Energy Saving in Cellular Networks by Dynamic RS–BS Association and BS Switching

Zexi Yang and Zhisheng Niu, *Fellow, IEEE*

***Abstract*—In this paper, we propose a dynamic scheme that reduces energy consumption in relay-assisted cellular networks by switching light-loaded base stations (BSs) into sleeping mode and by transferring the traffic load in these cells to neighboring cells using dynamic relay-station–BS (RS–BS) associations. We first investigate how to determine the RS–BS associations to guarantee the coverage when the BS modes are given. The problem is modeled as a generalized assignment problem (GAP) that is NP-hard, and a heuristic algorithm is then proposed. Based on the proposed RS–BS association algorithm, a tabu search approach, which finds the combination of BS modes, is presented. Simulation results show that our proposed scheme makes the active BS sets to dynamically change as the traffic load varies, and in the given scenario, the energy consumption of BSs can be reduced by nearly 40% when the traffic load is low.**

***Index Terms*—Cellular networks, energy saving, relay stations.**

I. Introduction

THE continuously increasing demand for ubiquitous information access triggers the rapid development of the information and communications technology (ICT) industry. Responsible for more than 2% of the energy consumption in the world [1], the ICT industry ranks among the top energy consumers and is expected to grow more rapidly in the future. As a result, energy-saving approaches are urgently required by both the government and network vendors.

Most existing cellular networks are designed to support a fixed traffic load (usually the peak traffic load). However, in many cases, the traffic load in practical systems dramatically varies [2]; therefore, in a low-traffic period, much capacity is wasted. Thus, one basic idea of energy saving is to reduce the redundant capacity by switching off some underutilized resources and by providing on-demand resource provisioning. For instance, switching the BSs in light-loaded cells into sleeping mode is a feasible network-level solution and has been studied in some existing works [1], [3]–[7]. These works show that, by dynamically switching BSs according to the traffic variations, the energy consumption of the network can be reduced significantly. However, most works assume a deterministic traffic variation pattern, and traffic randomness in both time and space domains is not considered.

In addition, when some BSs are switched into sleeping mode, the problem arises that the coverage of these BSs must be guaranteed by neighboring BSs. To solve the problem, our previous work [8] has proposed the concept of cell zooming, which adaptively adjusts the cell size according to the traffic load fluctuations. One of the techniques that implement cell zooming is relaying. Relaying has attracted much attention in recent years and has been adopted in Third-Generation Partnership Project Long-Term Evolution (LTE) Advanced [9]. In relay-assisted cellular networks, relay stations (RSs) are dedicated to forwarding data received from a BS to mobile users (MUs), and *vice versa*. The links between a BS and RSs, RSs and MUs, and a BS and MUs, are called the relay links, access links, and direct links, respectively. With the concept of cell zooming, in the network where RSs are available, RSs can be dynamically reassociated from a cell under a heavy load to a cell under a light load. In this case, the effective cell coverage is adjusted, and the former cell zooms in, whereas the latter zooms out. The RSs are usually deployed in the middle belt or near the edge of a cell area, and in different directions from the center of the cell. When a BS is switched into sleeping mode, for the MUs that were served by the BS, the nearest access node must be one of the RSs in the same cell. Thus, the RSs can be associated with neighboring BSs to guarantee the coverage, and the cell size of the sleeping BS shrinks to zero. Dynamic RS–BS association has not been included into consideration into standard [9]. In [10], an RS handover process that guarantees that the dynamic RS–BS is applicable in realistic networks is proposed. In our previous work [11], a distributed dynamic RS–BS association scheme is proposed to balance the traffic load among different BSs and to reduce system call-blocking probability. In [12], greedy RS reconfiguration algorithms are proposed to guarantee the quality of service (QoS) when some BSs are out of operation. In [13], the performance of relaying is evaluated when a BS is switched into sleeping mode and it is concluded that traffic transferring by relaying is robust to maintain the QoS. In [14], aiming at maximizing energy efficiency, a competitive power control scheme in relay-assisted cellular networks is proposed. In this paper, we assume that each BS has two modes: active and sleeping. We first investigate the RS–BS association problem when BS modes are given. By modeling it as a generalized assignment problem (GAP) $P_{\rm asso}$, we prove that the problem is NP-hard. Problem $P_{\rm asso}$ is an integer programming problem,

Manuscript received January 31, 2013; revised May 17, 2013; accepted May 25, 2013. Date of publication May 31, 2013; date of current version November 6, 2013. This work was supported in part by the National Basic Research Program of China under Grant 2012CB316001, by the National Natural Science Foundation of China under Grant 61201191, Grant 60925002, and Grant 61021001, and by Hitachi Ltd. The review of this paper was coordinated by Dr. I. Krikidis.

The authors are with Tsinghua National Laboratory for Information Science and Technology, Department of Electronic Engineering, Tsinghua University, Beijing 100084, China.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TVT.2013.2265403

and based on this, we construct a linear programming (LP) problem P^* and prove that the feasible solution of P^* can be converted to a feasible solution of P_{asso}. By comparing it with P_{asso}'s linear relaxation problem P_{LP}, we point out that, in some cases, problem P^* may not have feasible solutions, and we further propose an enhanced scheme to solve P_{asso}. Based on the proposed heuristic algorithm that determines the RS–BS associations, a tabu search approach, which identifies the combination of BS modes, is presented. Simulation results show that, compared with a greedy approach and the default case that all BSs are active all the time, our proposed approach can significantly improve the utilization of active BSs and can reduce the total energy consumption of the network.

The remainder of this paper is organized as follows. Section II provides the system model. A dynamic RS–BS association scheme is presented in Section III. Section IV presents a tabu search approach to determine the modes of BSs. Numerical results are given in Section V. Section VI concludes this paper.

II. System Model

Consider the forward links of a two-hop relay-assisted cellular network. The BS set is denoted by $\mathcal{B} = \{\text{BS}_1, \text{BS}_2, \ldots, \text{BS}_N\}$. The RS set is denoted by $\mathcal{R} = \{\text{RS}_1, \text{RS}_2, \ldots, \text{RS}_M\}$. Each RS_m is associated with one BS, which is denoted by $\text{BS}_{n(m)}$. The distance between BS_n and RS_m is denoted by $l_{n,m}$. Due to signal strength degradation, an RS cannot be associated with a BS unless the distance between them is smaller than a threshold denoted by $d^{(\text{th})}$. The set of RSs that are allowed to be associated with BS_n is denoted by $\mathcal{R}_n = \{\text{RS}_m | l_{n,m} < d^{(\text{th})}\}$. Matrix $\mathbf{A} = \{a_{m,n}\}_{M\times N}$, $a_{m,n} \in \{0, 1\}$, is used to indicate whether RS_m is associated with BS_n. When RS_m is associated with BS_n, $a_{m,n} = 1$; otherwise, $a_{m,n} = 0$. Each BS can be switched between two modes, i.e., active and sleeping modes, and RSs can be only associated with active BSs. Denote the set of active-mode BSs at time t by $\mathcal{B}_{\text{on}}(t)$.

The transmission power levels of BSs and RSs are fixed and are denoted by $P_t^{(b)}$ and $P_t^{(r)}$. For the channel model, we only consider large-scale channel gains, including path-loss fading and shadowing. The channel gains at time t between BS_n and an MU in location $\mathbf{x}$, between RS_m and an MU in location $\mathbf{x}$, and between BS_n and RS_m are denoted by $G_n^{(b)}(\mathbf{x}, t)$, $G_m^{(r)}(\mathbf{x}, t)$, and $G_{n,m}(t)$, respectively.

MUs randomly arrive and depart the network. Assume that the arrival process of MUs is a Poisson process, and denote the arrival rate at time t in location $\mathbf{x}$ by $\lambda(\mathbf{x}, t)$. Assume that the traffic size generated from an MU is an exponentially distributed random variable with mean $1/\mu$. Thus, the load density at time t in location $\mathbf{x}$ can be denoted by $\gamma(\mathbf{x}, t) = (\lambda(\mathbf{x}, t)/\mu)$. Each MU is associated with the BS or the RS that has the strongest signal strength. Denote the area where MUs are associated with BS_n at time t by $A_n^{(b)}(t)$ and the area where MUs are associated with RS_m at time t by $A_m^{(r)}(t)$. The traffic loads for BS_n and RS_m at time t can be expressed as $\int_{A_n^{(b)}(t)} \gamma(\mathbf{x}, t)d\mathbf{x}$ and $\int_{A_m^{(r)}(t)} \gamma(\mathbf{x}, t)d\mathbf{x}$, respectively. To avoid intracell interference, simultaneous transmissions in the same channel within one cell are not allowed. Assume that intercell interference (ICI) is well taken care of by a certain interference management tool and is randomized as noise. At time t, the ICI power at location $\mathbf{x}$ in the nth cell is denoted by $I(\mathbf{x}, n, t)$. When BS_n transmits to an MU at location $\mathbf{x}$, the receiving power is given by $G_n^{(b)}(\mathbf{x}, t)P_t^{(b)}$. Then, the signal-to-interference-plus-noise ratio (SINR) of this MU can be expressed as

$$\text{SINR}_n^{(b)}(\mathbf{x}, t) = \frac{G_n^{(b)}(\mathbf{x}, t)P_t^{(b)}}{I(\mathbf{x}, n, t) + n_0} \tag{1}$$

where n_0 is the noise power. Similarly, the SINR between RS_m and an MU at location $\mathbf{x}$ and between BS_n and RS_m are

$$\text{SINR}_m^{(r)}(\mathbf{x}, t) = \frac{G_m^{(r)}(\mathbf{x}, t)P_t^{(r)}}{I(\mathbf{x}, n, t) + n_0}$$

$$\text{SINR}_{n,m}(t) = \frac{G_{n,m}(t)P_t^{(b)}}{I(\mathbf{x}_m, n, t) + n_0} \tag{2}$$

respectively, where $\mathbf{x}_m$ is the location of RS_m. Based on the SINR, the transmission rates are approximately represented by Shannon's formula. The transmission rates between BS_n and an MU at location $\mathbf{x}$, between RS_m and an MU at location $\mathbf{x}$, and between BS_n and RS_m are represented by $r_n^{(b)}(\mathbf{x}, t)$, $r_m^{(r)}(\mathbf{x}, t)$, and $r_{n,m}(t)$, respectively. The average bandwidth requirement (ABR) for the nth cell is defined as

$$\rho_n(t) = \rho_n^{(b)}(t) + \sum_{\forall m, a_{m,n}=1} \rho_{m,n}^{(r)}(t) \tag{3}$$

where $\rho_n^{(b)}(t) = \int_{A_n^{(b)}(t)} (\gamma(\mathbf{x}, t)/r_n^{(b)}(\mathbf{x}, t))d\mathbf{x}$ is the ABR introduced by the MUs directly associated with BS_n, and $\rho_{m,n}^{(r)}(t) = \int_{A_m^{(r)}(t)} (\gamma(\mathbf{x}, t)/r_m^{(r)}(\mathbf{x}, t))d\mathbf{x} + \int_{A_m^{(r)}(t)} (\gamma(\mathbf{x}, t)/r_{n,m}(t))d\mathbf{x}$ is the ABR introduced by the MUs associated with RS_m. In [15], the authors have shown that the traffic load remains nearly constant within about 1 h. Thus, we divide time into periods with the same duration T, and without loss of generality, we assume that the traffic condition within each period are constant. Assume a central node exists in the network. Assume that, at the beginning of each period, the ABRs of each BS and RS can be predicted by measuring and monitoring previous bandwidth requirements, and all these ABRs are reported to the central node.[1] Then, the central node makes decisions on RS–BS associations and BSs modes, and spreads the results. The objective is to minimize the total energy consumption in this period. For ease of exposition, time subscript t is omitted in the following analysis. For an active BS, we adopt the model in EARTH project [16] and express its power consumption as

$$P = P_0 + \delta P_t^{(b)} \tag{4}$$

where P_0 is the fixed part, δ is the variable energy consumption slope, and $\delta P_t^{(b)}$ is the variable part related to transmission power. The typical values of δ for a macro BS and a micro

[1] For the sleeping BSs, the neighboring RSs can help gather traffic information and report to the central node.

BS are 4.7 and 2.6, respectively [16]. For BS_n, the bandwidth utility can be expressed as $\rho_n/\rho_{\mathrm{th}}$, where ρ_{th} is the available bandwidth of one BS. Thus, the energy consumption of BS_n in one time period, if and only it is active, can be expressed as

$$E_n = T\left(P_0 + \delta P_t^{(b)}\frac{\rho_n}{\rho_{\mathrm{th}}}\right). \tag{5}$$

For a sleeping BS, we assume that the power consumption is constant and denote it by P_s. In addition, we assume that, when a BS is switched from sleeping mode to active mode, an amount of energy E_s is consumed. Therefore, the total energy consumption in one period can be expressed as

$$E(\vec{d}) = T\sum_{n=1}^{N}\left[\left(P_0 + \delta P_t^{(b)}\frac{\rho_n}{\rho_{\mathrm{th}}}\right)d_n + P_s(1-d_n)\right] + E_s\sum_{n=1}^{N}\max\{d_n - d_n^-, 0\} \tag{6}$$

where $\vec{d} = \{d_1, d_2, \ldots, d_N\}$, $d_n \in \{0,1\}$ indicates whether BS_n is active in this period, and $d_n^- \in \{0,1\}$ indicates whether BS_n is active in the last period. To guarantee that the system is stable, the QoS is set so that the ABR of each BS ρ_n, $n \in \{1,2,\ldots,N\}$, must be smaller than or equal to the available bandwidth of each BS ρ_{th}. The optimization variables of the energy-saving problem include the RS–BS association matrix $\mathbf{A}$ and the BS modes $\vec{d}$, and the problem can be formulated as

$$\begin{aligned}
&\min_{\vec{d},\mathbf{A}} E(\vec{d})\\
&\text{s.t.} \sum_{n=1}^{N} a_{m,n} = 1 \qquad \forall m \in \{1,2,\ldots,M\}\\
&\quad a_{m,n} \le d_n \qquad \forall m \in \{1,2,\ldots,M\},\ n \in \{1,2,\ldots,N\}\\
&\quad (1-a_{m,n})l_{n,m} \le d^{(\mathrm{th})}\\
&\qquad\qquad \forall m \in \{1,2,\ldots,M\},\ n \in \{1,2,\ldots,N\}\\
&\quad \rho_n \le \rho_{\mathrm{th}} \qquad \forall n \in \{1,2,\ldots,N\}\\
&\quad d_n, a_{m,n} \in \{0,1\}\\
&\qquad\qquad \forall m \in \{1,2,\ldots,M\},\ n \in \{1,2,\ldots,N\}.
\end{aligned} \tag{7}$$

The first and second constraints imply that each RS must be associated with only one active BS. The third constraint guarantees that each RS can be only associated with a BS in the range of $d^{(\mathrm{th})}$. The fourth constraint guarantees the system QoS. Table I summarizes the main notations used in this paper.

III. Traffic Transferring by Relay-Station Association

Due to the high computational complexity, it is impractical to search across all the possible solutions of (7). Thus, here, we investigate the RS–BS association problem to guarantee the QoS when the BS modes are given. The proposed heuristic algorithm provides a method that measures the potential to have more BSs in sleeping mode, based on which a tabu search approach is presented in the following to find out the active BS set.

TABLE I
Notations

Notations	Descriptions
BS_n	The n^{th} BS, $n \in \{1,2,\cdots,N\}$
RS_m	The m^{th} RS, $m \in \{1,2,\cdots,M\}$
$\mathrm{BS}_{n(m)}$	The BS with which RS_m is associated
$\mathcal{R}_n$	The set of RSs that are allowed to be associated with BS_n
$\mathbf{A} = \{a_{m,n}\}_{M\times N}$	The indicator of whether RS_m is associated with BS_n
$\vec{d} = \{d_1,\cdots,d_N\}$	The indicator of whether BS_n is in active mode
$\mathcal{B}_{on}$	The set of BSs that are in active mode
ρ_{th}	The available bandwidth for one active mode BS
ρ_n	The ABR of the n^{th} cell
$\rho_n^{(b)}$	The ABR introduced by the MUs directly associated with BS_n
$\rho_{n,s}^{(b)}$	The ABR introduced by the MUs associated with RS_m
$c_{m,n}$	Defined in (11). Normalized ABR introduced by the MUs associated with RS_m

A. Problem Formulation and Hardness of the Problem

By denoting the number of active BSs by S, we have $S = |\mathcal{B}_{\mathrm{on}}|$. Without loss of generality, assume that $\mathcal{B}_{\mathrm{on}} = \{\mathrm{BS}_1, \mathrm{BS}_2, \ldots, \mathrm{BS}_S\}$. Since an RS must be associated with an active BS, for ease of exposure, in the following, the RS–BS association matrix is redefined as $\mathbf{A} = \{a_{m,s}\}_{M\times S}$, $a_{m,s} \in \{0,1\}$. All the RS–BS association matrices that satisfy the QoS constitute a feasible region $\mathcal{A}$, and the objective is to find out an RS–BS association scheme in the feasible region $\mathcal{A}$. From (7), the constraints that the RS–BS association schemes in $\mathcal{A}$ must satisfy include

$$\begin{aligned}
&\sum_{s=1}^{S} a_{m,s} = 1 \quad \forall m \in \{1,2,\ldots,M\}\\
&a_{m,s} = 0 \quad \forall \mathrm{RS}_m \notin \mathcal{R}_s\\
&\rho_s \le \rho_{\mathrm{th}} \quad \forall s \in \{1,2,\ldots,S\}\\
&a_{m,s} \in \{0,1\} \quad \forall m \in \{1,2,\ldots,M\},\ s \in \{1,2,\ldots,S\}.
\end{aligned} \tag{8}$$

The second constraint in (8) is equivalent to the third constraint in (7). Moreover, we also want to know whether an active BS pattern $\mathcal{B}_{\mathrm{on}}$ has the potential to switch more BSs into sleeping mode. We measure this potential by the aggregate remaining bandwidth of the active BSs. Thus, when looking for an RS–BS association scheme in the feasible region, the optimization objective is to minimize the sum of ABRs of all the BSs, i.e., to maximize the aggregate remaining bandwidth of all active BSs. Accordingly, the RS–BS association problem can be formulated as

$$\begin{aligned}
&\min_{A} \sum_{s=1}^{S} \rho_s\\
&\text{s.t.} \quad \text{constraints in (8)}.
\end{aligned} \tag{9}$$

From (3), we have $\rho_s = \rho_s^{(b)} + \sum_m \rho_{m,s}^{(r)} a_{m,s}$, based on which the third constraint in (8) can be rewritten as[2]

$$\sum_{m=1}^{M} \frac{\rho_{m,s}^{(r)}}{\rho_{\text{th}} - \rho_s^{(b)}}\, a_{m,s} \le 1 \qquad \forall s \in \{1, 2, \ldots, S\}. \tag{10}$$

Since we have $a_{m,s} \in \{0, 1\}$, it can be observed from (10) that, if $(\rho_{m,s}^{(r)}/\rho_{\text{th}} - \rho_s^{(b)}) > 1$, then a feasible RS–BS association scheme must have $a_{m,s} = 0$. Define matrix $\mathbf{C} = \{c_{m,s}\}_{M\times S}$ as follows:

$$c_{m,s} = \begin{cases} c, & \text{RS}_m \notin \mathcal{R}_s \text{ or } \frac{\rho_{m,s}^{(r)}}{\rho_{\text{th}} - \rho_s^{(b)}} > 1 \\ \frac{\rho_{m,s}^{(r)}}{\rho_{\text{th}} - \rho_s^{(b)}}, & \text{otherwise} \end{cases} \tag{11}$$

where c is a constant greater than 1. Thus, $c_{m,s}$ can be treated as the RS$_m$'s ABR normalized by the BS$_s$'s bandwidth available for MUs associated with RSs, and (10) and the second constraint in (8) can be equivalently rewritten as one constraint, i.e.,

$$\sum_{m=1}^{M} c_{m,s} a_{m,s} \le 1 \qquad \forall s \in \{1, 2, \ldots, S\}. \tag{12}$$

Further, the aggregate ABR of all the BSs can be divided into two parts, which are the ABRs introduced by the MUs directly associated with BSs and the ABRs introduced by the MUs associated with RSs as follows:

$$\sum_{s=1}^{S} \rho_s = \sum_{s=1}^{S} \rho_s^{(b)} + \sum_{s=1}^{S}\sum_{m=1}^{M} a_{m,s}\rho_{m,s}^{(r)}. \tag{13}$$

Since the ABRs introduced by the MUs directly associated with BSs $\rho_s^{(b)}$, $s \in \{1, 2, \ldots, S\}$, are not affected by the RS–BS association results, the objective of problem (9) is equivalent to minimizing the aggregate ABRs of MUs associated with RSs. Thus, by replacing the second and third constraints with (12) and by replacing the objective with minimizing the aggregate ABRs introduced by the MUs associated with RSs, problem (9) can be equivalently rewritten as

$$\begin{aligned} \min_{A} \; & \sum_{s=1}^{S}\sum_{m=1}^{M} a_{m,s}\rho_{m,s}^{(r)} && (P_{\text{asso}}) \\ \text{s.t.} \; & \sum_{s=1}^{S} a_{m,s} = 1 && \forall m \in \{1, 2, \ldots, M\} \\ & \sum_{m=1}^{M} c_{m,s} a_{m,s} \le 1 && \forall s \in \{1, 2, \ldots, S\} \\ & a_{m,s} \in \{0, 1\} && \forall m \in \{1, 2, \ldots, M\},\; s \in \{1, 2, \ldots, S\} \end{aligned} \tag{14}$$

and we name problem (14) as P_{asso}.

Theorem 1: The RS–BS association problem (9) is NP-hard. Judging the existence of a feasible solution for (9) is NP-complete.

[2]We assume that the ABR introduced by the MUs directly associated with BS$_s$ $\rho_s^{(b)}$ is smaller than the ABR threshold ρ_{th}. Otherwise, the RS–BS association feasible region $\mathcal{A}$ must be empty.

Proof: Based on the previous derivation, the RS–BS association problem can be equivalently transformed into problem P_{asso}. Problem P_{asso} is a standard format of the GAP [17]. GAP is known to be NP-hard, and the simpler problem of judging the existence of a feasible solution for the GAP problem is NP-complete [17]. Thus, this theorem is proven. ∎

B. Heuristic Algorithm

The following theorem gives a sufficient condition for the existence of feasible solutions of P_{asso}.

Theorem 2: Construct LP problem P^* as follows:

$$\begin{aligned} \min_{A} \; & \sum_{s=1}^{S}\sum_{m=1}^{M} a_{m,s}\rho_{m,s}^{(r)} && (P^*) \\ \text{s.t.} \; & \sum_{s=1}^{S} a_{m,s} = 1 && \forall m \in \{1, 2, \ldots, M\} \\ & \sum_{m=1}^{M} c_{m,s} a_{m,s} \le 1 - c^* && \forall s \in \{1, 2, \ldots, S\} \\ & a_{m,s} \ge 0 && \forall m \in \{1, 2, \ldots, M\},\; s \in \{1, 2, \ldots, S\} \end{aligned} \tag{15}$$

where c^* is defined as

$$c^* = \max_{c_{m,s} \le 1, s\in\{1,2,\ldots,S\}, m\in\{1,2,\ldots,M\}} c_{m,s}. \tag{16}$$

If LP problem P^* has feasible solutions and its optimum is ρ^*, then there exists a feasible RS–BS association solution for problem P_{asso}, and the corresponding objective value of P_{asso} is equal to or smaller than ρ^*.

Proof: By the definition of c^*, if $c_{m,s} > c^*$, then we must have $c_{m,s} > 1$, and then from the second constraint of P_{asso}, we must have $a_{m,s} = 0$. Thus, from (15), if LP problem P^* has a feasible solution and its optimum is ρ^*, then the following LP problem has a feasible solution:

$$\begin{aligned} & \sum_{s=1}^{S}\sum_{m=1}^{M} a_{m,s}\rho_{m,s}^{(r)} \le \rho^* \\ & \sum_{s=1}^{S} a_{m,s} = 1 \quad \forall m \in \{1, 2, \ldots, M\} \\ & \sum_{m=1}^{M} c_{m,s} a_{m,s} \le 1 - c^* \quad \forall s \in \{1, 2, \ldots, S\} \\ & a_{m,s} \ge 0 \quad \forall m \in \{1, 2, \ldots, M\},\; s \in \{1, 2, \ldots, S\} \\ & a_{m,s} = 0 \\ & \quad \text{if } c_{m,s} \ge c^*, \quad \forall m \in \{1, 2, \ldots, M\},\; s \in \{1, 2, \ldots, S\}. \end{aligned} \tag{17}$$

Then, from [18, Th. 2.1], we know that there exists a feasible RS–BS association scheme that has the objective of problem P_{asso} at most ρ^*. ∎

The proof of [18, Th. 2.1] also provides a method to convert a feasible solution of P^* to a feasible solution of P_{asso} that makes the objective of P_{asso} equal to or smaller than ρ^*. This method can be used to generate an RS–BS association scheme from a feasible solution of P^*, and is described as follow [18].

Construct a bipartite graph $\mathbf{B}(\mathcal{V}, \mathcal{W}, \mathcal{E})$. One side of the bipartite graph consists of RSs nodes, i.e.,

$$\mathcal{W} = \{w_m : m = 1, \ldots, M\}. \tag{18}$$

The other side consists of the nodes related to active BSs, i.e.,

$$\mathcal{V} = \{v_{s,k} : s = 1, \ldots, S, k = 1, \ldots, k_s\} \tag{19}$$

where $k_s = \lceil \sum_{m=1}^{M} a_{s,m} \rceil$. The k_s nodes $\{v_{s,k} : k = 1, \ldots, k_s\}$ correspond to BS_s, $s \in \{1, \ldots, S\}$. There will be one or two corresponding edges in the graph $\mathbf{B}(\mathcal{V}, \mathcal{W}, \mathcal{E})$ for each positive $a_{m,s}$. Denote the weight of edge $(v_{s,k}, w_m)$ by $x(v_{s,k}, w_m)$. The weight vector $\vec{x}$ has the following property:

$$a_{m,s} = \sum_{(v_{s,k}, w_m) \in E} x(v_{s,k}, w_m). \tag{20}$$

The construction of the bipartite graph's edges is the same as that in the proof of [18, Th. 2.1], and we omit the details. After constructing the bipartite graph $\mathbf{B}(\mathcal{V}, \mathcal{W}, \mathcal{E})$, the Hungarian algorithm [19] is adopted to solve the maximum weighted matching problem for graph $\mathbf{B}(\mathcal{V}, \mathcal{W}, \mathcal{E})$, and a maximum weighted matching that exactly matches all RSs nodes in the graph $\mathbf{B}(\mathcal{V}, \mathcal{W}, \mathcal{E})$ can be found. For each edge $(v_{s,k}, w_m)$ in the matching, make RS_m associated with BS_s, and we obtain the RS–BS association scheme. In the following, we refer to the given method that converts a feasible solution of P^* to a feasible solution of P_{asso} as Algorithm 1.

Theorem 2 presents a heuristic method to find feasible RS–BS association schemes for P_{asso} when P^* has feasible solutions, by first solving LP problem P^* and then executing Algorithm 1.

C. Analysis and Enhancement

Theorem 2 shows that, as long as a feasible solution of P^* can be found, a feasible RS–BS association solution for P_{asso} can be obtained. To investigate the relationship between problem P^* and problem P_{asso}, denote the linear relaxation problem of P_{asso} as P_{LP}, i.e.,

$$\min_{A} \sum_{s=1}^{S} \sum_{m=1}^{M} a_{m,s} \rho_{m,s}^{(r)} \qquad (P_{\mathrm{LP}})$$

$$\text{s.t.} \sum_{s=1}^{S} a_{m,s} = 1 \qquad \forall m \in \{1, 2, \ldots, M\}$$

$$\sum_{m=1}^{M} c_{m,s} a_{m,s} \le 1 \qquad \forall s \in \{1, 2, \ldots, S\}$$

$$a_{m,s} \text{ is relaxed from } \{0, 1\} \text{ to } [0, 1]$$

$$\forall m \in \{1, 2, \ldots, M\},\ s \in \{1, 2, \ldots, S\}. \tag{21}$$

Thus, the feasible region of P_{asso} is included in the feasible region of P_{LP}, and the optimal value of P_{LP} must be smaller than or equal to the optimal value of P_{asso}. It can be observed that P^* is the same as P_{LP}, except for the introduction of c^*.

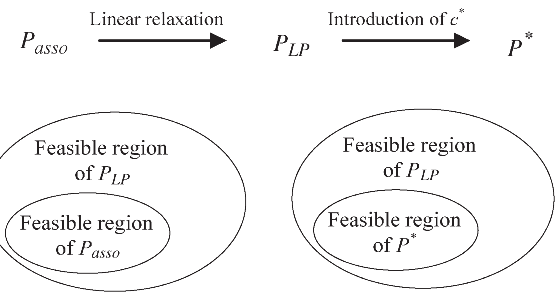


Fig. 1. Relationship of the problems and their feasible regions.

The introduction of c^* makes the feasible region of P^* to shrink compared with P_{LP}. Fig. 1 shows the relationship of problems P_{asso}, P_{LP}, and P^*.

Corollary 1: If $c^* = 0$, then the optimum of P_{asso} is the same as the optimum of P^*.

Proof: Assume $c^* = 0$. We have that problem P^* and problem P_{LP} are the same. Therefore, the optimum of problem P_{LP} is the same as that of problem P^*, which is denoted as ρ^*. Denote the optimum of P_{asso} by ρ^*_{asso}. Since the feasible region of P_{asso} is included in the feasible region of P_{LP}, we have

$$\rho^*_{\mathrm{asso}} \ge \rho^*. \tag{22}$$

On the other hand, from Theorem 2, we have

$$\rho^*_{\mathrm{asso}} \le \rho^*. \tag{23}$$

From (22) and (23), we have $\rho^*_{\mathrm{asso}} = \rho^*$. ■

From Theorem 2 and Corollary 1, we know that, if $c^* = 0$, the optimal solution for the RS–BS association problem P_{asso} can be obtained by solving P^*. Unfortunately, from (11), we know that c^* is the biggest $(\rho_{m,s}^{(r)}/\rho_{\mathrm{th}} - \rho_s^{(b)})$ that satisfies $(\rho_{m,s}^{(r)}/\rho_{\mathrm{th}} - \rho_s^{(b)}) < 1$ and $\mathrm{RS}_m \in \mathcal{R}_s$. Thus, we must have $c^* > 0$ as long as, for some RS_m, its ABR $\rho_{m,s}^{(r)}$ is positive. When each $\rho_{m,s}^{(r)}$, $s \in \{1, \ldots, S\}$, $m \in \{1, \ldots, M\}$, is much smaller than $\rho_{\mathrm{th}} - \rho_s^{(b)}$, from (16), the value of c^* is small, and the feasible region of P^* only shrinks a little compared with P_{LP}. In this case, a suboptimal solution for P_{asso} can be found according to Theorem 2. However, when the value of c^* is larger or even closer to 1, problem P^* may not have feasible solutions. In this case, by replacing c^* with c' that satisfies $c' < c^*$ in P^*, we can construct a new problem named P'. Since $c' < c^*$ and from (15), we know that the feasible region of P' is greater than or equal to the feasible region of P^*. Thus, problem P' may have feasible solutions even when P^* does not. The following corollary gives another sufficient condition that P_{asso} has feasible solutions.

Corollary 2: If a feasible solution $\mathbf{A}^*$ exists for P', and $\mathbf{A}^*$ satisfies the following condition:

$$a_{m,s} = 0 \qquad \forall c_{m,s} > c' \tag{24}$$

then problem P_{asso} has feasible solutions, and $\mathbf{A}^*$ can be converted into a feasible solution for problem P_{asso} using Algorithm 1.

The proof of Corollary 2 is similar to the proof of Theorem 2. The only difference is that c^* in Theorem 2 is replaced with c'. Thus, we omit the proof details of Corollary 2. Since Corollary 2 only provides a sufficient condition that problem P_{asso} has feasible solutions, then it is possible that, even when (24) cannot be satisfied, a feasible solution of problem P_{asso} still exists and can be obtained from a feasible solution of P' by Algorithm 1. Thus, we presents Algorithm 2 as an enhanced scheme to find out a feasible RS–BS association solution.

Algorithm 2: Enhanced scheme to solve problem P_{asso}

1: Construct LP problem P^* by making linear relaxation of P_{asso} and introducing c^*.
2: **if** *the problem* P^* *has a feasible solution* **then**
3: Execute Algorithm 1 to convert the feasible solution of P^* into a feasible solution of P_{asso}.
4: Return the feasible solution of P_{asso}.
5: **end**
6: **else**
7: Construct an array $\vec{c'} = [(Zc^*/Z+1), ((Z-1)c^*/Z+1, \ldots, (c^*/Z+1)]$, where Z is a positive integer and is the length of $\vec{c'}$. Thus, the values in $\vec{c'}$ are in a descending order and in the interval of $[0, c^*]$.
8: **for** each c' in $\vec{c'}$ **do**
9: By replacing c^* in the problem P^* with c', construct a new LP problem named P'.
10: **if** *the problem* P' *has a feasible solution* **then**
11: Execute Algorithm 1 to convert the feasible solution of P' into a solution of P_{asso}
12: **if** *the solution of* P_{asso} *is a feasible solution* **then**
13: Return the feasible solution of P_{asso}.
14: **end**
15: **end**
16: **end**
17: Cannot find a feasible solution of P_{asso}, return
18: **end**

In Algorithm 2, if a feasible solution of P^* is obtained, then it is guaranteed by Theorem 1 and Algorithm 1 that a feasible solution for P_{asso} can be obtained (lines 2–5). Otherwise, we replace c^* with c' that satisfies $0 < c' < c^*$ and solves P' (lines 7–16). From the definition of P^* (15), we know that, if c' approaches c^*, the feasible region of P' shrinks, and we may not be able to find a feasible solution for P'. On the other hand, when c' approaches 0, even if P' has a feasible solution, it may not be able to be converted into a feasible solution of P_{asso}. Fig. 2 shows an example of this dilemma. There may exist an interval for c' between 0 and c^*, within which the solution for P' can be converted into a solution of P_{asso}. In Algorithm 2, we search in $[0, c^*]$ and try to find out a c' that can result in a feasible solution of P_{asso}. The step size of the search is $(c^*/Z+1)$. Thus, the greater the Z, the higher the probability that a feasible solution for P_{asso} is found. However, as the value of Z increases, the complexity of the algorithm increases. In simulation results, we will show the impact of Z's value.

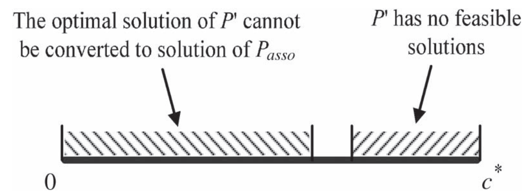


Fig. 2. Value of parameter c' that needs to be carefully searched in the range of $[0, c^*]$.

In Algorithm 2, the number of LP problems needed to be solved is at most $Z+1$, as well as the maximal execution number of Algorithm 1. There are algorithms which can solve the LP problem in polynomial time, including the ellipsoid algorithm with complexity of $O(K^4)$ [20] and the projective algorithm with complexity of $O(K^{3.5})$ [21], where K is the number of optimization variables. In our LP problems P^* and P', the number of optimization variables is $M \times S$. In Algorithm 1, we constructed a bipartite graph $\mathbf{B}(\mathcal{V}, \mathcal{W}, \mathcal{E})$. The number of first part nodes V is at most $M \times S$, and the number of second part nodes W is M. The computational complexity of the Hungarian algorithm, which is adopted to solve the maximum weighted matching problem, is $O(K^3)$, where K is the maximal number of the first part nodes and the second part nodes in the problem. Thus, in our weighted bipartite problem, the maximal number of K is $M \times S$. Therefore, the time complexity of our proposed enhanced RS–BS association scheme Algorithm 2 is polynomial with the number of RSs M and the number of active BSs S.

D. Extend to Multihop Transmissions Between BSs and RSs

In previous discussion, the transmissions between BSs and RSs are restricted to one hop. In practical networks, when deploying BSs and RSs, their locations cannot be accurately planned. It is possible that, when an RS tries to communicate with a BS in one neighboring cell, multihop transmission with the help of other RSs needs to be adopted. Thus, allowing multihop transmissions between RSs and BSs can improve the load transferring capability of RSs. We assume that, when considering multihop transmissions, some routing algorithm already exists. For the scenario where each RS is equipped with a single wireless interface and can be dynamically associated with different BSs, allowing multihop communications between RSs and BSs makes the problem much harder. The reason is shown as follows. Consider the multihop transmission between RS_m and BS_n, and denote the set of intermediate RSs by $\{\text{RS}_{m_1}, \ldots, \text{RS}_{m_{H-1}}\}$, where H is the number of hops between RS_m and BS_n. When RS_m is associated with BS_n, the intermediate RSs for the communication between them must also be associated with BS_n, and this can be expressed as

$$a_{m,n} \leq a_{m_h,n} \quad \forall h \in \{1, \ldots, H-1\}. \tag{25}$$

After incorporating (25) into constraints, problem (7) becomes intractable. However, for the following two scenarios, when considering multihop communications between RSs and BSs, the problem can be formulated without incorporating (25) as a constraint.

- Each RS is equipped with multiple interfaces and can be used for transmissions in different cells simultaneously.
- The RSs can be divided into two categories. The first category RSs are statically associated with BSs, and the remaining RSs can be dynamically associated with different BSs. The intermediate RSs for the multihop transmissions between the second-category RSs and BSs consist of only the first-category RSs.

In the given two scenarios, our proposed dynamic RS–BS association scheme can be extended to the case of multihop communications. According to Shannon's formula, the data rate between two RSs RS_{m_1} and RS_{m_2} can be calculated and denoted by $r^{(r)}_{m_1,\,m_2}$. Since it is assumed that, to avoid intracell interference, multiple transmissions cannot simultaneously happen in one cell, the ABR of the MUs associated with RS_m is the sum of ABRs on each hop from the MUs to BS_n and can be expressed as

$$\rho^{(r)}_{m,\,n}(t) = \int\limits_{A^{(r)}_m(t)} \frac{\gamma(x,\,t)}{r^{(r)}_m(\mathbf{x},\,t)}\, d\mathbf{x} + \sum_{h=0}^{H-2} \int\limits_{A^{(r)}_m(t)} \frac{\gamma(\mathbf{x},\,t)}{r^{(r)}_{m_h,\,m_{h+1}}(t)}\, d\mathbf{x} + \int\limits_{A^{(r)}_m(t)} \frac{\gamma(\mathbf{x},\,t)}{r_{n,\,m_{H-1}}(t)}\, d\mathbf{x} \quad (26)$$

where $m_0 = m$. In addition, $\mathcal{R}_n$, the set of RSs that are allowed to be associated with BS_n needs to be redefined. For example, $\mathcal{R}_n$ can be defined to restrict the maximal number of hops between RSs and BSs. By substituting $\rho^{(r)}_{m,\,n}(t)$ in (3) with the $\rho^{(r)}_{m,\,n}(t)$ expressed in (26) and by redefining $\mathcal{R}_n$, the RS–BS association problem considering multihop communications between BSs and RSs can be still formulated by (9), and the enhanced RS–BS association scheme can be still applied.

IV. Tabu Search Approach

Earlier, a heuristic algorithm is proposed to solve the RS–BS association problem. However, to minimize the total energy consumption in one period, it still requires high computational complexity to find out the optimum active BS set among all the $2^{|N|}$ sets. Here, we present a tabu search approach to determine the modes of BSs.

Tabu search, which is introduced in [22] and [23], is a metaheuristic algorithm that exploits a collection of principles for intelligent problem solving. It includes three primary themes [22]: 1) the use of flexible attribute-based memory structures designed to permit evaluation criteria and historical search information to be exploited; 2) an associated mechanism of control based on the interplay between conditions that constrain and free the search process; and 3) the incorporation of memory functions of different time spans, from short term to long term, to implement strategies to intensify and diversify the search.

Here, we briefly describe the short-term memory process, which is the core of tabu search. All the solutions constitute a solution set. By defining the structure of the solution set, each solution has several neighboring solutions. Given an initial solution, the short-term memory phase makes the search process by moving to one neighboring solution iteratively. To prevent from falling into local optimality, the process maintains an attribute list, which is called tabu list. The solutions that have any of these attributes are called tabus and are prevented. Meanwhile, to prevent from missing better solutions, an aspiration criterion is set, and a solution satisfying the aspiration criterion can be moved to even if it is a tabu. A commonly adopted rule to stop the short-term memory process is to set the number of iterations without improvement.

In our scheme, we only adopt a short-term memory process, and a long-term memory phase is not used. To apply the tabu approach to our BS switching problem, in the following, we first define the structure of the solution space and then develop the short-term memory phase.

A. Structure of Solution Space

In Section II, we denote a BS mode solution[3] by $\vec{d} = \{d_1, d_2, \ldots, d_N\}, d_n \in \{0, 1\}$. The set of all solutions is denoted by Π. For two arbitrary solutions $\vec{d}^{(0)} = \{d^{(0)}_1, \ldots, d^{(0)}_N\}$ and $\vec{d}^{(1)} = \{d^{(1)}_1, \ldots, d^{(1)}_N\}$, we define the distance between them by

$$\sum_{n=1}^{N} \left| d^{(0)}_n - d^{(1)}_n \right|. \quad (27)$$

Two solutions are said to be neighbor to each other if their distance is 1. Thus, for a given solution, there are two kinds of operations that can generate a neighborhood solution. The first kind is to switch a BS in sleeping mode to active mode, and we name this kind of operations as *add moves*. The second is to switch a BS in active mode to sleeping mode, and we name this kind of operations as *drop moves*.

In Section III, we propose an approach to find a feasible RS–BS association scheme to guarantee the QoS when the BS modes are given. However, each solution has no feasible RS–BS association schemes. If the RS–BS association approach returns a feasible RS–BS association for a solution, we name this solution a *feasible solution*. A move to a feasible solution is named as a *feasible move*. On the contrary, the solutions for which a feasible RS–BS association cannot be found are named *infeasible solutions*, and the moves to infeasible solutions are called *infeasible moves*.

It should be noted that the infeasible solutions cannot be selected as the final solution. However, when exploring the neighbors of a given solution in a heuristic-based method, restricting the search to feasible solutions tends to generate few

[3]In the following, we will name a BS mode solution as a solution for short.

trial solutions, forcing the process to quickly terminate, often with low-quality solutions. Allowing violation of feasibility when searching provides a higher level of flexibility, and much larger areas of the solutions' space can be explored. Thus, in our proposed tabu search approach, the violation of feasibility is allowed.

B. Short-Term Memory Phase

In the short-term memory phase, to proceed the search process, a neighborhood solution is selected from the neighborhood solution set of the current solution. Tabu restriction is implemented by defining a tabu list. Our scheme maintains two tabu lists. The first list records the BSs that are added in the recent l_{tb} moves and prevents the dropping of the BSs recorded in the next move. On the contrary, the second list records the BSs that are dropped in the recent l_{tb} moves and prevents the adding of the BSs recorded in the next move. The solutions that are restricted by the two tabu lists are said to be *unadmissible*, and other solutions are said to be *admissible*. The moves to admissible (unadmissible) solutions are said to be *admissible (unadmissible) moves*. The aspiration criterion is applied to a solution when it is feasible, and its energy consumption is smaller than the smallest energy consumption for the feasible solutions ever found.

Based on the concepts and the definitions of neighborhood structure, tabu list, aspiration criteria, and different kinds of moves, the following rules define the generation mechanism, from which the search process moves to the next solution. Each time the generation mechanism is executed, the rules are checked one by one until the next move is found.

1) Among all feasible moves, if there is a solution with total energy consumption smaller than the total energy consumption of solutions ever found, then move to this solution.
2) Among all admissible feasible drop moves, find the solution that returns the most aggregate remaining bandwidth, and make this solution as the next move.
3) Among all admissible feasible add moves, find the solution that returns the most aggregate remaining bandwidth, and make this solution as the next move.
4) Among all admissible infeasible add moves, find the solution that switches the sleeping-mode BS with the most traffic load into active mode as the next move.
5) Among all admissible infeasible drop moves, find out the solution that switches the active-mode BS with the least traffic load into sleeping mode as the next move.

The rationale behind the given rules is to switch BSs into sleeping mode as long as the QoS can be guaranteed and to switch BSs into active mode if the QoS cannot be satisfied. The fourth and fifth steps in the given rules show that the moves to infeasible solutions are also allowed. However, compared with the infeasible solutions, the feasible ones are inclined since the first three steps in the rule all aim at finding feasible solutions. The short-term memory phase is terminated after J times of iterations without improvement. Algorithm 3 summarizes the process of the proposed short-term memory phase.

Algorithm 3: Process of short-term memory phase

Given: the solution of the last period $\bar{d}_0$, the traffic load and ABR of each BS and RS, and the maximal iteration time without improvement for the short-term memory phase J.
Set: $j \leftarrow 0$, $k \leftarrow 0$.
while $j < J$ **do**
  $\Pi^{(k)} \leftarrow$ the set of neighbor solutions of $\bar{d}_k$.
  for each solution in $\Pi^{(k)}$ **do**
    Try to find out a feasible RS–BS association scheme for this solution by Algorithm 2, and record the aggregate remaining bandwidth if a feasible RS–BS association scheme can be found.
  end
  Find out the next move from $\Pi^{(k)}$ based on the five rules described earlier, and set it as $\bar{d}_{k+1}$.
  Update the tabu list, and update $k \leftarrow k + 1$.
  if $k = \arg\min_{w \in \{0,1,\ldots,k\}} E(\bar{d}_w)$ **then**
    $j \leftarrow 0$
  end
  else
    $j \leftarrow j + 1$
  end
end
Return the solution $\bar{d}_{k^*}$, where $k^* = \arg\min_{w \in \{0,1,\ldots,k\}} E(\bar{d}_w)$.

C. Implementation Issue

To implement our proposed dynamic RS–BS association scheme and the tabu search approach, the following arrangements and changes need to be applied to the current cellular networks.

- Each RS needs to record the BSs that are available to be associated with. For the scenarios that multihop transmissions between BSs and RSs are allowed, the route information needs to be stored in the intermediate RSs. Since BSs and RSs are fixedly deployed in the network, the available associated BS list and the route information can be treated as static information and set in the network initialization stage.
- A central node in the network is needed. At each execution time, the central node gathered information from all the BSs. The gathered information includes the modes of all BSs and the ABRs of each BS and RS. After calculation, the central node broadcasts the result to all the BSs and RSs. In cellular networks, a BS controller (BSC) is deployed to control a set of BSs. A typical BSC has tens or even hundreds of BSs under its control. A BSC and the BSs under its control are connected by wired lines and can conveniently exchange information. Thus, the central node in our proposed scheme can be deployed in BSCs.
- An RS handover mechanism is needed. Dynamic RS–BS association has not been incorporated into standards [9].

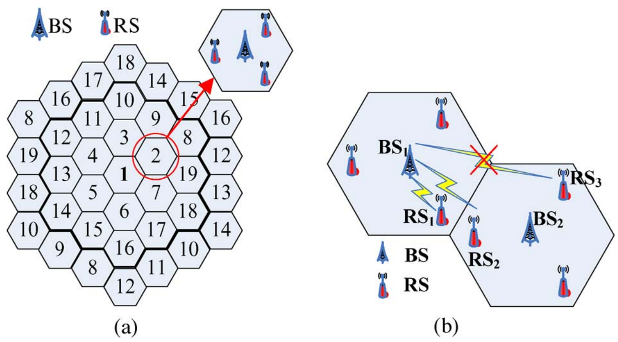


Fig. 3. Simulation topology. (a) Consider a 19-cell cellular network with wrap-around to avoid boundary effect. (b) RSs can be associated with BSs that are within the range of 600 m.

In [10], the LTE specification is extended to support the handover mechanism of RSs in cellular networks. In the handover process proposed in [10], the source BS sends information about the RS to the target BS that prepares radio resources to admit the RS based on the received information. Then, the RS is informed to make reassociation. In the scenario where the two BSs are synchronized with each other, the reassociation process is transparent to the MUs, and only the RS needs to detach from the source BS and synchronize with the target BS. After the reassociation of the RS, the resources used by the RS in the source BS can be released. This handover process guarantees that dynamic RS–BS association is practical in cellular networks.

V. Simulation and Discussion

A. Simulation Setup

Consider a 19-cell cellular network with wrap-around to avoid boundary effect, as shown in Fig. 3(a). The radius of one cell is 500 m. In each cell, three RSs are evenly deployed, and the distance between a BS and an RS in one cell is set as 300 m. Due to signal strength degradation, it is assumed that an RS can be associated with a BS within 600 m. That is, an RS can only be associated with the BS in the same cell or the nearest neighboring BS, as shown in Fig. 3(b). Each BS is allowed to be switched into sleeping mode to reduce energy consumption. Transmission power levels of BSs and RSs are set as 43 and 37 dBm. Assume that ICI is randomized as noise, and the noise plus interference power is -70 dB. We adopt the path-loss channel model proposed in [24], which has been verified by practical measurements in some urban areas in Japan. MUs randomly arrive in the network. We assume that the traffic is uniformly distributed in one original cell area.[4] The traffic load in each original cell area is described by an arrival rate. The file size of each MU is an exponentially distributed random variable, and the mean file size is 16 Mbits. We simulate the ABRs by Monte Carlo simulation, and the results are shown as follows. Take Fig. 3(b) as an example, and denote the arrival rate in the original cell area of BS$_1$ by λ. When BS$_1$ is active, the arrival rate of MUs directly associated with BS$_1$ is 0.45λ, and the bandwidth requirement of this part MUs is 1.88λ MHz. The arrival rate of MUs associated with RS$_1$ is 0.18λ, and the bandwidth requirement of this part MUs is 1.01λ MHz if RS$_1$ is associated with BS$_1$, and 1.12λ MHz if RS$_1$ is associated with BS$_2$. When BS$_1$ is sleeping, RS$_1$ is associated with BS$_2$ to transfer the traffic load from BS$_1$ to BS$_2$. In this case, the arrival rate of MUs associated with RS$_1$ is 0.33λ, and the ABR of this part MUs is 2.67λ MHz. For comparison, the scenario without RSs is also considered in simulation. In the scenario without RSs, when all BSs are active, the arrival rate of MUs associated with each BS is λ, the bandwidth requirement of these MUs is 5.74λ MHz. When a BS is sleeping, the MUs that were served by it need to be reassociated with the nearest neighboring BSs, and the bandwidth requirement of these MUs is 35.85λ MHz. It can be observed that, when some BS is sleeping, the bandwidth requirement for the MUs that are served by neighboring BSs is very high. The reason is that, when no RSs can help a relay signal, the communication between an MU and a neighboring BS suffers from severe path loss due to long transmission distance. These simulated bandwidth requirements are used in the following simulation. The bandwidth of each active BS is set as 5 MHz. To guarantee the QoS, the bandwidth consumed for one active BS must be smaller than or equal to 5 MHz. Assume that the traffic loads in different original cell areas are distinct. As authors of [25] has pointed out that the spatial statistic of cellular traffic can be described by a lognormal distribution, we want to generate traffic loads that satisfy the following conditions: 1) At any time, the traffic load of one original cell area should be a random variable with lognormal distribution; 2) at any time, the traffic loads of different original cell areas are independent of each other; 3) the traffic load of one original cell area should vary continuously. Denote the arrival rate in the original nth cell area at time t by $\Lambda_n(t)$. First, generate independent identically distributed normal distribution variables $\alpha_0^{(n)} \sim \mathcal{N}(\eta, \sigma^2)$ and $\alpha_{T_s}^{(n)} \sim \mathcal{N}(\eta, \sigma^2), n \in \{1, 2, \ldots, N\}$. Then, generate $\alpha_t^{(n)}, n \in \{1, 2, \ldots, N\}$, following $\alpha_t^{(n)} = (t'\alpha_{T_s}^{(n)} + (1-t')\alpha_0^{(n)} - \eta/ 2t'^2 + 1 - 2t') + \eta$, where $t' = (t/T_s)$, and T_s is the end time. It can be easily verified that $\Lambda_n(t) = e^{\alpha_t^{(n)}}, n \in \{1, 2, \ldots, N\}$, $t \in \{0, T_s\}$, satisfies the given conditions. The average arrival rate in each original cell area η is used to denote the intensity of the traffic load, and the variance coefficient (VC) defined as σ^2/η^2 is used to denote the variance of the traffic load.

B. Evaluation of the RS–BS Association Scheme

To verify the performance of the RS–BS association scheme proposed in Section III, we randomly generate all the BS modes and traffic load, and execute the RS–BS association scheme to check the probability that a feasible RS–BS association solution can be found. For all the BSs, the probabilities that they are active are the same, and they are independent of each other. For each simulated value of the probability that BSs are active, the BS modes and traffic load are randomly generated 10^4 times, as well as the executions of the scheme.

[4]The original cell area has a hexagonal shape, as shown in Fig. 3. When our scheme is applied, the real coverage of each cell will change as the associations of RSs change.

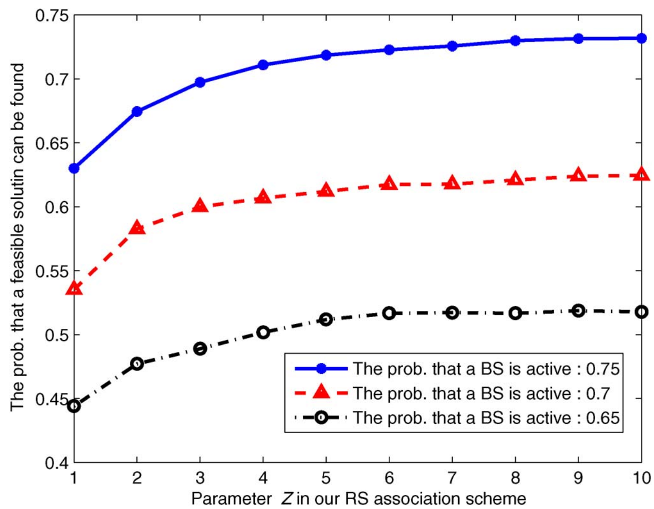


Fig. 4. Impact of parameter Z to the performance of our proposed RS–BS association scheme. The traffic intensity η is set as 0.2, and the VC of the traffic is set as 0.4.

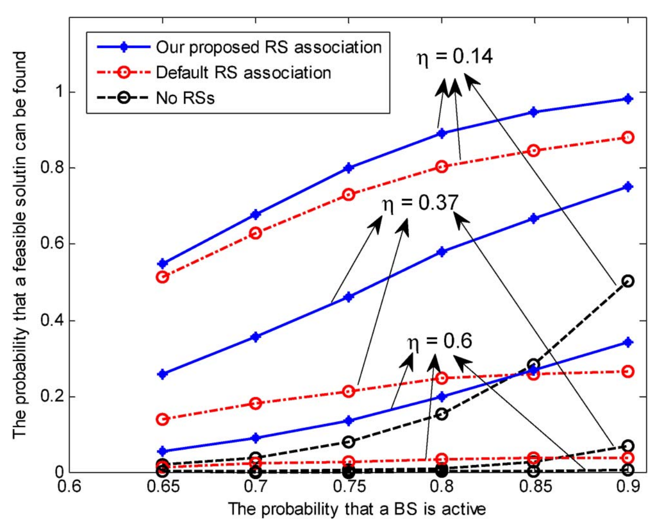


Fig. 5. Comparison of different RS–BS association schemes. The VC of the cell traffic is set as 0.4.

In our proposed scheme to identify a feasible RS–BS association solution, i.e., Algorithm 2, there is parameter Z that indicates the number of LP problems that we construct. As the value of Z increases, the probability that a feasible RS–BS association scheme exists increases, as well as the complexity of the algorithm. We compare the proposed RS–BS association schemes with different values of Z to evaluate its impact. The result is shown in Fig. 4. It shows that, when Z is greater than 6, the probability that a feasible solution can be found increases slowly. Thus, in the following simulations, the value of Z is set as 6. We also compare our proposed RS–BS association scheme with the scenario without RSs and the default association algorithm, with which each RS is associated with the nearest BS. We compare these three algorithms with different traffic intensity values η, and the VC of the traffic is set as 0.4. The simulation results are shown in Fig. 5. It can be observed that the performance of the default association scheme is apparently poorer than our proposed scheme, particularly when the traffic

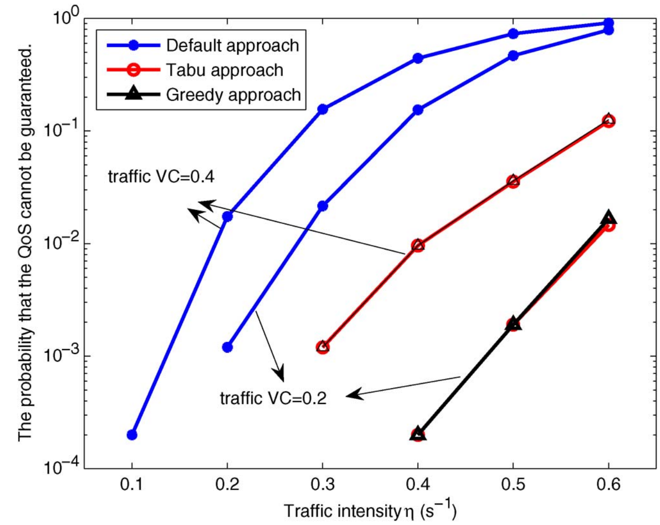


Fig. 6. Probability that the QoS cannot be satisfied.

load is heavier. This shows that, compared with the default association scheme, to serve the same amount of traffic, our proposed scheme needs fewer active BSs. Fig. 5 also shows, that in the scenario without RSs, the performance is much poorer than that of the scenario where RSs can help transfer traffic, particularly when traffic load is relatively heavy. This is because, when there are no RSs, the MUs associated with neighboring BSs will consume much bandwidth, as pointed out in the Monte Carlo simulation results earlier.

C. Evaluation of the Whole Approach

We compare our approach with two other approaches. The first approach is the default approach, which has all BSs to be active all the time and each RS to be associated with the nearest BS. The second approach is a greedy approach described as follows. The greedy approach contains two phases. In the first phase, the sleeping BS that has the most traffic load in its original cell coverage is switched into active mode until the QoS can be satisfied or until all BSs are active. In the second phase, the active BS that has the most remaining bandwidth capacity is switched into sleeping mode as long as the QoS can be satisfied. The rationale of the greedy approach is to switch BSs into sleeping mode on the condition that the QoS can be satisfied. For tabu search, the length of tabu list l_{tb} is set as 8, and the maximal iteration time without improvement for the short-term memory phase J is set as 10. Simulation time is set as 10 h. The time interval between two adjacent executions of the scheme, i.e., the duration of one period, is set as 1 h. Since the traffic varies in both time and space domains, it may happen that, in some period, the traffic is too heavy and the QoS cannot be guaranteed even when all BSs are active. Fig. 6 shows the probability that the QoS cannot be guaranteed for different approaches and traffic conditions. It can be observed that, compared with the default approach, our tabu approach and the greedy approach have much lower probability that the QoS cannot be satisfied. This demonstrates the load balancing effect of our tabu approach and the greedy approach.

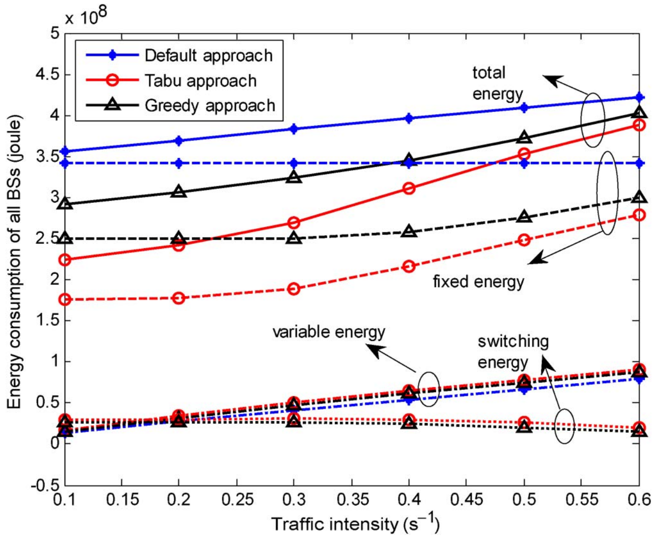


Fig. 7. Impact of different traffic conditions to the energy consumption.

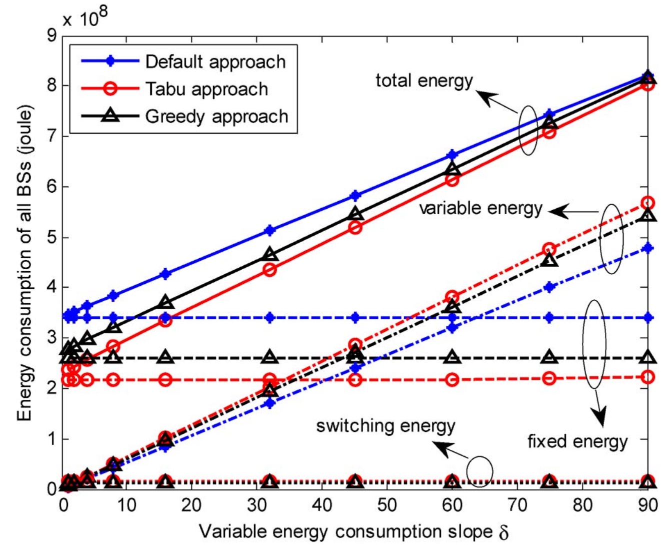


Fig. 9. Impact of the variable energy consumption slope δ to the energy consumption.

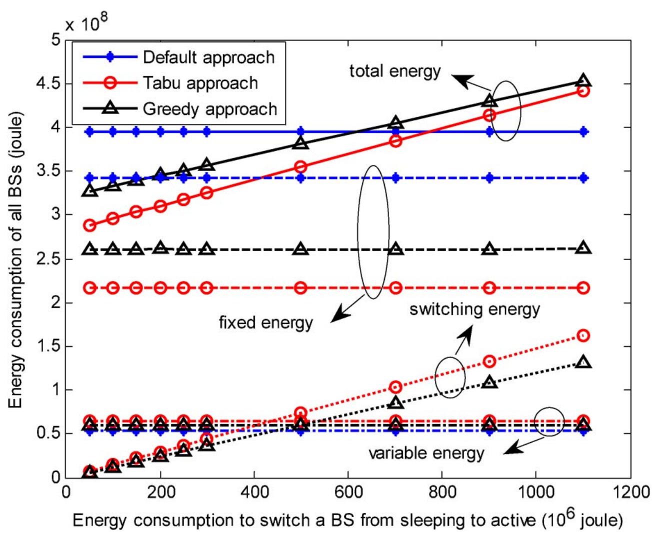


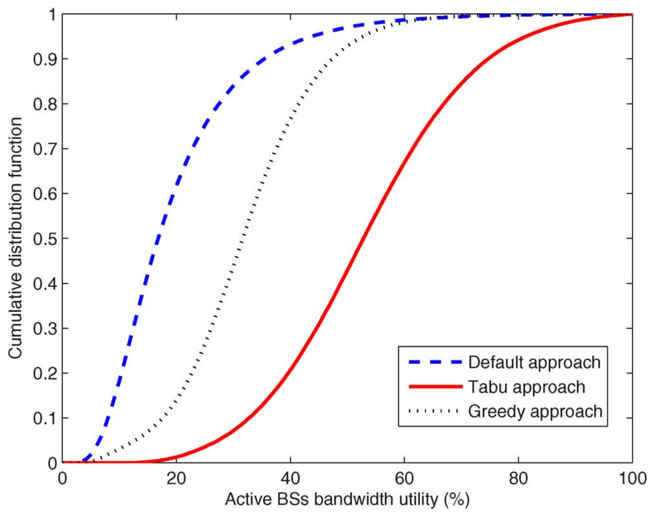
Fig. 8. Impact of the energy consumed when a BS is switched from sleeping mode to active mode to the energy consumption.

Figs. 7–9 compare the energy consumption of different approaches. To clearly observe the consisting of the energy consumption, we divide the energy consumption into three parts and plot them separately. The first part is the fixed energy, including the fixed energy for active BSs and sleeping BSs. The second part is the variable energy that is related to the bandwidth utility and transmission power. The third part is the switching energy that is used to switch BSs from sleeping mode to active mode. The default value of energy consumption to switch a BS from sleeping mode to active mode E_s is set as 3.6×10^5 J. The fixed power consumption for an active BS P_0 is set as 500 W. The power consumption for a sleeping BS P_s is set as 50 W. For the variable energy consumption slope δ, the default value is set as 10. The default value of traffic intensity η is $0.4s^{-1}$. The default value of VC of the traffic is set as 0.4.

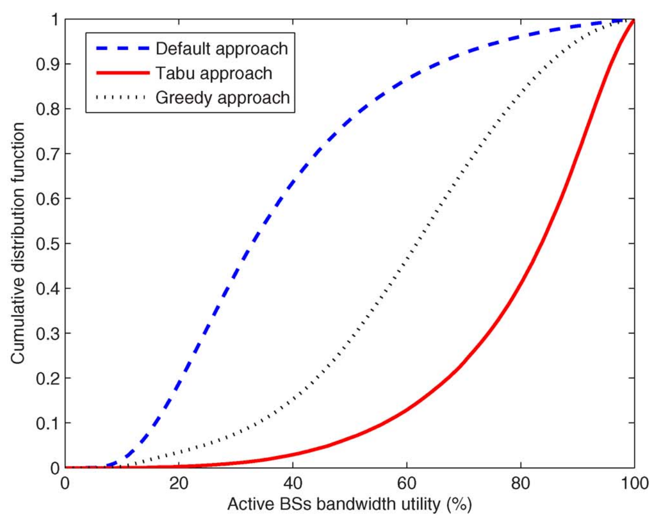
Fig. 7 shows the impact of traffic intensity. It can be observed that, when the traffic load is low, our proposed approach can reduce energy consumption by nearly 40% and 25%, compared with the default approach and the greedy approach, respectively. This energy-saving gain mainly comes from the fixed energy consumption part. As the traffic load increases, more BSs need to be active to guarantee the QoS, and the energy consumption gain of our proposed approach decreases. When the traffic load increases, load balancing is the main effect of our tabu approach, as shown in Fig. 6. It can be also observed in Fig. 7 that the variable energy increases as the traffic load increases. This is reasonable since, for an active BS, its variable energy is proportional to its bandwidth utility, which reflects the traffic load. The switching energy decreases as the traffic load increases. The reason is that, when traffic load becomes heavier, more BSs need to keep active to guarantee the QoS, and the switching of BSs becomes less frequent.

Fig. 8 shows the impact of the energy consumed when a BS is switched from sleeping mode to active mode. It can be observed that the fixed energy consumption and the variable energy consumption of our approach remain the same as the energy consumption of switching increases, which means that the decisions of our approach are almost independent to the switching energy consumption. The reason is that our approach only considers minimizing the energy consumption of the current period, thus would switch BSs into sleeping mode as long as the QoS can be satisfied, and will not consider the energy consumed to switch them into active mode in the future. Therefore, before applying our approach, it should be evaluated whether the energy consumption for switching a BS from sleeping mode into active mode is small enough.

Fig. 9 shows the impact of variable energy consumption slope δ. It can be observed that the energy consumption of our approach and that of the default approach become closer as δ increases. The typical value of δ is below 10 [16], and our approach can decrease energy consumption by at least 30%, compared with the default approach in that region.



(a)



(b)

Fig. 10. Active BS bandwidth utility comparison. (a) Traffic intensity is $0.2s^{-1}$. VC of the traffic is 0.2. (b) Traffic intensity is $0.4s^{-1}$. VC of the traffic is 0.4.

In Section II, we defined bandwidth utility for one BS as the ratio of the BS's bandwidth requirement and the BS's available bandwidth. High bandwidth utility indicates that the BS is fully utilized and that the energy is not wasted. To compare the bandwidth utility of BSs for different approaches, we compare the cumulative distribution function (cdf) of each active BS's bandwidth utility. The result is shown in Fig. 10. It can be observed that our tabu approach significantly moves the cdf curve to the right side, which shows that the bandwidth utility for active BSs is dramatically improved.

VI. Conclusion

In this paper, aiming at reducing energy consumption, we have investigated dynamic BS switching and RS–BS association in cellular networks. The RSs can help transfer data among different BSs. In the case where the traffic load is low, RSs help aggregate traffic to some BSs, whereas other BSs can be switched into sleeping mode to reduce energy consumption. We first investigate the RS–BS association problem when BS modes are given, and propose a heuristic algorithm that determines the RS–BS associations, and based on this, a tabu search approach that identifies the BS mode combination is presented. Simulation results show that, when the traffic load is low, our proposed approach can reduce the energy consumption of BSs by nearly 40%. When the traffic load is heavy, although the energy-saving effect is not obvious, our proposed approach can improve the QoS by allowing dynamic associations of BSs. In addition, we also check the impact of energy consumption parameters on the energy-saving effect of our proposed approach.

References

[1] M. A. Marsan, L. Chiaraviglio, D. Ciullo, and M. Meo, "Optimal energy savings in cellular access networks," in *Proc. IEEE ICC Workshops*, 2009, pp. 1–5.

[2] D. Willkomm, S. Machiraju, J. Bolot, and A. Wolisz, "Primary user behavior in cellular networks and implications for dynamic spectrum access," *IEEE Commun. Mag.*, vol. 47, no. 3, pp. 88–95, Mar. 2009.

[3] L. Chiaraviglio, D. Ciullo, M. Meo, M. A. Marsan, and I. Torino, "Energy-aware umts access networks," in *Proc. IEEE W-GREEN*, 2008, pp. 1–8.

[4] K. Samdanis, T. Taleb, D. Kutscher, and M. Brunner, "Self organized network management functions for energy efficient cellular urban infrastructures," *Mobile Netw. Appl.*, vol. 17, no. 1, pp. 119–131, Feb. 2012.

[5] G. Jie and N. Zhisheng, "A dynamic programming approach for base station sleeping in cellular networks," *IEICE Trans. Commun.*, vol. E95, no. 2, pp. 551–562, 2012.

[6] S. Zhou, J. Gong, Z. Yang, Z. Niu, and P. Yang, "Green mobile access network with dynamic base station energy saving," in *ACM MobiCom*, 2009, vol. 9, no. 262, pp. 10–12.

[7] L. Chiaraviglio, D. Ciullo, G. Koutitas, M. Meo, and L. Tassiulas, "Energy-efficient planning and management of cellular networks," in *Proc. 9th Annu. Conf. WONS*, 2012, pp. 159–166.

[8] Z. Niu, Y. Wu, J. Gong, and Z. Yang, "Cell zooming for cost-efficient green cellular networks," *IEEE Commun. Mag.*, vol. 48, no. 11, pp. 74–79, Nov. 2010.

[9] "Further advancements for e-utra physical layer aspects," Third-Generation Partnership Project, Sophia Antipolis, France, Mar. 2010.

[10] O. Teyeb, V. Van Phan, B. Raaf, and S. Redana, "Dynamic relaying in 3GPP LTE-advanced networks," *EURASIP J. Wireless Commun. Netw.*, vol. 2009, p. 6, Mar. 2009.

[11] Z. Yang and Z. Niu, "Load balancing by dynamic base station relay station associations in cellular networks," *IEEE Wireless Commun. Lett.*, vol. 2, no. 2, pp. 155–158, Apr. 2013.

[12] K. Samdanis and M. Brunner, "Self-organized network management functions for relay enhanced lte-advanced systems," in *Proc. IEEE 22nd Int. Symp. PIMRC*, 2011, pp. 227–231.

[13] D. Lee, S. Zhou, and Z. Niu, "Multi-hop relay network for base station energy saving and its performance evaluation," in *Proc. IEEE GLOBECOM*, 2011, pp. 1–5.

[14] A. Zappone, Z. Chong, E. Jorswieck, and S. Buzzi, "Energy-aware competitive power control in relay-assisted interference wireless networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 4, pp. 1860–1871, Apr. 2013.

[15] D. Willkomm, S. Machiraju, J. Bolot, and A. Wolisz, "Primary users in cellular networks: A large-scale measurement study," in *Proc. 3rd IEEE Symp. New Frontiers DySPAN*, 2008, pp. 1–11.

[16] M. Imran, E. Katranaras, G. Auer, O. Blume, V. Giannini, I. Godor, Y. Jading, M. Olsson, D. Sabella, P. Skillermark, and W. Wajda, "Energy efficiency analysis of the reference systems, areas of improvements and target breakdown," EARTH Project Deliverable D, vol. 2, 2011.

[17] S. Sahni and T. Gonzalez, "P-complete approximation problems," *J. ACM*, vol. 23, no. 3, pp. 555–565, Jul. 1976.

[18] D. B. Shmoys and É. Tardos, "An approximation algorithm for the generalized assignment problem," *Math. Program.*, vol. 62, no. 1–3, pp. 461–474, Dec. 1993.

[19] D. B. West, *Introduction to Graph Theory*. Englewood Cliffs, NJ, USA: Prentice-Hall, 2001.

[20] G. L. Nemhauser and L. A. Wolsey, *Integer and Combinatorial Optimization*. New York, NY, USA: Wiley, 1988.

[21] N. Karmarkar, "A new polynomial-time algorithm for linear programming," in *Proc. 16th Annu. ACM Symp. Theory Comput.*, 1984, pp. 302–311.

[22] F. Glover, "Tabu searchłpart i," *ORSA J Comput.*, vol. 1, no. 3, pp. 190–206, 1989.

[23] F. Glover, "Tabu searchłpart ii," *ORSA J. Comput.*, vol. 2, no. 1, pp. 4–32, 1990.

[24] Y. Oda, K. Tsunekawa, and M. Hata, "Advanced los path-loss model in microcellular mobile communications," *IEEE Trans. Veh. Technol.*, vol. 49, no. 6, pp. 2121–2125, Nov. 2000.

[25] U. Gotzner and R. Rathgeber, "Spatial traffic distribution in cellular networks," in *Proc. IEEE 48th VTC*, 1998, vol. 3, pp. 1994–1998.

Zexi Yang received the B.S. degree from Tsinghua University, Beijing, China, in 2007, where he is currently working toward the Ph.D. degree with the Department of Electronic Engineering.

His research interests include cooperative communications and green communications.

Zhisheng Niu (M'98–SM'99–F'12) received the B.S. degree from Northern Jiaotong University (currently Beijing Jiaotong University), Beijing, China, in 1985 and the M.E. and D.E. degrees from Toyohashi University of Technology, Toyohashi, Japan, in 1989 and 1992, respectively.

For two years, he was with Fujitsu Laboratories Ltd., Kawasaki, Japan. Since 1994, he has been with Tsinghua University, Beijing, where he is currently a Professor with the Department of Electronic Engineering, the Deputy Dean of the School of Information Science and Technology, and the Director of Tsinghua–Hitachi Joint Laboratory on Environmental Harmonious Information and Communications Technology. He is also a Guest Chair Professor with Shandong University, Jinan, China. His major research interests include queueing theory, traffic engineering, mobile Internet, radio resource management of wireless networks, and green communications and networks.

Dr. Niu is a Fellow of the Institute of Electronics, Information, and Communication Engineers (IEICE). He has been an active volunteer for various academic societies. He has served as the Director for Conference Publications (2010–2011) and Director for Asia–Pacific Board (2008–2009) of the IEEE Communications Society, the Membership Development Coordinator (2009–2010) of IEEE Region 10, a Councilor of IEICE Japan (2009–2011), and a council member of the Chinese Institute of Electronics (2006–2011). He is currently a Distinguished Lecturer (2012–2013) of the IEEE Communications Society, the Standing Committee Member of both the Communication Science and Technology Committee under the Ministry of Industry and Information Technology of China and the Chinese Institute of Communications (CIC), the Vice Chair of the Information and Communication Network Committee of CIC, the Editor of the IEEE WIRELESS COMMUNICATION MAGAZINE, and the Associate Editor-in-Chief of the IEEE/CIC joint publication *China Communications*. He received the Outstanding Young Researcher Award from the Natural Science Foundation of China in 2009 and the Best Paper Awards (with his students) from the 13th and 15th Asia–Pacific Conference on Communication in 2007 and 2009, respectively.