Distributed Base Station Activation for Energy-Efficient Operation of Cellular Networks*

Ali Abbasi
Department of Computer Science
University of Calgary
aabbasi@ucalgary.ca

Majid Ghaderi
Department of Computer Science
University of Calgary
mghaderi@ucalgary.ca

ABSTRACT

Dynamic base station activation (DBA) has recently emerged as a viable solution for reducing energy consumption in cellular networks. While most of the works on this topic focused on centralized decision making algorithms, in this paper we investigate distributive solutions. These solutions are particularly desirable due to importance of self-organization and self-optimization in future cellular networks. The goal of DBA is to achieve an optimal trade-off between network operator's revenue and operational cost while guaranteeing coverage for network users. The problem is posed as a network utility maximization aiming to find the optimal activation schedule of each base station. Using Lagrangian duality, the problem is decomposed into smaller subproblems, where each subproblem is solved locally at its associated base station. Controlled message passing among base stations ensures convergence to the global optimal solution. Moreover, this general solution is further extended to capture the combinatorial nature of DBA. Finally, numerical results are provided to demonstrate the behavior of our solution in terms of utility and cost trade-off and convergence in some example network scenarios.

Categories and Subject Descriptors

C.2.3 [**Network Operations**]: Network management

General Terms

Algorithms, Performance

Keywords

Energy Efficiency, Cellular Networks, Network Utility Maximization

*This work was supported in part by the NSERC Smart Applications on Virtual Infrastructure (SAVI) Strategic Network grant.

MSWiM'13, November 3–8, 2013, Barcelona, Spain.

ACM 978-1-4503-2353-6/13/11
10.1145/2507924.2507961 ...$15.00.

1. INTRODUCTION

With the proliferation of smartphones, cellular service providers have witnessed massive increase of data traffic in their networks. For example, one cellular operator in the US reported 8000% growth in data traffic from 2007 to 2010 [3]. It is expected that the cellular data traffic will continue to grow and will increase 13-fold over the next 5 years [4]. The increase in data traffic has put high pressure over cellular operators to enhance their network capacity by adopting new solutions and technologies [2]. Most notably, operators are moving towards denser deployment of base stations (BSs). In dense deployments, each base station covers a small geographical area and serves a small number of users which allows it to provide them with higher data rates.

While with the dense deployment of base stations, better data service is provided to users, more energy is also consumed to activate different parts of the radio access network. As a result, in addition to initial investment to expand the network capacity, higher operational expenditure, *e.g.*, electricity bill, is also a challenge that need to be considered by network operators. The issue is more highlighted to note that among various elements of a cellular network, base stations account for 60-80% of the total network energy consumption. With current base stations, 80-90% of peak energy[1] is consumed even in idle or low traffic state [17]. This energy is spent in cooling system and idle-mode signaling and processing as soon as the base station is powered on. In other words, current base stations are not *energy-proportional*[2]. However as we discuss later, by taking the characteristics of cellular traffic into consideration, energy-proportional base station operation can be emulated in the network, albeit at a coarser granularity.

Measurement studies have shown that cellular traffic exhibits periodic fluctuations both in time and space [20]. This behavior can be attributed to different usage patterns during days and nights, weekdays and weekends, and across residential and business areas. Nevertheless, cellular operators often deploy as many base stations as necessary to satisfy the peak traffic demand, while keeping them active (*i.e.*, in the On state) all the time. While it might seem that energy consumption issue could be addressed via dynamic transmission power control in base stations, the benefits of such mechanisms are marginal due to the mentioned sources of energy consumption in BS equipment (*e.g.*, cooling system).

[1]Energy consumed during the peak traffic load.
[2]Base stations in which energy consumption is proportional to the amount of traffic passing through.

Recent analysis of cellular traffic [19,22] has demonstrated that by dynamically activating base stations in a network, significant energy savings can be achieved. The idea is to completely power off underutilized base stations when their traffic load could be handled by nearby base stations. In turn, when the load in some parts of the network exceeds the capacity of the current active base stations, power on some inactive base stations so as to satisfy the demand. In this paper, we consider distributed base station activation problem in the context of a densely-deployed network. Specifically, we develop distributed mechanisms that explore the high coverage overlap in the network [11] to select an active set of base stations that results in the optimal matching between the provisioned network capacity and traffic load.

To this end, a measure is needed for evaluating the quality of a particular active set of base stations with respect to network users' satisfaction. A common approach to achieve such a measure, is to model user's satisfaction via the well-established notion of *utilities*. In this framework, a utility function is associated to each user, where the user utility is an increasing function of the amount of resources allocated to the user. Then, base station activation problem is modeled following the framework of *Network Utility Maximization* (NUM) [21]. There are however some subtle differences between our problem and classic NUM problems as follows. First, traffic requirements of cellular users are different from typical Internet users. Consequently the types of utility functions considered in our work are different from those considered in classic NUM problems. Second, due to the On/Off operation of base stations, this problem has a combinatorial nature, which adds to the complexity of the problem.

There are only a few recent works on this subject. In [20] a location-dependent traffic profiling study is conducted on real 3G network traces showing that 23-53% energy saving is possible via dynamic base station activation. Operators cooperation is investigated in [16], where optimal switch-off frequencies of base stations are computed in order to achieve balanced energy savings and roaming costs. Assuming a sinusoidal traffic profile, and following a threshold-based activation rule by each base station, an analysis of achievable energy savings is provided in [19]. The closest works to ours are presented in [25] and [22]. In [25], centralized and heuristic methods are presented for finding and deactivating the base station with the lowest load. The joint problem of base station activation and user association is studied in [22], where the objective is to minimize a joint energy and delay cost function. What distinguishes our work from the aforementioned works is that unlike them, we develop *distributed algorithms* that do not rely on a centralized controller.

Having a centralized decision maker, each base station is required to report its local state including load distribution in its coverage area and average channel gains to the central controller. These local states, collectively, define the state of the network. As network state changes over time, regular feedback is required to keep the controller informed about it. Thus, exchanging the state information between base stations and the controller imposes a high communication and computation overhead on the backhaul links, which is highly undesired in cellular networks [6]. A simpler scheme (in terms of the communication overhead) can be designed by delegating the role of the centralized controller to the base stations, as described in this paper. The distributed algorithm lets each base station decide on its activation/deactivation schedule by communicating only with its neighboring base stations without the need to communicate with a single centralized controller.

Our contributions in this paper can be summarized as follows:

- The DBA problem is formulated as a NUM problem with the objective of finding an optimal trade-off between operator revenue and network operational cost. Characteristics of cellular data traffic are considered in the formulation.

- The formulated problem features many local optima thus does not adhere to a distributed solution easily. Accordingly, an appropriate transformation is introduced in the objective function to make it strictly concave.

- A distributed algorithm is devised based on the Lagrangian dual decomposition, in which each base station computes its optimal activation probability through message exchange with its neighbors. In addition, each base station independently decides to activate or deactivate itself based on a locally computed activation probability.

- The DBA problem formulation is extended to more appropriately reflect the On/Off feature of the problem by considering general BS activation cost functions. The solution to the extended problem is obtained via difference of convex functions programming [24].

The rest of the paper is organized as follows. Section 2 describes the system model considered in the paper. The problem is formulated in section 3. Our distributed algorithm is presented in section 4. Sample numerical results are provided in section 6. Finally, section 7 concludes this paper.

2. SYSTEM MODEL AND ASSUMPTIONS

2.1 Network Model

The system considered in this paper includes the radio access part of a cellular network which consists of a number of base stations that collectively provide coverage for the network. The area under the coverage of the base stations is discretized to a set of locations. Let $\mathcal{B} = \{b_1, \ldots, b_n\}$ and $\mathcal{L} = \{l_1, \ldots, l_m\}$ denote the set of base stations and the set of locations respectively. Activating each base station b_i incurs a cost of c_i ($c_i \geq 0$) which is determined based on the total energy consumed at the base station while it is powered on. It is assumed that c_i is constant as the total power consumed at a base station is relatively constant (recall that base stations are not energy-proportional). Let $\mathcal{L}_i$ denote the subset of locations that can be covered by base station b_i (The notion of coverage will be clarified in the next section). Let $\mathcal{B}_j$ denote the set of all base stations that cover location l_j. Furthermore, let $\mathcal{S}_i$ denote the set of locations that are associated to b_i. By association of l_j to b_i, we mean that b_i is responsible for determining the amount of resources allocated to l_j from every BS in $\mathcal{B}_j$. We assume that a location might receive service from multiple BSs but it is associated to only one of them based on *e.g.*, proximity. The set of neighbors of base station b_i, denoted by $\mathcal{N}_i$, is defined as the set of all base stations that cover at least one common location with b_i, *i.e.*, $\mathcal{N}_i = \{b_j | \mathcal{L}_i \cap \mathcal{L}_j \neq \emptyset\}$.

Time is divided into scheduling epochs. At the end of each epoch, a subset $\mathcal{A}$ of base stations is selected for activation in the next scheduling epoch. Decisions are made only based on downlink traffic which is readily available to BSs[3]. Expectedly, $\mathcal{A}$ should match the traffic load of $\mathcal{L}$, while incurring the minimal total cost. The cost of activating the base station set $\mathcal{A}$ is given by $C(\mathcal{A}) = \sum_{b_i \in \mathcal{A}} c_i$.

Frequently switching base stations between On and Off states consumes energy, takes time, and generates huge amount of signalling traffic to re-associate users. To avoid unnecessary overhead due to transient network states, user association and resource allocation are carried out on a *different timescale* than the base station activation. The former is done at a *fast* timescale (*e.g.*, every few milli-seconds) based on current users' channel states, while the latter is done at a much *slower* timescale (*e.g.*, every few minutes) based on average network information, *i.e.*, capacity and demand, in the next scheduling epoch.

2.2 Resource Allocation

As implemented in the current 3GPP LTE systems [1], the radio access interface is assumed to be based on OFDMA. In OFDMA systems, available frequency bandwidth is partitioned into orthogonal subchannels which are the allocable resources to users. The total base station power P is divided between these subchannels. In our model, instead of dealing with individual users, we consider *locations*, where multiple users can be present in a single location. We further assume that channel variations across the channels allocated to a location are negligible as our model only considers *long-term* average rates that can be achieved over subchannels, as opposed to instantaneous rates that depend on short-term channel fluctuations.

Assume that base station b_i uses a subchannel to communicate with a user at location l_j. Let p_{ij} and g_{ij} denote the transmission power and long-term average power gain of the subchannel. Then, the average [4] received signal power at location l_j is given by $g_{ij} \cdot p_{ij}$. A location is considered covered by a base station if the received power of the pilot signal at that location is higher than a prespecified threshold. Let r_{ij} denote the received rate at location l_j. Using the Shannon capacity formula, we have $r_{ij} = \log(1 + \beta \frac{g_{ij} p_{ij}}{n_j + I_j})$, where β is the SINR gap due to limited modulation. Also, n_j and I_j denote background noise power and interference power at location l_j respectively. Similar to [18], to make the formulation tractable, we use an upper bound I on the interference power instead of using the exact I_j. I is the maximum multi-cell interference temperature or (maximum tolerable interference level) as mentioned in [9]. Doing so, we achieve a conservative rate function that does not need substantial amount of signaling to compute the actual value of interference at each location. We further borrow some simplifying assumptions from the literature [13, 15] given as follows:

- We assume that each subchannel can be fractionally shared among users [13]. This is indeed the case in OFDMA systems such as WiMax and LTE as each subchannel is shared among multiple users using TDMA, *i.e.*, resources are shared in frequency and time. The assumption is particularly true for our model as we consider the long-term system averages.
- We assume the total transmission power P is divided equally among all subchannels. Therefore, if there exists R subchannels then the allocated subchannel power is $p = P/R$. As shown in [13], this scheme is nearly-optimal.
- As provisioned in LTE networks [15], we assume that neighboring cells are able to coordinate allocation of resources to users in overlapping regions such that orthogonal resources are allocated from neighboring BSs to locations in overlapping regions.

[3]The model can be extended to take into consideration the uplink traffic as well.
[4]The term 'average' is omitted hereafter without ambiguity.

In the rest of the paper, we use the term *resource* to refer to the subchannels at a base station. Let γ_{ij} denote the fraction of resources allocated to location l_j from base station b_i. Following the third assumption, the total received rate at location l_j is given by:

$$r_j = \sum_{b_i \in \mathcal{B}_j} \gamma_{ij} \cdot R_{ij}, \quad (1)$$

where, R_{ij} is the rate received from base station b_i if all of its resources were to be allocated to location l_j, *i.e.*,

$$R_{ij} = R \cdot \log(1 + \beta \frac{p g_{ij}}{n_j + I}) \,. \quad (2)$$

Consider a set $\mathcal{A}$ of active base stations. Let $\boldsymbol{r} = [r_j]_{l_j \in \mathcal{L}}$ denote a vector of rates achievable at location set $\mathcal{L}$. Define the set of all rate vectors achievable at location set $\mathcal{L}$ by the active base station set $\mathcal{A}$ as the *rate region* of $\mathcal{A}$, which is denoted by $\mathcal{R}_\mathcal{A}$. We then have

$$\mathcal{R}_\mathcal{A} = \Big\{ \boldsymbol{r} = [r_j] : \sum_{b_i \in \mathcal{A}} \gamma_{ij} R_{ij} = r_j, \sum_{l_j \in \mathcal{L}_i : b_i \in \mathcal{A}} \gamma_{ij} \leq 1 \Big\} . \quad (3)$$

2.3 User Traffic and Utility

Cellular data traffic is a mix of elastic and inelastic traffic. Cellular networks support real-time audio and video streaming applications that require certain minimum rate guarantees to function properly. There are also applications such as file transfer that do not have this constraint. This mix of elastic and inelastic applications can be modeled as *rate-adaptive* applications with a minimum rate requirement. In this work, we consider α-*critical* functions to model utility of this type of applications. In α-critical functions [8], the function value is 0 before a certain threshold. Once the threshold is met, the increase of the utility value is based on a concave function. Moreover, we combine user utilities at each location and represent the aggregate utility as a single function. Specifically, at each location l_j, the utility function $U_j(\cdot)$ is defined as follows:

$$U_j(x_j) = u_j([x_j - d_j]^+), \quad (4)$$

where $u_j(\cdot)$ is an increasing concave function, x_j is the rate received at location l_j, d_j is the base demand at l_j, and $[x]^+ = \max(0, x)$. For an active base station set $\mathcal{A}$, the *system utility* of $\mathcal{A}$ is defined as follows:

$$U(\mathcal{A}) = \max_{\boldsymbol{x} \in \mathcal{R}_\mathcal{A}} \sum_{l_j \in \mathcal{L}} U_j(x_j) \,. \quad (5)$$

For an active set $\mathcal{A}$, the *net utility* $N(\mathcal{A})$ is defined as the difference between its system utility and cost, *i.e.*,

$$N(\mathcal{A}) = U(\mathcal{A}) - C(\mathcal{A})\,. \tag{6}$$

3. PROBLEM FORMULATION

To be able to make optimal activation decisions, each base station needs to know the state of its neighbors a priori. Due to inter-dependence of all BS decisions, this information cannot be made available to all BSs at the same time. Therefore, in our formulation, the next likely configuration of each base station is computed and propagated to its neighbors. To this end, an activation probability is assigned to each base station. The probabilities are assigned so that the expected net utility is maximized. To find the optimal activation probabilities without the need for a centralized controller, we design a distributed iterative message passing algorithm among base stations. Overall, the base station activation works as follows:

1. In each scheduling epoch, every base station iteratively solves a local problem to find its activation probability and exchanges its results with its neighbors until a stable solution is found.

2. At the end of a scheduling epoch, each base station decides to become active or inactive based on the optimal activation probabilities computed via the iterative algorithm of step (1).

Let α_i denote the activation probability associated with base station b_i. Accordingly, the net utility (6) maximization problem is given by

$$\max_{\boldsymbol{\alpha}} \sum_{b_i} U_i(\alpha_i,\, \{\alpha_{i'}\}_{i' \in \mathcal{N}_i}) - \alpha_i c_i, \tag{7}$$

where $U_i(\cdot,\cdot)$ is the utility of base station b_i, which depends on its activation probability as well as the activation probabilities of its neighbors, *i.e.*, $\{\alpha_{i'}\}_{i' \in \mathcal{N}_i}$. In (7), both system utility and system cost are expressed as functions of activation probabilities. We intend to express both system utility and cost as functions of allocated resources. Let y_{ij} denote the fraction of b_i resources allocated to l_j conditioned on the activation of b_i. Then, the expected fraction of resources allocated to l_j from b_i is given by $x_{ij} = \alpha_i y_{ij}$. It follows that

$$\sum_{l_j \in \mathcal{L}_i} \alpha_i y_{ij} = \sum_{l_j \in \mathcal{L}_i} x_{ij}\,. \tag{8}$$

When a base station is activated, all its available resources are allocated to users, *i.e.*, $\sum_j y_{ij} = 1$. Thus, the relation between α_i and x_{ij}'s is given by

$$\alpha_i = \sum_j x_{ij}\,. \tag{9}$$

Utility of base station b_i is defined as the sum of location utilities for all locations under its coverage assuming fixed rate contribution from its neighbors to each location. In other words, $U_i(.,.)$ is defined as follows

$$\begin{aligned} U_i(\alpha_i, \{\alpha_{i'}\}_{i' \in \mathcal{N}_i}) &= \sum_{l_j \in \mathcal{L}_i} u_j(x_{ij}, \{x_{i'j}\}_{b_{i'} \in \mathcal{B}_j \setminus b_i}) \\ &= \sum_{l_j \in \mathcal{L}_i} u_j(x_{ij} r_{ij} + \sum_{b_{i'} \in \mathcal{B}_j \setminus b_i} x_{i'j} r_{i'j}) \end{aligned} \tag{10}$$

Accordingly, we can express (7) as follows

$$\max_{\boldsymbol{x}} \sum_{l_j \in \mathcal{L}} u_j\big(\sum_{b_i \in \mathcal{B}_j} x_{ij} r_{ij} - d_j\big) - \sum_{b_i \in \mathcal{B}} \big(\sum_{l_j \in \mathcal{L}_i} x_{ij}\big) c_i\,. \tag{11}$$

In addition to maximizing the net utility, the minimum rate demand should be satisfied at each location as well. Thus, the optimization problem (11) has the following set of constraints

$$\sum_{b_i \in \mathcal{B}_j} x_{ij} r_{ij} \geq d_j,\ \forall l_j \in \mathcal{L}\,. \tag{12}$$

There is also another set of constraints on the resources available to BSs as

$$\sum_{l_j \in \mathcal{L}_i} x_{ij} \leq 1,\ \forall b_i \in \mathcal{B}\,. \tag{13}$$

Problem (11) is a convex optimization problem that can be solved efficiently using interior-point methods, however, applying dual decomposition techniques on (11) is not straightforward since these methods need the objective to be strictly concave. Although the objective function of (11) is strictly concave w.r.t resources allocated from a specific BS, *e.g.*, b_i to l_j, it is not w.r.t the total received rate (when $\sum_{b_i \in \mathcal{B}_j} x_{ij} r_{ij}$ is fixed the function is constant). This may result in indifferentiability of the dual of (11) at some points. To circumvent the issue, we apply the *proximal point* technique in which a quadratic term is added to the objective function for every variable x_{ij}. Doing so, the objective function is modified as follows

$$\begin{aligned} &\sum_{l_j \in \mathcal{L}} u_j\big(\sum_{b_i \in \mathcal{B}_j} x_{ij} r_{ij} - d_j\big) - \sum_{l_j \in \mathcal{L}} \sum_{b_i \in \mathcal{B}} \frac{e_j}{2}(x_{ij} - y_{ij})^2 \\ &+ \sum_{b_i \in \mathcal{B}} c_i\big(\sum_{l_j \in \mathcal{L}_i} x_{ij}\big) \end{aligned} \tag{14}$$

where $e_j > 0$. The optimal value of (14) coincides with the one from (11); that is if $\mathbf{x}^*$ is an optimal solution of (11), then $\mathbf{x} = \mathbf{x}^*, \mathbf{y} = \mathbf{x}^*$ is the optimal solution of (14). Problem (14) can be solved by Gauss-Seidel method in which the following steps are performed alternately. First, while keeping $\mathbf{y}$ fixed and considering constraints (12) and (13), $\boldsymbol{x}$ is found as follows

$$\begin{aligned} \mathbf{x(t+1)} = \max_{\boldsymbol{x}} &\sum_{l_j \in \mathcal{L}} u_j\big(\sum_{b_i \in \mathcal{B}_j} x_{ij} r_{ij} - d_j\big) \\ &- \sum_{l_j \in \mathcal{L}} \sum_{b_i \in \mathcal{B}} \frac{e_j}{2}(x_{ij} - y_{ij})^2 + \sum_{b_i \in \mathcal{B}} c_i\big(\sum_{l_j \in \mathcal{L}_i} x_{ij}\big)\,. \end{aligned} \tag{15}$$

Then $\boldsymbol{y}$ is updated according to

$$\mathbf{y(t+1)} = \mathbf{x(t+1)}\,.$$

It is easy to show that $\mathbf{x(t)} = \mathbf{x}^*$ as $t \to \infty$. Also, (15) can be solved using dual methods. Lagrangian of (15) is given by

$$\begin{aligned} L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = &\sum_{b_i \in \mathcal{B}} \Big(\sum_{l_j \in \mathcal{S}_i} \big(u_j\big(\sum_{b_k \in \mathcal{B}_j} x_{kj} r_{kj} - d_j\big) \\ &- \frac{e_j}{2}(x_{ij} - y_{ij})^2\big) - c_i\big(\sum_{l_j \in \mathcal{L}_i} x_{ij}\big)\Big) \\ &+ \sum_{l_j \in \mathcal{L}} \lambda_j\big(\sum_{b_i \in \mathcal{B}_j} x_{ij} r_{ij} - d_j\big) + \sum_{b_i \in \mathcal{B}} \nu_i\big(1 - \sum_{l_j \in \mathcal{L}_i} x_{ij}\big) \end{aligned} \tag{16}$$

where $\boldsymbol{\lambda} = [\lambda]_{|\mathcal{L}|}$ and $\boldsymbol{\nu} = [\nu]_{|\mathcal{B}|}$ are Lagrange multipliers for the sets of constraints (12) and (13) respectively. The associated Lagrange dual function for Lagrangian (16) is then expressed as

$$\max_{\boldsymbol{x}} L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = \sum_{b_i \in \mathcal{B}} \max_{\boldsymbol{x}} D_i(\boldsymbol{x}, \boldsymbol{\lambda}, \nu_i) - \sum_{l_j \in \mathcal{L}} \lambda_j d_j + \sum_{b_i \in \mathcal{B}} \nu_i \tag{17}$$

where D_i is given as follows

$$\begin{aligned} D_i(\boldsymbol{x}, \boldsymbol{\lambda}, \nu_i) = \sum_{l_j \in \mathcal{S}_i} \Big(& u_j\big(\sum_{b_k \in \mathcal{B}_j} x_{kj} r_{kj} - d_j \big) \\ & - \sum_{b_k \in \mathcal{B}_j} \frac{e_j}{2} (x_{kj} - y_{kj})^2 \\ & \sum_{b_k \in \mathcal{B}_j} \lambda_j x_{kj} r_{kj} - \sum_{b_k \in \mathcal{B}_j} x_{kj}(c_k + \nu_k) \Big) \end{aligned} \tag{18}$$

and represents the local subproblem that is solved at BS b_i. Then the dual of (15) is represented as

$$\min_{\boldsymbol{\lambda} \geq 0, \boldsymbol{\nu}} g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \, . \tag{19}$$

Since the objective function of (19) is differentiable at all locations based on Danskin's theorem [5], we have

$$\begin{aligned} \frac{\partial g}{\partial \lambda_j} &= \sum_{b_i \in \mathcal{B}_j} x_{ij} r_{ij} - d_j, \\ \frac{\partial g}{\partial \nu_i} &= 1 - \sum_{l_j \in \mathcal{L}_i} x_{ij} \, . \end{aligned} \tag{20}$$

Therefore, problem (15) can be solved by a gradient projection algorithm in which the dual variables are updated as follows

$$\begin{aligned} \lambda_j(t+1) &= \Big[\lambda_j(t) - \xi \big(\sum_{b_i \in \mathcal{B}_j} x_{ij} r_{ij} - d_j \big) \Big]^+, \\ \nu_i(t+1) &= \nu_i(t) - \xi \big(1 - \sum_{l_j \in \mathcal{L}_i} x_{ij} \big) \end{aligned} \tag{21}$$

where ξ is a sufficiently small step-size. In the next subsection, the message passing algorithm among base stations to implement the above solution is presented.

4. ALGORITHM DESIGN

The solution method presented in the previous section consists of an outer Gauss-Seidel loop and an inner gradient projection loop. In each iteration of the outer loop, before updating $\boldsymbol{y}$, it is assumed that the gradient projection algorithm has converged to its optimal solution. However, ensuring convergence of the inner loop in a distributed manner is cumbersome, so another algorithm is presented here in which the number of iterations of the inner loop is constant (in our case, it is limited to 1). The new algorithm is as follows

I. Fix $\boldsymbol{y} = \boldsymbol{y}(t)$. Assume $\boldsymbol{x}(t)$ is the primal variable that maximizes (17) given $\boldsymbol{y}$, $\boldsymbol{\lambda}(t)$, and $\boldsymbol{\nu}(t)$. Update dual variables $\boldsymbol{\lambda}(t+1)$ and $\boldsymbol{\nu}(t+1)$ according to (21).

II. Let $\boldsymbol{z}(t)$ be the point that maximizes (17) given the new dual variables $\boldsymbol{\lambda}(t+1)$ and $\boldsymbol{\nu}(t+1)$. Update $\boldsymbol{y}$ as follows

$$\boldsymbol{y}(t+1) = \boldsymbol{y}(t) + \tau(\boldsymbol{z}(t) - \boldsymbol{y}(t)) \tag{22}$$

where $\tau \in (0, 1]$. The algorithm is guaranteed to converge to the optimal solution. For details of the convergence, see [14]. Following the algorithm, each BS b_i operates as follows. Knowing r_{kj} for each neighboring BS b_k to all the locations $l_j \in \mathcal{S}_i$, it solves (18) and then sends the obtained primal variable x^*_{kj} to the corresponding BS b_k. Each BS b_k, updates ν_k according to (21) based on the received and locally-computed x^*_{kj}'s. Then it updates all its neighbors with the new ν_k. This allows b_i to compute $\boldsymbol{z}(t)$ and update $\boldsymbol{y}$. Following this procedure, the algorithm can be implemented in a completely distributed manner by message passing between neighboring base stations.

Now we would like to show how to find the optimal value of local optimization problem from the current set of Lagrange multipliers $\boldsymbol{\lambda}, \boldsymbol{\nu}$. According to Karush-Kuhn-Tucker (KKT) theorem, the stationary point of the Lagrangian, *i.e.*, the solution to $\nabla_{\boldsymbol{x}} L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = \mathbf{0}$, gives the unique solution to problem (17). Based on (18), we have

$$\begin{aligned} \frac{\partial D_i}{\partial x_{kj}} &= r_{kj} u'\big(\sum_{b_k \in \mathcal{B}_j} x_{kj} r_{kj} - d_j \big) - e_j (x_{kj} - y_{kj}) + \lambda_j r_{kj} \\ & - (c_k + \nu_k) \end{aligned} \tag{23}$$

then by setting $\frac{\partial D_i}{\partial x_{kj}} = 0$, for every x_{kj}, we obtain x^*_{kj}. Motivated by the definition of *weighted proportional fairness*, let us assume that the utility of each l_j is defined as

$$u_j\big(\sum_{b_k \in \mathcal{B}_j} x_{kj} r_{kj} - d_j \big) = w_j \ln\big(\sum_{b_k \in \mathcal{B}_j} x_{kj} r_{kj} - d_j \big) \, .$$

Then we have

$$\begin{aligned} \frac{\partial D_i}{\partial x_{kj}} = & \frac{w_j r_{kj}}{\sum_{b_k \in \mathcal{B}_j} x_{kj} r_{kj} - d_j} - e_j (x_{kj} - y_{kj}) \\ & + \lambda_j r_{kj} - (c_k + \nu_k) = 0 \end{aligned} \, . \tag{24}$$

Note that all the BSs that could provide service to l_j are related according to (24). Therefore, to compute the corresponding x^*_{kj}'s, both sides of (24) are multiplied by r_{kj} then summed over all $b_k \in \mathcal{B}_j$. This gives the following relation

$$\begin{aligned} & \frac{w_j \sum_{b_k \in \mathcal{B}_j} r^2_{kj}}{\sum_{b_k \in \mathcal{B}_j} x_{kj} r_{kj} - d_j} - e_j \sum_{b_k \in \mathcal{B}_j} x_{kj} r_{kj} \\ & + \sum_{b_k \in \mathcal{B}_j} \big(e_j y_{kj} r_{kj} + \lambda_j r^2_{kj} - (c_k + \nu_k) r_{kj} \big) = 0 \, . \end{aligned} \tag{25}$$

Define s_j, θ_j and μ_j as follows:

$$\begin{aligned} s_j &= \sum_{b_k \in \mathcal{B}_j} x_{kj} r_{kj} \\ \theta_j &= \sum_{b_k \in \mathcal{B}_j} \big(e_j y_{kj} r_{kj} + \lambda_j r^2_{kj} - (c_k + \nu_k) r_{kj} \big) \\ \mu_j &= w_j \sum_{b_k \in \mathcal{B}_j} r^2_{kj} \end{aligned}$$

Then, (25) is simplified as follows

$$\frac{\mu_j}{s_j - d_j} - e_j s_j + \theta_j = 0$$

which results in the following quadratic equation

$$-e_j s_j^2 + (e_j d_j + \theta_j) s_j + (\mu_j - \theta_j d_j) = 0$$

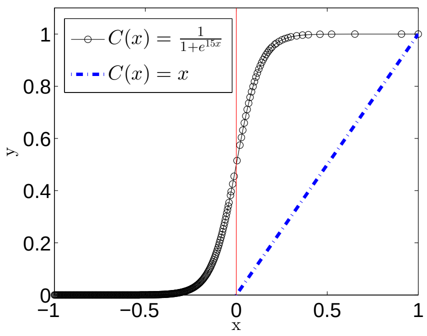


Figure 1: Comparison of different cost functions.

Solving the above equation, s_j is given by

$$s_j = \frac{(e_j d_j + \theta_j) + \sqrt{(e_j d_j + \theta_j)^2 + 4e_j(\mu_j - \theta_j d_j)}}{2e_j}. \quad (26)$$

After finding s_j, every x^*_{kj} can be computed by replacing the corresponding value of s_j in (25) as follows

$$x^*_{kj} = \Big[\frac{1}{e_j}\big(\frac{w_j r_{kj}}{s_j} + e_j y_{kj} + \lambda_j r_{kj} - (c_k + \nu_k)\big)\Big]_{[0,1]},$$

where $[x]_{\mathcal{S}}$ is the projection of x onto the set $\mathcal{S}$.

5. GENERAL CONCAVE COST

In previous section, activation probability of a base station is shown to be equal to the fraction of resources allocated to users from that base station. Also, cost of activating a BS was considered linearly dependent on its probability of activation (fraction of allocated resources). However, this model may not appropriately reflect the energy consumption at base stations which is highly dependent on energy consumed for cooling, *etc.* and not the transmission power. This base energy cost can be included in the model by considering concave cost functions instead of linear ones. Such functions feature steep slope at lower activation values and become less steep as they move toward higher values. This behavior approximately resembles the zero-one nature of the cost function. Specifically, *sigmoidal cost functions*[5] are considered which are defined as follows

$$C(\alpha) = \frac{c}{1 + e^{-d(\alpha - \alpha_0)}} \quad (27)$$

These functions are concave where $c, d > 0$ and $\alpha > \alpha_0$. A sigmoidal function along with a linear one are illustrated in Figure 1. Employing a concave cost function changes the problem (7) as follows

$$\max_{\alpha} \sum_{b_i} U_i(\alpha_i, \{\alpha_{i'}\}_{i' \in \mathcal{N}_i}) - C_i(\alpha_i), \quad (28)$$

where each $C_i(.)$ is an increasing, concave and continuously differentiable function in $[0, 1]$. Problem (28) is in fact the difference of two concave functions, so it is not a convex optimization problem. However, it can be tackled using convex-concave procedure (CCCP) [24]. In CCCP, instead of dealing with (28) directly, a sequence of convex programs is formulated. At each step the following problem is solved

$$\alpha_i^{l+1} \in \arg\max U_i(\alpha_i, \{\alpha_{i'}\}_{i' \in \mathcal{N}_i}) - \alpha \nabla C_i(\alpha_i^l), \quad (29)$$

The premise behind (29) is to replace the convex part of the objective, *i.e.*, $-C_i(\alpha_i)$, with an affine approximation to make the problem locally convex. To do so, first-order Taylor approximation of the cost function around the current solution, *i.e.*, α^l, is utilized. Then, the new one, *i.e.*, α^{l+1}, is obtained by solving (29). Since $C_i(.)$ is concave, we have

$$C_i(x) \le C_i(y) + (y - x)\nabla C_i(x),$$

for all $x, y \in [0, 1]$. Therefore, we have

$$U_i(\alpha_i^{l+1}) - C_i(\alpha_i^{l+1}) \ge U_i(\alpha^{l+1}) - C_i(\alpha^l) - (\alpha^{l+1} - \alpha_i^l)\nabla C_i(\alpha^l) \quad (30)$$

Since $\alpha^{l+1} \in \arg\max U_i(\alpha) - \alpha \nabla C_i(\alpha^l)$, we also have

$$U_i(\alpha^{l+1}) - \alpha^{l+1}\nabla C_i(\alpha^l) \ge U_i(\alpha^l) - \alpha^l \nabla C_i(\alpha^l)$$

which results in

$$\begin{aligned} U_i(\alpha_i^{l+1}) - C_i(\alpha_i^{l+1}) &\ge U_i(\alpha^l) - C_i(\alpha^l) - (\alpha^l - \alpha_i^l)\nabla C_i(\alpha^l) \\ &\ge U_i(\alpha^l) - C_i(\alpha^l) \end{aligned} \quad (31)$$

The above argument shows that the sequence of α's is ascent, therefore, (29) converges to a local maximum. See [23] for detailed discussion of the convergence of CCCP.

Adapting our previous solution to the new sigmoidal cost function is straightforward. The only change that is needed to be made is the value of c_k. In this case, c_k is not constant anymore but varies in each iteration. Specifically, the value of $c_k(t)$ depends on the fraction of resources allocated in the previous iteration as follows

$$c_k(t) = \nabla_x C(x(t)).$$

Assuming sigmoidal cost function (27) with inflection point of 0, c_k is given by

$$c_k(t) = \frac{cde^{dx(t)}}{(e^{dx(t)} + 1)^2} \quad (32)$$

in problem (18). In this case, each base station b_i, in addition to updating its neighbors with the current value of ν_i, needs to update them with the current value of $c_i(t)$ in each iteration.

6. NUMERICAL RESULTS

6.1 Simulation Parameters

We consider a network of size $1200m \times 1200m$. Base stations are placed on a regular grid. The distance between each two neighboring BSs is $200m$. There are a total of 25 base stations in the network. A base station is able to cover users which are up to $150m$ away from it.

The parameter values are adopted from [7] as the assumptions regarding channel gains are consistent with the standard 3GPP propagation models. The power gain between the sender and a receiver is $g = f(d)$ where d is the distance from the sender to the receiver in (km). $f(d) = 10^{h_0} d^{-\kappa}$ with path loss exponent $\kappa = 3.5$ and $h_0 = -14.4$. The background noise is $N_0 = -174$ dbm (Hz^{-1}). The bandwidth is 1 MHz and maximum power P varies from $P = 4W$, to $P = 16W$.

[5] An increasing function $f(x)$ is called a sigmoidal function, if it has one inflection point x_0, and $f''(x) > 0$ for all $x < x_0$ and $f''(x) < 0$, for all $x > x_0$.

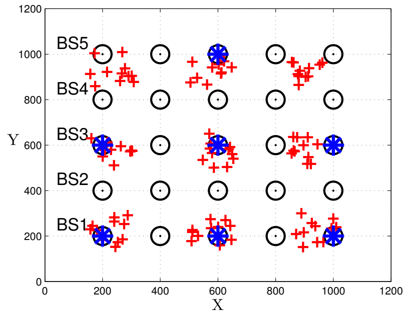


Figure 2: Snapshot of activated base stations ('$*$') in a network with non-uniform user distribution .

Our goal is to study the behavior of the distributed algorithm. Two scenarios for distribution of users in the network are considered: *uniform* and *non-uniform*. In the uniform case, location of a user is chosen uniformly at random in the network. To distribute users non-uniformly, nine crowded regions are considered in the network. Each crowded region is a circle of radius $160m$. Users are divided equally among theses crowded regions and distributed uniformly within each region.

6.2 Convergence of Distributed Algorithm

We have implemented the distributed Algorithm in MATLAB. In the first set of results, we would like to demonstrate convergence of the algorithm to the optimal solution. Users are distributed non-uniformly in the network. In Figure 2, snapshot of base station locations, user distribution and activated base stations is demonstrated. As it is apparent from the figure, activated base stations are located at more populated areas (higher utilities). All BSs maximum transmission power is $P = 4W$. Also, cost of BS activation is assumed to be the 250. Demand at all locations is 0, so some of them are not covered.

In 3, variation of activation probabilities of five nearby base stations (BSs in the left vertical line of the network snapshot) are shown during the course of the algorithm. Two of the aforementioned base stations are located in crowded regions, one of them is in a lightly-loaded region, and the other two in regions with no users. Activation probabilities of the BSs in crowded regions converge to 1, the base station

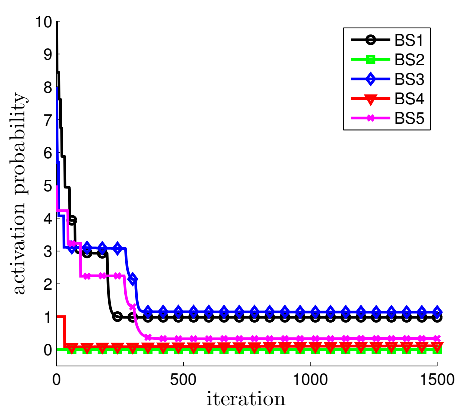


Figure 3: Convergence of BS activation probabilities with step-size $\xi = 0.02$.

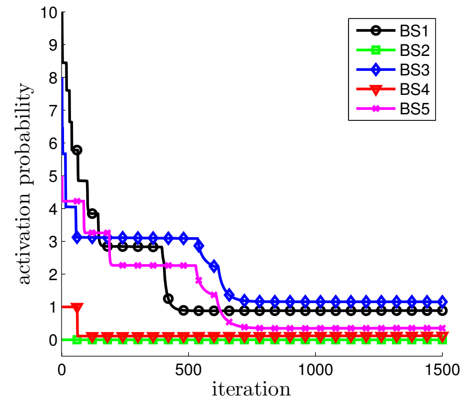


Figure 4: Convergence of BS activation probabilities with step-size $\xi = 0.01$.

in the lightly loaded region to almost 0, and the other two to 0. As demonstrated, the algorithm converges in about 350 iterations with step-size $\xi = 0.02$. To demonstrate the sensitivity of the algorithm to step-size selection, the convergence of probabilities are shown for another step-size $\xi = 0.02$ in figure 4. In this case, convergence happens in about iteration 700. This shows that the algorithm can be slow for lower values of ξ.

6.3 Effect of Utility and Activation Cost

To see the behavior of the distributed algorithm, both the activation cost and utilities have been varied. To increase the utilities, we increase the BS transmission power from $P = 4W$ to $P = 16W$. Figure 5 and 6 shows the number of activated base stations for different values of P and activation costs. As can be seen in the figure, increasing the utilities allows more base stations to be activated while increasing the activation cost reduces the number of active base stations. In addition, uniform user distribution associates approximately the same number of locations to base stations which results in similar utilities for them. In comparison to the non-uniform user distribution, this allows activating more base stations when the cost is low. The opposite is true when the cost is high.

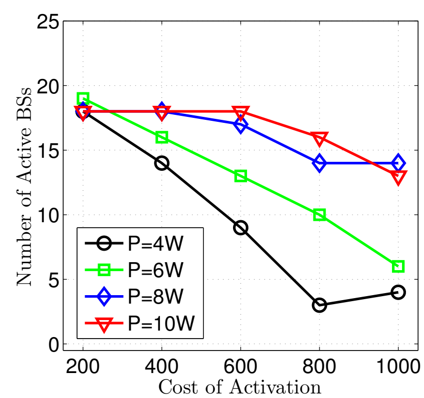
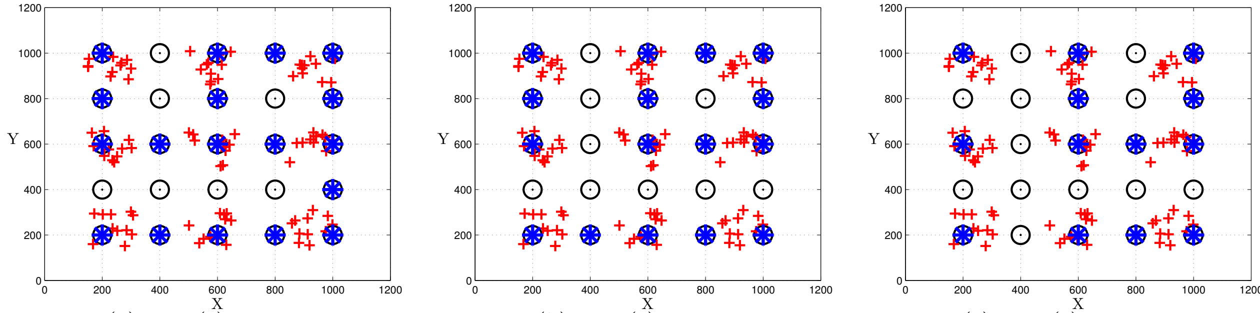


Figure 5: Number of active BSs versus transmission power and activation cost in a network with uniform user distribution.



(a) cost (c) = 600. (b) cost (c) = 800. (c) cost (c) = 1000.

Figure 7: Snapshots of network and activated base stations ('$*$').

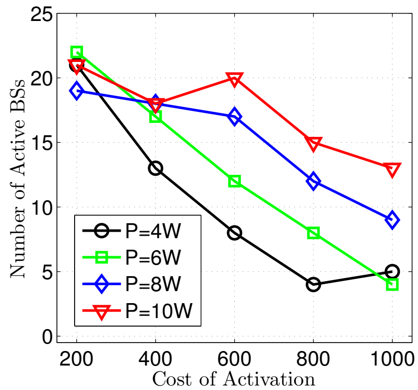


Figure 6: Number of active BSs versus transmission power and activation cost in a network with uniform user distribution.

What is interesting is the matching between base stations and crowded places which is demonstrated in Figure 7. The figure demonstrate snapshots of the network and the corresponding active base stations while increasing the activation cost. The snapshots are taken from the above results where $P = 16W$. As the snapshot highlights, although the algorithm is randomized, it results in a good match between populated areas and activated base stations.

7. CONCLUSION AND FUTURE WORK

Distributed algorithms to the base station activation problem are investigated in this paper. The presented algorithm is designed to find the optimal operating point of the network in terms of utility and cost while guaranteeing network coverage. In the algorithm, finding appropriate step-size to ensure fast convergence of the distributed algorithm is difficult. As a future work, we plan to investigate distributed algorithms based on resorting KKT conditions [12], which do not use step-size and seem to result in faster convergence. Also, in this work, a randomized rounding procedure is used to activate base stations after finding the optimal activation probabilities. Other rounding methods [10] might be considered to guarantee coverage of inelastic demands.

8. REFERENCES

[1] Long term evolution of the 3gpp radio technology. `http://www.3gpp.org/LTE/`.

[2] LTE-Advanced. `http://www.3gpp.org/article/lte-advanced`.

[3] AT&T SXSW Press Release. `http://www.att.com/Common/docs/SXSW_Network%20Fact_Sheet.doc`, Mar. 2011.

[4] Cisco visual networking index: Global mobile data traffic forecast update. `http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-520862.pdf`, Feb. 2013.

[5] D. P. Bertsekas. *Nonlinear Programming.* Athena Scientific, 1999.

[6] P. Bhat, S. Nagata, L. Campoy, I. Berberana, T. Derham, G. Liu, X. Shen, P. Zong, and J. Yang. LTE-Advanced: An Operator Perspective. *IEEE Commun. Mag.*, 50(2):104–114, 2012.

[7] S. Borst, M. Markakis, and I. Saniee. Distributed Power Allocation and User Assignment in OFDMA Cellular Networks. In *In Proc. Allerton Conference on Communication, Control, and Computing*, pages 1055–1063, Urbana, USA, Sep 2011.

[8] M. H. Cheung, A. H. Mohsenian-Rad, V. W. Wong, and R. Schober. Utility-Optimal Random Access for Wireless Multimedia Sensor Networks. *IEEE Commun. Lett.*, 1(4):340–343, 2012.

[9] FCC Spectrum Policy Task Force. Report of the Spectrum Ffficiency Working. `http://www.fcc.gov/sptf/reports.html`, Nov. 2002.

[10] R. Gandhi, S. Kuller, S. Parthasarathy, and A. Srinivisan. Dependent Rounding and its Applications to Approximation Algorithms. *Journal of the ACM*, 53(3):324–360, 2006.

[11] J. Gong, S. Zhou, Z. Niu, and P. Yang. Traffic-Aware Base Station Sleeping in Dense Cellular Networks. In *Proc. IEEE IWQoS*, pages 1–2, Beijing, China, Jun 2010.

[12] J. Huang, R. A. Berry, and M. L. Honig. Distributed Interference Compensation for Wireless Networks. *IEEE J. Sel. Areas Commun.*, 24(5):1074–1084, 2006.

[13] J. Huang, V. G. Subramanian, R. Agrawal, and R. A. Berry. Downlink Scheduling and Resource Allocation for OFDM Systems. *IEEE Trans. Wireless Commun.*, 8(1):288–296, 2009.

[14] X. Lin and N. B. Shroff. Utility Maximization for Communication Networks with Multipath Routing. *IEEE Trans. Automat. Contr.*, 51(5):766–781, 2006.

[15] D. Lopez-Perez, I. Guvenc, G. D. L. Roche, M. Kountouris, T. Q. S. Quek, and J. Zhang. ENHANCED INTERCELL INTERFERENCE COORDINATION CHALLENGES IN HETEROGENEOUS NETWORKS. *IEEE Wireless Commun. Mag.*, 18(3):22–30, 2011.

[16] M. A. Marsan and M. Meo. Energy Efficient Management of two Cellular Access Networks. *ACM SIGMETRICS Performance Evaluation Review*, 37(4):69–73, 2010.

[17] M. A. Marsan and M. Meo. Green Wireless Networking: Three Questions. In *Proc. IFIP Med-Hoc-Net*, pages 41–44, Sicily, Italy, Jun 2011.

[18] D. W. K. Ng and R. Schober. Resource Allocation and Scheduling in Multi-Cell OFDMA Systems with Decode-and-Forward relaying. *IEEE Trans. Wireless Commun.*, 10(7):2246–2258, 2011.

[19] E. Oh and B. Krishnamachari. Energy Savings Through Dynamic Base Station Switching in Cellular Wireless Access Networks. In *Proc. IEEE GlobeCom*, pages 1–5, Miami, USA, Dec 2010.

[20] C. Peng, S.-B. Lee, S. Lu, H. Luo, and H. Li. Traffic-Driven Power Saving in Operational 3G Cellular Networks. In *Proc. ACM MobiCom*, pages 121–132, Las Vegas, USA, Sep 2011.

[21] S. Shakkottai and R. Srikant. Network Optimization and Control. *Foundations and Trends® in Networking*, 2(3), 2007.

[22] K. Son, H. Kim, Y. Yi, , and B. Krishnamachari. Base Station Operation and User Association mechanisms for Energy-Delay Tradeoffs in Green Cellular Networks. *IEEE J. Sel. Areas Commun.*, 29(8):1525–1536, 2011.

[23] B. Sriperumbudur and G. Lanckriet. On the convergence of the concave-convex procedure. In *Proc. NIPS*, pages 1759–1767, Vancouver, Canada, Dec 2009.

[24] A. Yuille and A. Rangarajan. The Concave-Convex Procedure (CCCP). In *Proc. NIPS*, volume 2, pages 1033–1040, Vancouver, Canada, 2002.

[25] S. Zhouy, J. Gongy, Z. Yangy, Z. Niuy, and P. Yang. Green Mobile Access Network with Dynamic Base Station Energy Saving. In *ACM MobiCom posters*, Beijing, China, Sep 2009.