

Online Algorithms for Energy Cost Minimization in Cellular Networks

Ali Abbasi and Majid Ghaderi

Department of Computer Science, University of Calgary
{aabbasi, mghaderi}@ucalgary.ca

Abstract—Dynamic base station activation and transmission power control are the key mechanisms to reduce energy consumption in cellular networks. In this work, we consider employing these methods for the purpose of minimizing long-term energy cost in cellular networks. Based on the two-timescale Lyapunov optimization technique, we formulate an online control problem to ensure achieving minimal energy cost while stabilizing use queues. While the control problem can be solved in a centralized manner, we limit our attention to distributed solutions which are highly attractive in the design of next generation mobile networks. Due to the combinatorial nature of the problem and the complex relation of achievable rates to interfering signals, the problem is non-convex. Consequently, conventional duality methods cannot be employed to achieve the distributed solution. Thus, we design a distributed solution for the problem based on Gibbs sampling method. The proposed algorithm can be implemented in a fully distributed manner, does not depend on the convexity or continuity of the energy cost functions, and guarantees solution optimality. Numerical results are provided to demonstrate the behavior of the solution in some example network scenarios.

I. INTRODUCTION

A. Motivation

Over the past few years, the volume of cellular data traffic has increased rapidly [1] forcing network operators to adopt new technologies in order to cope with the increased wireless traffic demand. A promising trend is to employ small cells such as microcells and picocells to increase capacity and provide more fine-tuned coverage across the network. Dense cell deployment, however, creates new challenges for the operation and management of cellular networks. Firstly, cellular providers face higher operational expenditure due to increased number of BSs if the BSs are not operated the right way. For example, increasing the number of BSs may result in a substantially higher electricity bill. Therefore, energy efficiency has emerged as a critical performance metric for cellular network operation [2]. Secondly, with a large number of network elements spread in a large geographic area, monitoring and optimizing their operation become a cumbersome task. Hence, among the most important objectives

of next generation mobile networks are self-organization and self-optimization [3].

Previous studies [2] have shown that base stations account for 60-80% of the total cellular network energy consumption. A notable observation is that while cellular traffic exhibits periodic behavior, the energy consumption approximately remains the same [4]. This can be attributed to the fact that cellular operators often deploy as many base stations as necessary to satisfy the peak traffic demand, while keeping them active all the time. In addition, due to various sources of energy consumption in BS equipment (*e.g.*, cooling system), transmission power control mechanisms cannot compensate for BSs being always active. Dynamic base station activation has been considered as a viable solution to solve this problem [5].

In this paper, we investigate this idea with the aim of minimizing the *long-term energy cost* of operating a cellular network. This is a challenging problem as solving it requires knowledge of future network conditions, *e.g.*, traffic load. Since this information is not available a priori, we seek *online* control mechanisms which do not rely on such information. To this end, we model the problem as a Lyapunov optimization problem [6]. The presented solution only relies on the information that is readily available in current cellular networks, *e.g.*, user queue backlogs and average power price. In addition, due to importance of self-optimization, we focus on distributed solutions that enable base stations (*e.g.*, eNBs in LTE) to autonomously decide about their configuration without relying on a central control entity. By distributed, we mean mechanisms in which communications are confined within neighboring BSs. This requirement is necessary for the scalability of the proposed solution.

As will be shown later, the formulated control problem is not convex. The non-convexity is due to combinatorial nature of BS activation and user association as well as complex interaction of interfering signals. Thus, the problem does not conform to conventional duality methods used typically to obtain distributed algorithms for communication problems. Therefore, to tackle the problem we employ a randomized method namely *Gibbs Sampling* [7]. The method is attractive as it does not limit the type of objective functions to convex and continuous ones. In addition, convergence to optimal solution is guaranteed if the algorithm runs for a sufficiently long time. Our contributions in this paper can be summarized as follows:

This work is supported in part by the Smart Applications on Virtual Infrastructure (SAVI) project funded under the National Sciences and Engineering Research Council of Canada (NSERC) Strategic Networks grant number NETGP394424-10.

- Dynamic base station activation and transmission power control is formulated with the objective of minimizing long-term energy cost of a cellular network.
- Based on two-timescale Lyapunov optimization approach, a control decision problem is derived which solving it and configuring system parameters based on it ensures system stability as well as minimal energy consumption.
- As the control problem is nonconvex, a distributed algorithm called Gibbs Sampling-based Activation (GSA) is proposed according which each base station decides independently about its configuration, knowing only the configurations of its neighbors.

There has recently been much attention paid to greening cellular networks through base station activation. The most relevant works to this paper in addition to some related works on Lyapunov optimization and Gibbs sampling are mentioned next.

B. Related Work

In [4], a location-dependent traffic profiling study is conducted on real 3G network traces showing that 23-53% energy saving is achievable via dynamic base station activation. In [8], centralized and heuristic methods are presented for deactivating the base station that has the lowest load in the network. The joint problem of base station activation and user association is studied in [5], 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. In [9], distributed base station activation is posed as a network utility maximization problem to find the optimal BS activation probabilities. As opposed to our work, the proposed solution in [9] is unable to provide performance guarantees on the system cost and does not optimize the long-term energy cost.

There have been several recent works that modeled energy cost reduction in data centers following the framework of two-timescale Lyapunov optimization [6], [10]. In [6], an online algorithm is presented to distribute the load among a set of data centers, activate enough servers in each data center, and scale their speeds so as to satisfy the load and stabilize the system queues. In [10], employing multiple power sources including long-term and real-time power market, renewable green energy, and local power storage for reducing long-term power cost are investigated. However, unlike [6], [10], the problem in our work has a combinatorial nature (due to On/Off behavior of base stations) which makes it different and more challenging to solve.

Designing distributed algorithms using Gibbs sampling has been recently investigated in wireless literature. User association schemes based in this approach are discussed in [11]. More relevant to our work, [12] considered distributed power control in OFDMA-based systems using Gibbs sampling. A similar radio access system model as [12] is adopted in our

work whereas in addition to power control, we also consider BS activation and user scheduling which changes the problem.

C. Paper Structure

The rest of the paper is organized as follows. In section II, system model is described and the problem is formulated. In section III, the control problem is derived following two-timescale Lyapunov optimization approach. In section IV, GSA algorithm is presented. Sample numerical results are provided in section V. Finally, section VI concludes this paper.

II. SYSTEM MODEL AND PROBLEM FORMULATION

We consider downlink of an OFDMA-based cellular network, *e.g.*, downlink in an LTE network. The network consists of the set $\mathcal{B} = \{b_1, \dots, b_n\}$ of base stations which collectively provide service to the set $\mathcal{U} = \{u_1, \dots, u_m\}$ of users. The frequency bandwidth used for communication between BSs and users is divided into set $\mathcal{H} = \{h_1, \dots, h_k\}$ of subcarriers. The system operates on a discrete-time basis. BS activation is performed at a timescale that is different from other network operations. Therefore, time is divided into frames of size T timeslots. The activation/de-activation decisions are made at the beginning of each time frame. Other decisions (power allocation, user scheduling) might change at each timeslot.

Notation. Bold letters are used to represent vectors, *e.g.*, $\mathbf{v} = [v_i]_d$ denotes a d-dimensional vector. $[v'_i, \mathbf{v} \setminus v_i]$ shows the vector obtained by replacing i -th element of vector \mathbf{v} with v'_i . Sets are shown by calligraphic letters. The cardinality of a set \mathcal{S} is denoted by $|\mathcal{S}|$.

A. Energy Cost Model

Assume that base station b_i consumes the total power $P_{b_i}(t)$ at timeslot t . Energy cost incurred by b_i at t is given by

$$C_i(t) = C_p(t) \cdot P_{b_i}(t), \quad (1)$$

where $C_p(t)$ is the energy price at t . $C_p(t)$ changes according to an exogenous random process which is assumed to have a stationary distribution. There exists a bound on the incurred operational cost of BS b_i such that $C_i(t) \leq C_i^{max}$. The total power consumption P_{b_i} consists of two parts [4] as follows

$$P_{b_i}(t) = P_{tx}(t) + P_{misc}(t), \quad (2)$$

where $P_{tx}(t)$ is the transmission power used to communicate with users. The term $P_{misc}(t)$ accounts for the power spent in cooling and power supply. Clearly, $P_{tx}(t)$ depends on the carried traffic load and can be approximated as follows [4]

$$P_{tx}(t) = P_\alpha \cdot \mu(t) + P_\beta, \quad (3)$$

where $\mu(t)$ is the traffic load factor, of the base station. Overall there is a base cost for activating a base station due to residual factors P_{misc} and P_β and a traffic dependent part due to P_α . As reported in [13], the base activation cost can take up to 50% of the total base station power consumption.

B. Power Allocation

The transmission power *i.e.*, P_{tx} of an active base station varies between minimum power P_{min} and maximum power P_{max} . This power is distributed among the set of subcarriers \mathcal{H} . Let $\mathbf{P}_i(t) = [P_{ik}(t)]_{|\mathcal{H}|}$ denote the power vector of b_i , where $P_{ik}(t)$ is the power allocated to k -th subcarrier from b_i at t . As the total transmission power of the BS does not exceed P_{max} , we have the following constraint on every vector $\mathbf{P}_i(t)$

$$\sum_{h_k \in \mathcal{H}} P_{ik}(t) \leq P_{max}, \quad \forall b_i \in \mathcal{B}. \quad (4)$$

Activation of BSs in timeslot t is captured via binary vector $\mathbf{Y}(t) = [y_i(t)]_{|\mathcal{B}|}$.

C. User Association

While include user association in the model is straightforward as shown in [12], for the sake of clarity, we make some practical assumptions regarding user association that simplifies the model. First, we consider a standard OFDMA system without carrier aggregation capability presented in LTE-Advanced [14]. In these systems, each mobile user is served by at most one base station at a time. Second, as in traditional deployments every mobile user associates to the BS from which it receives the strongest signal strength, indicated by the *Reference Signal Received Power* (RSRP). At frame intervals, user associations may change. User association is denoted by vector $\mathbf{A}(t) = [a_{ij}(t)]_{|\mathcal{B}| \times |\mathcal{U}|}$.

D. Achievable Rates

Assume BS b_i at timeslot t allocates power $P_{ik}(t)$ to subcarrier h_k on which it communicates with user u_j . Let g_{ijk} denote the power gain between b_i and u_j on h_k which is a function of distance and propagation environment. Power gain is assumed to be fixed over the duration of a frame. Then, the received power at u_j is $P_{ik}(t) \cdot g_{ijk}$. User u_j is under the coverage of b_i if the received power of the pilot signal from b_i at u_j is higher than a pre-specified threshold. For future reference, let \mathcal{U}_i denote the subset of users that are under the coverage of BS b_i and \mathcal{B}_j denote the set of all base stations that cover user u_j . According to our assumption regarding the received power, the serving BS for each user is the active BS with the closest distance to the user.

The achievable rate of a user has direct relation to the received Signal-to-Noise-and-Interference Ratio (SINR). Following the above notation, the SINR of user u_j when served by BS b_i on subcarrier h_k is given by

$$\text{SINR}_{ijk}(t) = \frac{P_{ik}(t) \cdot g_{ijk}}{\sum_{b_{i'} \neq b_i} P_{i'k}(t) \cdot g_{i'jk} + \eta}, \quad (5)$$

where η is the background noise power. The received rate of a user is obtained from a rate function $R(\cdot)$ which is generally increasing w.r.t the received SINR. A common choice is the Shannon capacity formula which gives the following rate function

$$R(\text{SINR}_{ijk}(t)) = \Gamma \log(1 + \text{SINR}_{ijk}(t)), \quad (6)$$

where Γ is the width of downlink channel. We assume that subcarriers can be fractionally (*i.e.*, in time) shared among users using a technique like TDMA. Let $0 \leq \tau_{ijk}(t) \leq 1$ denote the fraction of time that subcarrier h_k of BS b_i is allocated to user u_j . Then, the received rate at user u_j is given by

$$r_j(t) = \sum_{b_i \in \mathcal{B}_j} \sum_{h_k \in \mathcal{H}} a_{ij}(t) \tau_{ijk}(t) R(\text{SINR}_{ijk}(t)), \quad u_j \in \mathcal{U}. \quad (7)$$

Moreover, the time fractions should satisfy the following constraint

$$\sum_{u_j \in \mathcal{U}} \tau_{ijk}(t) \leq 1, \quad \forall b_i \in \mathcal{B}, h_k \in \mathcal{H}. \quad (8)$$

E. Problem Formulation

The traffic intended for users is first received and stored in user queues at base stations. Queue $Q_j(t)$ is kept for each user u_j in its associated base station. We denote the amount of workload arrived at timeslot t for u_j by $w_j(t)$ and the total arrival vector by $\mathbf{w}(t) = [w_1(t), \dots, w_m(t)]$. We assume the arrival for user u_j follows an i.i.d. distribution throughout the whole frame while the average rate \bar{w}_j is known to the base station (the base station can estimate this over each frame). In addition, there exist bounds w_{min} and w_{max} such that $w_{min} \leq w_j(t) \leq w_{max}$ for all $u_j \in \mathcal{U}$.

The data stored in queue Q_j will be transmitted to user u_j according to the associated BS's resource allocation policy. Let $r_j(t)$ denote the amount of u_j 's workload transmitted to the user at timeslot t . Let r_{max} denote a limit on the rate that can be provided to each user such that the inequality $r_j(t) \leq r_{max}$ holds for all users $u_j \in \mathcal{U}$. Queues evolve in consecutive timeslots according to the following queuing dynamic,

$$Q_j(t+1) = \max[Q_j(t) - \sum_{b_i \in \mathcal{B}} y_i(t) a_{ij}(t) r_j(t), 0] + w_j(t). \quad (9)$$

We say that the system is stable if the following condition holds on queue backlogs,

$$\bar{Q} \triangleq \limsup_{t \rightarrow \infty} \frac{1}{t} \sum_{\tau=0}^{t-1} \sum_{j=1}^m \mathbb{E}\{Q_j(\tau)\} < \infty. \quad (10)$$

The energy cost of the system at timeslot t is the sum of the energy cost of all base stations, *i.e.*,

$$\text{Cost}(t) = \sum_{b_i \in \mathcal{B}} C_i(t). \quad (11)$$

We aim to design an online control mechanism to determine the set of active base stations $\mathbf{Y}(t)$, and allocate resources (power $\mathbf{P}(t)$ and time fractions $\boldsymbol{\tau}(t)$) to users so as to minimize the long-term cost of the system defined as follows,

$$\begin{aligned} \mathbf{P1:} \quad & \text{Minimize } \limsup_{t \rightarrow \infty} \frac{1}{t} \sum_{\tau=0}^{t-1} \mathbb{E}\{\text{Cost}(\tau)\} \\ & \text{subject to: (4), (8), (10).} \end{aligned} \quad (12)$$

In the next section, we describe the steps to derive the control algorithm to solve (12).

III. ONLINE CONTROLLER DESIGN

Throughout this section, we focus on the changes in the system over the span of a specific frame which starts from timeslot t to timeslot $t+T-1$. To develop the online control algorithm, we first define the Lyapunov function $L(t)$ as a scalar measure of queue backlog in the system as follows,

$$L(t) = \sum_{j=1}^m \frac{1}{2} [Q_j(t)]^2. \quad (13)$$

It is desirable for our algorithm to push the system towards a lower backlog state. Therefore, to observe the expected change in the Lyapunov function over T timeslots, we define the T -slot Lyapunov drift as follows,

$$\Delta_T(t) = \mathbb{E}\{L(t+T) - L(t) | \mathbf{Q}(t)\}. \quad (14)$$

In addition, we would like to minimize the long-term energy cost of the system as defined in (12). Hence, following the *drift-plus-penalty* approach [15], we add the expected energy cost of the system to (14), which results in the following drift-plus-penalty expression,

$$\Delta_T(t) + V \cdot \mathbb{E}\left\{ \sum_{\tau=t}^{t+T-1} \text{Cost}(\tau) \right\}, \quad (15)$$

where the parameter V is chosen so as to control the trade-off between energy cost and congestion (reflected in queue backlogs) in the network. The next derivation step in Lyapunov optimization is to find an upper bound on this expression. We show that the following theorem holds.

Lemma 1. Let $V > 0$ and $t = kT$ for some $k \in \mathbb{Z}_+$. For any set of possible BS activation, user association, and resource allocation decisions that conform to the constraints of (12), we have

$$\begin{aligned} \Delta_T(t) + V \cdot \mathbb{E}\left\{ \sum_{\tau=t}^{t+T-1} \text{Cost}(\tau) \right\} &\leq B_1 T - \\ &\mathbb{E}\left\{ \sum_{\tau=t}^{t+T-1} \sum_{u_j \in \mathcal{U}} Q_j(t) \left[\sum_{b_i \in \mathcal{B}} y_i(t) a_{ij}(t) r_j(\tau) - w_j(\tau) \right] | \mathbf{Q}(t) \right\} \\ &+ V \mathbb{E}\left\{ \sum_{\tau=t}^{t+T-1} \sum_{b_i \in \mathcal{B}} y_i(t) C_i(\tau) \right\}. \end{aligned} \quad (16)$$

where, $B = \frac{1}{2} m (w_{max}^2 + r_{max}^2) + \frac{T-1}{2} m r_{max} [r_{max} - w_{min}]$.

Proof: The result is obtained by squaring the queuing dynamic (9), using inequalities $\sum_{b_i \in \mathcal{B}} y_i(t) a_{ij}(t) r_j(\tau) \leq r_{max}$ and $w_j(\tau) \leq w_{max}$, and finally adding the cost based in the definition of drift plus penalty (15). ■

Our goal is to minimize the right-hand side of (16) or equivalently we seek to maximize

$$\mathbb{E}\left\{ \sum_{\tau=t}^{t+T-1} \sum_{b_i \in \mathcal{B}} y_i(t) \left[\sum_{u_j \in \mathcal{U}_i} Q_j(t) a_{ij}(t) r_j(\tau) - V C_i(\tau) \right] | \mathbf{Q}(t) \right\}. \quad (17)$$

This is the joint problem of base station activation at the beginning of each frame in addition to resource allocation at each timeslot. Since we considered that power gains are fixed, there exists an optimal solution to (17) in which the allocated power and time-fractions remain fixed during the frame which simplifies (17) as

$$\mathbb{E}\left\{ \sum_{\tau=t}^{t+T-1} \sum_{b_i \in \mathcal{B}} y_i(t) \left[\sum_{u_j \in \mathcal{U}_i} Q_j(t) a_{ij}(t) r_j(t) - V C_i(\tau) \right] | \mathbf{Q}(t) \right\}. \quad (18)$$

This expression is a function of $\mathbf{Y}(t)$, $\mathbf{P}(t)$, $\boldsymbol{\tau}(t)$, and $\mathbf{Q}(t)$ which we call it the *net utility* of the system and denote it by $N(\cdot)$. Problem (18) is a non-convex integer optimization program that is generally hard to solve. Hence, devising a distributed solution would be a challenging task. Therefore, to tackle the problem, we present a Gibbs Sampling-based Activation algorithm to solve (18) in the following section.

IV. GSA ALGORITHM

In this section, we present our Gibbs Sampling-based Activation (GSA) algorithm to achieve maximum net utility (18). At the end of each time frame, along with active base stations, each inactive BS becomes active to examine its surrounding network condition, participate in the algorithm, and decide about its activation in the next frame. Note that we assume that GSA runs in the background and until its completion and computation of a new network configuration, only previously active base stations provide service to users.

A. Power Allocation

Similar to [12], we assume that the transmission power of a BS can vary from 0 to P_{max} in small discrete steps of size δ , i.e., $0, \delta, 2\delta, \dots, P_{max}$. Given a BS power vector, time fractions of the associated users should be chosen so that the maximal net utility is achieved to avoid wasting resources. Therefore, the time fraction vector is specified by the power vector. Accordingly, the state of a base station is defined as its transmission power vector. For instance, the state of BS b_i is given by \mathbf{P}_i . The set of all possible power vectors of BS b_i is represented by \mathcal{P}_i and the set of possible configurations (states) of the system is then given by $\mathcal{F} = \prod_{b_i \in \mathcal{B}} \mathcal{P}_i$.

Two base stations b_i and $b_{i'}$ are considered neighbors, denoted by $b_i \sim b_{i'}$, if there exists a user u_j that is under the coverage of both of them. The set of all one-hop neighbors of b_i is shown by \mathcal{N}_i^1 . To be able to compute the effect of changing the power vector on the achievable rates of itself and its neighbors, each BS needs to know the power vectors of all its two-hop neighbors. Thus, we consider system graph $G = (\mathcal{B}, \mathcal{N})$ constructed such that each base station is connected to all of its two-hop neighbors, that is, $\mathcal{N} = \{\mathcal{N}_i\}$

where $\mathcal{N}_i = \cup_{b_i' \sim b_i} \mathcal{N}_i^1$. This way, the neighborhood of each node forms a clique [12]. In addition, the net utility (17) can be expressed as the summation of a net utilities over the set of cliques $\mathcal{Q}(G)$, That is, for $f \in \mathcal{F}$ we have

$$N(f) = \sum_{q \in \mathcal{Q}(G)} N_q(f) \quad (19)$$

where

$$N_q(f) = \sum_{b_i: \mathcal{N}_i=q} \mathbb{E} \left\{ \sum_{\tau=t}^{t+T-1} y_i(t) \times \left[\sum_{u_j \in \mathcal{U}_i} Q_j(t) a_{ij}(t) r_j(t) - VC_i(\tau) \right] | \mathbf{Q}(t) \right\}, \quad (20)$$

is the net utility of clique q .

The base station whose state is to be changed is selected randomly and independently. The base station goes to a new state with the following transition probability

$$\mathbb{P} \{ \mathbf{P}_i \rightarrow \mathbf{P}'_i \} = \frac{\exp(N_q([\mathbf{P}'_i, \mathbf{P} \setminus \mathbf{P}_i]) / \Theta)}{\sum_{\mathbf{P}''_i \in \mathcal{P}_i} \exp(N_q([\mathbf{P}''_i, \mathbf{P} \setminus \mathbf{P}_i]) / \Theta)}. \quad (21)$$

where q is the clique that contains b_i , \mathbf{P} is the vector of all power vectors, and Θ is a constant called the temperature.

Any change to the local state is carried out in two steps. First, the power vector is updated. This may also include activating an inactive BS with the minimum power δ or shutting down a BS with the minimum power. Second, the time fraction vector that results in the optimal net utility for the new power vector is chosen as the new time fraction vector. Allocating time fractions to users of a BS does not affect users associated to the neighboring BSs. Thus it is a local problem that can be solved locally and efficiently by each BS. After updating its transmission power, every base station informs all its two-hop neighbors about its new power vector.

Following the above procedure, the system will eventually reside in one of the maximum net utility states with high probability when $\Theta \rightarrow 0$ [7]. The algorithm converges geometrically fast to the optimum solution [7].

V. NUMERICAL RESULTS

A. Simulation settings

The wireless parameters are adopted consistent with the standard 3GPP propagation models [3]. 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). Specifically, $f(d) = 10^{h_0} d^{-\kappa}$ with path loss exponent $\kappa = 3.5$ and $h_0 = -14.4$. The background noise power is $N_0 = -174$ dbm (Hz^{-1}). We consider a network of size 1.2 by 1.2 km. 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 350m away from it. Users are distributed non-uniformly in the network. To do so, nine crowded regions are considered in the network where each one is a circle of radius 160m. Users are divided equally among these crowded

regions and distributed uniformly within each region. There are in total 100 users in the network.

Transmission power of each BS varies from 0W to 16W in steps of $\delta = 4W$. The base power used for activating each BS is assumed to be 16W. We assume that there are only 4 subcarriers and bandwidth per frequency is 1 MHz. The system parameters are selected so as to allow exhaustive search of all possible configurations and find the optimal one in a timely manner as required for the comparison of GSA and the optimal solution. However, The GSA algorithm can handle much larger instances of the problem.

Our goal is to study the behavior of GSA algorithm. In each iteration of the algorithm, one of the BSs is randomly selected. The BS can choose from the following set of options: (1) Activation by allocating the minimum transmission power δ to some frequency band. (2) De-activation if its total transmission power is equal to the minimum value δ . (3) Increasing transmission power by δ . (4) Decreasing transmission power by δ . Moreover, in the numerical results, a new state f' is accepted with the following probability

$$\min(1, \exp(\frac{N(f') - N(f)}{\Theta})), \quad (22)$$

where, (22) represents the Metropolis-Hastings sampler that shows characteristics very similar to that of the Gibbs sampler.

B. GSA performance

In the first set of results, we compare the performance of GSA versus the optimal solution in terms of the achieved net utility. We consider only four base stations and 11 users are associated to these base stations. Even for this small set of BSs with the aforementioned set of parameters, there are 71^4 combinations for BS activation and power allocation that should be considered to find the optimal solution. Queue backlogs for the users are considered the same and set to 40Kb. Distribution of the energy price C_p has the expected value of 6.

Figure 1 demonstrates the behavior of GSA algorithm when temperature Θ is set to 50 and 0.1 respectively. The Optimal line shows the optimal net utility of the considered set of BSs achieved by exhaustive search. As seen in the figure, the GSA result becomes very close to the optimal when using a small value for Θ . Although the same performance cannot be achieved when the temperature is high, convergence happens faster in this case. Another notable observation is that the algorithm shows more fluctuations in this case. Both observations can be attributed to the higher temperature.

C. Energy cost vs delay

In this section, the behavior of GSA is studied in terms of energy cost and delay trade-off. These parameters can be approximated indirectly via the number of active BSs and average queue sizes respectively. Each frame is assumed to consist of 10 timeslots. Arrival traffic for each user follows a Bernoulli process. Associated to each user u_j is a probability of acceptance p_j . This probability is determined randomly and

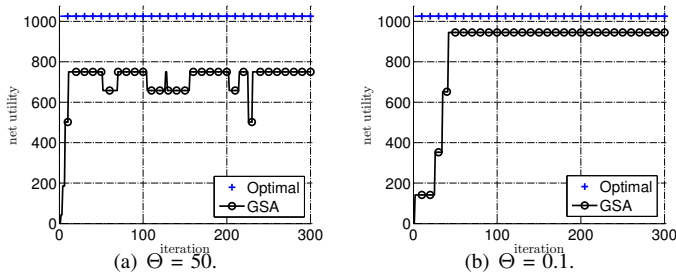


Fig. 1. Convergence of net utility.

independently for each user. With probability p_j , $40Kb$ is added to the queue of user u_j in each timeslot. To demonstrate how the algorithm responds to variation in the arrival traffic by changing the set of active BSs, we divide frames into framesets. In even-numbered framesets arrived data enters the user queues as normal, whereas there is no arrival in odd-numbered framesets. Each frameset consists of 5 frames. For different values of control parameter V , average queue sizes and the number of active BSs are depicted in Figure 2. As can be seen in the figure, initially the queues are empty and all BSs are inactive. During the first frame, the queues start to grow in size as the arrival traffic enters and does not get served. This lasts until the beginning of the next frame at which some base stations are activated and queues start to get served. Overall, during even-numbered framesets some BSs are active and queues have some data. In odd-numbered framesets, as there is no new arrival, queues gradually become empty and BSs are turned off. A notable observation is that by increasing V , the average queue sizes increase as well. On the other hand, average number of active BSs is decreased. In addition, queue sizes almost remain the same during a frameset which indicates that GSA could stabilize the system. In this set of results on average lower than 10 BSs are needed to satisfy the aforementioned arrival traffic demand. This shows great improvement in terms of energy consumption compared to activating all 25 BSs in the network.

VI. CONCLUSION AND FUTURE WORK

In this paper, designing a distributed solution for base station activation and power allocation with the aim of minimizing long-term energy cost of cellular network is investigated. The problem is modeled following the framework of two-timescale Lyapunov optimization and a provably-efficient online control algorithm is proposed based on Gibbs sampling method. In this paper, we assumed that users are static and power gains are fixed which eliminates the need for fine-tuned power allocation and scheduling at each timeslot. Extending the framework by considering time-varying channel conditions remains as a future work

REFERENCES

[1] "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.

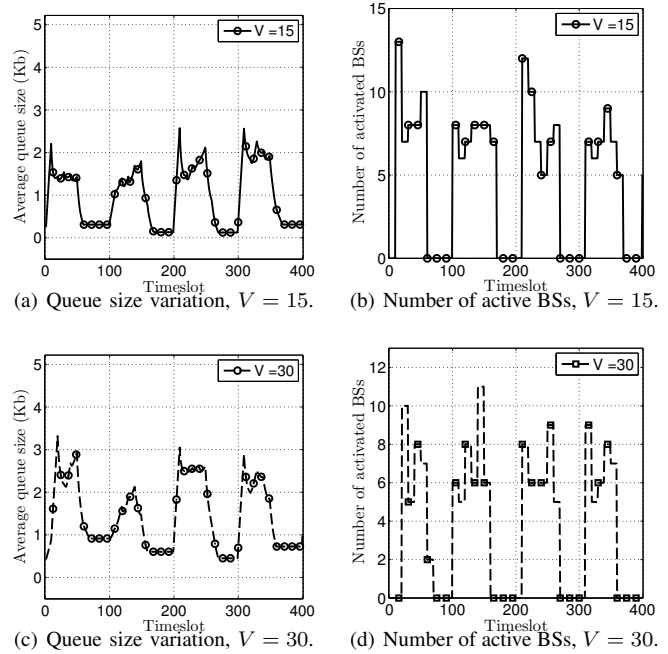


Fig. 2. Variation of queue sizes and active number of BSs versus V .

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

[3] "Long term evolution of the 3gpp radio technology," <http://www.3gpp.org/LTE/>.

[4] 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*, Las Vegas, USA, Sep 2011, pp. 121–132.

[5] 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.*, vol. 29, no. 8, pp. 1525–1536, 2011.

[6] Y. Yao, L. Huang, A. Sharma, L. Golubchik, and M. Neely, "Data Centers Power Reduction: A two Time Scale Approach for Delay Tolerant Workloads," in *Proc. IEEE INFOCOM*, Orlando, USA, Mar 2012, pp. 1431–1439.

[7] P. Bremaud, *Markov chains: Gibbs fields, Monte Carlo simulation, and Queues*. New York, USA: Springer, 1999.

[8] S. Zhou, 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.

[9] A. Abbasi and M. Ghaderi, "Distributed base station activation for energy-efficient operation of cellular networks," in *Proc. ACM MSWiM*, Barcelona, Spain, Nov 2013, pp. 427–436.

[10] W. Deng, F. Liu, H. Jin, C. Wu, and X. Liu, "MultiGreen: Cost-Minimizing Multi-source Datacenter Power Supply with Online Control," in *Proc. ACM energy*, Berkeley, USA, May 2013, pp. 149–160.

[11] B. Kauffmann, F. Baccelli, A. Chaintreau, V. Mhatre, K. Papagiannaki, and C. Diot, "Measurement-Based Self Organization of Interfering 802.11 Wireless Access Networks," in *Proc. IEEE INFOCOM*, Anchorage, Alaska, USA, May 2007, pp. 1451–1459.

[12] S. C. Borst, M. G. Markakis, and I. Saniee, "Nonconcave Utility Maximization in Locally Coupled Systems, With Applications to Wireless and Wireline Networks," *IEEE/ACM Trans. Neww.*, 2013.

[13] G. Fettweis and E. Zimmermann, "ICT Energy Consumption - Trends and Challenges," in *Proc. WPMC*, LapLand, Finland, Sep 2008.

[14] G. Yuan, X. Zhang, W. Wang, and Y. Yang, "Carrier Aggregation for LTE-Advanced Mobile Communication Systems," *IEEE Commun. Mag.*, vol. 48, no. 2, pp. 88–93, 2010.

[15] M. J. Neely, "Stochastic Network Optimization with Application to Communication and Queuing Systems," *Synthesis Lectures on Communication Networks*, vol. 3, no. 1, pp. 1–211, 2010.