro cells in the order of increasing cell size. Our proposed method will be explained based on the HetNet where macro cells provide ubiquitous basic coverage for a certain area and small cells are overlaid over the macro cells using separate frequency spectrums to meet the additional capacity demand at peak user activity times. To quantify the energy consumption for an operating region of an operator's network, the power consumption models of small cell and macro cell and the traffic distribution model based on which the power consumption for a given time period and an area can be calculated are introduced in detail in the subsections II.A and II.B. To evaluate the ES gains of our proposed method and its impact on the QoS delivered to network users, a number of assessment metrics are presented in subsection II.C.

A. *Power consumption models*

We adopted the state-of-the-art generic power consumption models of various LTE BS types published by the EARTH project [7] to calculate the energy consumption of a BS with the variation of traffic load. According to [7], a BS's power consumption model, regardless of the type of a BS, can be approximated by a linear mathematical model and expressed as:

$$P_{in} = \begin{cases} N_{TRX} \cdot P_0 + \Delta_p P_{out}, & 0 < P_{out} \le P_{\max} \\ N_{TRX} \cdot P_{sleep}, & P_{out} = 0 \end{cases} \quad (1)$$

where the meanings of parameters are: P_{in} is the consumed input power to attain a certain output power P_{out}, $P_{\max}$ denotes the maximum RF output power at maximum load, P_0 is the power consumption calculated at the minimum possible output power, N_{TRX} is the number of transceiver chains of a BS and Δ_p is the slope of load-dependent power consumption. In an LTE downlink, the BS load $P_{out}/P_{\max}$ is proportional to the utilized resources. Therefore P_{out} is determined by:

$$P_{out} = L \cdot P_{\max} \quad (2)$$

where L is the traffic load of a BS in terms of the percentage of consumed resources. Using equation (2), the power consumption model of a BS can be rewritten as:

$$P_{in} = \begin{cases} N_{TRX} \cdot P_0 + g_p L, & 0 < P_{out} \le P_{\max}, g_p = \Delta_p P_{\max} \\ N_{TRX} \cdot P_{sleep}, & P_{out} = 0 \end{cases} \quad (3)$$

Equation (3) shows that the power consumption of a BS essentially varies with the supported traffic load, which is measured by the percentage of consumed resources by the served users.

Table 1 presents the power model parameters for different BS types [7], from which we can see that there is a significant difference in power consumption when a cell is activated but has no traffic (with minimum transmit power for broadcast control channels) compared to when a cell is deactivated (i.e. zero transmit power). Therefore deactivating a cell during low traffic load periods has the potential for reducing the energy consumption of the network.

Additionally the power consumption profiles (based on equation (3)) of various BSs w.r.t. the traffic load are shown in Figure 1 (assuming $N_{TRX} = 2$ for each BS type). By comparison, it is obvious that the gradient of the macro type is the steepest, and hence for per unit load increase, the power consumption increment of a macro BS is much larger than that of a smaller BS, which is an important property that is used by our proposed ON/OFF control method.

TABLE I. POWER MODEL PARAMETERS FOR DIFFERENT BS TYPES

BS type	P_{max} [W]	P_0 [W]	Δ_p	P_{sleep} [W]
Macro	20	130	4.7	75
Micro	6.3	56	2.6	39
Pico	0.13	6.8	4	4.3
Femto	0.05	4.8	8	2.9

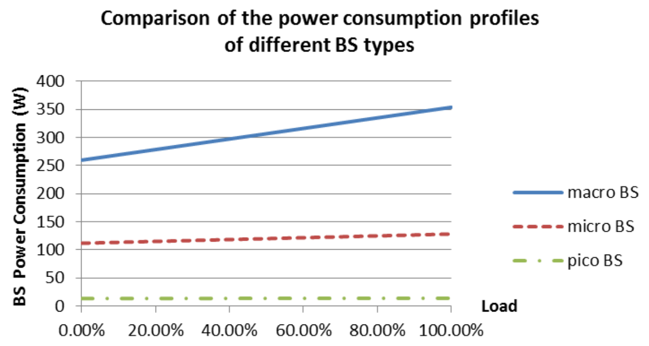

Figure 1 Power consumption profiles of different BS types

B. *Traffic distribution models*

The traffic distribution of a certain geographical area has both temporal and spatial variations. Depending on the type of hotspot regions where small cells are deployed, the daily variation of data traffic has different profiles. For example, during weekdays the data traffic at business hotspots mainly distributes between 9am and 5pm, but at transportation hotspots the data traffic peaks before and after the working hours.

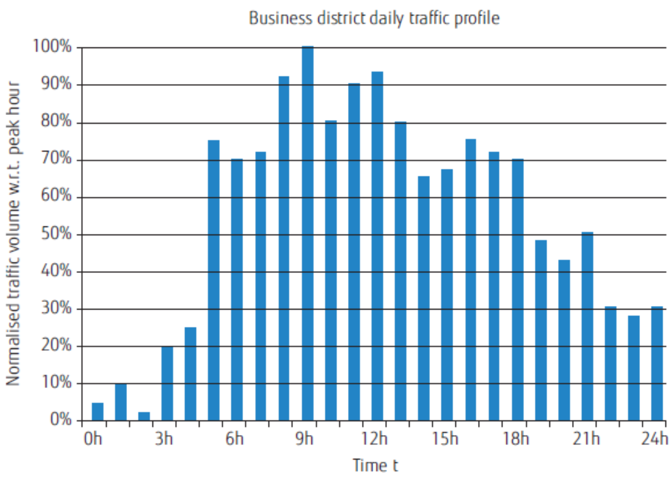

Figure 2 Daily traffic profile of business district

In order to embody the temporal variation in the traffic volume of a designated area, we sampled the continuous time domain into discrete time points and assign traffic volume values to the sampling points with each representing the aggregated traffic volume over one sampling interval. Figure 2 illustrates the daily traffic profile that we defined for business district type of hotspot regions, with one hour sampling interval, which is used in this paper for the performance evaluation of the proposed method.

To reflect the spatial distribution of traffic volume over a sampling interval within a designated area, we associated the distribution of active network users in terms of number and location with the clutter map of a certain geographical region. The clutter map stores the information on how the earth is covered and hence can indicate the distribution of population density within a certain region. The detailed procedures are:

1) A specific population density (in users/km^2) is configured for each of the clutter types;

2) Based on the statistical figures published in [8] regarding the percentage of active users in the peak hour of a day, for each clutter type, the number of active users per squared kilometer is determined for the peak hour, and used as the baseline to derive the active users/ km^2 for the other hours of a day based on the hourly scaling factor of traffic volume presented in Figure 2;

3) For a given operating region, the area that each clutter type occupies is calculated, based on which the number of active users distributed at the geographical locations belonging to each clutter type is obtained for different hours of the day;

4) For a given operating region and for each hour, the active users associated with each clutter type are allocated to the corresponding physical locations belonging to each clutter type using uniform random distribution;

5) For each active user, a requirement on the transmission data rate is configured to mirror the required data rates by different types of data services, for example:

a) High demand profile: configure 2 Mbps/user for HDTV services;

b) Medium demand profile: configure 0.5 Mbps/user for SDTV services;

c) Low demand profile: configure 100 Kbps/user for HQ internet radio services;

Based on the five steps above, our proposed traffic model can emulate the distributions of active users as well as network traffic in both the spatial and temporal domain taking into account the variation properties and randomness of network traffic in two dimensions. In this paper, to simply the emulation process, each simulation scenario adopts a single traffic demand profile for all active users by setting each active user's requirement on data rate with the same value.

C. Network performance metrics

The energy consumption of a BS k over a time period T is defined as:

$$E_k = \int_0^T P_{in}^{(k)}\left(L_k(t)\right)dt \tag{4}$$

where the consumed input power for BS k $P_{in}^{(k)}$ is a function of the time-variant traffic load $L_k(t)$.

The traffic load of a BS is measured by the percentage of consumption in resource blocks (RB) owned by a BS. For a BS k, the traffic load per unit time (e.g. per second) $l(k)$ is determined by the number of active users served $N_{user}(k)$, each active user's required data rate, each active user's perceived signal to interference plus noise ratio (SINR) and the total number of RB owned by BS k $N_{total}(k)$. The relationships between these parameters are:

$$RBPerUserPerSecond_x = \frac{BitRatePerUser_x}{BitRateOfferedPerRB_x} \tag{5}$$

$$N_{req}(k) = TotalRequiredRB = \sum_{x=1}^{N_{user}(k)} RBPerUserPerSecond_x \tag{6}$$

$$l(k) = N_{req}(k) / N_{total}(k) \tag{7}$$

where $BitRateOfferedPerRB_x$ is determined by the perceived SINR of user x, and LTE specifies the relationship between SINR and supported bit rate per RB. Users' serving BSs are determined based on a user's SINR measurements w.r.t. a number of neighboring BSs and the principle of maximizing each BS's capacity with their available RB.

The time-variant traffic load used in equation (4) is derived from equation (7), consequently $P_{in}^{(k)}\left(L_k(t)\right)$ can be derived based on equation (3).

The total energy consumption of an operating region consisting of N_m macro BSs and N_s small cell BSs is calculated as:

$$E_{total} = \sum_{k=1}^{N_m+N_s} E_k \tag{8}$$

Additionally the consumed energy to run the wireless network in a designated area over time T with and without using our proposed method are denoted by E_{total}^{W} and E_{total}^{WO} respectively, and the energy reduction gain (ERG) obtained from our proposed method is calculated as:

$$E_R = \left(E_{total}^{WO} - E_{total}^{W}\right) / E_{total}^{WO} \tag{9}$$

To compare the QoS delivered to end users before and after applying our proposed method, two evaluation metrics are considered, namely the data rate and service dissatisfaction rate (SDR). Assuming ideally each user is allocated sufficient RB by the corresponding serving BS to assure the provision of required data rate, but in reality when a BS is overloaded provisioning the exact required RB to all its users is not guaranteed, and the users that are not allocated enough RBs are experiencing degradation in QoS. Hence the SDR for a designated region during a time period T is defined as:

$$SDR = \frac{totalNoOfUnsatisfiedUsers}{totalNoOfActiveUsers} \tag{10}$$

III. DYNAMIC OPERATION MODE CONTROL

The dynamic operation of the cell switching algorithm is based on continuously monitoring the load conditions of the network and identifying small cells which can be switched

OFF for a particular time interval and identifying small cells which needs to be turned back on. For the switch OFF process, the smaller cells are ranked as per a novel criterion we introduce. This ranking process is detailed below in section III.A. In section III.B we detail how active users can be transferred from macro cells to small cells remaining to be ON, to achieve further power savings.

A. Small cell ranking process

First of all, a threshold load is derived and used to identify the candidate cells for switch OFF during the next time interval, whose value depends on the power consumption models of small cells. As an example, the threshold load of a micro cell is calculated as follows. Assuming N_1 RB is required by a micro cell to support its load, the power saving by switching OFF the micro cell is:

$$P_s = N_{TRX}^{micro} P_0^{micro} + g_p^{micro} \frac{N_1}{N_{total}^{micro}} - P_{sleep}^{micro} \quad (11)$$

To support the micro cell's load by the underlay macro cell(s), presumably the required RB is $N_1 + \delta$, as these are not the best serving cells. The power consumption increase for the underlay macro cell(s) is:

$$P_c = g_p^{macro} \frac{N_1 + \delta}{N_{total}^{macro}} \quad (12)$$

The g_p terms refer to the respective gradients of the linear power models of the two cells (as given by equation (3)). For power saving by switching OFF the micro cell, we need $P_c \le P_s$, therefore if we assume $N_{total}^{macro} = N_{total}^{micro} = N_{total}$, then

$$N_1 \le \frac{N_{total}\left(N_{TRX}^{micro} P_0^{micro} - P_{sleep}^{micro}\right)}{g_p^{macro} - g_p^{micro}} \quad (13)$$

As an approximation, we assumed δ is zero, and used the upper bound value of equation (13) as the threshold load of micro cell. If a small cell's traffic load is below its corresponding threshold, the small cell is designated as a candidate cell for switching OFF.

The candidate small cells are ranked to gain priority for switch OFF, during the next time interval. This ranking is based on two factors. First, we calculate the power efficiency of the specific small cell, i.e. the amount of power that can be saved by switching it OFF. In some deployments, different makes of small cells can be employed, hence it makes sense to rank these small cells in terms of power efficiency. The power efficiency term (P_e) is calculated as follows;

$$P_e(k) = P_{out}^{(k)} \big/ P_{in}^{(k)} \quad (14)$$

The ranking of the small cell in terms of its P_e is carried out by the following equation, where N_c denotes the number of candidate small cells:

$$P_{rank}(k) = \frac{1}{P_e(k)} \Bigg/ \sum_{i=1}^{N_c} \frac{1}{P_e(i)} \quad (15)$$

The second parameter involves calculating the resource efficiency of switching OFF a particular small cell. This reflects the amount of radio resources the macro cell has to employ in order to support the few remaining active users in the small cell. Usually for most users, the small cell provides better SINR than the underlay macro cell, hence switching OFF the small cell would require more resources from the macro cell to provide the same data rates to the users. Also the resource availability at the macro cell is an essential factor to look at. The ranking for resource efficiency (R_{rank}) is executed as follows, where M_i is the total available RB for the i^{th} macro cell and N_i is the required number of RB from the i^{th} macro cell. The number of macro cells involved, K=1 in many cases and K>1 only if multiple macro cells are providing coverage for the small cell.

$$R_{rank}(k) = \sum_{i=1}^{K} \frac{M_i}{N_i} \quad (16)$$

The normalized resource efficiency ranking, amongst N_c candidate small cells is calculated as follows;

$$R'_{rank}(k) = R_{rank}(k) \Bigg/ \sum_{i=1}^{N_c} R_{rank}(i) \quad (17)$$

The two ranking parameters are combined with appropriate weighting factor α as follows;

$$C_{rank}(k) = \alpha \cdot P_{rank}(k) + (1-\alpha) \cdot R'_{rank}(k) \quad (18)$$

Once the small cells are ranked, the deactivation process should be carried out, starting from the top ranked small cell. The process should continue as resources are available in the macro cells to support the users who will be handed over from the deactivating small cells.

B. Transferring users from macro cells

Once the small cell deactivation process is complete, a further process can be executed to increase the ES in the HetNet. Looking at the slopes of linear power models of Figure 1, it is clear that the macro cells have the steepest slope while the smallest cells have very low slopes. From this feature, we can derive that if some of the macro cell users can be transferred to active small cells, even if they report lower SINR and need more RB from the small cell, there can be a net ES. The condition for a net ES is given by the following equation:

$$N_4 \le \frac{N_3 \cdot g_p^{macro}}{g_p^{small}} \quad (19)$$

where N_3 is the number of RB used by the macro cell, while N_4 is required number of RB by the small cell.

IV. PERFORMANCE EVALUATION

A. Introduction to the secnario setting

A sophisticated software simulator written in C# was built up in-house to evaluate the performance of our proposed ES solution, namely the green wireless (GW) simulator. The system architecture of the GW simulator is illustrated in Figure 3. During a sampling interval T_N, the two-dimensional (2D) traffic map of a designated area is emulated and used in conjunction with the information on the operation states of BSs to calculate the traffic load of each active BS, considering the coverage map and the available RB of each active BS. Subsequently the energy consumption of the region during T_N can be calculated based on the BSs' operation states and their traffic loads. In the meanwhile, the proper operation modes of BSs during the next interval T_{N+1} are analyzed based on our proposed method, the output of which is used to update the operation states of BSs during the next sampling interval.

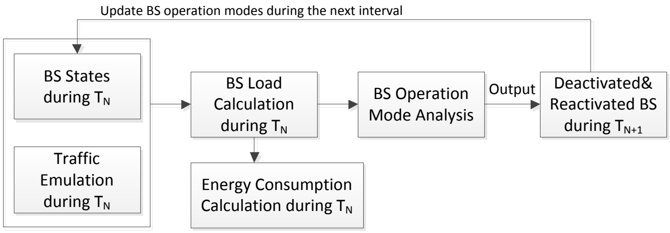

Figure 3 System architecture of the green wireless simulator

We based our simulations on a densely populated metropolitan area of London, i.e. the Oxford and Piccadilly Circus region as denoted by the blue box in Figure 4, which represents a typical application scenario of small cells. To add to the practicality of the simulations, the underlay macro cells and overlay small cells (in our case micro cells are used) are properly planned using Atoll [9] to handle the peak hour traffic demands of the investigated area. Figure 4 shows the clutter map of London.

The specifications of BS parameters and scenario settings are summarized in Table 2.

According to the network planning result, the investigated region is deployed with 12 macro BSs and 138 micro BSs to meet the traffic demand at peak hours assuming each user's requirement on data rate is 1 Mbps. The reasonability of the network planning is assessed by MATLAB simulations that are designed to assign the peak hour traffic in the region to the BSs deployed based on the criteria of assuring each user's required data rate while maximizing the capacity of each BS. Figure 5 shows the validation results, looking at the investigated region, at peak hours, most BSs are fully loaded with few having free capacity, and 99% of users can perceive satisfied QoS, which proves the validity of the network planning.

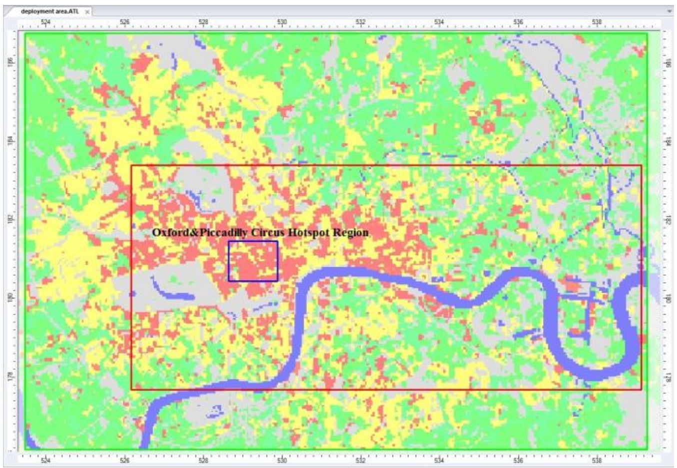

Figure 4 Illustration of the investigated region and clutter map

TABLE II. SYSTEM PARAMETERS

Macro BS Tx Power	40 W
Macro BS operating band and bandwidth	2.11 GHz, 10 MHz
No.of Tx Antennas per Macro BS	2
Micro BS Tx Power	6.3 W
Micro BS operating band and bandwidth	2.13 GHz, 10 MHz
No.of Tx Antennas per Micro BS	2
Maximum data rate per user	1 Mbps

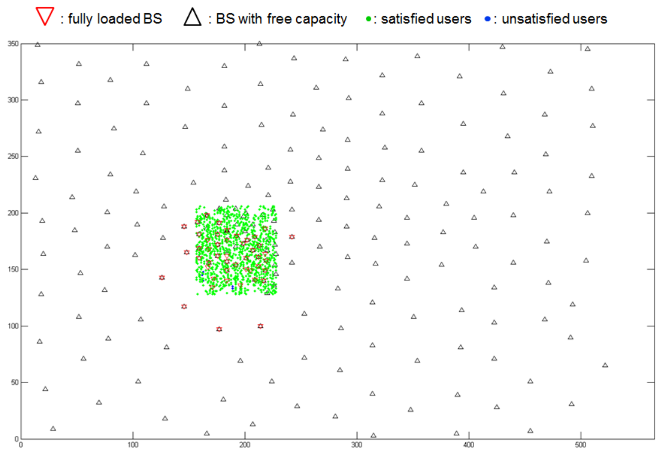

Figure 5 Validation results of network planning

B. *Simulation results*

Our proposed solution is evaluated against the reference operation mode in which all the small cells and macro cells are kept active throughout the day. Comparing the daily energy consumption of different operation modes with variant data rates (Figure 6), we can see that the total energy consumption varies with the overall traffic load of the network, and when the network is operated in the ES mode, the energy consumption is significantly reduced for variant data rate cases. By calculation the ERGs for the two data rate cases 1000 kbps/user and 600 kbps/user are 20% and 27% respectively (Figure 7).

The ERG is obtained without compromising the perceived QoS of users but merely by adapting the operation modes of small cells to the traffic load variation of the network. Figure 8 shows the comparison of SDR of different scenarios. For both data rate cases, the SDR of the ES mode is similar to that of the reference mode.

The number of deactivated small cells at different times of the day is demonstrated for the two data rate cases in Figure 9. The profile of the number of deactivated small cells has an inverse relationship to that of the daily traffic, and a very large number of small cells can be deactivated during the night time based on the daily traffic profile of business district.

From these simulation results, we can observe that there is an obvious trade-off between the QoS delivered to users and the ERG for the network, i.e. as the traffic demand of the network increases, the ERG reduces but the SDR of the network increases.

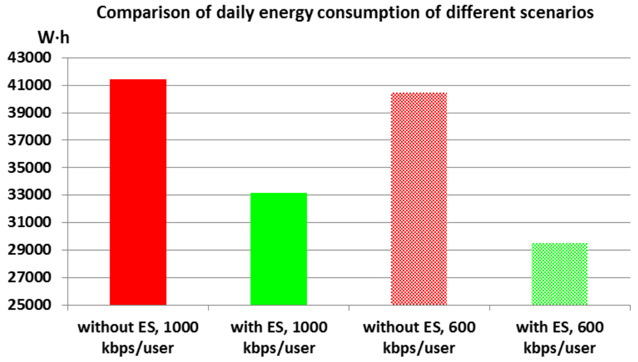

Figure 6 Comparison of daily energy consumption

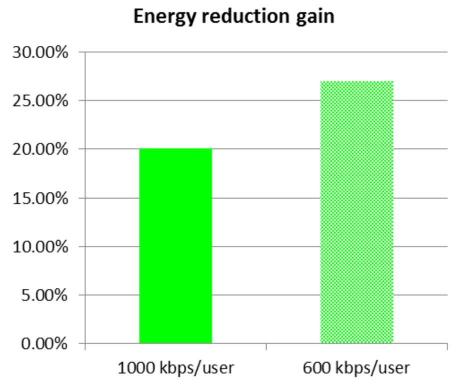

Figure 7 Comparison of energy reduction gain

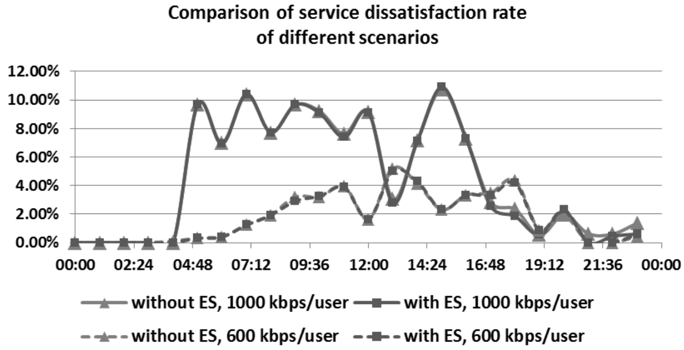

Figure 8 Comparison of service dissatisfaction rate

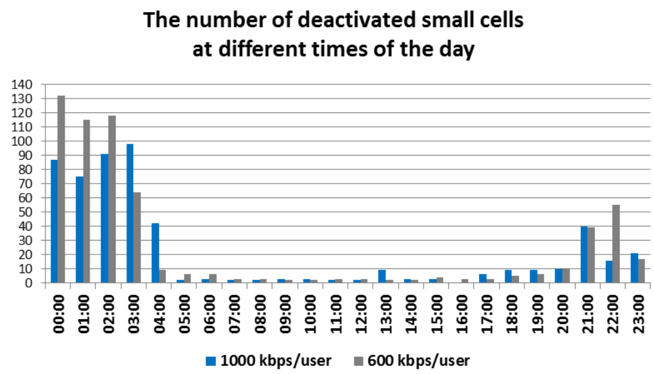

Figure 9 Daily profile of the number of deactivated small cells

V. Conclusions

In order to satisfy the demands for data, mobile operators are expected to roll out increasing numbers of small cells in hotspot areas based on the peak demands in those areas, adding to the costs in energy consumption for running the network.

Exploiting the times when the hotspots aren't operating at full capacity is the underlying principle on which we have developed an innovative method to control the active states of small cells within the network. Our proposed method features:

1) Adaptive operation mode switching based on the 2D traffic distribution trends;

2) Flexibility to handle various power models derived from different kinds of BS equipment specifications;

3) Capability to deal with a range of radio technologies that are deployed to provide capacity enhancements for high traffic hotspots.

Based on a reference network configuration, industry agreed equipment power models and realistic traffic models we were able to demonstrate an energy saving of around 20% in the selected hotspot area. Our proposed method enables operators to effectively optimize the energy consumption of their networks without compromising the perceived QoS of users.

References

[1] Wireless Braodband Alliance whitepaper "Dealing with Density: The Move to Small-Cell Architectures" available at: http://www.wballiance.com/wba/wpcontent/uploads/downloads/2013/07/Dealing-with-Density_RuckusWireless.pdf

[2] J. G. Andrews, S. Singh, Q. Ye, X. Lin and H. S. Dhillon, "An Overview of Load Balancing in HetNets: Old Myths and Open Problems", *IEEE Wireless Communications*, July 2013.

[3] 3GPP Technical Report, "Potential solutions for energy saving for E-UTRAN", TR 36.927, v11.0.0., Sept. 2012.

[4] Green Touch consortium website: http://www.greentouch.org/index.php?page=home

[5] Blume et. al., "Energy Savings in Mobile Networks Based on Adaptation to Traffic Statistics", Bell Labs Technical Journal, 15(2), 2010, pp.77-94.

[6] A. De Domineco et.al., "Enabling Green cellular networks: A survey and outlook", Computer Communications (Elsevier), Vol. 37, Jan, 2014.

[7] Gunther Auer et. al., "How much energy is needed to run a wireless network?", *IEEE Wireless Communications*, Vol.18, Issue 5, 2011.

[8] Energy Aware Radio and neTwork tecHnologies (EARTH) project deliverables D2.3, "Energy efficiency analysis of the reference systems, areas of improvements and target breakdown".

[9] http://www.forsk.com/atoll/