On Balancing Energy Efficiency for Network Operators and Mobile Users in Dynamic Planning

Muhammad Ismail, Mohamed Kashef, Erchin Serpedin, and Khalid Qaraqe

ABSTRACT

The high energy consumption of network operators and mobile users has raised environmental, financial, and quality-of-experience concerns. These concerns have renewed the research efforts in developing green communication strategies for energy efficient wireless network operation. Network operators employ dynamic planning to save energy at low call traffic load by switching off some of their base stations (BSs), and mobile users are served by the remaining active BSs. The existing research investigates energy efficiency of dynamic planning approaches only from the network operator perspective. Dynamic planning, if not carefully designed, can lead to higher energy consumption for the mobile users in the uplink, which in turn degrades the uplink service quality due to mobile terminals' battery depletion. In this article we propose a dynamic planning framework with balanced energy efficiency that accounts for the energy consumption of the mobile users in the uplink as well as that of the network operators in the downlink. We discuss the associated challenges and implementation issues. A dynamic planning approach based on a multi-time scale decision process is proposed to achieve the balanced energy efficiency framework. Numerical results demonstrate the improved energy efficiency performance for the uplink mobile users as compared with the traditional dynamic planning approach.

INTRODUCTION

The great advancement in wireless communication services has resulted in high energy consumption for network operators and mobile users. Overall, there exist three million base stations (BSs) and approximately three billion mobile terminals (MTs) worldwide that consume 4.5 GW and 0.2 GW to 0.4 GW of power, respectively. The high energy consumption of network operators and mobile users has resulted in environmental, financial, and quality-of-experience (QoE) concerns.

From an environmental point of view, the telecommunication industry is responsible for about 2 percent of the total CO_2 emissions worldwide, and the percentage is expected to double by 2020 [1]. From a financial standpoint, it has been estimated that the energy bills of service providers cost about 18 to 32 percent of their operating expenditures [2]. From a QoE consideration, it has been estimated that almost 60 percent of mobile users suffer from a limited battery capacity, a problem that is further complicated by the slow advance in battery technology. The aforementioned concerns have motivated research in *green* communications for energy efficient wireless network operation.

In general, the green solutions that can be deployed by network operators can be classified based on the call traffic load condition. At a low call traffic load, network operators switch off some of their BSs to save energy and MTs are served by the remaining active BSs, which is referred to as *dynamic planning* [1]. In this context, one limitation associated with the existing research is that it investigates energy efficiency only from the network operator perspective. Dynamic planning approaches, if not carefully designed, can lead to higher energy consumption for the MTs in the uplink due to larger transmission distances. In such a case, dynamic planning would only shift the energy consumption burden from the BSs to the MTs, which results in battery drain for MTs at a faster rate. Consequently, this will degrade the service quality perceived by the mobile users because of, for example, lower throughput, higher call dropping rate, and so on. Hence, the future design of dynamic planning should capture and balance the trade-off in energy efficiency among network operators and mobile users.

In this article we shed the light on such a trade-off to motivate research in this direction. We first review the fundamentals of the traditional dynamic planning approaches. Then we present a dynamic planning framework with balanced energy efficiency and discuss the associated challenging design and implementation issues. Finally, we propose a dynamic planning approach

Muhammad Ismail, Mohamed Kashef, and Khalid Qaraqe are with Texas A&M University at Qatar.

Erchin Serpedin is with Texas A&M University, College Station.

 0163-6804/15/$25.00 © 2015 IEEE

with balanced energy efficiency based on a multi-time scale decision process, and we present numerical results to illustrate the performance of the proposed approach.

Dynamic Planning in Green Networks

In this section we discuss the motivation for adopting dynamic planning for energy efficient network operation along with design fundamentals.

Motivations for Dynamic Planning

In cellular network planning, the cell size and capacity are traditionally designed to satisfy the peak call traffic load. However, it has been shown that the call traffic load exhibits temporal and spatial fluctuations [2]. Consequently, networks are over-provisioned at a low call traffic load, which leads to wasted energy. Thus, BS on-off switching is proposed to adapt network energy consumption according to the call traffic load condition, an approach that is referred to as dynamic planning. Hence, at a low call traffic load, lightly loaded BSs are switched off and MTs are connected to the remaining active BSs.

Two fundamental issues must be tackled while designing a dynamic planning mechanism, namely MT association and BS operation, as discussed in the following.

MT Association

This issue deals with associating each MT to a given BS for service. In the following, we discuss the design objectives of the MT association mechanism in the context of dynamic planning. Then we review the MT association mechanism decision criteria.

Design Objectives: In literature, the MT association serves two objectives for dynamic planning deployment. The first objective is related to concentrating the MTs' traffic in a few BSs to switch-off the remaining BSs. The second objective aims to balance the trade-off between network energy efficiency and flow level performance of the MTs (e.g. data rate, time delay, and so on.) [3]. The rationale behind such an objective is not to jeopardize the target quality-of-service (QoS) of MTs while reducing the network energy consumption.

Decision Criteria: The simplest decision criterion for the MT association problem is MT-BS distance based, where MTs with downlink traffic are associated with the nearest BSs [4]. As a result, low transmission power is consumed due to the short transmission distances. The network impact is another decision criterion introduced in [5] where MTs' traffic is concentrated in the BSs that lead to lower inter-cell interference. Furthermore, coverage hole avoidance is a decision criterion that aims to concentrate the call traffic load in a subset of BSs that can provide acceptable network coverage [6].

BS Operation

The BS operation specifies which BSs are switched off, when to wake up a given BS, and how to implement the switching decisions. The associated design issues are discussed in the following.

Prediction of Future Traffic Demands: The BS mode (on or off) lasts for a long duration (i.e. in hours) to avoid frequent BS on-off switching. Hence, the BS on-off switching decision should consider not only the current traffic load (through MT association) but also future demands so as to guarantee an acceptable QoS [1]. As a result, extra resources should be reserved at the active BSs to satisfy future traffic demands [7]. Information regarding future demands can be inferred from historical call traffic load pattern [1] or via prediction using an online stochastic game [8].

BS Wake-Up Design: When the call traffic load served by the active BSs increases beyond their capacity limitation, some of the switched off BSs have to be turned on. Hence, in dynamic planning, it is necessary to specify the wake-up instants for the inactive BSs. For instance, in [9] N-based and V-based wake-up schemes are proposed. Specifically, in the N-based scheme an inactive BS wakes up only when N MTs request service. On the other hand, in the V-based scheme, an inactive BS wakes up after a vacation time V. Similarly, different wake-up schemes are proposed for femto-cell BSs, which can be BS, MT, or network controlled [10].

Switching off Mode Entrance and Exit: An MT may not be able to complete its handoff procedure to an active BS and suffers from call dropping if its associated BS is switched off too quickly. This is due to the low received signal from the neighboring BSs and the limited signaling channel capacity. In [11] BS wilting is proposed for a smooth switch-off mode entrance, where the BS transmit power is progressively halved until the BS is turned off. Similarly, MTs in service can suffer from strong interference if a BS is switched on too quickly. A BS blossoming process is proposed in [11] for smooth switch-off mode exit, where the BS transmit power is progressively doubled until the BS is turned on.

An MT may not be able to complete its handoff procedure to an active BS and suffers from call dropping if its associated BS is switched off too fast. This is due to the low received signal from the neighboring BSs and the limited signaling channel capacity.

Dynamic Planning with Balanced Energy Efficiency

The main goal of dynamic planning is to reduce BS energy consumption while ensuring an acceptable downlink service quality for mobile users. For instance, in [3] the objective is to balance BS energy consumption performance with the MTs' downlink flow-level performance. However, no attention is paid to the relation between the mobile users' perceived service quality and their incurred uplink energy consumption. For instance, if an MT is consuming high energy in the uplink, it is expected to drain its battery at a fast rate, eventually leading to call dropping. Hence, dynamic planning approaches with balanced energy efficiency among network operators and mobile users should be investigated. In the following, the motivation for dynamic planning with balanced energy efficiency is presented. Then challenging design and implementation issues are discussed.

Motivation

When the main scope of dynamic planning is to enhance the energy efficiency of network operators, which is the case for the existing research, the BSs' switch-off decisions can result in energy

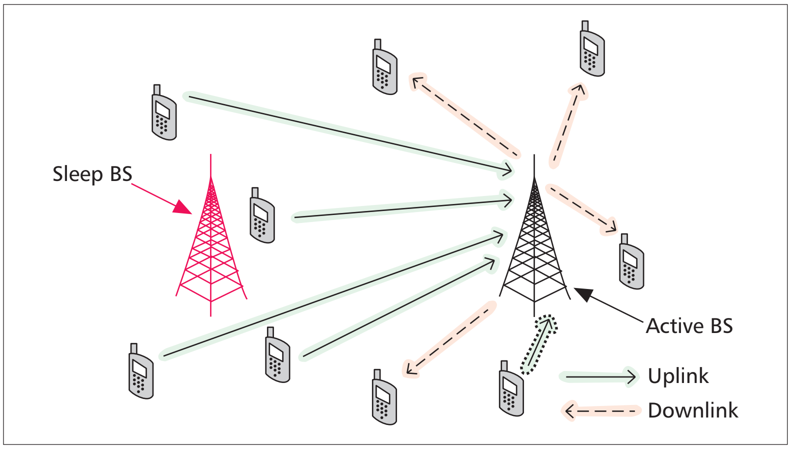

Figure 1. Dynamic planning with unbalanced energy saving. MTs with uplink traffic are associated with faraway BSs.

inefficient user association from the mobile users' standpoint. As shown in Fig. 1, accounting only for the downlink performance, MTs with uplink traffic can be associated with a faraway BS, due to a switched off nearby BS. Because of the long transmission distance, high energy consumption of MTs in the uplink is expected, which leads to energy depletion for MTs at a faster rate. Although energy consumption for MTs is not that much compared with BS energy consumption, a fast rate battery depletion for MTs still results in a high rate of dropped services in the uplink, which jeopardizes the mobile users' perceived service quality. Hence, the dynamic planning approach should be designed to capture and balance the trade-off in the resulting energy efficiency for network operators and mobile users.

Challenging Issues

In this subsection we discuss the challenging design and implementation issues toward developing a dynamic planning approach with balanced energy efficiency performance between network operators and mobile users.

The Coupling Between MT Association and BS Operation: One challenge with dynamic planning is that the switching decisions of BSs are coupled with MT associations. Specifically, when a BS is switched off, the MTs associated with it need to perform a handover process to another BS. Similarly, when a BS is turned-on, the nearby MTs can perform a handover process to this BS. Also, newly incoming MTs are associated with a subset of active BSs to obtain service. However, the BS operation (i.e. on-off switching) does not occur at the same rate as MT association. Hence, dynamic planning is a two time scale problem. At a high level, the BS operation occurs at a slow rate (with a scale of hours) that depends on the call traffic load density. At a low level, the MT association takes place at a faster rate (with a scale of minutes) based on user arrivals and departures. When only downlink traffic is considered, as in the existing research, the decisions at both levels are determined based only on BS energy consumption. With the coexistence of uplink and downlink traffic, the decisions at both levels are determined based on the expected energy consumption at the BSs and the MTs.

New Switch-off and Wake-up Decision Criteria: When uplink traffic is considered, the BS switch-off decision criteria should be revised. Specifically, the switch-off decision criteria should capture the impact of MTs' battery drain on uplink service degradation, for example, lower throughput, higher latency, higher call dropping rate, and so on. Hence, the uplink service degradation is due to two factors, namely unavailability of radio resources at the BSs (due to BS switch-off) and MTs' battery drain (due to communicating with faraway BSs). The BS switch-off decision metric should balance BS energy consumption with uplink service quality due to MTs' battery drain. Similarly, the existing mechanisms employ the call traffic load increase as a wake-up decision criterion for a switched off BS [5]. However, in the presence of uplink traffic, the wake-up decision criteria should include, besides the call traffic load measure, a measure of MTs' service degradation due to battery drain. As a result, if the MTs' service quality is degraded due to battery drain, a nearby inactive BS should be turned on to avoid MTs' battery depletion and hence dropping of uplink calls.

A Two Time Scale Dynamic Planning Approach with Balanced Energy Efficiency

In this section we propose a two time scale dynamic planning approach with balanced energy efficiency.

System Model

For illustration purposes, we consider a geographical region that is covered by two BSs from different networks, as shown in Fig. 2. Let s denote the BS index, with $s \in \{1, 2\}$. The BSs operate in separate frequency bands, hence no interference is considered. Interference management schemes (e.g. frequency reuse [12]) are employed for interference mitigation among BSs within each network. The distance between the two BSs is given by D and each BS has a height of H_s, as shown in Fig. 2. The network operators cooperate with each other for energy saving by alternately switching on and off their BSs according to the call traffic load condition [1]. The active BSs carry the call traffic load in the geographical region. Each BS can control its coverage area through antenna tilting [7]. For simplicity, two tilting angles are considered per BS, which corresponds to two cell coverage areas for each BS, namely A_{s,k_s}, where $k_s = 1$ and $k_s = 2$ denote the cell coverage area corresponding to the first and second tilting angles, respectively, as shown in Fig. 2.

In this article both downlink and uplink video traffic loads are considered. The uplink traffic includes mobile users who capture videos on their MTs and transmit them for online posting, while the downlink traffic includes mobile users performing video streaming. The user arrival rates, λ_{UL} and λ_{DL} for the uplink and downlink traffic loads, respectively, vary through the day [1]. The user service time has an average duration of μ_{UL} and μ_{DL} for the uplink and downlink traffic loads, respectively. The minimum required data rate is R_{UL} and R_{DL} for uplink and downlink users, respectively. A frequency division duplex (FDD) technique is considered at the BSs with available bandwidth of B_s^{UL} and B_s^{DL} at BS s for uplink and downlink traffic, respectively. Hence, BS s can support in total N_s and M_s users simultaneously on the uplink and downlink, respectively.

The channel average power gain, G_{UL} and G_{DL} in the uplink and downlink, respectively, is determined based on the path loss. To determine the average path loss, for BS s with coverage area $A_{s,1}$, the average transmission distance is approximated[1] by H_s, while for coverage area $A_{s,2}$, the transmission distance is approximated by

$$\sqrt{H_s^2 + D^2}.$$

The power consumption of a given BS s in the downlink is determined according to its mode of operation, i.e. active or inactive. For an inactive BS, the power consumption is $P_{L,s}$, which is much smaller than the active power consumption. For an active BS, the power consumption has two parts. The first is a fixed power component $P_{F,s}$, which captures the power consumed by the power supply, cooling, backhaul, and other circuits. The second component is proportional to the call traffic load, and is given by the multiplication of Δ_s (a slope of the load-dependent power consumption of BS s) by the BS transmission power. The BS average transmission power consumption, $P_{s,tx}$, is a function of the average channel power gain (and hence the tilting angle index k_s) and the number of mobile users associated with BS s in the downlink $m_s (\leq M_s)$. Specifically, to support m_s MTs by BS s each with minimum required data rate of R_{DL}, the BS downlink transmission capacity should at least be $m_s R_{DL}$. Using the Shannon formula, the minimum average transmission power $P_{s,tx}(k_s, m_s)$ can be computed. The total average power consumption of an BS s is denoted by $P_{s,DL}(k_s, m_s)$, which has a maximum of $P_{s,mx}$. The power consumption of a given MT in the uplink has two components. The first is a circuit power consumption part, while the second is the average transmission power component. For $n_s (\leq N_s)$ MTs supported by BS s each with minimum required data rate of R_{UL}, the BS uplink transmission capacity should be $n_s R_{UL}$. Using the Shannon formula, the minimum average transmission power consumption per MT associated with BS s, $P_{UL,tx}(k_s, n_s)$, can be computed. The total average power consumption of an MT supported by BS s is denoted by $P_{UL,s}(k_s, n_s)$, which has a maximum of $P_{UL,mx}$.

The number of MTs in the geographical region is denoted by $m = \Sigma_s m_s$ and $n = \Sigma_s n_s$ in the downlink and uplink, respectively. The spatial distributions of MTs in the geographical region follow probability mass functions (PMFs) of $\rho_{DL}(A_{s,1})$ and $\rho_{UL}(A_{s,1})$ for MTs with downlink and uplink traffic, respectively. Table 1 summarizes important mathematical symbols.

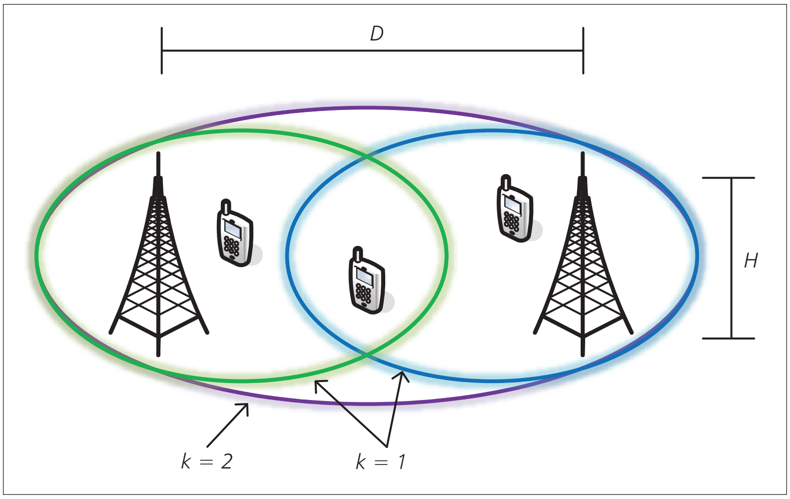

Figure 2. An example of dynamic planning cluster consisting of two BSs. For simplicity, two tilting angles are assumed per BS, leading to two coverage areas per BS.

The Two Time Scale Decision Problem Formulation

Time is divided into two scales, namely slow and fast scales, as shown in Fig. 3. At the slow scale, time is partitioned into a set of periods, $\mathcal{T} = \{1, 2, \ldots, T\}$, with fixed duration τ, that covers the 24 hours of the day. The slow time scale captures the variation in the call traffic density. During one period $t \in \{\mathcal{T}\}$, the uplink and downlink arrival rates $\lambda_{UL}(t)$ and $\lambda_{DL}(t)$ are fixed, and vary from one period to another. At the fast scale, time is partitioned into a set of periods, $\mathcal{I} = \{1, 2, \ldots, I\}$, of equal duration ω, $I = \lceil \tau/\omega \rceil$. The fast time scale captures the MT arrivals and departures. Hence, during one period $i \in \mathcal{I}$, the number of mobile users in the uplink and downlink in the geographical region, $n(i)$ and $m(i)$, are fixed, and may vary from one period to another.

The decision problem model has five elements: decision epochs, states, actions, transition probabilities, and cost. These are discussed for the two time scales in the following.

Slow Time Scale: At the beginning of every period t, the network operators make a decision regarding their BS operation mode and the tilting angle. The system state $\Upsilon(t) = (\gamma_{UL}(t), \gamma_{DL}(t))$ is given by the call traffic load density in the uplink and downlink, $\gamma_{UL}(t) = \lambda_{UL}(t)/\mu_{UL}$ and $\gamma_{DL}(t) = \lambda_{DL}(t)/\mu_{DL}$. The set of system states are pre-known from historical load patterns [1]. Given the uplink and downlink call traffic load densities, the actions specify the BS operation mode and tilting for the current period, that is, $W(t) = (k_1(t), k_2(t))$, where an inactive BS has $k_s(t) = 0$, otherwise $k_s(t) \in \{1, 2\}$. Given the system transmission capacity and users' minimum required data rate, the chosen action $W(t)$ should result in an acceptable service quality within t in the uplink and downlink, for example, $W(t)$ satisfies η_{UL} and η_{DL} target upper bounds on call blocking probabilities in the uplink and downlink, respectively. To provide radio coverage guarantees in the geographical region, both BSs are not allowed to be switched off simultaneously. Furthermore, if one BS is switched off, the

[1] More accurate expressions can be obtained using the distance distribution between BSs and mobile users [13].

other BS must provide radio coverage for the whole geographical region. For the system state $\Upsilon(t)$, the next state transition probability is deterministic and derived from the historical load patterns.

Fast Time Scale: At the beginning of every period i, the network operators decide the BSs' transmission power in the downlink and control the MTs' transmission power in the uplink. The system state $X(i, t)$ gives the number of uplink and downlink MTs in the geographical region, that is, $X(i, t) = (n(i, t), m(i, t))$. For a discrete-time decision problem, the fast time scale evolves as a discrete queuing system. Specifically, within one period i, the arrivals of MTs in the uplink and downlink are described by Bernoulli processes with probability $\lambda_{\mathrm{UL}}(t)\omega/\tau$ in the uplink and $\lambda_{\mathrm{DL}}(t)\omega/\tau$ in the downlink. The service processes follow geometric distributions with parameter $\mu_{\mathrm{UL}}\omega/\tau$ in the uplink and $\mu_{\mathrm{DL}}\omega/\tau$ in the downlink. Hence, the number of MTs being served in uplink and downlink are described by $Geo/Geo/N/N$ and $Geo/Geo/M/M$ queues, respectively [14]. The steady state probabilities of having $n(i, t)$ and $m(i, t)$ MTs in the uplink and downlink and the system transition probabilities can be found from the analysis of the discrete queues in [14]. The fast scale actions are to set the BSs' and MTs' transmission powers, that is, $J(i, t) = (P_{\mathrm{DL},s}(k_s(t), m_s(i, t)), P_{\mathrm{UL},s}(k_s(t), n_s(i, t)))$ for all s. Given the actions taken at period i, the BSs' expected energy consumption, $E_{\mathrm{DL}}(m(i, t))$, and the total expected energy consumption for the MTs with uplink traffic, $E_{\mathrm{UL}}(n(i, t))$, can be obtained. In order to account for the MT energy efficiency in the dynamic planning problem, the fast time scale cost function is modeled as a weighted function of the total downlink and uplink energy consumption, with a weighting factor β, that is,

$$C_{\mathrm{f}}(X(i,t), J(i,t)) = E_{\mathrm{DL}}(m(i,t)) + \beta E_{\mathrm{UL}}(n(i,t)). \quad (1)$$

The weighting factor β captures the impact of the MT energy consumption on the uplink service quality degradation (in terms of call dropping, throughput, etc.). Large values of β imply high impact of MT uplink energy consumption on uplink service quality degradation.

Optimal Solution: Let π_t denote the policy of the fast time scale decision problem in time slot t, that is, π_t is the set of actions taken for all $i \in \mathcal{I}$ at a given t, with a policy space of Π_t. In time slot t with system state $\Upsilon(t)$ and action $W(t)$, let $V_{\pi_t}(X_0(t))$ denote the total value function given initial system state $X_0(t)$ for the fast time scale decision problem. Hence, $V_{\pi_t}(X_0(t))$ is given by averaging $C_f(X(i, t), J(i, t))$ given some action $W(t)$ over system states and time. For the slow time scale decision problem, the immediate cost in time slot t, $C_s(\Upsilon(t), W(t), \pi_t)$, is determined by finding the expectation of $V_{\pi_t}(X_0(t))$ over the initial state $X_0(t)$. The slow time scale policy is denoted by $\pi = \{W(1), \ldots, W(T)\}$, with a policy space of Π. The dynamic planning approach with balanced energy efficiency follows the policies π and π_t for all $t \in \mathcal{T}$ that minimize the total uplink and downlink expected energy cost, that is,

$$\min_{\pi \in \Pi} \min_{\pi_1, \pi_2, \ldots \pi_t} \mathbb{E} \sum_{t=1}^{T} C_s(\Upsilon(t), W(t), \pi_t) \quad (2)$$

where $\mathbb{E}$ denotes the expectation, which is taken over the states $\Upsilon(t)$. In order to solve Eq. 2, we first find the optimal fast time scale policy π_t that minimizes the expected total energy consumption $C_f(X(i, t), J(i, t))$ given the slow time scale action $W(t)$. Then we find the optimal action $W(t)$ that minimizes the expected total energy consumption $C_s(\Upsilon(t), W(t), \pi_t)$.

Symbol	Definition
$B_s^{\mathrm{UL/DL}}$	Uplink/downlink available bandwidth at BS s
$E_{\mathrm{UL}}(n(i, t))$	Uplink total energy consumption at $n(i, t)$
$E_{\mathrm{DL}}(m(i, t))$	Downlink total energy consumption at $m(i, t)$
$J(i, t)$	Fast time scale action
k_s	Tilting angle index for BS s
m_s	Number of mobile users associated with BS s in the downlink
n_s	Number of mobile users associated with BS s in the uplink
$P_{s,\mathrm{DL}}(k_s, m_s)$	Total average power consumption of a BS s
$P_{\mathrm{UL},s}(k_s, n_s)$	Total average power consumption of an MT supported by BS s
$R_{\mathrm{UL/DL}}$	Uplink/downlink minimum required data rate
s	BS index
$W(t)$	Slow time scale action
$X(i, t)$	Fast time scale system state
β	Weighting factor
$\lambda_{\mathrm{UL/DL}}$	Uplink/downlink user arrival rate
$\mu_{\mathrm{UL/DL}}$	Uplink/downlink service time average duration
$\rho_{\mathrm{UL/DL}}(A_{s,1})$	PMF of spatial user distributions for MTs with uplink/downlik traffic
$\Upsilon(t)$	Slow time scale system state

Table 1. Summary of important symbols.

Numerical Results and Discussions

In this section we evaluate the performance of the proposed dynamic planning approach with balanced energy efficiency, by solving Eq. 2, as compared with a traditional dynamic planning approach that does not account for the energy consumption of the mobile users (i.e., with $\beta = 0$), which resembles existing mechanisms, for example, [1] and [7]. The system model is given in

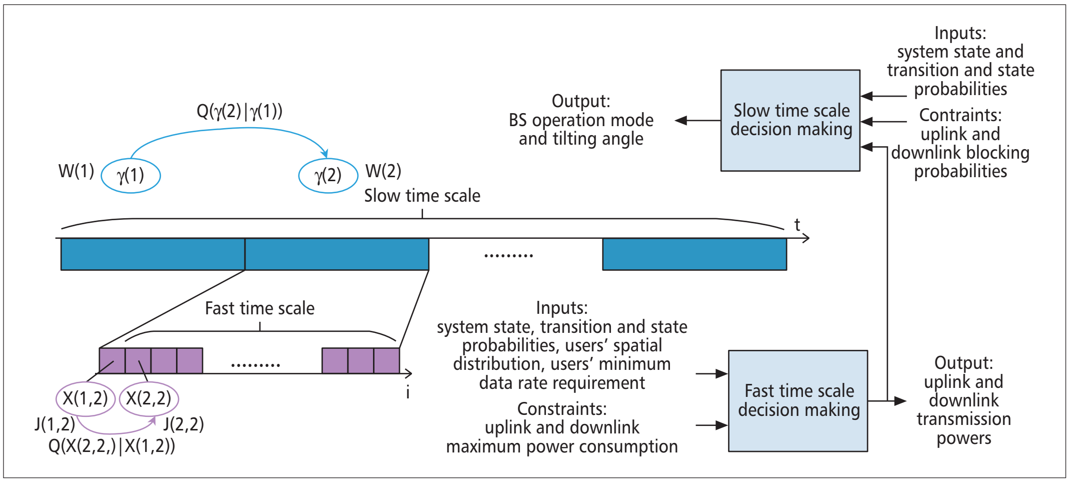

Figure 3. An illustration of the fast and slow time scales under consideration, the system states, actions, transition probabilities, and the decision making process.

Even if the unbalanced approach follows a random or round-robin BS switching off policy to decide which BS should be switched off, the unbalanced approach still will lead to higher expected uplink energy consumption compared with the balanced approach.

Fig. 2. The two BSs are identical and the system parameters are given by H_s = 100 meter, D = 150 meter, $\eta_{\mathrm{UL}} = \eta_{\mathrm{DL}}$ = 0.01, $M_s = N_s$ = 7, $R_{\mathrm{UL}} = R_{\mathrm{DL}}$ = 5 Mb/s, $B_s^{\mathrm{UL}} = B_s^{\mathrm{DL}}$ = 5 MHz, $P_{\mathrm{F},s}$ = 390 watts, Δ_s = 4.7, and $P_{\mathrm{L},s}$ = 75 watts, τ = 1 hour, ω = 5 minutes, and β = 50. The fast time scale arrival rate in the downlink is 0.5, and the average service duration is 0.2 for both the uplink and downlink.

Figures 4a and 4b show the expected downlink energy consumption for both balanced and unbalanced dynamic planning. The unbalanced dynamic planning energy consumption performance does not vary with the weighting factor β since it does not account for the MTs' incurred energy consumption. It is affected only by the arrival rate. At low arrival rates [0.3, 0.6], only one BS is kept active to serve the MTs, while at higher arrival rates greater than 0.6, both BSs are switched on to satisfy the target service quality in terms of minimum required data rates (and hence upper bound on call blocking probabilities). On the other hand, the balanced dynamic planning energy consumption performance is affected by both β and the arrival rate. For low arrival rates and low β, a single BS is kept active to serve the MTs. As the arrival rate increases, a second BS is switched on to satisfy the users' target service quality (in terms of minimum required data rate and call blocking probabilities). In addition, large β values force the second BS activation to avoid uplink service degradation (e.g. higher call dropping rate, lower throughput, and so on) due to MTs' battery depletion. At low arrival rate values, the second BS activation is dominated by large β values, since the uplink service degradation due to MTs' battery depletion is more pronounced than users' call blocking due to limited radio resources, and the opposite is true at high arrival rate values. In Figures 4c and 4d, when a single BS is switched on at arrival rates [0.3, 0.6] and β = 5550 to 1400, respectively, the balanced approach decides which BS should be kept active based on the spatial distribution of the uplink users. Hence, for the balanced approach the second BS is kept active while the first BS is switched off. However, for the unbalanced approach, the expected energy consumption of the uplink users is not accounted for and hence the first BS is kept active while the second BS is switched off. Even if the unbalanced approach follows a random or round-robin BS switching off policy to decide which BS should be switched off, the unbalanced approach still will lead to higher expected uplink energy consumption compared with the balanced approach.

Figure 5 shows the expected uplink energy consumption versus the spatial distribution of uplink users near the proximity of the first BS. The arrival rate for the uplink users is fixed at 0.4. Due to the low arrival rate, only a single BS is kept active (Fig. 4a). As shown in Fig. 5, with more uplink users concentrated around the second BS ($\rho_{\mathrm{UL}}(A_{1,1}) \in [0.1, 0.5]$), the balanced dynamic planning approach keeps the second BS active and switches off the first BS, resulting in low expected energy consumption for the uplink users, unlike the unbalanced approach, which keeps the first BS active and switches off the second BS. As uplink users become more concentrated around the first BS ($\rho_{\mathrm{UL}}(A_{1,1}) > 0.5$), the balanced approach switches off the second BS and keeps the first BS active to keep the expected energy consumption of the uplink users as low as possible.

Conclusions and Future Research

In this article we have proposed a dynamic planning approach with balanced energy efficiency between network operators and mobile users. The proposed approach decouples the MT association and BS operation phases based on a two time scale decision problem. It introduces a new BS switch off metric based on the uplink mobile users' spatial distribution. Also, it accounts for the mobile users' uplink energy consumption in BS wake up, which avoids service quality degradation in the uplink. Our future research will focus on modeling the impact of uplink energy consumption on service quality degradation (e.g. in terms of higher call dropping, lower throughput, and so on) to provide a better representation of the weighting factor β. In addition,

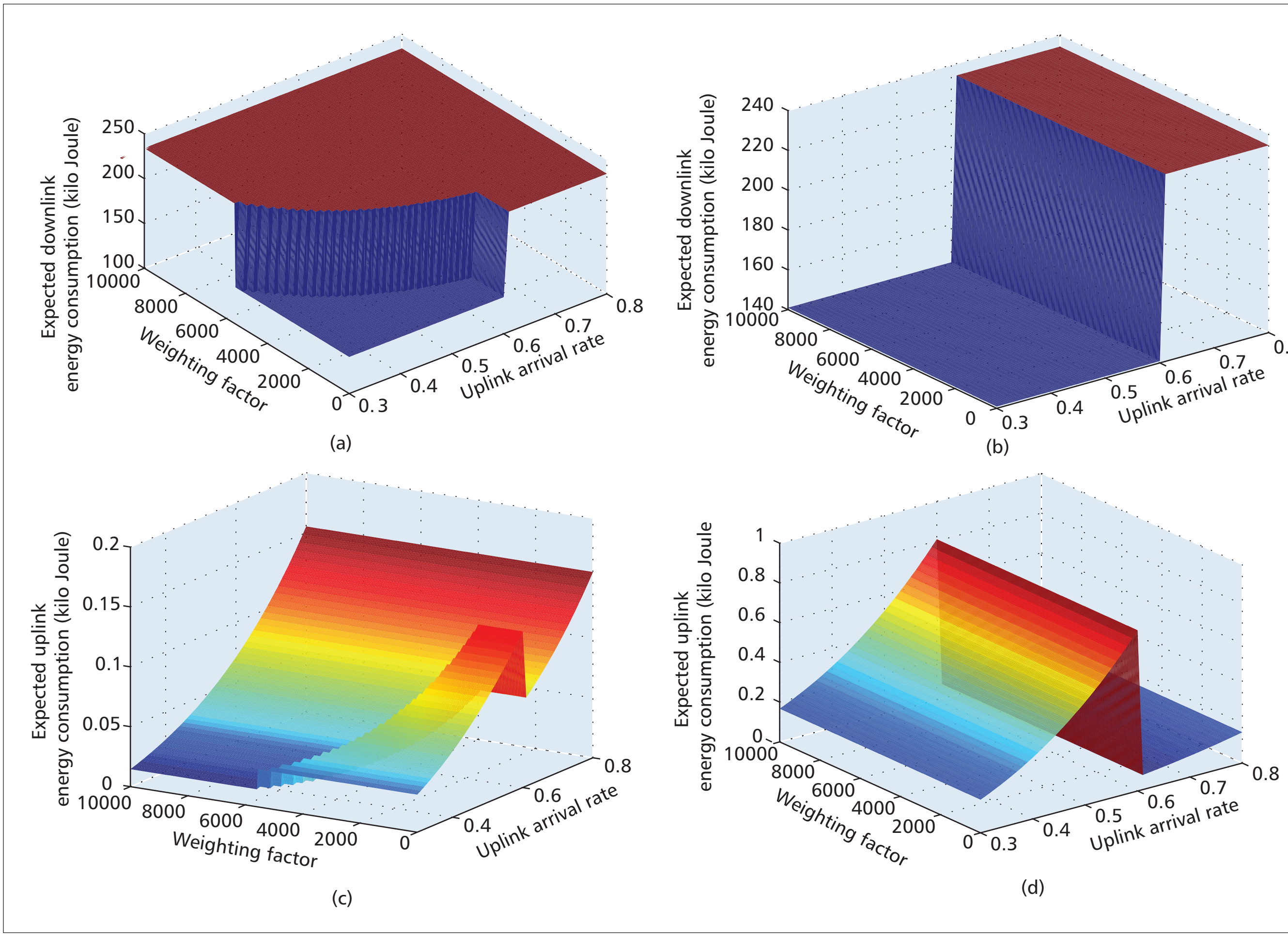

Figure 4. The expected energy consumption versus the arrival rate of uplink users and the weighting factor β: a) downlink expected energy for balanced approach; b) downlink expected energy for unbalanced approach; c) uplink expected energy for balanced approach; d) uplink expected energy for unbalanced approach. The spatial distributions are $\rho_{UL}(A_{2,1}) = 0.7$ and $\rho_{DL}(A1,1) = 0.8$ for uplink and downlink users, respectively.

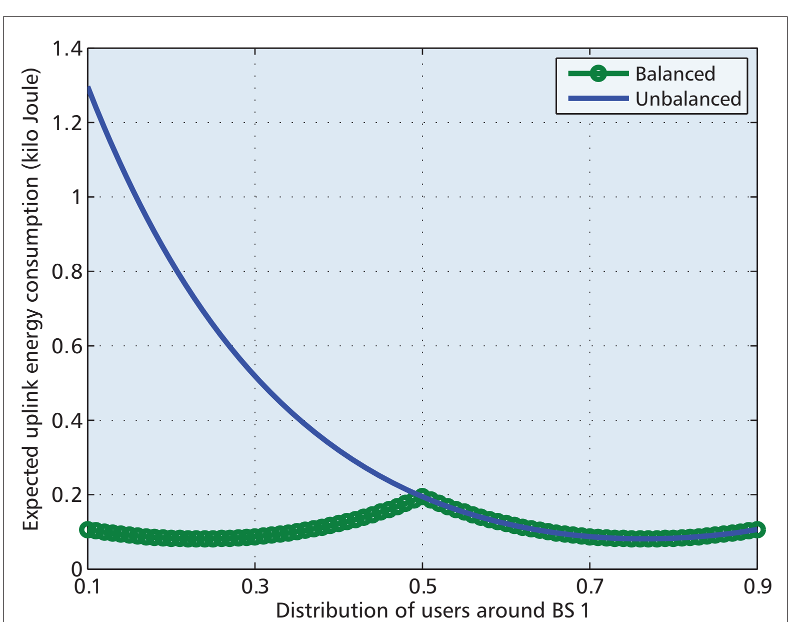

Figure 5. The expected energy consumption of uplink users versus the spatial distribution of the uplink users near the proximity of the first BS. The uplink users' arrival rate is 0.4.

cooperation incentives will be investigated to motivate different networks' cooperation in dynamic planning. Finally, future work will aim to balance backhaul links for improved energy efficiency of cellular networks [15].

Acknowledgment

This publication was made possible by NPRP grant # NPRP 6-415-3-111 from the Qatar National Research Fund (a member of the Qatar Foundation). The statements made herein are solely the responsibility of the authors.

References

[1] M. Ismail and W. Zhuang, "Network Cooperation for Energy Saving in Green Radio Communications," *IEEE Wireless Commun.*, vol. 18, no. 5, Oct. 2011, pp. 76–81.

[2] L. Budzisz *et al.*, "Dynamic Resource Provisioning for Energy Efficiency in Wireless Access Networks: A Survey and an Outlook," *IEEE Commun. Surveys & Tutorials*, vol. 16, no. 4, 2014, pp. 2259–85.

[3] K. Son *et al.*, "Base Station Operation and User Association Mechanisms for Energy-Delay Trade-Offs in Green Cellular Networks," *IEEE JSAC*, vol. 29, no. 8, Sep. 2011, pp. 1525–36.

[4] T. Han and N. Ansari, "On Greening Cellular Networks via Multicell Cooperation," *IEEE Wireless Commun.*, vol. 20, no. 1, Feb. 2013, pp. 82–89.

[5] E. Oh, K. Son, and B. Krishnamachari, "Dynamic Base Station Switching-On/Off Strategies for Green Cellular Networks," *IEEE Trans. Wireless Commun*., vol. 12, no. 5, May 2013, pp. 2126–36.
[6] C. Y. Chang *et al*., "On Optimal Cell Activation for Coverage Preservation in Green Cellular Networks," *IEEE Trans. Mobile Computing*, to appear.
[7] Z. Niu *et al*., "Cell Zooming for Cost-Efficient Green Cellular Networks," *IEEE Commun. Mag*., vol. 48, no. 11, Nov. 2010, pp. 74–79.
[8] N. Saxena, B. J. R. Sahu, and Y.S. Han, "Traffic-Aware Energy Optimization in Green LTE Cellular Systems," *IEEE Commun. Lett*., vol. 18, no. 1, Jan. 2014, pp. 38–41.
[9] J. Wu, S. Zhou, and Z. Niu, "Traffic-Aware Base Station Sleeping Control and Power Matching for Energy-Delay Trade-Offs in Green Cellular Networks," *IEEE Trans. Wireless Commun*., vol. 12, no. 8, Aug. 2013, pp. 4196–4209.
[10] S. Navaratnarajah, A. Saeed, M. Dianati, and M. A. Imran, "Energy Efficiency in Heterogeneous Wireless Access Networks," *IEEE Wireless Commun*., vol. 20, no. 5, Oct. 2013, pp. 37–43.
[11] A. Conte *et al*., "Cell Wilting and Blossoming for Energy Efficiency," *IEEE Wireless Commun*., vol. 18, no. 5, Oct. 2011, pp. 50–57.
[12] D. L. Perez *et al*., "OFDMA Femtocells: A Roadmap on Interference Avoidance," *IEEE Commun. Mag.*, vol. 47, no. 9, Sept. 2009, pp. 41–48.
[13] Y. Zhuang *et al*., "A Geometric Probability Model for Capacity Analysis and Interference Estimation in Wireless Mobile Cellular Systems," *IEEE Globecom*, Dec. 2011, pp. 1–6.
[14] V. Goswami and U. C. Gupta, "Analyzing the Discrete-Time Multiserver Queue Geom/Geom/M Queue with Late and Early Arrivals," *Information and Management Sciences*, vol. 9, no. 2, June 1998, pp. 55–66.
[15] X. Ge *et al*., "5G Wireless Backhaul Networks: Challenges and Research Advances," *IEEE Network*, vol. 28, no. 6, Nov. 2014, pp. 6–11.

Biographies

Muhammad Ismail (S'10, M'13) is an assistant research scientist in the Electrical and Computer Engineering Department at Texas A&M University at Qatar. He received his Ph.D. degree in electrical and computer engineering from the University of Waterloo, Canada in 2013, and the M.Sc. and B.Sc. degrees with honors in electrical engineering (electronics and communications) from Ain Shams University, Cairo, Egypt in 2007 and 2009, respectively. Dr. Ismail's research interests include distributed resource allocation, green wireless networks, cooperative networking, smart grid, and biomedical signal processing. He is a co-recipient of the best paper awards at IEEE ICC 2014, IEEE Globecom 2014, and SGRE 2015. He is a co-author of two research monographs by Springer and Wiley/IEEE Press. He served as a TPC member for the ICWMC in 2010–2014 and IEEE ICC 2014 and 2015. He served on the IEEE INFOCOM 2014 organizing committee. He has been an associate editor with *IET Communications* since 2014. He was an editorial assistant for *IEEE Transactions on Vehicular Technology* from January 2011 to July 2013.

Mohamed Kashef (M'14) is a postdoctoral research associate with Texas A&M University at Qatar. He obtained his Ph.D. in electrical engineering from the University of Maryland at College Park in 2013. He received the B.Sc. and the M.S. degrees with honors in electronics and electrical communications engineering from Cairo University, Cairo, Egypt, in June 2006 and August 2009, respectively. His research interests generally include wireless communication systems and networks, visible light communication networks, and optimization of stochastic systems.

Erchin Serpedin (F'13) is a professor in the Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX, USA. He received the specialization degree in signal processing and transmission of information from Ecole Superieure D'Electricite (SUPELEC), Paris, France, in 1992, the M.Sc. degree from the Georgia Institute of Technology, Atlanta, USA, in 1992, and the Ph.D. degree in electrical engineering from the University of Virginia, Charlottesville, USA, in January 1999. Dr. Serpedin is the author of two research monographs, one textbook, nine book chapters, 110 journal papers, and 180 conference papers. His research interests include signal processing, biomedical engineering, bioinformatics, and machine learning. Dr. Serpedin is currently serving as an associate editor of *IEEE Signal Processing Magazine* and as the Editor-in-Chief of *EURASIP Journal on Bioinformatics and Systems Biology*, an online journal edited by Springer. He has served as an associate editor of dozens of journals, such as *IEEE Transactions on Information Theory*, *IEEE Transactions on Signal Processing*, *IEEE Transactions on Communications*, *IEEE Signal Processing Letters*, *IEEE Communications Letters*, *IEEE Transactions on Wireless Communications*, *Signal Processing* (Elsevier), *Physical Communications* (Elsevier), *EURASIP Journal on Advances in Signal Processing*, and as a Technical Chair for five major conferences. He received numerous awards and research grants.

Khalid Qaraqe (M'97, SM'00) received the B.S. degree in EE from the University of Technology, Iraq, in 1986, with honors. He received the M.S. degree in EE from the University of Jordan, Jordan, in 1989, and he earned his Ph.D. degree in EE from Texas A&M University, College Station, TX, in 1997. From 1989 to 2004 Dr. Qaraqe held a variety of positions in many companies, and he has more than 12 years of experience in the telecommunication industry. Dr. Qaraqe has worked for Qualcomm, Enad Design Systems, Cadence Design Systems/Tality Corporation, STC, SBC, and Ericsson. He has worked on numerous GSM, CDMA, and WCDMA projects, and has experience in product development, design, deployments, testing, and integration. In July 2004 Dr. Qaraqe joined the Department of Electrical and Computer Engineering at Texas A&M University at Qatar, where he is now a professor. Dr. Qaraqe's research interests include communication theory and its application to design and performance, analysis of cellular systems and indoor communication systems. Particular interests are in mobile networks, broadband wireless access, cooperative networks, cognitive radio, diversity techniques, and beyond 4G systems.

The proposed approach decouples the MT association and BS operation phases based on a two time scale decision problem. It introduces a new BS switch off metric based on the uplink mobile users' spatial distribution. Also, it accounts for the mobile users' uplink energy consumption in BS wake up, which avoids service quality degradation in the uplink.