Low-Carb: A Practical Technique for Improving Energy Efficiency in Operational Cellular Networks

Muhammad Saqib Ilyas[a,*], Ghufran Baig[a], Mubashir Adnan Qureshi[a], Qurat Ul Ain Nadeem[a], Ali Raza[a], Munaf Arshad Qazi[a], Adnan Ahmed[a], Zartash Afzal Uzmi[a], Bilal Rassool[b]

[a] *School of Science and Engineering, LUMS, Lahore, Pakistan*
[b] *Warid Telecom, Lahore, Pakistan*

Abstract

Electricity costs are a significant fraction of a cellular network's operations costs. We present Low-Carb, a practical scheme to decrease electrical energy consumption in operational cellular networks by coupling *Base Transceiver Station (BTS) power savings* with call *hand-off*—two features commonly used by cellular operators. Motivated by the practical observation that most callers are in the vicinity of multiple BTSs, Low-Carb presents and solves an optimization problem, allowing calls to *hand-off* from one BTS to another so that *BTS power savings* can be applied to a maximal number of BTSs throughout the cellular network.

We use BTS locations and traffic volume data from a large live GSM network to evaluate the power savings possible using our proposed approach called *Low-Carb*. Our results indicate that performing coordinated call hand-off and BTS power-savings, a GSM 1800 network operator with 7000 sites in an urban setting can reduce annual electricity consumption by up to 35.36 MWh. This is at least 9.8% better than the energy savings achievable by using BTS power savings alone.

Keywords:
Green communication, BTS power-saving, energy conservation, energy efficiency

*Corresponding author
Email addresses: `{saqibm, 13100162, 13100152, 13100094, 14100214, 14100212, 13100204, zartash}@lums.edu.pk` (Bilal Rassool), `bilal.rassool@waridtel.com` (Bilal Rassool)

1. Introduction

Cellular networks are quite energy inefficient and consume several tens of TWhs of electrical energy every year worldwide [1]. This is a source of concern not only due to ecological concerns of global warming but also due to rising operational expenditure resulting from rising fuel and electricity prices.

In the present work, we focus on the second generation (2G) cellular network technology, Global System for Mobile communication (GSM). While most recent research focuses on later generations of cellular networks, we focused on GSM for the several reasons. First, it has a large consumer base. Second, upgrade to later generations of cellular networks is prohibitively expensive especially for operators in the developing world where profits are low due to cut-throat competition and slow return on investment. Our proposed energy efficiency improvement framework is generic and should be applicable to other types of cellular networks as well, but we make no claim to its effectiveness in such scenarios.

The major sink of power in a cellular network are Base Transceiver Stations (BTSs), accounting for 50% to 90% of the total power consumption [2, 1]. Our focus in the present work is to reduce the energy consumption of BTSs to improve the overall cellular network's energy efficiency.

In a GSM network, every BTS is equipped with several transceivers (TRXs), each of which is allocated a single frequency band for transmission and reception of radio signals. Each TRX further uses time multiplexing to handle up to 8 full-rate voice calls over its assigned frequency band in GSM systems. A typical configuration is "6+6+6" depicting a BTS serving three *sectors* each with six TRXs. Thus, a BTS offers a *fixed* capacity, as determined by the total number of TRXs installed. Sites are deployed such that this fixed BTS capacity can handle the peak traffic load. However, traffic peaks only for a short duration dropping off to a much lower trough each day, which means that the GSM networks are over-provisioned during low-traffic regimes.

Over-provisioned BTSs would be fine if they also consumed little power at no traffic load. However, according to [3] the no-load power consumption can be as high as 95% of that at full load. With fixed BTS capacity that is over-provisioned for low traffic loads, today's cellular networks are highly energy inefficient.

There are generally two approaches to improve cellular network energy efficiency. First is a clean-slate redesign which includes innovations in communication systems, circuits and components. This approach is not attractive for existing GSM operators, which are the most prevalent in the developing world and are expected to stay as such for several years to come, primarily due to the required expensive upgrades. A second approach is to make optimizations to the existing system and equipment to get an improvement in overall energy efficiency. Our present work is aligned with this latter philosophy.

One can improve the energy efficiency of a cellular network by adapting its "online" capacity to changes in traffic load. Recent work has proposed turning off base stations to reduce energy consumption during times of low traffic load [2, 1, 3, 4]. However, our conversations with multiple network operators indicate that they are reluctant to employ such techniques citing three reasons:

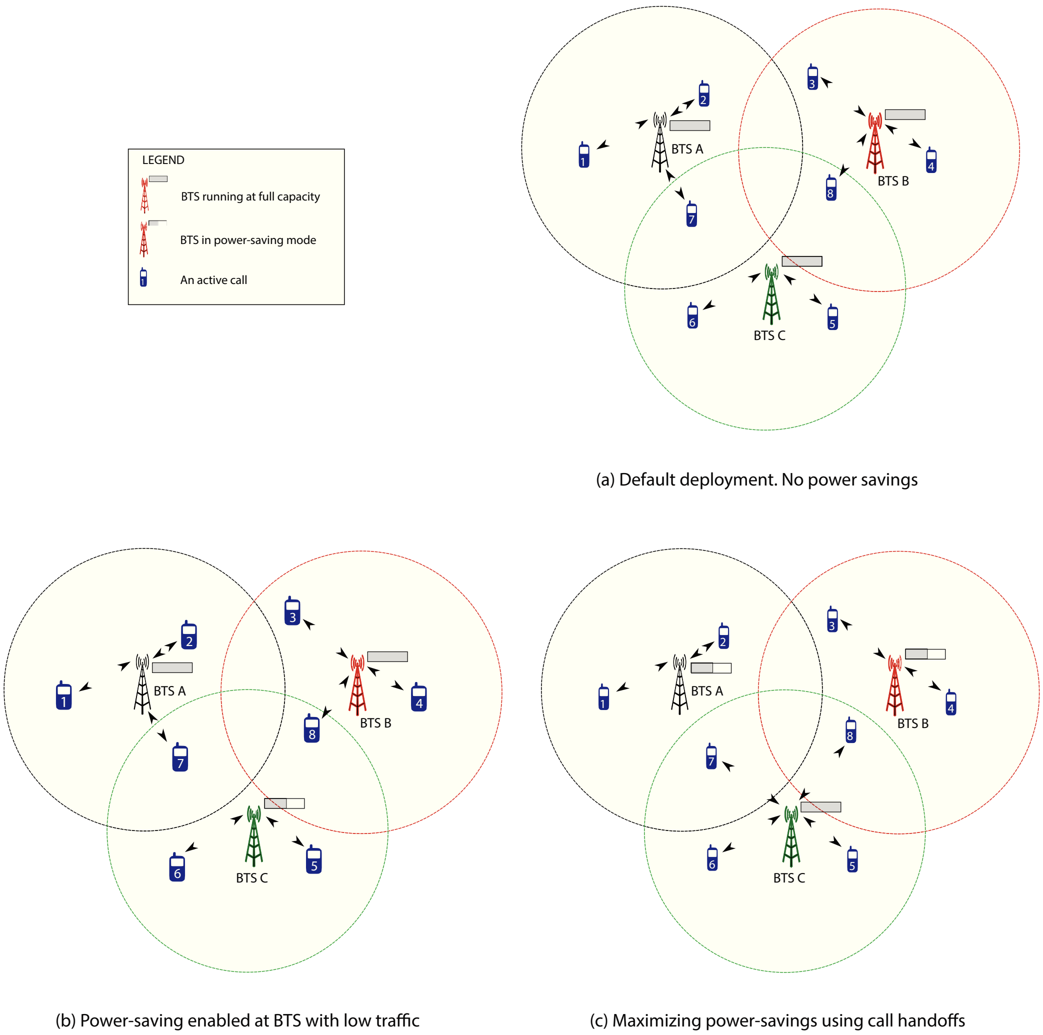
- Power cycling of entire base stations is expected to reduce equipment life time.
- Turning off some BTSs may require an increased uplink power which may not be handled by many low-cost/power-limited mobile stations (MSs). This raises a risk of customer churn and is not acceptable to the operators in cut-throat competition prevalent in today's market.
- These techniques of turning off BTSs may underestimate the increase in power needed for indoor MSs.

Our conversations with wireless providers reveal that during low traffic periods, they often use a feature available in most vendor's equipment that power-gates TRX circuits at locations that serve very few customers. Huawei calls this feature *TRX shutdown* while Ericsson calls it *BTS power saving*. We use the latter term generically in this paper. Since BTS power consumption has a traffic-independent component [3] that depends, among other factors, on the number of active TRXs, deactivating TRXs reduces the BTS power consumption. For instance, turning off one TRX cuts down BTS power consumption anywhere from $20W$ to $100W$, depending upon the frequency band (900 or 1800) and deployed equipment [5, 6]. Thus, scaling a "6+6+6" to a "2+2+2" configuration, by deactivating 12 TRXs will result in a saving of 240W to 1200W on a single site. The decision to use *BTS power saving* feature is generally local to the BTS without any coordinated effort at the network level.

Let us illustrate the main idea behind the energy-saving approach proposed in this paper with the help of the illustration in Figure 1. The figure shows three nearby BTSs collectively serving eight active calls. The association of call to the serving BTS is indicated by means of a dashed line, while the boundary of the potential service area of a BTS is indicated by means of a dashed circle around it. By default, each call is served by the BTS from which the mobile station receives the strongest signal. In this example, we assume that the call handling capacity of each BTS is six simultaneous calls. Furthermore, we assume that the power-saving threshold is three calls, i.e., if a BTS is serving less than three calls, it may be put into power-saving mode. In Figure 1, a BTS in it's default configuration, i.e., all TRXs enabled is indicated with a solid bar next to the BTS symbol. Meanwhile, a BTS in power-saving mode is shown by a half-filled bar next to the BTS symbol.

In Figure 1, note that calls 1 through 6 each have only one candidate serving BTS, while call 7 may be served either by BTS A or C and call 8 may be served either by BTS B or C. Figure 1 (a) shows the default deployment with default call routing whereas no power-saving is used. Since BTS C is serving only two calls, it may be placed in power-saving mode. Figure 1 (b) shows this network state, which indicates the current practice in operational cellular networks, whereby BTS C is put into power-saving mode because it's current call volume is below the power-saving threshold. However, this is not the optimal call routing strategy in terms of energy-savings. We may handoff calls 7 and 8 to BTS C, thereby reducing the call volume at both BTS A and B below the power-saving threshold. This results in the energy-optimal call routing strategy, proposed in this paper, shown in Figure 1 (c).

This paper presents Low-Carb which combines the *BTS power saving* with *hand-off*, another commonly used feature in cellular networks that facilitates user movement from one location to another. Low-Carb proposes to hand-off calls from one BTS to another, without making a negative impact on the network quality of service, such that the *BTS power savings* can be applied to a maximal number of base stations throughout the cellular network. As compared to the use of uncoordinated *BTS power savings*, Low-Carb offers additional power savings as it may allow a larger number of TRXs to be deactivated. In present day deployments, this is possible since most callers receive sufficiently strong signal from several nearby BTSs [3].

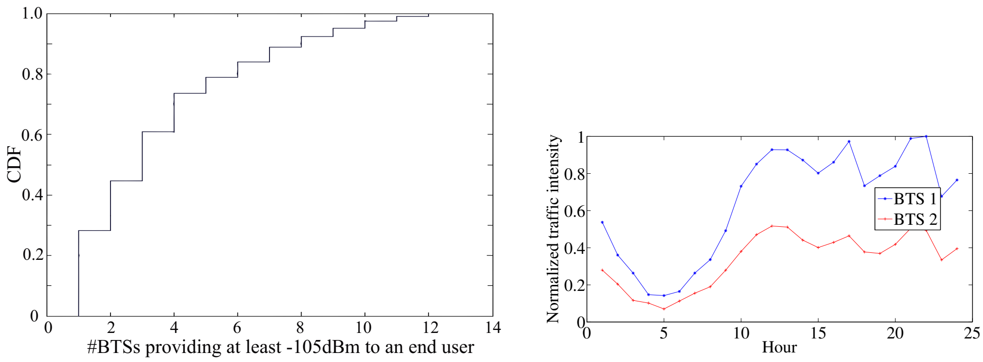


(a) Default deployment. No power savings

(b) Power-saving enabled at BTS with low traffic

(c) Maximizing power-savings using call handoffs

Figure 1: A toy example to illustrate the main idea behind Low-Carb. Three BTSs (A, B and C) are shown alongwith eight active calls. The serving BTS for each call is shown using a dashed line. A solid completely filled bar alongside a BTS symbol indicates that it is running in the default configuration where all TRXs are enabled. A BTS in power-saving mode is indicated with a half-filled bar next to it. Assume that power-saving mode may be enabled at a BTS if it has less than three active calls. (a) shows the default configuration where power-saving mode is not used. (b) shows the approach currently used by operators whereby power-saving mode is enabled on a BTS with low-traffic. (c) shows our proposed approach, whereby calls may be handed-off to nearby BTSs, thus maximizing the number of BTSs in power-saving mode.



(a) (b)

Figure 2: Characteristics of our dataset: (a) Empirical CDF of the number of potential serving BTSs for a call in our dataset (large metropolitan area), (b) Normalized traffic intensity at two neighboring sites in our dataset

Fig. 2(a) shows coverage diversity evident in the urban data from a large cellular provider that we used in our evaluations; one can see that about half of the callers have 3 or more candidates for serving BTS. Furthermore, neighboring sites can have different traffic loads at a given time. Fig. 2(b), for instance, shows normalized traffic at two neighboring sites in our dataset for a 24 hour period. Thus, some calls may be handed off from busy BTSs to nearby BTSs with lower traffic volume to increase the number of BTSs in power-saving mode.

We formulate an optimization problem to minimize the power consumption in a GSM network by shuffling active calls between nearby BTSs while keeping in check the MS uplink budget and without dropping any active calls. By constraining the uplink budget, we ensure that call quality is not adversely affected.

In a shorter version of this paper [7], we made the following contributions:

1. We formulated, Low-Carb, a mathematical optimization problem to maximize energy savings in a cellular network such that call quality is not compromised. To maximize energy savings, Low-Carb relies on two features that are implemented in typical BTS hardware and are commonly used by cellular operators, albeit in a non-systematic manner.
2. Since our formulation of Low-Carb is NP-Hard, we provided a heuristic algorithm for solving the Low-Carb problem.

3. We used real data sets from a large GSM network operator in Pakistan to evaluate Low-Carb's utility.

In this paper, we extend our prior work to make the following additional contributions:

1. Let δ be the traffic capacity of the BTS in low-power mode. If the BTS is placed in power-saving mode as soon as the traffic reaches $\delta - 1$ calls, then there may be several oscillations in and out of low-power mode due to short-term traffic variations. During these oscillations, some calls may be blocked as well, which is undesirable. Thus, traffic must be allowed to fall to at least $\delta - \epsilon$ calls before the BTS is put into low-power mode. If we pick a high value for ϵ, the amount of energy savings is expected to drop. We experiment with various values for ϵ and study the dependence of the amount of energy savings on the value of ϵ.
2. In [7], we had considered that a BTS could be put into one of two possible states, namely low-power and high-power mode. In this paper, we consider that a BTS may operate in one of six modes of operation in terms of its power consumption. We show that a higher granularity of number of BTS operational states results in greater energy savings.
3. We provide another heuristic algorithm for solving Low-Carb which is different from the one presented in the shorter version of this paper.

The rest of the paper is structured as follows. The formulation of Low-Carb optimization problem is given in section 3. Experimental setup and the results are presented in sections 4 and 5, respectively. In section 6, we draw the conclusions highlighting the power saving strategy for service providers.

2. Related Work

The problem that we investigated in this work falls into the broad category of resource scheduling problem under constraints. Resource scheduling problem occurs in many different domains whereby similar solution strategies are applicable. Some examples of prior work in other domains on similar problems include resource provisioning in data centers [8, 9, 10, 11, 12], scheduling in compute clusters [13], System on Chip (SOC) [14], electric power systems and smart grid [15, 16, 17, 18], WiFi access points [19], wide area networks [20], high performance computing [21, 22, 23, 24, 25, 26] as well as cellular networks [27, 3].

Our focus in the present work is to develop techniques that are practically usable in operational networks instead of taking a clean-slate approach. In this sense, it differs from much prior work in energy efficiency in cellular networks. A key benefit of this approach is that it does not require any additional hardware and works within the GSM specifications. Our work is very similar in spirit to the concept of *frequency dimming* in [28] albeit at a different level of abstraction. A similar approach is also proposed in [29] with some rough estimates of expected savings. We, on the other hand, use site locations and traffic traces from a large cellular network with more than 13 million subscribers to run a simulation study assessing the benefits of dynamic equipment scaling coupled with call hand-offs.

3. Problem Formulation

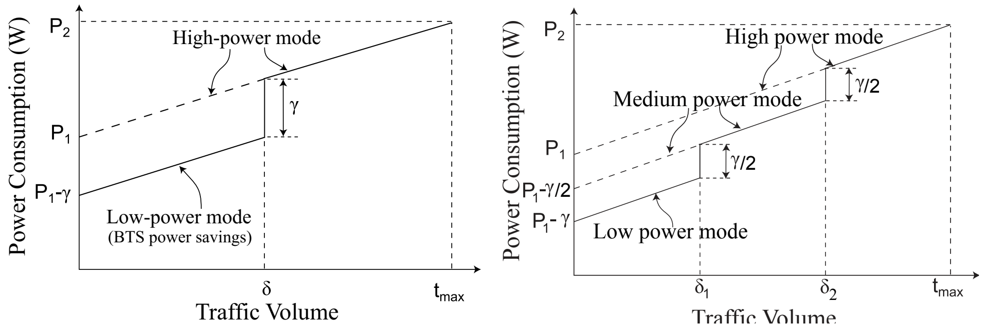
3.1. Single Base Transceiver Station (BTS)

Power consumed by a BTS may be well-approximated as an affine function of it's traffic load [3], given as $P_1 + l(P_2 - P_1)/t_{max}$. Here P_1 and P_2 are the power consumption at no load and full load, respectively, l is the number of calls presently being handled, and t_{max} is the maximum number of calls that can be handled.

Let δ be the traffic threshold at which the *BTS power savings* is applied (i.e., when BTS deactivates some TRXs moving into low-power mode). Since all TRXs are identical, the per call increase in power consumption, and hence the slope of the power consumption profile in Fig. 3(a), remains the same whether or not some TRXs are deactivated. As also indicated in Fig. 3(a), the no-load power consumption drops to $P_1 - \gamma$ in the low-power mode, where γ is a constant that depends on the equipment type and the number of TRXs deactivated.

If x is an indicator variable which is 1 when *BTS power savings* is applied, and 0 otherwise, then the BTS power consumption may be given by $P_1 + l(P_2 - P_1)/t_{max} - (1 - x)\gamma$, also indicated in Figure 3(a) by the piecewise linear solid line.

The BTS equipment may be configurable in more than two-states, depending on the available granularity of TRX deactivation. for instance, a three-state BTS power consumption model is shown in Figure 3(b). If the traffic is above threshold δ_2, the BTS operates in the default high-power



(a) (b)

Figure 3: BTS power consumption model. (a) Low-power (BTS power savings) mode is optional and kicks in at low loads. (b) BTS power saving is applied in a more granular way.

mode. When the traffic drops below threshold δ_2, two TRXs per sector are power-gated, and the BTS enter the medium-power mode (4+4+4 configuration). When the traffic falls below threshold δ_1, two more TRXs are power-gated and the BTS enters the low-power mode. The drop in power consumption for each of the deactivation steps is $\gamma/2$, because compared to Figure 3(a), half as many TRXs are deactivated in each step.

The more granular model of Figure 3(b) should offer greater energy savings. If traffic load is $\delta + \epsilon$ then the less granular model of Figure 3(a) does not offer any energy savings but the more granular model ofFigure 3(b) offers an energy saving of $\gamma/2$. The more granular model of Figure 3(b) should, therefore, offer greater energy savings overall. We will show later that this greater energy saving is at the cost of increased complexity of the optimization problem.

3.2. Multi-BTS Cellular Setting

Consider an area with n active callers being served by m BTSs. We introduce indicator variable $w_{i,j}$, which is 1 if call i *is being* handled at BTS j and 0 otherwise. We assume availability of an $n \times m$ matrix whose entry $c_{i,j}$ is 1 if caller i *can be* served through BTS j without exceeding the uplink or downlink budgets. This information can be extracted by the data periodically transmitted by each MS comprising the received signal strength from nearby BTSs during a call. We also introduce indicator variable x_j, which is 1 if BTS j is operating in high-power mode (i.e., without *BTS power savings*) and 0 otherwise. Using these variables and parameters, we can formulate an optimization problem to minimize the total power consumption over the network as:

$$minimize \quad \sum_{j=1}^{m} \left[P_1 + \sum_{i=1}^{n} \frac{w_{i,j}(P_2 - P_1)}{t_{max}} - (1 - x_j)\gamma \right] \tag{1}$$

subject to the following constraints:

$$\sum_{j=1}^{m} w_{i,j} = 1 \qquad \forall i \tag{2}$$

$$w_{i,j} \leq c_{i,j} \qquad \forall i, j \tag{3}$$

$$\sum_{i=1}^{n} w_{i,j} - \delta \leq M x_j \qquad \forall j \qquad\qquad \sum_{i=1}^{n} w_{i,j} \leq t_{max} \qquad \forall i \tag{4}$$

$$w_{i,j}, x_j \in 0, 1 \qquad \forall i, j \tag{5}$$

We call this optimization problem, the two-step Low-Carb problem. The objective function is a simple generalization from the case of one BTS. The first constraint ensures that no active call is dropped just to save on power. The second constraint secures the uplink budget by ensuring that no call is routed to a BTS that can not handle it. The third constraint picks the correct value for the decision variable x_j. The fourth constraint is the capacity constraint on all BTSs, while the last constraint is the binary value constraint on the decision variables.

Note that the first term in the objective function is a constant and may be dropped from the optimization. Similarly, the second term ($\frac{P_2 - P_1}{t_{max}} \sum_{j=1}^{m} \sum_{i=1}^{n} w_{i,j}$) is also a constant which equals $\frac{n(P_2 - P_1)}{t_{max}}$ because of the first constraint, and hence may be omitted from the optimization. Similarly, multiplication by a scalar does not affect the optimal so γ may also be omitted from the optimization. This leaves the objective function as:

$$minimize \quad \sum_{j=1}^{m} -(1 - x_j) \tag{6}$$

This may alternatively be specified as:

$$maximize \quad \sum_{j=1}^{m} x_j \tag{7}$$

which makes intuitive sense, too, as maximizing the number of BTSs in power-saving mode would minimize the energy consumption.

For the three-step BTS power-saving model of Figure 3(b), the optimization problem can be stated as:

$$\textit{maximize} \quad \sum_{j=1}^{m} [x_j + y_j] \tag{8}$$

subject to the following constraints:

$$\sum_{j=1}^{m} w_{i,j} = 1 \qquad \forall i \tag{9}$$

$$w_{i,j} \leq c_{i,j} \qquad \forall i, j \tag{10}$$

$$\sum_{i=1}^{n} w_{i,j} - \delta_1 \leq M x_j \qquad \forall j \tag{11}$$

$$\sum_{i=1}^{n} w_{i,j} - \delta_2 \leq M y_j \qquad \forall j \tag{12}$$

$$\sum_{i=1}^{n} w_{i,j} \leq t_{max} \qquad \forall i \tag{13}$$

$$w_{i,j}, x_j, y_j \in 0, 1 \qquad \forall i, j \tag{14}$$

In the above statement, we've added an indicator variable y_j, which is 1 if BTS j is in medium-power mode, 0 otherwise. The fourth constraints ensures that y_j takes on the proper value depending on the current traffic volume at BTS j. We call this optimization problem the three-step Low-Carb problem.

Adding a granularity level introduces m additional binary variables causing the solution space to grow by a factor of 2^m. At the same time, $O(m)$ constraints are also added to the problem. Binary Integer Programming problems such as ours are NP-Hard. While solutions to small-sized problems may be obtained in a reasonable period of time, but it lacks scalability. Therefore, increasing the number of BTS operating modes causes an exponential increase in the execution time to obtain an optimal solution.

For ease of modeling and implementation, we assume that all sectors will be configured to have the same number of TRXs active at any given time. Since a typical BTS has a 6+6+6 configuration, we could have upto six possible states 1+1+1, 2+2+2, ..., 6+6+6 for a BTS. To develop the optimization model for a six-state BTS, consider an integer variable z_j which represents the number of TRXs active in each sector of BTS j. The power consumption at BTS j may be given by $P_1 + p\sum_{i=1}^{n} w_{i,j} + \gamma z_j$, where p is the slope of the power consumption vs traffic volume curve. The RED-BL optimization problem for the multi-BTS setting in this case is:

$$minimize \quad \sum_{j=1}^{m} z_j \tag{15}$$

subject to the following constraints:

$$\sum_{j=1}^{m} w_{i,j} = 1 \qquad \forall i \tag{16}$$

$$w_{i,j} \leq c_{i,j} \qquad \forall i, j \tag{17}$$

$$\sum_{i=1}^{n} w_{i,j} - \delta_1 + \epsilon \leq M z_j \qquad \forall j \tag{18}$$

$$\sum_{i=1}^{n} w_{i,j} \leq t_{max} \qquad \forall i \tag{19}$$

$$w_{i,j} \in \{0, 1\}; \quad z_j \in \{1, 2, 3, 4, 5, 6\} \qquad \forall i, j \tag{20}$$

The first constraint ensures that each call is handled at exactly one BTS. The second one ensures that a call is mapped to a BTS that may handle it. The third constraint ensures that z_j picks a value at least equal to the minimum number of TRXs required at BTS j. Since the optimization is a minimization problem, the solver will pick z_j to be exactly equal to the minimum number of TRXs required for the current traffic load. The fourth constraint is the capacity constraint on the BTS. The last set of constraints specify the domains of the decision variables.

3.3. Heuristic solutions to Low-Carb

Both of the above optimization problems are Binary Integer Programs (BIP), which is NP-Hard. It is intractable to solve it for an operator's entire network, but solving it for a subset of the network will provide some estimates of the amount of energy savings possible using Low-Carb. Deployment to large operator networks would require approximation algorithms. In this paper, we present two heuristics to solve the problem approximately and compare the results with the optimal solution.

3.3.1. Heuristic 1

We describe here, the first heuristic for the two-step Low-Carb problem. The pseudo-code for the algorithm is given in Algorithm 1. On lines 1 and 2, our heuristic divides the set of BTSs (B) into two disjoint subsets B_1 and B_2. Subset B_1 consists of those BTSs that have traffic above threshold δ. Subset B_2 consists of all other BTSs. On line number 3, our heuristic shuffles the members of B_1 and then iterates over the set on lines 4 through 24. In every iteration, our heuristic computes the current load on a particular BTS, thereby determining the minimum number of calls to be handed-off from that BTS to enable power-saving mode at it (line number 5). The heuristic then iterates over the calls mapped to that BTS and for each such call, tries to identify a BTS in B_2 that this call may be handed off to. This heuristic can be invoked multiple times, using a different shuffled order of BTSs in B_2 to potentially improve the heuristic's performance. At the end of this process, members of B_2 are placed in low-power mode.

For the three-step Low-Carb problem, members of B_1 are those BTSs that have traffic above threshold δ_2, while B_2 contains all other BTSs. We shuffle some calls from BTSs selected in a random order attempting to reduce the cardinality of set B_1. Once we are done iterating over all members of B_1, we re-initialize B_1 to contain BTSs that have traffic above threshold δ_1 and B_2 to contain all other BTSs and repeat the same call hand-off process as described earlier. In the end, those BTSs that have traffic above δ_1 but below δ_2 are placed in the medium-power mode while those with traffic below the δ_1 threshold are palced in the low-power mode. Once again, this heuristic can be invoked multiple times to improve the probability of finding a near-optimal solution.

input : B (the set of BTSs) $=\{b_1, b_2, ..., b_m\}$,
C (the set of calls) $=\{c_1, c_2, ..., c_n\}$,
W (current call association) $=\{w_{i,j} | 1 \leq i \leq n, 1 \leq j \leq m\}$,
A (Adjacency matrix) $= \{a_{i,j} = 1$ if c_i can be served through b_j, 0 otherwise$\}$
output: A new and potentially more energy efficient mapping of calls to BTSs

$B_1 = \{b_j | \sum_{i=1}^{n} w_i^j > \delta\}$;
$B_2 = B - B_1$;
$B_1 =$ random_shuffle(B_1);
forall the $b_j \in B_1$ **do**
    a $= \sum_{i=1}^{n} w_i^j - \delta$;
    d = 1;
    shuffled = 0;
    while $d < n$ *AND shuffled* $\leq a$ **do**
        if $w_d^j = 1$ **then**
            e = 1;
            mapped = 0;
            while $e \leq m$ *AND mapped = 0* **do**
                if $e \in B_2$ *AND* $a_{d,e} = 1$ **then**
                    $w_k^l = 1$;
                    $w_k^j = 0$;
                    shuffled = shuffled+1;
                end
                e = e + 1;
            end
            d = d + 1;
        end
    end
end

Algorithm 1: Energy-saving heuristic 1

3.3.2. Heuristic 2

We first describe the second heuristic for the two-step Low-Carb problem. In contrast to the first heuristic, the second one iterates over calls first. It first assigns all calls that only have one candidate BTS to the only BTS that can handle them.

4. Experimental Setup

Our dataset is obtained from a cluster of 26 BTSs operated by a large network operator with more than 7000 sites. These sites are spread over a 31.25 km^2 urban terrain. We obtained each site's coverage prediction using a tool popular amongst the operators called Forsk Atoll. With this information, alongwith a caller's location, we can determine the candidate set of BTSs for the corresponding call (the c_i^j parameters). Note that in this work, we do not incorporate user mobility into our model, since we are only interested in instantaneous optimization. Furthermore, according to our conversations with multiple cellular operators in Pakistan, almost 90% or more cellular calls originate indoors with negligible mobility.

Also available to us are the hourly cumulative traffic, in Erlang, for each of the sites, spanning two consecutive weekdays. The traffic remained remarkably similar across both days for each site. We have, therefore, only used one day's traffic data in our experiments.

Using the above datasets, we conducted a set of experiments mimicking a 24-hour operation of a subset of a cellular network. Each experiment is a discrete event simulation of the arrival and placement of calls. Since our dataset does not include the arrival times and duration of calls, we synthetically generated this information using the assumption of Poisson call arrivals and exponentially distributed call duration with a mean of 180 seconds [30].

For every hour, the simulator determines the Poisson call arrival rate for each BTS, using Little's Law and the BTSs traffic intensity for that hour. Using the resulting Poisson process, calls are generated such that it is equally likely for a call to be anywhere in the serving BTSs coverage area.

Our simulator tracks the call volume at every BTS on a minute's granularity. This enables us to calculate the power consumption level (in Watts) of the BTS during each minute. To calculate the baseline energy consumption, i.e., in the absence of any optimization, we may simply aggregate the power consumption for every minute. Accumulating these numbers over the 24 hour period leads to the daily amount of energy consumed (in kWh) if no optimization is used in the network. Our simulator also monitors each BTSs call volume every minute and places the ones with sufficiently low traffic into power-saving mode. This enables us to calculate the possible energy savings using only the BTS power-saving feature. In addition, our simulator also periodically determines the instantaneous optimal call placement configuration that minimizes the power consumption level by handing-off some calls, thereby placing a maximal number of BTSs in power-saving mode. This allows us to determine the energy savings possible by combining call hand-offs with BTS power-saving.

The call placement re-optimization may be done at various frequencies. A very aggressive re-optimization regime would keep the network in an optimal state more often than a conservative one, thereby enabling greater energy savings. In order to study the scaling of energy saving with re-optimization frequency, we experimented with a range of intervals between successive optimizations, ranging from a minute to an hour. A more frequent re-optimization is expected to bring greater energy savings compared to a less frequent one, but some cost would be traded-off. Let us now consider such costs.

First and foremost, a computational cost is incurred with each optimization. In our case, an optimization run to determine the optimal state over 26 BTSs required an average running time of about 50 seconds on a Core i3 laptop with 4 GB of RAM. An optimization requiring 50 seconds would not be practical to use every minute but may be fine if used less often. For a practical deployment the computational time can be reduced by using a combination of a more powerful machine, distributed optimization and approximation algorithms.

In addition to the computational overhead, for every unit of energy saved some extra energy may be consumed in the network to perform call hand-offs or entering and leaving BTS power-saving mode. Call hand-offs and TRX (de)activation involve signaling between a Base Station Controller (BSC), BTSs and MSs. The additional energy incurred thus, should be small, because it has been observed that variation in power consumption of network equipment with changes in traffic volume (data or control) is quite small [31]. As far as increased power consumption on MSs due to a greater number of call hand-offs is concerned, we opine that it may be negligible because the MSs energy consumption is far outweighed by that of BTSs.

4.1. Site Characteristics

All sites in our dataset had three sectors, each equipped with 6 TRXs, for a maximum of 132 simultaneous voice calls[1]. The GSM standard includes a provision for half-rate calls, which enables handling greater traffic at the expense of reduced voice quality by allowing a single voice channel to be shared amongst two calls, each using a half-rate codec. In this paper, we only shuffle full-rate calls around, which may be, in reality, two half-rate calls. We do not foresee any significant error arising from using this convention.

When considering the two-step BTS scaling model, a "$6+6+6$" site may be scaled down to a "$2+2+2$" site. In this case, δ should be strictly less than $t_{max}/3$ to avoid quick oscillations into and out of BTS power-saving mode due to short-term traffic variations. We have set δ equal to $\lceil t_{max}/3 \rceil - \epsilon$ and experimented with all possible values of ϵ.

For the three-step BTS scaling model, whereby a "$6+6+6$" site may be scaled down to "$4+4+4$" site or a "$2+2+2$" site depending on current traffic conditions, two traffic thresholds, namely δ_1 and δ_2 must be specified. In this paper, we set δ_2 to $2t_{max}/3 - 5$ and δ_1 to $t_{max}/3 - 5$, respectively. Also, due to the NP-hard nature of the Binary Integer Program, experiments using the three-step model and all 26 BTS sites take excessively long to run. We, therefore, used a 20 site subset of the global dataset in our experiments when using the three-step model.

The BTS power consumption model parameters may vary from one BTS model to another. In this paper, we use three different sets of model parameters as listed in table 1. We now describe the sources and methods from which we obtained these models.

4.1.1. Model 1

For the first model, we have used $1.5kW$ as the maximum power consumption [32], a $20W$ per TRX saving when scaling the BTS down [6] and a 5% swing in power consumption between no-load and full-load [3].

4.1.2. Model 2

Lorincz et. al reported the single sector DC power consumption for a GSM 900 BTS [5]. The sector under consideration had 7 TRXs, as opposed

[1]Each TRX's frequency is shared in time-domain by 8 calls for a total of $3\times6\times8 = 144$ channels. Four channels in each sector were reserved for control and broadcast purposes.

Parameter	Value		
	Model 1	Model 2	Model 3
P_1	1425	2401.8	2341.5
P_2	1500	3887.5	2973.9
γ	20	50	100

Table 1: BTS model parameter values

to 6 TRXs in our case. To approximate the DC power consumption for a site with 3 sectors, each with 6 TRXs, we scaled the power consumption by a factor of $3 \times 6/7$. The DC power consumption does not include the AC power consumed in the power supply units and in air-conditioning. We must, therefore, also compensate for those, to obtain the overall site power consumption. Power supply unit load is negligible compared to air-conditioning, which has a typical power consumption of 1 kW [32]. We applied this scaling and addition to the minimum reported DC power consumption for the GSM 900 site to obtain an approximate value of P_1 for a site comparable to ours. Similarly, we used the maximum reported DC power consumption and applied the scaling and AC load correction to approximate the value of P_2. Furthermore, the authors measured a drop of $50W$ in power consumption when a TRX is disable, which gives us the value of γ as listed in table 1.

4.1.3. Model 3

Using the same method as for model 2 in 4.1.2, we derived the values for P_1 and P_2 using the measurements for the GSM 1800 BTS in [5]. As for the value of γ, the paper reported a $100W$ cut in power consumption when deactivating a single TRX. The parameter values for this model are given in Table 1

5. Results

5.1. BTS with two possible power states

The following results were obtained through simulation experiments driven by real traffic traces and deployment geography. The experiments perform a combination of activation of BTS power-saving mode on BTSs alongwith a periodic update of serving BTS for each active call, such that the instantaneous energy consumption in the network is minimized.

Energy saving	Model 1	Model 2	Model 3
Percentage	4.73%	5.43%	12.89%
Daily absolute saving over 26 BTSs (in kWh)	43.28	109.68	217.12
Country-wide daily saving over 31000 sites (in MWh)	51.6	130.77	258.87

Table 2: Energy savings by using BTS power savings only

First, we consider the benefit of BTS power-saving alone, resulting from traffic diversity at each BTS compared to running the network in the default configuration. The percentage reduction in energy consumption is listed in table 2. The results indicate that a saving of between 4% and 12% can be achieved in a network just by activating BTS power savings. We note here that some of these results are in agreement with Ericsson's claim of saving 10-20% energy by using BTS power-saving on Germany's Vodafone network [33].

In absolute terms, this represents a cumulative saving of between 43 kWh and 217 kWh per day on 26 BTSs. Now, consider that there are five cellular oeprators in Pakistan: Mobilink with more than 8500 sites [34], Ufone with more than 8000 sites [35], Zong with more than 5500 sites [35], Telenor with more than 7000 sites [36] and Warid with more than 4500 sites [35]. Overall, there were more than 31000 sites in Pakistan at the end of 2011. We extrapolated the daily energy savings number over 26 BTSs to calculate the daily energy savings possible for a country like Pakistan with over 31000 BTSs (see the last row of table 2). The results indicate that mere activation of BTS power saving option itself can save quite a bit of electrical energy, a critical resource, especially in a developing country. As we shall see next, greater energy savings are possible if we couple periodical call shuffling with BTS power savings in the network.

If periodic optimization of call placement is coupled with BTS power-saving, the energy saving improves, as shown in Figure 4(a). For all three BTS models, we see an almost linear increase in power saving as the duration of the re-optimization interval is decreased. Recall that the three models are significantly different in terms of power consumption (see Table 1). We can not directly say that because model 3 BTS offers the highest percentage reduction in energy consumption, it also saves the most energy (in kWh).

To *compare* the three BTS models in terms of energy saving potential,

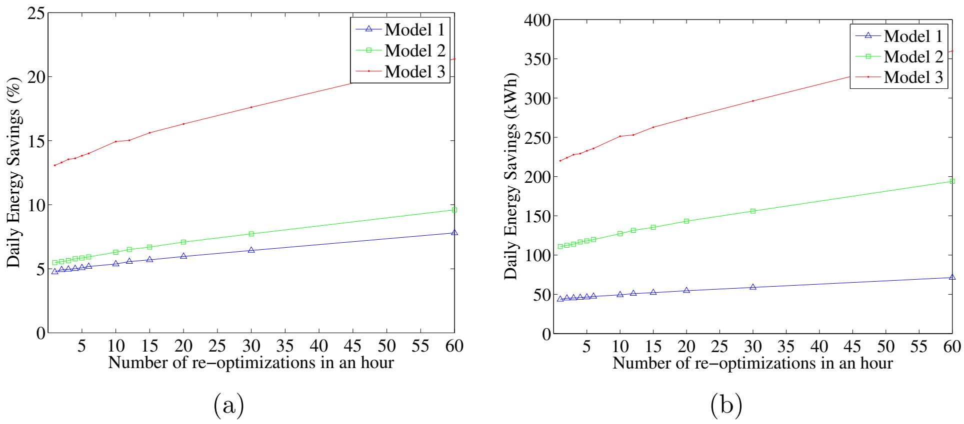


Figure 4: (a) Percent reduction in energy consumption vs re-optimization interval, (b) Reduction in energy consumption vs re-optimization interval

we also present the absolute reduction in energy consumption for the three BTS models in Figure 4(b). We see the same linear trend alongwith the same relative order of the three models in terms of amount of saved energy, as in Figure 4(a).

Re-optimizing at an interval less than the mean call duration should offer greater savings than a less frequent re-optimization, because the former regime has the opportunity to optimize more by handing off most of the calls. This is confirmed in our results. For instance, for model 1 BTS, the gain in energy savings going from a 60 minutes inter-optimization interval to 30 minutes gains an energy saving of only 0.0506 kWh per minute, while decreasing the inter-optimization interval from 2 minutes to 1 minute gains 12.5421 kWh.

Let us now interpret what these results mean physically in terms of ecological impact. If we extrapolate our results, the total energy saving for Pakistan are projected to be 60.72 MWh, 156.84 MWh and 301.61 MWh daily, respectively, according to the three BTS models. These savings in energy are significant, especially for small and developing countries. Since network deployments and traffic patterns are similar in different countries, we also expect that similar savings should be achievable in many other countries as well.

In the above extrapolation, we have assumed that the same amount of energy saving would be applicable in rural as well as urban settings. While this may not necessarily be true because the deployments are sparse in rural

Granularity	BTS Model 1	BTS Model 2	BTS Model 3
2-State	5.38%	6.29%	14.94%
3-State	6.81%	7.73%	18.62%
6-State	14.69%	25.33%	33.69%

Table 3: Percentage electricity savings for different granularity of resource pruning

settings, resulting in reduced potential to save energy by means of call hand-off to neighboring sites, the potential to save energy merely by BTS power-saving should be higher in a rural setting because traffic loads are typically lower.

5.2. Multi-state BTS

In our experimental results discussed so far, we have observed that going from a 6+6+6 configuration to a 2+2+2 configuration can save a significant amount of energy. Intuition suggests that going to a finer granularity of resource pruning should enable greater energy savings. We now present two cases that are different from the configuration considered so far. In the first case, we consider the ability to (de)activate TRXs in pairs, i.e., a site may be in one of three configurations at a given time: 6+6+6, 4+4+4 or 2+2+2. In the second case, we consider the ability to (de)activate each TRX on a site independently, i.e., at a given time, a site may be in one of six possible configurations.

In this scenario, we conducted simulation experiments where a re-optimization was performed every six minutes using model 1 BTS. The results of these experiments are given in Table 3. For all three BTS models, we see that going from a 2-state model to a 3-state model gives a relatively small increase in energy savings compared to the jump from 3-state to 6-state model.

5.3. Performance of heuristic algorithms

We also ran experiments for each BTS model in which the electricity cost for the optimal as well as the heuristic algorithms (Algorithm 1 and 2) was computed. We assessed the performance of our heuristics by computing the difference (error) in the electricity cost of the two solutions. For statistical significance, we computed the error in our heuristic relative to the optimal solution over 48 different experiment runs for each BTS model. The resulting CDF of the heuristic error (in Wh) is plotted in Fig. 5. We can see in Fig. 5 that our heuristic algorithm 1 is quite close to the optimal solution most of

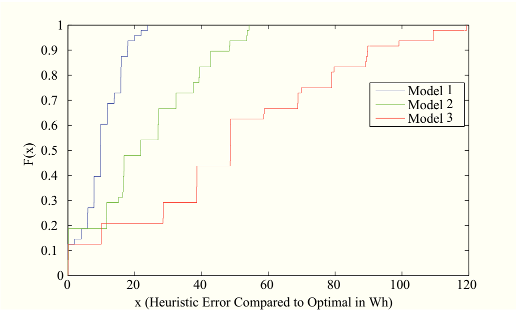


Figure 5: Empirical CDF of the difference between the cost offered by our heuristic compared to the optimal

the time, especially for the Model 1 and Model 2 BTS. For Model 3 BTS, although the error is comparatively larger, but since the amount of savings with the optimal solution is quite high (Fig. 4(a)), the heuristic will still result in significant energy savings.

5.4. Sensitivity to the value of ϵ

If the value of ϵ in our optimization is set too aggressively, a BTS that is placed in low-power mode may have to be moved back to high-power mode soon afterwards due to short time scale variations in call volume. Theoretically, it is even possible that there may be several such back and forth transitions at a BTS. Such rapid state oscillations may be undesirable and to avoid these, the value of ϵ must be set at a safe value. Furthermore, if ϵ is set too aggressively, a BTS placed in low-power mode would be operating very close to it's *new* and lower traffic capacity. If several calls arrive in a short time window, the BSC may not have sufficient time to bring the BTS back into high-power mode and, thus, some calls may be blocked. However, if ϵ is set too conservatively, the energy savings would be smaller.

We carried out experiments to assess the impact of the value of ϵ on the energy savings achievable through RED-BL. For this purpose, we fixed the inter-optimization interval at 6 minutes and carried out RED-BL optimizations for all three BTS models. Furthermore, we considered a two-state BTS model, i.e., a BTS may be placed in either a 6+6+6 or a 2+2+2 configuration. The range of possible values for epsilon were 5, 10, 15 and 25. Since the traffic capacity of a 2+2+2 BTS is 44[1], any larger value for ϵ did not

[1]The capacity of the 2+2+2 BTS is $3\times2\times8 = 48$, but 4 channels were reserved by the

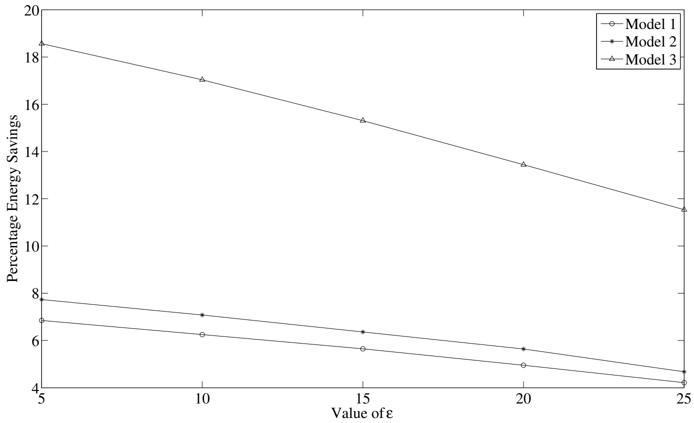


Figure 6: The percentage energy savings for all three BTS models considered in this paper vs the value of ϵ, with a six minute inter-optimization interval

make sense. Figure 6 shows the results. As expected, the percentage savings deplete almost linearly with increasing values of ϵ.

6. Conclusions

BTSs account for most of a cellular network's energy consumption. Motivated by the non load-proportionality of BTS energy consumption, prior work proposed shutting down some BTSs when traffic is low. However, network operators are reluctant to do so for a variety of reasons.

To reduce energy consumption, we propose using a commonly available and used feature called BTS power savings that deactivates some TRXs at BTSs that have low traffic. Furthermore, calls may be handed-off from BTSs with higher load to neighboring ones with lighter load to increase the benefits of BTS power-saving.

Using real network topology and traffic traces in a simulation study , we found that merely using BTS power saving in an urban setting can result in considerable energy savings. Moreover, our results also indicate that periodic call-shuffling between BTSs can further reduce energy consumption in existing large GSM networks.

---

operator for control and broadcast channels.

References

[1] E. Oh, B. Krishnamachari, X. Liu, Z. Niu, Toward dynamic energy-efficient operation of cellular network infrastructure, in: IEEE Communications Magazine, 2011.

[2] J. T. Louhi, Energy efficiency of modern cellular base stations, in: IEEE INTELEC '07, IEEE, New York, NY, USA, 2007, pp. 475–476.

[3] C. Peng, S.-B. Lee, S. Lu, H. Luo, H. Li, Traffic-driven power saving in operational 3g cellular networks, in: ACM MobiCom '11, 2011, pp. 121–132.

[4] H. He, J. Sun, Y. Wang, Q. Liu, J. Yuan, Cluster analysis based switching-off scheme of base stations for energy saving, Journal of Networks 7 (3) (2012) 494–501.

[5] J. Lorincz, T. Garma, G. Petrovic, Measurements and modelling of base station power consumption under real traffic loads, Sensors 12 (4) (2012) 4281–4310.

[6] N. S. Networks, Flexi BSC (BSC14), `http://bit.ly/QMD8O5` (June 2011).

[7] M. Ilyas, G. Baig, M. Qureshi, Q. Nadeem, A. Raza, M. Qazi, B. Rassool, Low-carb: Reducing energy consumption in operational cellular networks, in: Global Communications Conference (GLOBECOM), 2013 IEEE, 2013, pp. 2568–2573. doi:10.1109/GLOCOM.2013.6831461.

[8] R. Jeyarani, N. Nagaveni, R. Vasanth Ram, Design and implementation of adaptive power-aware virtual machine provisioner (apa-vmp) using swarm intelligence, Future Gener. Comput. Syst. 28 (5) (2012) 811–821. doi:10.1016/j.future.2011.06.002.
URL `http://dx.doi.org/10.1016/j.future.2011.06.002`

[9] Y. Chen, A. Das, W. Qin, A. Sivasubramaniam, Q. Wang, N. Gautam, Managing server energy and operational costs in hosting centers, in: ACM SIGMETRICS, ACM, New York, NY, USA, 2005, pp. 303–314.

[10] M. Mazzucco, D. Dyachuk, R. Deters, Maximizing cloud providers revenues via energy aware allocation policies, CoRR abs/1102.3058.

[11] F. Y.-K. Oh, H. S. Kim, H. Eom, H. Y. Yeom, Enabling consolidation and scaling down to provide power management for cloud computing, USENIX HotCloud, USENIX Association, Berkeley, CA, USA, 2011, pp. 14–14.

[12] J. S. Chase, D. C. Anderson, P. N. Thakar, A. M. Vahdat, R. P. Doyle, Managing energy and server resources in hosting centers, SIGOPS Oper. Syst. Rev. 35 (5) (2001) 103–116.

[13] H. Al-Daoud, I. Al-Azzoni, D. G. Down, Power-aware linear programming based scheduling for heterogeneous computer clusters, Future Generation Computer Systems 28 (5) (2012) 745 – 754, ¡ce:title¿Special Section: Energy efficiency in large-scale distributed systems¡/ce:title¿. doi:10.1016/j.future.2011.04.001.
URL `http://www.sciencedirect.com/science/article/pii/S0167739X11000525`

[14] Z. Fang, L. Zhao, R. R. Iyer, C. F. Fajardo, G. F. Garcia, S. E. Lee, B. Li, S. R. King, X. Jiang, S. Makineni, Cost-effectively offering private buffers in SoCs and CMPs, in: Proceedings of the international conference on Supercomputing, ICS '11, ACM, New York, NY, USA, 2011, pp. 275–284.

[15] F. Javed, N. Arshad, On the use of linear programming in optimizing energy costs, in: 3rd IWSOS, Springer-Verlag, Berlin, Heidelberg, 2008, pp. 305–310.

[16] T. Logenthiran, D. Srinivasan, A. M. Khambadkone, Multi-agent system for energy resource scheduling of integrated microgrids in a distributed system, Electric Power Systems Research 81 (1) (2011) 138 – 148.

[17] G. Celli, F. Pilo, Optimal distributed generation allocation in mv distribution networks, in: 22nd IEEE PICA 2001, 2001, pp. 81 –86.

[18] F. Javed, N. Arshad, Adopt: An adaptive optimization framework for large-scale power distribution systems, IEEE International Conference on Self-Adaptive and Self-Organizing Systems (2009) 254–264.

[19] M. A. Marsan, L. Chiaraviglio, D. Ciullo, M. Meo, A simple analytical model for the energy-efficient activation of access points in dense wlans,

in: Proceedings of the 1st International Conference on Energy-Efficient Computing and Networking, e-Energy '10, ACM, New York, NY, USA, 2010, pp. 159–168.

[20] C. Cavdar, A. Yayimli, L. Wosinska, How to cut the electric bill in optical wdm networks with time-zones and time-of-use prices, in: Optical Communication (ECOC), 2011 37th European Conference and Exhibition on, 2011, pp. 1 –3.

[21] S. Lee, S. Sahu, Efficient server consolidation considering intra-cluster traffic, in: GLOBECOM, 2011, pp. 1–6.

[22] E. Pinheiro, R. Bianchini, E. V. Carrera, T. Heath, Load balancing and unbalancing for power and performance in cluster-based systems, in: Compilers and Operating Systems for Low Power, 2001.

[23] Y. Yao, L. Huang, A. Sharma, L. Golubchik, M. Neely, Data centers power reduction: A two time scale approach for delay tolerant workloads, in: INFOCOM, 2012 Proceedings IEEE, 2012, pp. 1431 –1439.

[24] H. Herodotou, H. Lim, G. Luo, N. Borisov, L. Dong, F. B. Cetin, S. Babu, Starfish: A self-tuning system for big data analytics, in: 5th Biennial Conference on Innovative Data Systems Research (CIDR) 2011, 2011, pp. 261 –272.

[25] H. Herodotou, F. Dong, S. Babu, No one (cluster) size fits all: automatic cluster sizing for data-intensive analytics, in: 2nd ACM SOCC, ACM, New York, NY, USA, 2011, pp. 18:1–18:14.

[26] D. Aikema, R. Simmonds, Electrical cost savings and clean energy usage potential for hpc workloads, in: Sustainable Systems and Technology (ISSST), 2011 IEEE International Symposium on, 2011, pp. 1 –6.

[27] C. Peng, S.-B. Lee, S. Lu, H. Luo, H. Li, Traffic-driven power saving in operational 3G cellular networks, in: 17th MobiCom, ACM, New York, NY, USA, 2011, pp. 121–132.

[28] D. T. A. Rezgui, Krishnamurthy, P. Pacharintanaku, Dimming cellular networks, in: IEEE GLOBECOM'10, 2010.

[29] O. Blume, Energy savings in mobile networks based on adaptation to traffic statistics, Bell Labs Technical Journal 15 (2) (2010) 77–94.

[30] M. Gerla, J. T.-C. Tsai, Multicluster, mobile, multimedia radio network, Wireless Networks 1 (3) (1995) 255–265.

[31] J. Chabarek, J. Sommers, P. Barford, C. Estan, D. Tsiang, S. Wright, Power awareness in network design and routing, in: In Proc. IEEE INFOCOM, 2008.

[32] S. Mbakwe, M. T. Iqbal, A. Hsaio, Design of a 1.5kw hybrid wind/photovoltiac power system for a telecoms base station in remote location of benin city nigeria, in: IEEE NECEC, 2011.

[33] Ericsson, Vodafone Germany first to launch Ericsson's power-saving feature to reduce energy consumption and cut CO_2 emissions, `http://bit.ly/Sls1v1` (December 2007).

[34] Mobilink, Mobilink Network, `http://bit.ly/UWXYzB` (September 2012).

[35] Pakistan Telecommunication Authority, Pakistan Telecommunication Authority, 2011 Annual Report, `http://bit.ly/S8YtWS` (December 2011).

[36] Wikipedia, Telenor Pakistan - Wikipedia, `http://bit.ly/R1Uyqm` (September 2012).