Fluid Capacity for Energy Saving Management in Multi-Layer Ultra-Dense 4G/5G Cellular Networks

Stephen S. Mwanje, Janne Ali-Tolppa
Nokia Bell Labs, Munich, Germany
Email: {stephen.mwanje; janne.ali-tolppa}@nokia-bell-labs.com

***Abstract*—There is a major demand for reducing energy consumption in mobile networks and it is expected become even more vital in the future (5G) multi-layer Ultra Dense Networks (UDNs), in which the number and density of cells in the different layers will grow dramatically. In these networks, multiple geographically overlapping layers are deployed to increase the capacity and throughput, but also increasing the energy consumption. In this paper we present an end-to-end solution that manages energy saving mechanisms in order to scale the provided capacity to the traffic. Assuming a Heterogeneous Network (HetNet) deployment, the solution dynamically selects cells to activate and/or deactivate considering the prevailing network load and the expected spectral efficiency of those cells. Evaluation in a small HetNet scenario showed that the proposed solution is able to reduce the energy consumption by more than 30%.**

I. Introduction

Cellular network traffic varies over the course of a day. In a typical traffic profile shown in Fig. 1, the maximum busy hour load can be many-fold compared to the quiet hours [1]. Networks are typically dimensioned to be able to guarantee the desired Quality of Service (QoS) and avoid congestion during peak traffic. Consequently, this results in over-provisioning and unnecessary energy consumption outside the busy hours, for which Energy Saving (ES) solutions have been designed.

In a Heterogeneous Network (HetNet) additional layers, often small cells, are deployed to enhance the capacity, while a base layer is responsible for the coverage. Obviously, network energy consumption can be reduced, if the capacity enhancing cells are deactivated whenever they are not needed. Significant savings can be achieved, since the radio network makes up to 90% of the total energy consumption of a mobile network [2]. This is also the case even if only the power amplifier is switched off, given its consumption is up to 50% to 65% [3]

Many publications have justified the need for Energy Saving Management (ESM) (e.g. [4]–[8]). However, only a few [9] delivered concrete solutions for distributed HetNet environments and even fewer for the problem of optimal switch-on order.

For ESM in a Multi-layer Ultra Dense network (UDN), three problems must be solved:

1) How to group the cells in order to monitor the load and the free capacity in the network and how to select the cell(s) that can be put to energy saving state?
2) It must be ensured that the coverage layer has sufficient capacity to compensate for the cells that are switched off.
3) How to select the cells to activate if multiple cells are inactive?

The last problem is more difficult than choosing the cells to be switched off, since when the cells are off, their potential load is not known (without using a beacon feature to listen for User Equipments (UEs) within range). A simple solution is to activate all inactive cells. This is very inefficient, since unnecessary capacity is provisioned, as showed in Fig. 1, even with a subsequent action to again deactivate any cells with low load. Alternative approaches include, for example, activating cells based on a fixed time sequence or using historical data and traffic profiling to determine the switch-on order. Although better than the first approach, they are also inefficient in a dynamic HetNet environment.

The best deactivation/reactivation decision should, as we propose in this paper, be based on the aggregated traffic in the area and should aim to maximize the network's spectral efficiency. Our proposed solution, called the Fluid Capacity Engine (FCE), evaluates the traffic in all cells in a Power Saving Group (PSG) and their spectral efficiency and decides, which cells should be activated or deactivated while ensuring that the active cells are able to serve the traffic.

II. Fluid Capacity Engine (FCE)

A PSG is a group of cells, which cover a given area and among which some cells can compensate for the others that can be deactivated, when there is low demand. FCE groups cells in PSGs and categorizes them to two types of cells: 1) reference cells that ensure the coverage and 2) helper cells that

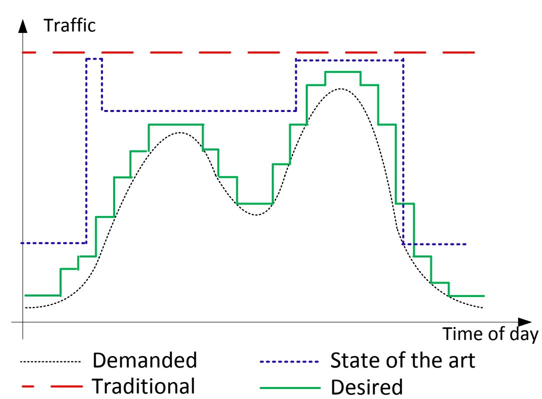


Fig. 1. A typical diurnal load pattern and the actual provided network capacity with different energy saving methods

978-3-901882-85-2 © 2016 IFIP

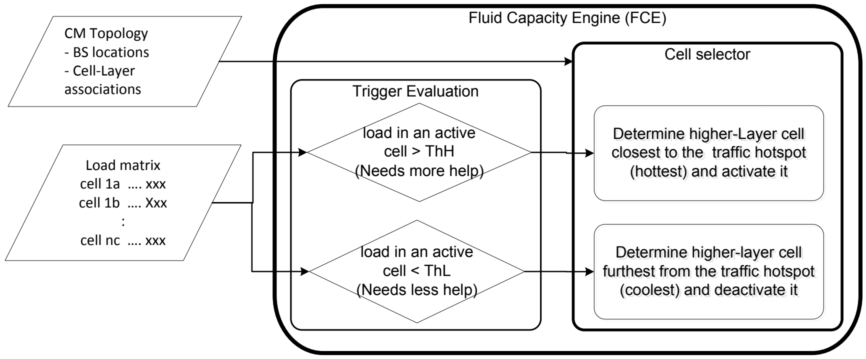


Fig. 2. The ESM Fluid Capacity engine (FCE)

provide extra capacity and throughput, but can be deactivated and reactivated as needed.

Assuming that load information is available for all cells, the FCE triggers evaluation module, shown in Fig. 2, which evaluates the load of the cells in a PSG to determine, if there is need for (de)activating some cells. By looking at the aggregated load, it can trigger cell activation or deactivation even if no single cell has very high or low load.

Following the triggers, the cell selection module chooses the appropriate candidate cells to deactivate or reactivate. Using an analogy of heat flow to model the load in reference and helper cells, it triangulates the heat flow to determine the small cell that is best suited to switch on or off so as to maximize the spectral efficiency of the network. The logical reasoning is that UEs act as primary heat sources that induce an amount of heat at the coverage cell. The coverage cells then act as a secondary heat sources towards the capacity cells. Consequently, the selector does not need to directly consider or know the UE locations, since the effect of their heat can be deduced from the secondary heat sources.

A. Power Saving Groups (PSGs)

The FCE assumes availability of location data for all Base Stations (BSs)

1) Selection of Power Saving Groups (PSGs): First, we define reference cells as those cells that offer full coverage in the considered area (usually macro). A PSG is then defined per reference cell to be the list of all the neighbor cells to it that are not themselves reference cells. The other cells are considered as helpers to the reference cell and are candidates for deactivation (and reactivation). Note from this definition that a helper cell can belong to more than one PSG.

2) PSG load and FCE triggering: The solution considers load in all the cells in a PSG to trigger the FCE. A weighted average of the load that gives more weight to the reference cell can be considered. Thus, the FCE is activated if the weighted average of the load in the cells of the considered network scope increases above a threshold ThH, or reduces below a threshold ThL. If the UEs are always pushed to the small cell layer via Traffic Steering (TS), the activation decision can be based on the reference cell load as an indication of the PSG load. This would be a special case of the default solution in which the reference cell weight is set to one while the helper cells weights are all set to zero.

Note that: 1) The appropriate values of the threshold ThL and ThH can be determined subjectively and operationally. 2) The FCE can alternatively be periodically activated to calculate the induced heat at all active and inactive cells to decide if some cells could be activated or deactivated.

B. Switch-On Switch-Off Order

Candidates for deactivation and reactivation are selected based on their expected spectral efficiency i.e. always ensuring to retain those helpers that result in the highest spectral efficiency for the network/area. The idea is to activate starting with the small cells which are furthest, in radio terms, from the reference cell , i.e., the cells which are closest to the cell edge. A user nearest such a small cell would have the worst spectral efficiency at the reference cell. So if that user is transferred to the small cell more resources are freed at the reference cell. Deactivation then goes in the reverse direction starting with the small cells that are closest to the reference.

The cell selector is based on the triangulation of heat floor, as shown in Figure 3. It depicts a reference macro cell with seven helper small cells in it. Consider the load at each reference cell j as an amount of secondary heat generated in the cell, with the mobile devices in the cell as the primary distributed heat sources. Maximum load (heat) is generated at the edge of the cell, i.e. maximum load is transferred from cell j if a new small cell i is activated at or closer to the edge of cell j.

1) Heat Intensity: Consider a reference cell j with cell range R_j and having a set of helper cells $i \in I$ (i=1, 2, 3) as shown in Fig. 3. For the helper i, given distance d_{ij} to each cell j with unit load, induced heat intensity (from hotspot near j), is

$$h_{ij} = \begin{cases} (r_{ij})^r & ; \quad r_{ij} < R_j \\ (r_{ij})^r \cdot \left[1 - \left(\frac{r_{ij}-R_j}{R_j}\right)^{\frac{1}{r}}\right] & ; \quad r_{ij} > R_j \end{cases} \tag{1}$$
$$with \quad r_{ij} = \frac{d_{ij}}{cos(\alpha^{\tau})}$$

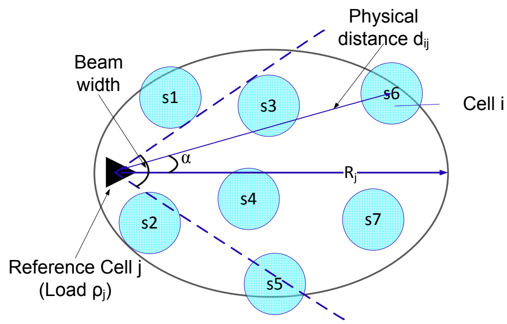


Fig. 3. Cell switch on/off candidate selection with a single reference cell

d_{ij} is the distance between i and j while R_j is the radius or range of cell j. α is the angle between the direction of cell j and the line between cells j and i, which for an Omni-directional cell would be $\alpha = 0$, since the line between the two cells lies along the path of maximum gain. r is the coefficient of heat floor with an assumed default value of 1. We use it as a general parameter but extra studies would be required to justify other values of r. τ is the beamwidth factor that accounts for how much, for a given distance d, the received signal changes as a function of the antenna's beamwidth. Specifically, for the different antenna beamwidths of 60^o, 90^o, 120^o, or 360^o (omni), τ= 0.3; 0.45; 0.5; 1 respectively.

Note that the heat, i.e. the cell load, can be measured in terms of carried data, e.g., Mbps, or in terms of used cell resources. e.g. LTE Physical Resource Blocks.

2) Cell Activation - Singe Reference Cell: Considering only 1 reference cell j with load ρ_j, FCE activates the helper cell with the highest heat intensity, h_i due to the load ρ_j, i.e.

$$Candidate = \arg\max_i \{\rho_j \cdot h_{ij}\} \tag{2}$$

3) Cell Deactivation - Single Reference Cell: The basic idea is to disable those helper cells first, whose induced heat on the reference cell is the lowest, i.e. cells with low load and/or closer to the reference cell center. The helper cell users, who are moved to reference cell through process, will be served well, since they are close to it. Assuming the reference cell j and small cell i, the heat intensity h_{ji} will be as defined in equation 1. The FCE deactivates helpers with the lowest induced heat (see equation 3), but for which the expected total load added to the reference cell is $\Delta\rho_j < (TMhigh - TMlow)$.

$$Candidate = \arg\min_i \{\rho_i \cdot h_{ij}\} \tag{3}$$

4) Multiple Reference Cells: In cell activation, when some helper cells belong to several reference cells, as shown in Figure 4, the FCE determines the helper cell that will assist as many reference cells as possible. i.e., it selects the helper cells that will take the most combined load from the reference cells. The FCE uses the ranking equation 1 and, for each helper cell i, it aggregates the induced heat h_{ij} from each of the reference cells j. It then activates the helper cell with the highest total

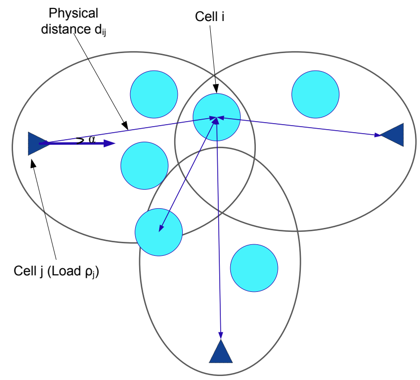


Fig. 4. Cell switch on/off candidate selection with multiple reference cells

induced heat h_i, considering the different reference cell loads, i.e.

$$Candidate = \arg\max_i \left\{ \sum_{\forall j} (\rho_j \cdot h_{ij}) \right\} \tag{4}$$

Similarly, for cell deactivation, the FCE aggregates, for each helper cell i, the induced heat h_{ij} from each of the reference cells j. It then activates the helper cell with the lowest total induced heat h_{ij} as per equation 5, but which together maintain the load in all reference cells below the maximum threshold i.e., $\Delta\rho_j < (TMhigh - TMlow), \forall j$.

$$Candidate = \arg\min_i \left\{ \sum_{\forall j} (\rho_i \cdot h_{ij}) \right\} \tag{5}$$

C. Practical Considerations

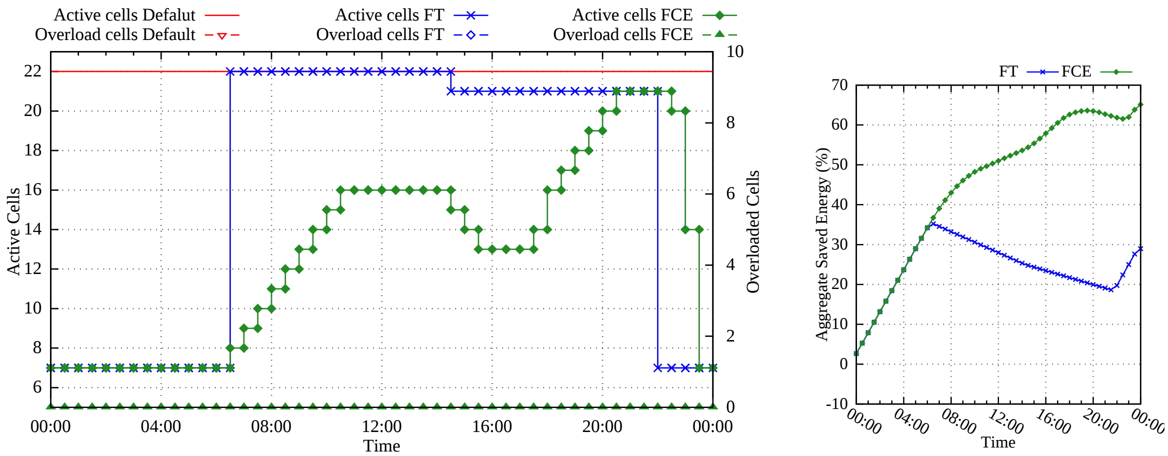
1) Sub-Optimal Static Solution: Without considering the dynamic instantaneous load in the cells (i.e., by taking $\rho_i = \rho_j = 1$), the solution changes into static version that can be applied at network planning time. The obtained ordering of the cells can be used as a fixed sequence for cell activation. Although not optimal, it can still serve as a reasonable first approximation for the cell switch on/off order in a PSG.

2) Hybrid Implementation: The FCE is best suited for a hybrid centralized/distributed-SON implementation, where the heat intensities are calculated centrally, requiring knowledge of BS locations, and provisioned to the BSs as part of the PSG definitions. Monitoring the cell loads and switch on/off decisions based on the induced heat are done in a distributed way, e.g. in the reference cell BSs, where instantaneous load information is available (e.g. for the neighbors via the X2 interface in Long Term Evolution (LTE)).

III. Simulation Scenario and Results

A. Study Scenario

The proposed solution has been evaluated in a simulation environment that models an UDN deployment. It was developed for network-level studies, as described in [10]. Here we highlight only the aspects that are critical for ESM. Assuming a macro cell radius of 450m, we consider an area of 300 x 450m in the simulation, i.e. an area slightly larger than the coverage area of a single macro cell.



(a) Number of active cells

(b) Aggregate Energy saving

Fig. 5. Comparing the number of active cells and amount of saved energy

1) Macro cell coverage: Here, without modeling the low-level radio properties, we assume that each macro cell can cover up to a distance of 400 m. Meanwhile, for complete coverage with 3 sectors (cells) per macro eNB, we apply a beamwidth of 120^o. Effectively, the macro cell can be modeled as an ellipse with a major axis of 400m and a minor axis of 250m.

2) Pico cell coverage: Pico cells are modeled as omni directional with a coverage of up to 50 m, i.e. as circles of 50 m radius. They are deployed at locations that are shifted by small random values off a uniform grid. The locations are randomly selected so as to ensure a small cell coverage of about 75 % of the simulated area.

3) Traffic modeling: The coverage area is divided into a grid of 5 m square pixels, each modeled to have varying traffic throughout the day as profiled in Fig 1. Then, each cell is allocated a maximum data handling capacity that scales with the size (radius) of the cell, but with the small cells having a higher spectral efficiency. In other words, small cells are allocated a higher maximum data rate per unit radius, although the total aggregate maximum rate for the small cells is less than that of a macro cell.

During the simulation, each pixel evaluates all available cells and choses the cell providing the best coverage - much in the same way as in UE cell selection procedures. The induced load to the cell is modulated by the distance of each pixel from it. This accounts for the cell-specific spectral efficiency in determining, how much of the cell's resources are used for the load presented by the given pixel.

B. Results

Figure 5 summarizes the observations. We compare the performance of the FCE against two other solutions - the default and a Fixed-Time energy saving (FT) solution. The default solution keeps all available cells always on, while the FT activates all helper cells at a fixed time every day (6:30 am), but deactivates any helper cells, when their traffic falls below the set threshold of ρ_i <0.1.

We observe that the FCE reacts to the traffic variations in a more dynamic way. Although it activates the first cells at the same time as FT, it only activates a few cells and then gradually more as the load increases. Furthermore, it activates first those small cells, which have the highest estimated spectral efficiency, i.e. highest induced heat considering the heat intensity and the traffic distribution. This enables it to activate a much lower number of cells compared to FT. Interestingly, FCE even deactivates some cells in the early afternoon, when the load reduces and only reactivates more, when the load increases again leading up to the day's peak at about 22:00 hours. Finally, although not as sudden as FT, FCE quickly reduces the number of active cells as the load drops after the peak. Once again the shutdown sequence is optimized such that the spectrally least efficient small cells, the ones offloading the macro the least, are shut down first. The result of all this is that FCE can achieve up to 30% more reduction in small cells' energy consumption over the course of the day, as shown in Fig. 5b.

IV. Conclusion

In this paper we presented a network-scale solution for dynamically controlling the available network capacity in multi-layer ultra-dense Heterogeneous Networks (HetNets), in order to minimize the energy consumption. Our solution, called Fluid Capacity Engine (FCE), models the load presented by each User Equipment (UE) as heat that is induced to the coverage layer cells. The coverage cells propagate this heat to the small cells in an overlaid capacity layer, which we call helper cells. With this analogy, we evaluate the induced heat (or load) intensity at the helper cells to rank them and determine, if and which particular helper cells should be activated or deactivated. As a result, the network can effectively adapt to changes in the traffic and we can always choose optimal helper cells for activation or deactivation, as shown in our simulations. We estimate that the FCE is able to lower energy consumption of a HetNet by more than 30%.

REFERENCES

[1] 3GPP, "Telecommunication management: Study on energy saving management (esm) (release 10)." 3GPP, Tech. Rep. TS32.826 V10.0.0, (2010).

[2] D. A. R. Dilupa, R.Withanage, *Handbook of Research on Green ICT: Technology, Business and Social Perspectives*. IGI Global, 2011, ch. Infrastructure Sharing & Renewable Energy Use In Telecommunication Industry for Sustainable Development.

[3] G. Micallef, L. Saker, S. E. Elayoubi, and H.-O. Scheck, "Realistic energy saving potential of sleep mode for existing and future mobile networks," *JOURNAL OF COMMUNICATIONS*, vol. VOL. 7,, no. NO. 10,, OCTOBER 2012.

[4] 3GPP, "Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Self-configuring and Self-Optimizing Network (SON) use cases and solutions," 3GPP, Technical Report 36.902 version 9.3.1 Release 9, May 2011. [Online]. Available: http://www.3gpp.org

[5] J. Fang, "Method, apparatus and system for realizing coverage compensation," China Patent WO2 014 110 910 A1 / CN103 945 506A, 2014.

[6] T. K. Elena Voltolina, "Energy-saving mechanisms in a heterogeneous radio communication network," Sweden Patent WO2 011 021 975 A1 / US12 790 055, 2011.

[7] H. P. Karl Heinz Nenner, Dieter Jacobsohn, "Method for saving energy in operating a first and second mobile communication networks, a mobile communication networks," Germany/USA Patent EP2 676 491 A1 / US20 130 344 863, 2013.

[8] E. Roth-Mandutz and A. Mitschele-Thiel, "Lte energy saving son using fingerprinting for identification of cells to be activated." in *Future Network & Mobile Summit*, 2013.

[9] Eunsung Oh, Kyuho Son and Bhaskar Krishnamachari, "Dynamic Base Station Switching-On/Off Strategies for Green Cellular Networks," in *IEEE Transactions on Wireless Communications, vol. 12*, May 2013.

[10] S. S. Mwanje, J. Ali-Tolppa, and H. Sanneck, "On the limits of pci auto configuration and reuse in 4g/5g ultra dense networks," in *The 11th International Conference on Network and Service Management*, 2015.