Changing paradigms for green cellular networks: the case of delay-tolerant users

Samantha Gamboa, Alexander Pelov and Nicolas Montavont
Institut Mines-Telecom / Telecom Bretagne / IRISA / Université Européenne de Bretagne
Department RSM, 2 rue de la Châtaigneraie, 35576 Cesson Sevigné, France
firstname.lastname@telecom-bretagne.eu

***Abstract*—In this paper, we propose to reduce the energy consumption of cellular networks by leveraging coordinated cell switching algorithms with user cooperation. Cell switching consists in deactivating some base stations in periods of low load in order to reduce the power consumption, while the coverage is provided by remaining active base stations. We propose to maximize these periods of low power consumption by delaying the start of some user services. We present two different strategies to control the cell switching considering the users delay tolerance and we evaluate their performance in the well known Network Simulator 3 (NS-3). We show how appropriate thresholds selection leads to bound the user waiting time, and we highlight how the system load estimation is important for this purpose. Finally, we show that power reductions from 18% to 72% are possible for different offered load levels in the network when using the proposed algorithms.**

I. Introduction

The number of cellular base stations (BSs) is expected to augment with the evolution of the technologies such as Long Term Evolution (LTE) and high capacity dense deployments, increasing considerably the power consumed by the networks, traditionally designed to operate using the Always On paradigm. The new hardware generation is designed to be more energy efficient, adapting the radio resource configuration depending on the traffic level [1]. One of the techniques to exploit this feature is the coordinated cell switching [2].

Some BSs can deactivate part of their hardware to turn into a low energy consumption mode (e.g. sleep mode) at the cost of reducing the system capacity. In this paper, we study a complementary approach to maximize the low consumption periods of the BSs, by considering the user awareness and cooperation. We consider that a set of BSs is enabled to use a coordinated cell switching algorithm. The BSs exchange load information and some of them turn to sleep mode when possible. Users located in this part of the network, when the BSs are in sleep mode, could be asked to collaborate with the network, delaying the start of their services for a bounded delay. Thus, the BS can remain for longer periods in sleep mode. Such cooperative users are called Delay Tolerant Users (DTU).

This work has received a French government support granted to the CominLabs excellence laboratory and managed by the National Research Agency in the "Investing for the Future" program under reference ANR-10-LABX-07-01

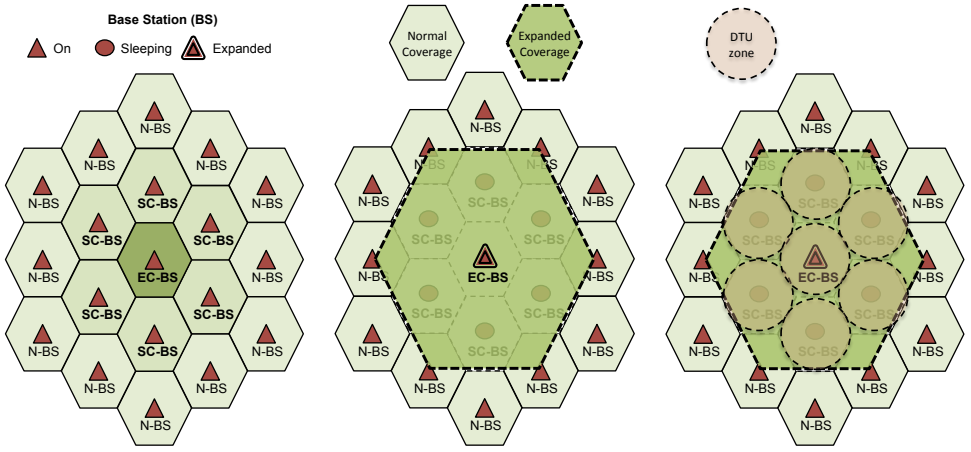
In previous work, we evaluate the theoretic impact of such concept in the standalone cell sleeping [3], dynamic sectorization and capacity adaptation techniques [4]. In this paper we extend this concept to the coordinated cell switching algorithm and validate it using system level simulations in the LTE module of the well known Network Simulator 3 (NS-3). Thus, we consider a more realistic system model and we account for the impact of the BS switching in the user satisfaction.

The rest of the paper is organized as follows. We overview the related work in Section II. We describe the DTU-aware sleep mode mechanisms in Section III and we present their evaluation using system level simulations in Section IV. Finally, Section V concludes the paper.

II. Related Work

1) Dynamic access network management: The purpose of these approaches is to adapt the amount of active infrastructure resources to the traffic demand fluctuations. One important classification criteria is the time window in which they are applied. This time window involves three processes: the load measurement, the reconfiguration and the steady state (time the system remains unchanged). In the *Offline* techniques, the decisions are taken based in historical load information and the resource provisioning is made based in predefined schedules, e.g. the cell switching daily planning [5]. The *Online* algorithms estimate the load based on measurement of the system state. The measurement periods normally depends on the reconfiguration and steady state times. An algorithm is considered having *Slow reaction* when the reconfiguration process is time consuming, e.g., in the order of minutes. Thus, the measurement periods have to be large enough and statistically significant to ensure the reconfiguration is needed/worth (i.e. large enough steady state periods). The early works in cell switching considered these aspects [6]. *Fast reaction* algorithms can adapt the infrastructure in the time order of seconds. This allows to very short measurement and steady state periods, tracking the load variations in an almost real time fashion. Most of the cell-switching algorithms are assumed to have *fast reaction* [2] [7]. When this is not directly applicable to all current deployed networks, this assumption is justified by the constant hardware innovation. For example, adaptive power amplifiers [1], antenna muting [8], beamforming

978-3-9018-8273-9/15/ ©2015 IFIP



(a) *all-On* state (b) *min-On* state (c) *min-On* state, localized DTU zone

Fig. 1: System configurations and identification of the DTU zones

[9], together with the standardization efforts of the SON (Self Organizing Network) strategies [10], make fast adaptive infrastructure attractive to operators. The application of the different types of algorithms depends on the available technology, the component dynamism, the control scheme and the metric to optimize. For a detailed overview of the existing literature in this field, the reader can refer to some of the recent surveys, e.g. [11]. In this paper we consider a coordinated cell switching online algorithm with fast reaction. We rely on the delay tolerance of the users to extend the periods in which the cells can remain switched off, further decreasing the power consumption of the network.

2) User awareness and cooperation: New interactive infrastructures will allow the user to actively interact with the network operator to achieve mutual benefits. The possible actions the operator could expect the user perform are deployment, location and time dependent. For example: move to another location, wait a certain time to use the service, use a downgraded version of it or even not use the service at all. In [12] a survey is performed to evaluate the willingness of the users to some of these possibilities if some incentives or deterrents are applied. Several expressions for the user cooperation are derived from the empirical data. The impact of the movement of the user to a location with better SINR (signal-to-interference-plus-noise ratio) to improve the spectral efficiency of the network is studied as well.

Several approaches rely on the user cooperation to avoid congestion, using dynamic pricing depending on the network status. Thus, the arrival rate of new sessions are controlled depending of the congestion level, and the demand is fitted to the capacity of the system. This technique is evaluated analiticaly in the context of cellular networks by Schoenen et al. [13]. In this study, the user is persuaded (using pricing surcharge) to not use the service in congestion periods. However, the impact of the session deferral is not studied. Ha et al. [14] focused in elastic data services and present the architecture, implementation and field tests of the strategy involving all entities of the cellular network. The user equipment front-end allows the user to choose to use the service right away for a given price (dynamically calculated by the operator, higher in congestion periods) or schedule it for later use with a discount in the price.

When these approaches profit of the user cooperation to avoid congestion, the approach proposed in this paper relies on it to dynamically adapt the network resources to save energy. The user is asked to wait in order to extend the low power consumption periods. The waiting time is bounded and the service is ensured within this time. Thus, the session deferral impact is considered in the design of the strategy. This concept was analytically evaluated in prior work for the standalone cell sleeping [3] and for the dynamic sectorization and capacity adaptation techniques [4]. In this paper we extend this concept to the coordinated cell switching algorithm and validate it using system level simulations.

III. DTU-aware Coordinated Sleep Mode

A. System model

We consider a cellular access network enabled to use fast reaction cell switching, also known as sleep mode, in some of the base stations (BSs). In this paper we focus in a hexagonal deployment of micro BSs to show the potential gains of the proposed strategy. Although regular hexagonal BSs scarcely exist in real networks, the practical deployment of operator-side macro and micro BSs is always well planned and never random. Moreover, the usage of traffic statistics facilitate the operator to identify potential groups of BSs, in which cooperative cell switching can be applied satisfying coverage constraints.

Particularly, we consider a system composed of 19 LTE omnidirectional micro BSs depicted in Fig 1. Three types of BSs are defined. A BS enabled to enter in sleep mode is denoted as *Sleep-capable base station* (SC-BS). The BS in charge of compensate the coverage of the sleeping BSs is called *Expand-capable base station* (EC-BS). Finally, a BS which do not participate in the sleep mode algorithm is denoted as *Normal base station* (N-BS). The execution of the cell switching algorithm is controlled by the EC-BS. The SC-BSs report periodically their load levels to the controller using the Resource Status Update procedure via the X2 interface [15].

Two system states are differentiated as well. The system is in *all-On* state when all the BSs are operational (Fig 1(a)). In this state the system has an available maximum capacity denoted by $C_{\max}$. The system is in the *min-On* state when the SC-BSs are sleeping. In this state the system has a capacity $C_{\min} < C_{\max}$, provided by the coverage compensation made by the EC-BSs (Fig. 1(b)). We assume that the radio planing is done adequately in order to avoid coverage holes in the access network when in *min-On* state.

B. Cell switching transitions

The transition between system states is performed progressively. When the cell switching controller decides to switch to *min-On* state, each SC-BS decreases its transmission power by X dBm every t_x seconds. Simultaneously, the EC-BS increases its transmission power by Y dBm every t_y seconds until reaching the maximal transmission power. Once each SC-BS has no more users associated, it turns to sleep mode. In each transmission power change, a batch of users may need to handover to the EC-BS. Too long power steps (i.e. X and Y) or too short time steps (i.e. t_x and t_y) could produce handover failures. The handover protocol could be compromised, in the first case due to bad signal condition of the source BS (i.e. too late handover) and in the second case due to signalling overhead of a big number of user performing handover. On the other hand, too long time steps or too short power steps will produce unnecessary long reconfiguration periods. Thus, a trade off between power steps and time step is needed to minimize the reconfiguration periods and to let the users successfully perform handover if needed. When the objective is to switch to *all-On* state, the opposite procedure is performed. Each SC-BS increases the transmission power by Z dBm every t_z seconds while the EC-BS shrinks concurrently, decreasing the transmission power by W dBm each t_w seconds.

C. Delay Tolerant Zone

Users accepting to offset the start of their services are called *Delay tolerant users* (DTUs). The area covered by the EC-BS and SC-BSs when the system is *all-On* state is defined as the *DTU zone*. When the system is in *min-On* state, users located in this area could be asked to collaborate with the network, delaying the start of their services. For simplicity, we approximate the coverage area of the BSs using a circular model, with center in the BS position and with a radius $r = \frac{\text{ISD}}{2}$, were ISD is the Inter Site Distance in the hexagonal deployment (Fig 1(c)). The system is designed to support a maximal delay (D) for users in the DTU zone. In case that some delay of the users service is required, the network informs them that the waiting time should not be longer than D, and the cell switching algorithm is configured to satisfy this constraint.

D. DTU-aware cell switching

The cell switching strategies adapt the resources depending on the load level (L). When the system is in *all-On* state and the load decreases, closer to the switching off threshold L_1, the reconfiguration to *min-On* state is triggered. L_1 should be chosen appropriately so that the current system load can be absorbed by the EC-BS along with the new (estimated) arrivals. When the system is in *min-On* state and the load increases, surpassing the switching on threshold L_2, the reconfiguration to *all-On* state is triggered. The choice of L_2 is usually done assuming that a new arrival will be blocked if the resources in *min-On* state are exhausted. Thus, $L_2 < C_{\min}$ in order to trigger the reconfiguration before a blocking situation could arrive [16] [2].

Algorithm 1: Strategy One algorithm

```
Data: L_1, L_DTU, State, Event, UE
if State is all-On and Event is Load Notification then
    Calculate number of active users in the DTU zone (L);
    if L < L_1 then
        Perform reconfiguration to min-On state ;
if State is min-On and Event is Arrival then
    if UE is in the DTU zone and UE cooperates then
        n_DTU += 1 ;
        if n_DTU < L_DTU then
            Push the UE in the waiting queue;
        else
            Push the UE in the waiting queue;
            Perform reconfiguration to all-On state ;
            Serve all UEs in the waiting queue;
            n_DTU = 0
    else
        Start UE service ;
```

When a DTU-aware strategy is employed, we consider that a part of the users are willing to cooperate with the network, delaying the start of their services. When the system is in *min-On* state, some users will be served instantaneously and some others will be put on hold, so none of them will be blocked, as long the system capacity ($C_{\max}$) is not reached. Thus, when using DTU-aware strategy, we introduce a different wake up reconfiguration threshold L_{DTU}. When the number of active users in the system (including waiting users) surpass this threshold, the reconfiguration is triggered to *all-On* state. It is critical to notice that L_{DTU} should be appropriately selected so that the waiting time of the users is bounded and inferior to D.

Two DTU-aware cell switching strategies are considered. The first strategy delays DTU services if they arrive to the DTU zone when the system is in *min-On* state. The second strategy only delays DTU requests when there are not enough resources to serve them (i.e. $L > C_{\min}$), serving them otherwise. Each strategy adapts the threshold configuration depending on the system conditions and D. Afterwards, the dynamic cell switching algorithm is put into action, tracking the load variations (user session arrival or departure) and reacting accordingly. The controller EC-BS identifies three kind of events (*Load notification*, *Arrival*, *Departure*) and may react differently depending on the used

strategy.

In both strategies, when the system is in *all-On* state, the only triggering event is the *Load Notification*. The controller EC-BS collects the load information from the SC-BSs, calculates the aggregated load (L) and performs the switching to *min-On* state when it is appropriated ($L < L_1$). When the system is in *min-On* state the strategies react differently. The EC-BS controller has complete and real time information about the load in the DTU zone as it is under its coverage. Thus, the EC-BS reacts to each *Arrival* or *Departure* event. The Strategy One follows Algorithm 1. In this algorithm, all DTU arrivals located in the DTU zone will always be delayed if the system is in *min-On* state. Once the number of waiting DTUs (n_{DTU}) reaches the strategy threshold (L_{DTU}), the system switches to *all-On* state and serve all waiting users.

Strategy Two follows Algorithm 2. In this strategy, DTU arrivals are delayed only if the DTU zone is congested. This happens when the number of active users in the DTU zone (N) surpass the capacity of the EC-BS (C_{min}). Departure events are proper to serve waiting users in a FIFO manner. As in Strategy One, if the number of waiting users surpass the strategy threshold the system reconfiguration is triggered. The system state changes when the reconfiguration is finished. Thus, arrivals during reconfiguration periods are treated by the algorithm as if the system were in the previous state.

Algorithm 2: Strategy Two algorithm

```
Data: L_1, L_DTU, State, Event, UE
if State is all-On and Event is Load Notification then
    Calculate number of active users the DTU zone (L);
    if L < L_1 then
        Perform reconfiguration to min-On state ;
if State is min-On then
    Calculate number of active users in the DTU zone (N);
    if Event is Arrival then
        if UE is in the DTU zone then
            N+ = 1;
            if N > C_min and UE cooperates then
                n_DTU+ = 1 ;
                if N ≤ L_DTU then
                    Push the UE in the waiting queue;
                else
                    Push the UE in the waiting queue;
                    Perform reconfiguration to all-On state ;
                    Serve all UEs in the waiting queue;
                    n_DTU = 0
            else
                Start UE service ;
        else
            Start UE service ;
    if Event is Departure then
        N− = 1;
        if UE is in the DTU zone and n_DTU > 0 then
            Start service UE in the front of the queue ;
            n_DTU− = 1 ;
```

E. Threshold Selection

The strategy thresholds are selected depending on the estimated offered load. This estimation can be done in a static or dynamic fashion. In the static case, statistical models derived from previously collected data are used. In the dynamic case, the system collects information about the served users for a given period, then it calculates the needed metrics for the configuration. Finally, the DTU-aware strategy thresholds are selected depending on the estimated offered load (L) and the maximal tolerated delay (D) proposed in the DTU zone.

In our previous work [3] [4] we developed a general model for the system user dynamics when using a DTU-aware strategy. The model corresponds to a set of ergodic and homogeneous discrete-state Markov Chains (*MCs*). Each strategy and set of parameters (L,L_1,L_{DTU},C_{min},C_{max}) have a distinctive MC. Their analysis allows to calculate the system state probabilities, the users waiting time distribution (W) as well as the blocking and dropping probabilities (p_{block} and p_{drop}) of the system using the DTU-aware strategy.

The parameters used to model the dynamic are the following: the interarrival time follows an exponential distribution with parameter λ. The service time is exponentially distributed with parameter μ. The system capacity C_{max} is fixed and the method of service is FIFO. The offered load is given by $L = \frac{\lambda}{\mu}$. We assume that during all the service time a user consumes a fixed number of resources (e.g. a given target downlink throughput) which is the same for all users. Thus, we refer to the load and the system capacity in terms of number of users. Even when the actual interarrival and service time distribution could differ from the model, the mean values give a reference for the threshold estimation. For a given C_{min}, C_{max} (fixed by the system characteristics), L (estimated) and D (DTU zone specific), we use exhaustive search to select the strategy thresholds L_1 and L_{DTU} producing the MC that solves the following optimization problem:

$$\underset{L_1, L_{\text{DTU}}}{\text{maximize}} \quad p_{\text{min-On}} \tag{1a}$$

$$\text{subject to} \quad P(W > D) \leq \gamma \tag{1b}$$

$$\beta p_{drop} + (1-\beta) p_{block} \leq \delta \tag{1c}$$

where $p_{\text{min-On}}$ is the probability the system is in *min-On* state, calculated as the aggregated steady-state probabilities of the MC states representing the system in *min-On* state. The parameter γ is the system tolerance to exceed the maximal waiting time (D). The parameter β is the weight of the dissatisfaction metrics (p_{block} and p_{drop}) and δ is the system dissatisfaction tolerance.

IV. Performance Evaluation

A. Simulator setup

The performance evaluation of the DTU-aware cell switching algorithms has been conducted using the LTE module of the well-know Network Simulator 3 (NS-3) [17]. The general architecture of the model implemented in the NS-3 LTE module is depicted in Fig. 2 and is divided in two major parts. The LTE model comprises the LTE Radio Protocol stack implemented in the BS nodes and in the User Equipment (UE) nodes. The EPC (Evolved Packet Core) model includes the core network stack, which is implemented in the SGW (Serving Gateway), the PGW (Packet data network Gateway) and MME (Mobility Management Entity) nodes, and partially in the BS nodes. The BSs are interconnected with each other by means of the X2 interface and to the EPC by means of the S1 interface. The principal parameters used for the evaluations are summarized in Table I. Further details about the evaluation implementation are given in this section.

1) Scenario: We simulated the 19 micro BSs in the system model described in Section III-A and depicted in Fig. 1. We consider static UEs, sparsely and uniformly distributed in a grid fashion with 50 meters of separation between them. We focus the evaluation in the dynamic part of the access network, i.e. the SC-BSs and the EC-BS. Thus, we only consider the UEs positioned in the area covered by them.

2) Traffic generation: At the beginning of the simulation the inter arrival time and the duration of the desired service is generated for all the simulated UEs, according to their respective distributions. In this paper we used the exponential distribution for both cases. When a UE finishes its service, a new arrival time and service duration is generated. This allows to generate traffic with a given intensity no matter the length of the simulated time. It is important to notice that the arrival time is the time in which the UE attempts to use the service. The actual starting time of the service may differ from this value, depending on the DTU-aware algorithm.

TABLE I: Evaluation Parameters

Scenario (NS-3 LTE module [17])	
Deployment type	Hexagonal Micro
Inter Site Distance $[m]$	250
Number of BS sites	19
Antenna model	Isotropic
Path Loss model	Friis
Bandwidth [MHz]	5
Transmission Power [dBm]	38 (Expanded) 32 (Normal)
Scheduler	Round Robin
Handover Algorithm	A3 Rsrp
User distribution	Uniform grid
User density	330 users/Km^2
Mobility	Constant position
Algorithm parameters	
DTU zone session capacity	$C_{\max} = 30$ $C_{\min} = 10$
γ	0.05
δ	0.05
β	0.9
Base station power consumption [18]	
N_{TRX}	2
$P_{\max}$ [W]	6.3 (Expanded) 1.6 (Normal)
P0 [W]	56
Δ_P	2.6
P_{sleep} [W]	0

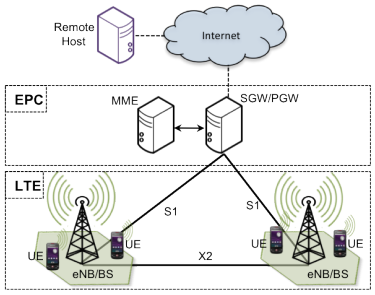


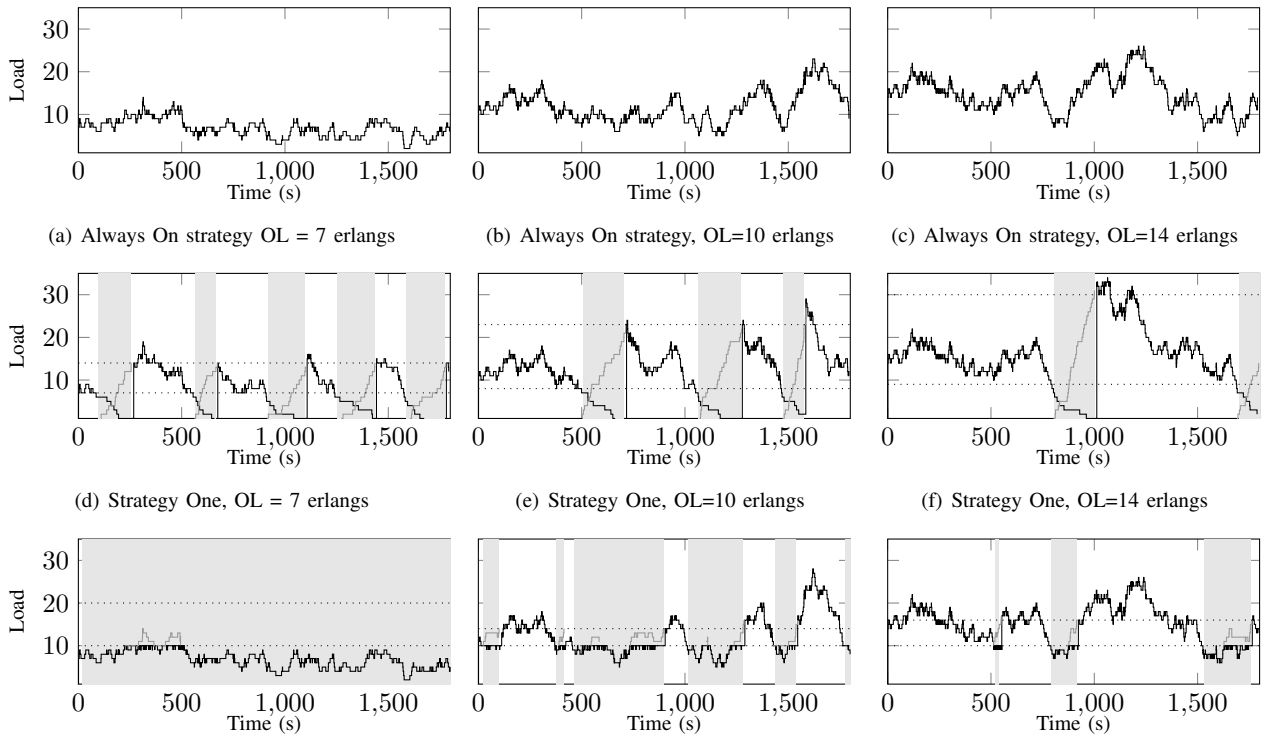
Fig. 2: NS-3 LTE module architecture

In this paper the service required by the UEs is a voice call. No implementation of voice services over LTE is given in NS-3. Thus, we implemented a pseudo voice service as follows. For each call we configure two UDP applications. One of them sends packets from the UE to a remote host (RH) in the internet and the other from the RH to the UE. This emulates a two-way communication. Each UDP client sends packets of a specified size periodically during the service time. We consider the end nodes are using the Codec G.711 [19] to encode the voice packets, which generates a payload of 160 Bytes each 20 ms. Considering the protocol overhead associated to the voice service, the UDP packet size of the pseudo voice call is 179 Bytes and it is sent every 20 ms during the service time. A Packet Sink application is configured in each of the end nodes to receive and consume the packets, avoiding overflow.

3) Traffic monitoring: The DTU-aware cell switching decisions are taken based on a load measurement function implemented over the top of the module. In practice, the cell switching controller, i.e. the EC-BS, could receive this information via the X2 interface, using the Resource Status Update procedure [15]. However, this procedure is not fully implemented in the used version of NS-3 [1]. We used a simplified traffic monitoring function in which the EC-BS is aware of each arrival or departure in the BS cluster, creating each time a *Load Notification* event. Thus, when the system is in *all-On* state, the EC-BS has perfect information about the aggregated load of the participating BSs to decide if a switch to *min-On* state is suitable. Perfect information about the position of the UEs is assumed as well. Thus, the EC-BS can detect if the UE is in the DTU-zone testing a simple geometric condition for the corresponding BS positions: $(\text{UE}_x - \text{BS}_x)^2 + (\text{UE}_y - \text{BS}_y)^2 < (\frac{\text{ISD}}{2})^2$

4) Coordinated cell switching: The coordinated cell switching algorithm described in Section III-B was implemented with tunable time and power steps. The mobility of UEs between cells due to the cell switching is handled using the event based handover algorithms defined in the LTE standard and implemented in the NS-3 module. After an evaluation process, we selected the values for the time and power steps in order to

[1]We use the ns-allinone-3.21 version, which at the time of writing is the latest release



(a) Always On strategy OL = 7 erlangs (b) Always On strategy, OL=10 erlangs (c) Always On strategy, OL=14 erlangs

(d) Strategy One, OL = 7 erlangs (e) Strategy One, OL=10 erlangs (f) Strategy One, OL=14 erlangs

(g) Strategy Two, OL = 7 erlangs (h) Strategy Two, OL=10 erlangs (i) Strategy Two, OL=14 erlangs

Fig. 3: DTU zone dynamic examples. Three different offered load scenarios, each one simulated applying the different strategies: Always On and no delay, Strategy One D=150s and Strategy two D=60s. Gray lines: users waiting. Black lines: users being served. White periods: the system is in *all-On* state. Light gray periods: the system is in *min-On* state (the SC-BSs are sleeping)

avoid handover failures. When the system switches to *min-On* state: the SC-BS decreases the transmission power (P_{max}) 1 dBm every 0.5 seconds, while the EC-BS increases P_{max} in 0.5 dBm every second. When the system switches to *all-On* state: the SC-BS increases P_{max} in 1 dBm every 0.5 seconds, while the EC-BS decreases P_{max} in 3 dBm every second.

5) Base station power consumption: The baseline BS power consumption model used in this paper was introduced by Auer et al. [18]. This model relates P_{out} (the output power radiated at the antenna) and P_{in} (the total power needed by the BS to operate) for each type of LTE BS. The energy model is well approximated by the linear model given by:

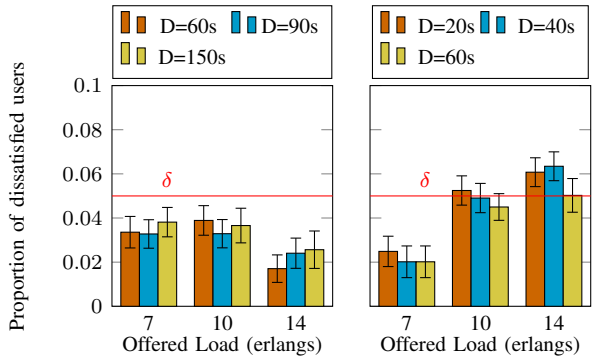
$$P_{in} = \begin{cases} N_{TRX}(P_0 + \Delta_P P_{out}) & 0 < P_{out} < P_{max} \\ N_{TRX} P_{sleep} & P_{out} = 0 \end{cases} \quad (2)$$

where N_{TRX} is the number of transceiver chains (depending on the number of active sectors), P_0 represents the power consumption of an empty BS, Δ_P is the slope of the load dependent power consumption, P_{max} represents the maximum transmission power achievable by the BS and P_{sleep} represents the power consumption of the BS in sleep mode. In LTE systems the downlink transmission scheme uses orthogonal frequency-division multiplexing (OFDM). Thus, P_{out} depends on the BS physical resource allocation in the downlink.

B. Evaluation

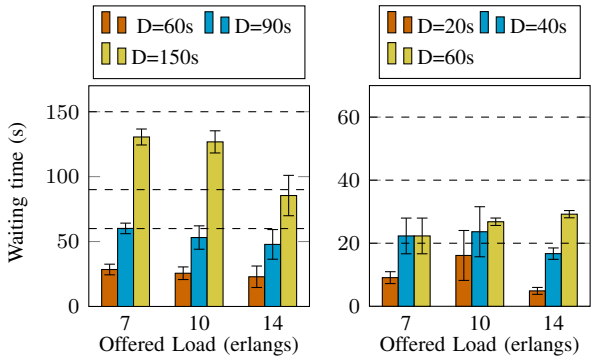
We focus the evaluation on the execution of the DTU-aware cell switching algorithm. Thus, we assume that for each simulation, the algorithm thresholds were selected depending on accurate load estimations. We assume complete user cooperation. We consider that a user call is satisfied if the average packet delay is inferior to 50 ms. Because we use the round robin scheduler, the more users in the system, the larger the packet delay. Thus, the system capacities (C_{max} and C_{min}) are calculated as the maximum number of simultaneous calls the system can have, while satisfying 95% of them. The simulated time was 30 minutes for each scenario. Each simulation was repeated 50 times using different seeds and the mean along with the 95 percent confidence interval is plotted for every parameter.

1) User dynamics: In Fig. 3 we present some examples of the load and system dynamics during the simulated time. For Strategy One, the frequency of entering in sleep mode is reduced when the load increases. However, the length of the sleeping periods is relatively uniform for the same maximal tolerated delay in the DTU zone. The simple dynamic of the waiting queue without service until turning on the SC-BSs, makes Strategy One more predictable and suitable for less dynamic hardware. The dynamic of the strategy Two is more complex, as waiting users can be served by the system in *min-On* state, causing diversity in the lengths of the sleeping periods. Strategy Two is more efficient in relatively low loads. For example, when the system is experiencing an offered load of 7 erlangs, the users can be served without the need of turning On the SC-BSs, and the amount of users experiencing delays is reduced (Fig. 6).



(a) Strategy One (b) Strategy Two

Fig. 4: Proportion of dissatisfied users in the system



(a) Strategy One (b) Strategy Two

Fig. 5: 95th percentile of the call waiting time

2) Quality of Service: The DTU-aware strategies aim to reduce the power consumption of the access network while controlling the user dissatisfaction. Two situations can cause user dissatisfaction: the system is in congestion and there are not enough resources to serve the user (call blocking), or the communication is interrupted due to the cell switching transition (call dropping). For both strategies the user dissatisfaction is controlled in the simulated scenarios, as presented in Fig. 4. The dynamic of Strategy Two, makes it more likely to experience call dropping, as the cells switch more frequently, following closely the load variations. However, the proportion of dissatisfied users in congestion periods and cell switching transitions, is in most of the cases below the target of 5% set for the algorithm dimensioning ($\delta = 0.05$).

3) Waiting times: The system tolerance to delay surpassing was set to 5% for the dimensioning of the algorithms ($\gamma = 0.05$). The 95th percentile of the call waiting time in the system during the simulated time is depicted in Fig. 5, showing that in all the considered scenarios the constraint was respected: 95% of the users experienced a waiting time inferior to the one proposed to them in the DTU zone. The remaining 5% of the users experience slightly higher delays as can be seen in Fig. 6. In this last figure we can also see that higher the delay proposed, higher the proportion of users that actually experiment delay. However, in most of the evaluated scenarios, around 80% of the users did not experienced any delay in the start of their calls. Thanks to the dynamic of Strategy One, higher delays can be proposed without causing congestion. When the offered load increases (e.g. 14 erlangs), the strategy is less active, resulting in lower delays for the users (Fig. 5).

4) Power consumption: Considerable power consumption reductions are observed when using the DTU-aware cell switching strategies, as presented in Fig. 7. The results show up to 78% of power reduction when the system is experiencing an offered load of 7 erlangs and using Strategy Two. The dynamic of Strategy One limits the possible power reductions as the SC-BSs have to always turn on to serve the users. However, an important reduction of 35% is observed for 7 erlangs of offered load when using this strategy. For 10 and 14 erlangs, reductions up to 45% and 17% are achieved respectively.

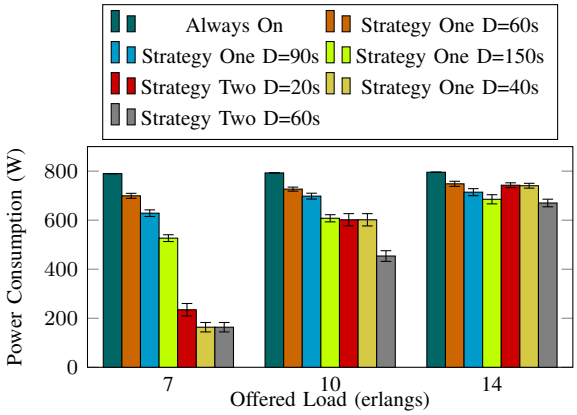


Fig. 7: Average power consumption of the dynamic part of the access network in study (SC-BSs plus EC-BS)

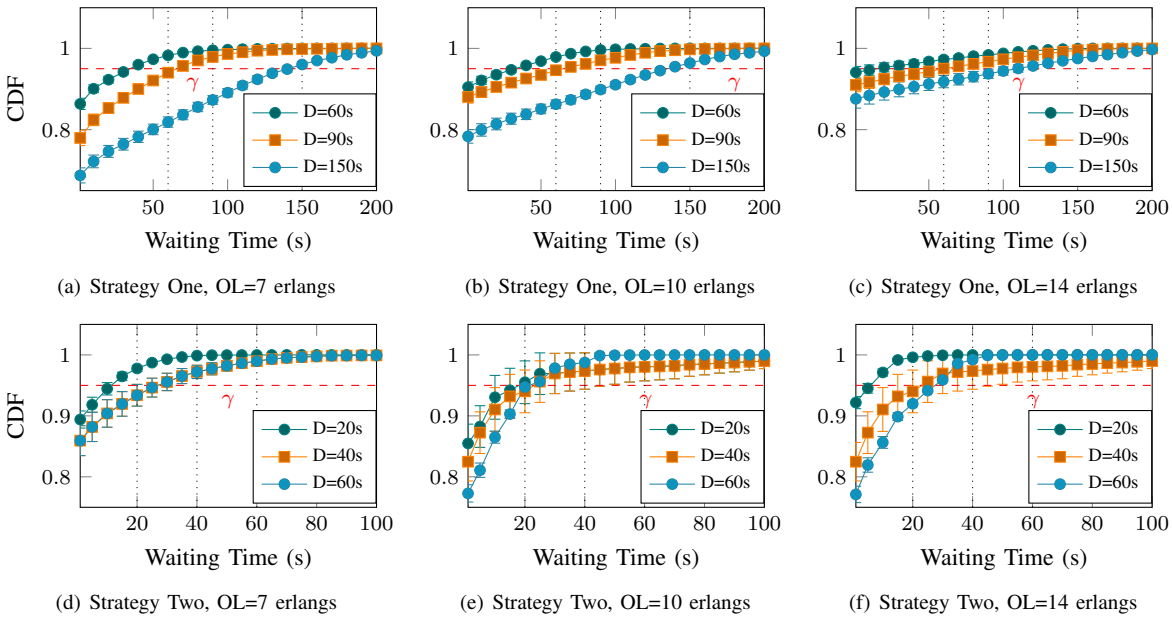
5) Importance of the load estimation: In this paper we focus on the execution of the DTU-aware cell switching algorithm, assuming a correct estimation of the offered load experienced by the system. In Table II we show the impact of the load estimation and threshold selection in satisfying the delay constraints. For example, using Strategy One and for an offered load of 10 erlangs, the thresholds are selected to support a maximum delay of 90 seconds in the DTU zone. If, instead, the offered load experienced in the system is 9 erlangs, the users are susceptible to wait up to 30 seconds more than what is proposed by the operator, as the thresholds were selected for a different offered load. Thus, accurate load estimation and strategy adaptation mechanism are critical aspects for the implementation of dynamic cell switching algorithms with delay constraints.

TABLE II: Threshold selection for different offered loads and the impact in the maximum waiting time (D). The same thresholds cause different D if the experienced offered load differs

	Offered Load (erlangs)	**Thresholds**		D **(s)**
		L_1	L_{DTU}	
Strategy One	10	7	14	90
	9	7	14	120
	11	7	14	70
Strategy Two	10	10	20	40
	9	10	20	25
	11	10	20	55

V. Conclusion

In this paper we presented an approach that further reduce the power consumption of cellular networks



(a) Strategy One, OL=7 erlangs (b) Strategy One, OL=10 erlangs (c) Strategy One, OL=14 erlangs

(d) Strategy Two, OL=7 erlangs (e) Strategy Two, OL=10 erlangs (f) Strategy Two, OL=14 erlangs

Fig. 6: Cumulative distribution function of the waiting time

combining coordinated cell switching and user awareness. Our strategy takes into consideration the user cooperation to extend the periods when the cellular access network can remain in low power consumption. In particular, we propose to offset the start of a requested service for a given bounded delay when required. We presented two threshold-based cell switching strategies enhanced with the proposed user cooperation approach. We evaluated the performance of the strategies using system level simulations, employing the LTE module of NS-3. We show that the strategies are able to control the waiting time and the impact of the cell switching in the user satisfaction, while providing power reductions. Finally, we show that if the users are able to occasionally tolerate some modest delays, the base stations can remain in low power consumption mode for longer periods, decreasing the total power consumed by the network.

References

[1] D. Ferling *et al.*, "Power efficient transceivers to enable energy-efficient mobile radio systems," *Bell Labs Technical Journal*, vol. 15, no. 2, pp. 59–76, Aug. 2010.

[2] W. Guo *et al.*, "Dynamic Cell Expansion with Self-Organizing Cooperation," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 5, pp. 851–860, May 2013.

[3] S. Gamboa *et al.*, "Energy efficient cellular networks in the presence of delay tolerant users," in *2013 IEEE Global Communications Conference (GLOBECOM)*. IEEE, Dec. 2013, pp. 2574–2580.

[4] ——, "Exploiting User Delay-Tolerance to Save Energy in Cellular Network : an Analytical Approach," in *2014 IEEE 25rd International Symposium on Personal, Indoor and Mobile Radio Communications - (PIMRC)*, 2014, pp. 1426–1431.

[5] M. A. Marsan *et al.*, "Switch-Off Transients in Cellular Access Networks with Sleep Modes," *2011 IEEE International Conference on Communications Workshops (ICC)*, pp. 1–6, Jun. 2011.

[6] L. Saker *et al.*, "Sleep mode implementation issues in green base stations," in *21st Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*. IEEE, Sep. 2010, pp. 1683–1688.

[7] K. Samdanis *et al.*, "Self-organized energy efficient cellular networks," in *21st Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*. IEEE, Sep. 2010, pp. 1665–1670.

[8] P. Skillermark *et al.*, "Enhancing Energy Efficiency in LTE with Antenna Muting," in *2012 IEEE 75th Vehicular Technology Conference (VTC Spring)*. IEEE, May 2012, pp. 1–5.

[9] F. Cardoso *et al.*, "Energy efficient transmission techniques for LTE," *IEEE Communications Magazine*, vol. 51, no. 10, pp. 182–190, Oct. 2013.

[10] D. Laselva *et al.*, "LTE SelfOrganising Networks (SON)," in *LTE SelfOrganising Networks (SON)*. John Wiley & Sons, Ltd, 2011, pp. 135—-234.

[11] L. Budzisz *et al.*, "Dynamic Resource Provisioning for Energy Efficiency in Wireless Access Networks: a Survey and an Outlook," *IEEE Communications Surveys & Tutorials*, vol. PP, no. 99, pp. 1–1, 2014.

[12] R. Schoenen *et al.*, "First Survey Results of Quantified User Behavior in User-in-the-Loop Scenarios for Sustainable Wireless Networks," in *2012 IEEE Vehicular Technology Conference (VTC Fall)*. IEEE, Sep. 2012, pp. 1–5.

[13] ——, "Erlang analysis of cellular networks using stochastic Petri nets and user-in-the-loop extension for demand control," in *2013 IEEE Globecom Workshops (GC Wkshps)*. IEEE, Dec. 2013, pp. 298–303.

[14] S. Ha *et al.*, "Tube: time-dependent pricing for mobile data," in *ACM SIGCOMM conference on Applications, technologies, architectures, and protocols for computer communication*, 2012, pp. 247–258.

[15] 3GPP, "36.902 version 9.2.0 Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Self-configuring and self-optimizing network (SON) use cases and solutions," vol. 0, pp. 0–22, 2010.

[16] M. F. Hossain *et al.*, "An eco-inspired energy efficient access network architecture for next generation cellular systems," in *2011 IEEE Wireless Communications and Networking Conference*. IEEE, Mar. 2011, pp. 992–997.

[17] NS-3 Consortium, "NS-3 LTE Module," *Retrieved Nov 2014, from http://www.nsnam.org/docs/models/html/lte.html*.

[18] G. Auer *et al.*, "How much energy is needed to run a wireless network?" *IEEE Wireless Communications*, vol. 18, no. 5, pp. 40–49, Oct. 2011.

[19] Cisco, "Voice Over IP - Per Call Bandwidth Consumption," *Online: http://www.cisco.com/c/en/us/support/docs/voice/voice-quality/7934-bwidth-consume.html*, 2014.