

Received November 8, 2019, accepted December 5, 2019, date of publication December 10, 2019, date of current version January 15, 2020.

Digital Object Identifier 10.1109/ACCESS.2019.2958654

Energy-Efficient Switching ON/OFF Strategies Analysis for Dense Cellular Networks With Partial Conventional Base-Stations

ZHANG JIAN^{ID}, WU MUQING^{ID}, AND ZHAO MIN^{ID}

School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China

Corresponding author: Zhang Jian (zhang_jian@bupt.edu.cn)

This work was supported by the Beijing Laboratory of Advanced Information Networks.

ABSTRACT Focusing on the energy efficiency (EE) improving problem in a densely deployed cellular network with partial conventional base stations (BSs) that cannot be frequently switched on/off, this paper jointly optimizes the switching on/off strategy and user association policy with the consideration of quality-of-service (QoS). By applying a modified switching cost model, we formulate the EE improving problem as an energy efficiency maximization problem (EEMP). To solve the EEMP efficiently, we first concentrate on the EEMP with a constant user distribution where switching cost is omitted temporarily, i.e., static-EEMP, and then the EEMP can be settled by using the results of the static-EEMP. After analyzing the hardness of solving the static-EEMP directly, several sub-optimal strategies are proposed for the static-EEMP to optimize the BSs working states and user association policy. With the results obtained from solving the static-EEMP, we propose a QoS First Maximum EE (QFMEE) scheme to solve the EEMP in two steps. In the first step, necessary BSs are switched on when the user distribution changes to ensure user QoS. In the second step, we judge the extra cost and the energy saving caused by the switching off operations to improve the network EE. Simulation results verify the effectiveness of our proposed strategies for static-EEMP and show that QFMEE can improve the network EE and reduce switching cost significantly when the switching cost occupies a considerable proportion in the total energy consumption.

INDEX TERMS Switching strategy, switching cost, dense cellular networks, energy efficiency.

I. INTRODUCTION

With the exponentially increasing of the mobile devices such as smart phones, tablets and hand-held terminals, the densification of small base stations (BSs) is one of the key approaches to meet the exponentially growing demand in high data rate and thus energy consumption is growing tremendously in wireless networks [1]–[3]. Naturally, improving the energy efficiency (EE) for wireless communication has become a burning issue from now on. It is worth noting that most of the energy resources in wireless networks are consumed by BSs and it has been revealed that BSs consume approximately 58% of the total power consumption in a typical wireless cellular network [4], [5]. Further, according to the report from China Mobile, a BS at the idle state with no traffic load consumes about 50%-60% of the energy that cost at the laden state with the maximum

traffic load, and 40% of the energy consumption of a BS can be saved if it is switched off (i.e., in the sleeping state) [6]. Therefore, an effective way to achieve energy saving and EE improvement in wireless cellular networks is to dynamically switch off BSs, especially for low traffic load cases where a small part of BSs can guarantee the service of all users. Since nowadays more and more small BSs are densely deployed in the urban area to meet the peak traffic demand and the coverage is always overlapped, switching off some of them will not affect the coverage of the entire network [7].

Many works have studied the switching on/off (or called sleeping) strategies for different network architectures and verified the resulting benefit in terms of energy saving or EE improving [8]–[10]. In reference [11], the minimal energy consumption problem is formulated as an integer programming (IP) model to minimize the number of active BSs while the quality-of-service (QoS) can be preserved. Then two low-complexity approximation algorithms are proposed for solving the formulated problem, but the switching cost is

The associate editor coordinating the review of this manuscript and approving it for publication was Carol Fung.

not considered. In reference [12], coalition-based sleeping strategy and power allocation are jointly optimized to improve the dense small cell network EE while the QoS, i.e., the minimum rate for a user, is satisfied, however, the switching cost is omitted as well. Reference [13] minimizes the energy cost with the consideration of switching cost by jointly determining the set of BSs to be active and the levels of transmit power with a predictable traffic flow, however, the switching cost of a BS is modeled as linear with its workload, which is not accurate enough.

Some methods or techniques are commonly used for analyzing the performance in switching on/off strategies designing. Stochastic geometry tools, as one of them, are used in several references to model the BS and/or user distribution and derive tractable analytical expressions for the network performance analyzing under simple sleeping strategies [14]–[16]. Authors in [14] investigate the optimal percentage of deep-sleep BSs to maximize network EE without sacrificing the performance on network converge probability, BS overload probability and transmit power. Four working states for BS (i.e., on, standby, sleep, and off) are considered in [15] and the percentage of BSs in each mode under random sleeping strategy is optimized on the basis of the tractable expression of average achievable rate and coverage probability. After the analysis of random sleeping, strategic sleeping approach is proposed to improve the EE further more. In reference [16], the reinforcement learning technique is applied to the sleep/wake-up algorithm designing. By using a stochastic geometry approach, analytical expressions for the coverage probability and service success probability are derived to formulate the network EE and a Fuzzy Q-Learning based energy-efficient sleep/wake-up mechanism is proposed for energy saving and EE improving.

Then, queuing theory is also frequently used for the analysis of the network performance with sleeping mode [17]–[19]. In reference [17], each BS in the cellular network is modeled as an $M/M/1/k$ -Processor Sharing queue with vacations, where the network power consumption is optimized. Performances are evaluated for different sleeping schemes in terms of the tradeoff between energy saving and user QoS (including delay and blocking probability). In reference [18], the transition of BS state is modeled as a two-dimensional Markov model and the users are modeled as a discrete-time queue. After that, different sleeping strategies along with power control are analyzed to optimize the energy saving under the constraints of QoS and fronthaul capacity, and a Lyapunov method is applied to guarantee the queue stability. In reference [19], BSs are equipped with energy harvesting devices, batteries and interfaces of external power grid at the same time. Energy scheduling and sleep control are jointly optimized to minimize the system cost while the stability of workload queues and battery queues are ensured by using a Lyapunov optimization based two-timescale approach.

Energy harvesting is considered in quite a few references besides [19], and making best of the renewable energy may help with minimizing the system cost. In reference [20], BSs'

work modes, and the allocation of harvested energy as well as subcarriers are jointly optimized to minimize the weighted energy consumption. A two-stage dynamic programming algorithm is proposed to solve the formulated problem with low computational complexity. In reference [21], small BSs are classified into three different types, i.e., conventional small BSs, renewable small BSs and hybrid small BSs. An energy efficient scheme that jointly considers BS switching strategy and user association is proposed which aims at making the best of the harvested energy to serve more users and reducing the on-grid power consumption.

Problems are also analyzed from unique perspectives. In references [22] and [23], BSs are clustered and the on/off strategies are closely related with the clustering results. In reference [22], switching on/off operation for hotspot areas is analyzed. Active remote radio heads (RRHs) are located close to the hotspot areas by being mapped to the cluster centroids of a clustering process. Different strategies under different infrastructure conditions and information availabilities are proposed to find the cluster centroids on the basis of a complete framework. In reference [23], BSs are clustered based on the work load and distance between them and then an intra-cluster switching on/off strategy is proposed for energy saving. Authors in [24] consider the instants of time at which the BS switching operations must be executed to adjust to the changed user distribution. A risk-aware probabilistic manner is adopted to determine the states updating time for the BS and then the BS configuration to minimize the total power consumption of the network is determined while ensuring the QoS with high probability. In reference [25], the direction of arrival (DOA) estimation is creatively used for BSs to accurately get the directions of areas with high traffic levels which help with the switching decisions to be made appropriately.

In general, most of these existing works either omit the switching cost during the switching operation or consider the switching cost with inaccurate models. Actually, switching on/off operations will cause extra handovers and delay for users besides the intrinsic energy cost for a BS to power on/off the corresponding components. Thus, the switching cost should be an important consideration. In addition, predictable traffic flows are adopted for most of these algorithms designing which may impact the accuracy.

In this paper, we study the problem of maximizing the EE of dense cellular networks with partial conventional BSs by proposing a dynamically switching on/off strategy. The energy efficiency maximization problem (EEMP) is formulated as a complex combinational optimization problem which is hard to solve directly, and a two-step sub-optimal algorithm is proposed to solve the EEMP efficiently. Firstly, the EEMP under a constant user distribution case is studied. Then, as the user distribution changes, switching cost is considered. The main contributions of this paper are listed as follows:

- 1) With the developing of wireless communications, millions of new BSs are built and quite a part of the conventional BSs are not updated yet. Thus, a theoretical dense cellular

network is developed where BSs are classified into two kinds: one is new type BSs that can be switched on/off as needed; the second is conventional BSs that cannot be switched off/on frequently. Different with previous works, switching cost in this paper is remodeled with two parts: a constant part for BS energy cost itself and a linear part related with the BS load. After the analysis of transmission model and power consumption model, the network EE is obtained. Then the EEMP is formulated and the problem hardness is discussed as well.

2) To solve the EEMP efficiently, this paper deals with the NP-hard problem in two stages. In the first stage, the EEMP with a constant user distribution, i.e., static-EEMP, is considered. In this case, switching cost can be omitted when calculating the network EE and only an ideal BSs working state under a constant user distribution is concerned. Several strategies are proposed for the static-EEMP where different user association policies are applied. A Two Distance Based Strategy (TDBS) is proposed as a modification scheme for distance based algorithms proposed in previous works where users access the closest BS and then an Energy Saving Based Strategy (ESBS) is proposed where users access the BS offering the best SINR. In these two strategies, the distance between a user and the nearest BS except its serving BS is also considered. Then two random switching strategies are proposed as comparison. In the second stage, a QoS First Maximum EE (QFMEE) scheme is proposed to solve the EEMP on the basis of the results from static-EEMP. The QFMEE is executed in two steps: firstly, the necessary BSs are switched on to ensure the user QoS; secondly, some of the unnecessary BSs are switched off in specific order to improve the network EE. These necessary and unnecessary BSs can be found from solving static-EEMP. At last the performance of the proposed strategies is evaluated. Simulation results show that the ESBS performs best and can be chosen for solving the EEMP. Then the effectiveness of QFMEE is shown in improving EE and reducing switching cost.

The remainder of this paper is organized as follows. In Section II, we describe the system model. In Section III, the energy efficiency maximization problem is formulated and the problem hardness is analyzed. Algorithms design and analysis are given in Section IV. In Section V, the simulation results are shown and discussed. Finally, conclusions are given in Section VI.

II. SYSTEM MODEL

In this section, the system model adopted in our work is presented and the transmission model and power consumption model are introduced separately. Table 1 shows the main notations to be used in the following sections.

A. NETWORK MODEL

In this paper, a downlink dense cellular network with two kinds of BSs deployment is considered. Denote $\mathbf{bs} = \mathbf{b}_1 \cup \mathbf{b}_2$ as the set of all BSs, where $\mathbf{b}_1 = \{1, 2, \dots, N_1\}$ is the set of N_1 new type BSs that can be switched on/off as needed for

TABLE 1. Notations.

Notation	Definition
\mathbf{bs}	The set of all BSs
\mathbf{b}_1	The set of new type BSs
\mathbf{b}_2	The set of conventional BSs
\mathbf{U}	The set of users.
s_j	BS state indicator
$x_{k,j}$	User-BS association indicator
R_{min}	The minimum requirement of data rate for users.
N_{max}	The total number of RBs available for a BS
B	The bandwidth per sub-channel
$p_{tx,j}$	The transmit power per RB of BS j .
$n_{k,j}$	The RBs required to support R_{min} for UE k served by BS j .
l_j	The load of BS j
P_j	The power consumption of BS j under a constant state.
$P_{t,j}$	The transmit power of BS j .
$P_{c,j}$	The constant power for BS j to maintain the active state
$P_{s,j}$	The power consumption when BS j is in sleep mode.
P_j^{sw}	The switching power consumption for BS j .
p_c^{sw}	The constant energy cost for switching on/off a BS
ϕ	The extra re-association cost per user for switching.
T	The length of a time slot in which BSs keep their states.
κ	Impact factor for the tradeoff between energy saving and extra consumption lightening
γ	Impact factor for the influence of the variety of transmit power consumption

energy saving, and $\mathbf{b}_2 = \{N_1 + 1, N_1 + 2, \dots, N_1 + N_2\}$ is the set of N_2 conventional BSs which are not suitable for switching on/off frequently due to the lagging of manufacturing technology. Then we denote $\mathbf{S} = \{s_j | j \in \mathbf{bs}\}$ as the state of these BSs, where $s_j = 1$ means BS j is active and $s_j = 0$ otherwise. Obviously, $s_j = 1$ for $j \in \mathbf{b}_2$. Let $\mathbf{U} = \{1, 2, \dots, K\}$ denote the user set of the system with K users which are randomly distributed, and each user has a minimum requirement of data rate R_{min} as a QoS constraint. Binary variable set $\{x_{k,j}\}$ describes the user-BS association, which takes the value of 1 if user k is associated with BS j and 0 otherwise. Here we consider the full frequency reuse case in the downlink orthogonal frequency division multiple access (OFDMA) system. The spectrum used by the BSs is overlaid and all BSs share the same bandwidth with N_{max} sub-channels where the bandwidth per sub-channel is B kHz. We denote the one sub-channel along with a unit slice of time as a unit time-spectrum resource block (RB) and thus the number of RBs available for a BS is N_{max} . We assume that a user can only be served by one BS but may be allocated with multiple RBs to satisfy its QoS requirement.

B. TRANSMISSION MODEL

For simplicity, we assume that the transmit power of a BS is constant and equally allocated to each RB [26]. Then the received signal to interference and noise ratio (SINR) per RB of user k that is served by an active BS j can be modeled as:

$$SINR_{k,j} = \frac{p_{tx,j} h_{k,j}}{\sum_{i \in \mathbf{bs} \setminus j} p_{tx,i} s_i h_{k,i} + \sigma^2}, \quad (1)$$

where $p_{tx,j}$ is the transmit power per RB of BS j , $h_{k,j}$ is the channel fading gain (including path loss and shadow fading)

of user k with respect to BS j and σ^2 denotes the thermal noise power. The denominator part in (1) indicates that switching off necessary BSs will not only save the cost of the sleeping BS itself, but also mitigate the interference to other active BSs. Thus, we can roughly get that switching off BSs in \mathbf{b}_1 as many as possible with the premise of guaranteeing user QoS will contribute to the energy saving and then help with improving the network EE.

According to Shannon capacity formula, the received data rate per RB of user k from BS j is:

$$r_{k,j} = B \log(1 + SINR_{k,j}). \quad (2)$$

Then the number of RBs required to support R_{min} for UE k from BS j can be computed as follows:

$$n_{k,j} = x_{k,j} \lceil R_{min}/r_{k,j} \rceil. \quad (3)$$

Thus, we can get the transmit power of BS j by accumulating the transmit power of all used RBs in this BS:

$$P_{t,j} = p_{tx,j} \sum_{k \in U} n_{k,j} = p_{tx,j} l_j, \quad (4)$$

where $l_j = \sum_{k \in U} n_{k,j}$ denotes the load of BS j i.e., the number of RBs that have been occupied in this BS. The definition of l_j considers the resources occupation and user's SINR, which is much more reasonable and practical than a simple definition of serving users number.

C. POWER CONSUMPTION MODEL

To evaluate the energy consumption during different states for a BS, we refer to [27] to build the power consumption model within a constant state as:

$$P_j = s_j(P_{c,j} + P_{t,j}) + (1 - s_j)P_{s,j}, \quad (5)$$

where $P_{c,j}$ is the constant power consumption (i.e., circuit cost, air conditioner consumption and so on) for BS j to maintain the active mode, $P_{s,j}$ is the energy cost when BS j is in sleep mode.

Except the power consumption of BS under a constant state during a period of time, the BS switching on/off operation will cause extra energy cost. In most cases, switching cost in the previous work is considered to be constant. Actually, other than the energy cost that is needed to power on/off the components of a BS, the change of the BS state will also cause users' handovers and re-associations or even extra resources cost for some new architecture such as Mobile Cloud Computing (MCC) system and Cloud-Radio Access Network (C-RAN). Here we roughly think that the latter cost is linear with the number of users associate with the BS. Based on the above analysis, we introduce a novel switching power consumption model as follows:

$$P_j^{sw} = |s_j - s_j^{new}|(p_c^{sw} + \varphi) \left| \sum_{k \in U} x_{k,j} - \sum_{k \in U} x_{k,j}^{new} \right|, \quad (6)$$

where s_j and $x_{k,j}$ are the state mode and user association for BS j and user k before the switching occurs, s_j^{new} and $x_{k,j}^{new}$ are the new state mode and user association for BS j and

user k after switching. p_c^{sw} is the constant energy cost for powering on/off the components of a BS and φ is the extra re-association cost per user.

III. PROBLEM FORMULATION

In this section, the EEMP is formulated in which variables are intricately coupled and thus it is hard to solve directly. Then a static-EEMP is introduced to help with solving the EEMP efficiently. After that the hardness of static-EEMP is analyzed.

A. EEMP FORMULATION

Switching on/off frequently too much is not conducive to energy saving and may damage the components inside BSs either. Here, we uniformly divide a period of time into several time slots where the length of each time slot is T . BSs are allowed to change their states at a specific time clock, specifically, the BSs in \mathbf{b}_1 could change their states at the beginning of every time slot to suit the changed user distribution which is also varied every time slot in this paper. In addition, we assume that the switching on/off operations will be done within a sufficiently small time which can be omitted.

Similar with [14], the overall network EE is measured in terms of bits/joule like this:

$$EE = \frac{R_{total}}{P_{total}}, \quad (7)$$

where R_{total} and P_{total} denote the total network throughput the whole network energy consumption in a given period of time, respectively.

Under a short term predictable traffic flow case, we think that not only the user distribution in the current time slot is known but also the user distribution in the next time slot can be accurately predicted. Here we consider an energy efficiency maximization problem (EEMP) in a typical period of L time slots and $m = 1, 2, \dots, L$ is the time slot index, thus the time slot m means the period of $[(m-1) \times T, m \times T)$. At the beginning of each time slot, i.e., the moment $t = (m-1) \times T$, BSs choose to keep the states or execute the switching on/off operations. Typically, we think the executions at the moment $t = 0$. The on/off decisions are made based on the present BSs states, i.e., the BSs states during $[-T, 0)$, the changed user distribution, i.e., the user distribution during $[0, T)$ and the predicted user distribution during $[T, 2T)$, to maximize the network EE. The rest on/off decisions on other time slots are similarly made. Hence, from (7) the EEMP in a typical L time slots can be formulated as:

$$\begin{aligned} \max EE &= \frac{T \left(\sum_{m=1}^L \sum_{k \in U} \sum_{j \in \mathbf{bs}} x_{k,j}^{(m)} r_{k,j}^{(m)} \right)}{T \sum_{m=1}^L \sum_{j \in \mathbf{bs}} P_j^{(m)} + \sum_{m=1}^L \sum_{j \in \mathbf{b}_1} P_j^{sw(m)}} \\ \text{s.t. C1} &: x_{k,j}^{(m)} \in \{0, 1\}, \\ \text{C2} &: s_j^{(m)} \in \{0, 1\}, \forall j \in \mathbf{b}_1, \\ \text{C3} &: s_j^{(m)} = 1, \forall j \in \mathbf{b}_2, \end{aligned}$$

$$\begin{aligned}
C4: & \sum_{k \in \mathbf{U}} n_{k,j}^{(m)} \leq N_{\max}, \forall j \in \mathbf{bs}, \\
C5: & \sum_{j \in \mathbf{bs}} x_{k,j}^{(m)} = 1, \forall k \in \mathbf{U},
\end{aligned} \quad (8)$$

where the variables with the superscript (m) indicate their corresponding representations in the time slot m .

Constraint C1 can be got from Section II.A. Constraints C2 and C3 are state constraints for new type BS and conventional BS, respectively. Constraint C4 is the RB capacity constraint for a BS. Constraint C5 specifies that a user can only associate with one BS.

Consider that the user-BS association variable set $\{x_{k,j}^{(m)}\}$ is coupled with the working states of BSs, thus the network throughput is directly related with $\{s_j^{(m)}\}$ under a given QoS requirement. As for $P_j^{(m)}$, the static power consumption is coupled with working states of BSs, and the transmit power consumption that depends on the BS load, is coupled with the configurations of its neighboring that impact the SINR of its associated users. The switching operations directly decide the working states of BSs, so the switching energy cost is strongly coupled with $\{s_j^{(m)}\}$ either. All in all, the operations of different BSs are strongly coupled and the EEMP is hard to be solved directly.

B. DEFINITION OF STATIC-EEMP

To solve the EEMP efficiently, we omit the switching energy cost temporarily to get an ideal state set and user association set under a constant user distribution in a typical time slot (i.e., $m = 1$ and m is omitted in this sub-section for simplicity). Then, we use the static result as a reference to go a step further. The static-EEMP is formulated as:

$$\begin{aligned}
\max EE_{static} &= \frac{\sum_{k \in \mathbf{U}} \sum_{j \in \mathbf{bs}} x_{k,j} r_{k,j}}{\sum_{j \in \mathbf{bs}} P_j} \\
s.t. & C1 - C5.
\end{aligned} \quad (9)$$

Given the solutions obtained from solving the static-EEMP, we can get that some BSs is necessary to keep on for guaranteeing the QoS, and some others could be switched off in terms of energy saving. After switching on the BSs that must be active, the user QoS would be satisfied. In other words, the network throughput could be considered constant. And thus, we can concentrate on the switching operations, i.e., the tradeoff between promoting the energy saving and lightening the extra switching cost.

C. PROBLEM HARDNESS DISCUSSION

Theorem 1: The static-EEMP is NP-hard.

From (1), we can see that users suffer the interference from all active BSs, thus each BS can determine the set of associated users. Here we denote the user set of BS j as \mathbf{U}_j , then $\mathbf{U} = \{\mathbf{U}_1, \mathbf{U}_2, \dots, \mathbf{U}_j, \dots, \mathbf{U}_{N1+N2}\}$. We assume that the RBs of BSs are sufficient for satisfying the QoS of all users. Then with a given R_{min} , the throughput of the network

is constant and the static-EEMP will convert to a minimum energy consumption problem. In other words, the optimizing objective is to find a set cover $\mathbf{U}^* = \sum_{j \in \mathbf{bs}} \mathbf{U}_j^*$, and the cost of \mathbf{U}^* is minimized. Hence, the static-EEMP can be equivalent to a minimal set cover (MSC) problem, and the MSC problem is a known NP-hard problem [28]. Thus, the static-EEMP is NP-hard.

Since the static-EEMP can be considered as a special case of EEMP when the reference time T is long enough, it is easy to get that EEMP is NP-hard either. Thus, it is necessary and significant to propose an efficient algorithm with low-complexity and relatively good performance to solve the formulated EEMP.

IV. ALGORITHM DESIGN

In this section, we solve the EEMP in two steps. In the first step, we analyze the static-EEMP separately with omitting the switching cost and propose several heuristic algorithms as candidates for comparison. In the second step, necessary BSs which are obtained from the solutions of the static-EEMP should be switched on to ensure the user QoS requirement, and after that we can maximize the total network EE by concentrating on the tradeoff between the extra switching off cost and the energy saving, both of which are related to the switching off operation directly.

A. STATIC-EEMP ALGORITHM

Although static-EEMP is simplified from the EEMP, user association and BSs states are strongly coupled with the network energy consumption or even network throughput. Different user association and BSs switching on/off strategies impact the network EE and the user QoS directly. Switching off a BS will affect not only its own serving users but also the neighboring BSs and their serving users. For a given user association policy, switching off a BS should improve the network EE while maintaining the user QoS. Taking this as the basis for strategy designing, we choose two user association policies and propose four heuristic BSs switching on/off strategies.

1) TWO DISTANCE BASED STRATEGY

In this algorithm, a user k chooses the closest BS who has enough RBs to satisfy the user's QoS as the target BS for accessing, this candidate user association policy is motivated by Algorithm 2 in [12]. As the user distribution is known, a BS j in \mathbf{b}_1 not only has the distance between it and its serving users but also knows the distance between its serving users and other BSs. Once BS j is switched off, its serving users will re-associate to the neighboring BSs, thus the distance between them will impact the power consumption of these neighboring BSs directly. Because of the path loss in the channel gain, users that re-associate to farther BSs will likely cause extra transmit power and weaken the energy saving of the switching off operation. Thus, switching off a proper BS can lighten the extra transmit power consumption and contribute to the improvement of the network EE. A simple

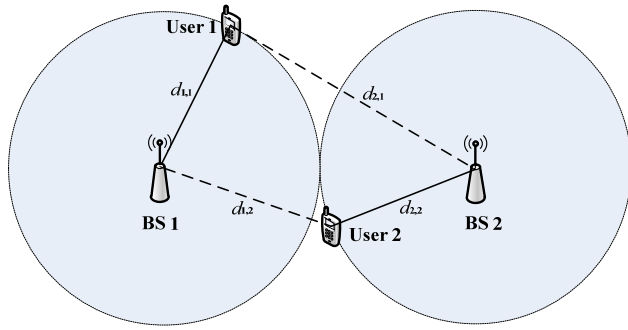


FIGURE 1. A sketch for Algorithm 1.

sketch is shown in Fig.1. From the sketch we can see *User1* accesses *BS 1* and *User 2* accesses *BS 2* where $d_{1,1} = d_{2,2}$ and $d_{1,2} < d_{2,1}$. When one of these two BSs is switched off, the two users will access the rest active BS. Based on above analysis, in which shadowing and fading are omitted and only the path loss is considered, switching off *BS 1* will probably cause more extra transmit power than that of switching off *BS 2*, thus *BS 2* is the better choice to be switched off. To find the proper BS, we define the distance impact factor of a user k for its associated BS j as:

$$\omega_{k,j} = x_{k,j}(PL'_{k,j} - PL_{k,j}), \quad (10)$$

where $PL_{k,j}$ is the path loss between user k and its associated BS j , $PL'_{k,j}$ is the path loss between user k and the nearest BS except BS j . Thus, the weighted distance load for BS j can be calculated as:

$$\Omega_j = \sum_{k \in \mathbf{U}} \omega_{k,j}. \quad (11)$$

A small $\omega_{k,j}$ means little extra transmit power consumption after the re-association and thus a BS with the smallest weighted distance load is preferred to be switched off. The main steps of TDBS are shown in Algorithm 1. All

BSs are set to be active in the initialization and user accesses the closest BS who has enough RBs to satisfy its QoS requirement. Then we concentrate on the BS in \mathbf{b}_1 . To be more specific, BSs with no users are switched off and the weighted distance load is calculated in each active BS in \mathbf{b}_1 . In each iteration for the loop of step 13-19, the BS with minimal weighted distance load in \mathbf{b}_1 is switched off and its associated users are assigned to these neighboring BSs just like step 3-6. This loop terminates when no more BSs in \mathbf{b}_1 can be switched off to achieve further promotion of network EE or user QoS cannot be satisfied.

2) ENERGY SAVING BASED STRATEGY

Path loss is only a part of the channel fading and a user accesses the closest BS will not always have the best SINR. In this algorithm, a user accesses the BS with the best SINR who has enough RBs to satisfy the user's QoS. Similar with Algorithm 1, we try to find the proper BS that the switching off operation can save as much energy as possible and lighten

Algorithm 1 Two Distance Based Strategy (TDBS)

- 1: **Input:** $\mathbf{bs} = \mathbf{b}_1 + \mathbf{b}_2, \mathbf{U}, R_{min}, P_{c,j}, p_{tx,j}, P_{s,j}$
- 2: **Initial:** $s_j = 1, x_{k,j} = 0, N_j = N_{max}, \forall j \in \mathbf{bs}, \forall k \in \mathbf{U}$
- 3: **For** each user k do
- 4: Find BS $j^* = \arg \min d_{k,j}$ that satisfies $n_{k,j^*} < N_{j^*}$;
- 5: $x_{k,j^*} = 1, N_{j^*} = N_{j^*} - n_{k,j^*}$;
- 6: **End for**
- 7: **For** each BS $j \in \mathbf{b}_1$ do
- 8: **If** no users are associated to j
- 9: $s_j = 0$;
- 10: **Else**
- 11: Calculate the weighted distance load Ω_j ;
- 12: **End if**
- 13: **End for**
- 14: **Repeat**
- 15: Find an active BS $j \in \mathbf{b}_1$ that has minimum weighted distance load and switch it off;
- 16: Assign the users that are served by BS j to neighboring BSs with enough RBs;
- 17: Update $x_{k,j}, N_j, \Omega_j$;
- 18: Check the network EE and user QoS;
- 19: **Until** user QoS cannot be satisfied or the network EE won't increase anymore

the extra transmit power consumption at the same time, and thus contribute to the improvement of the network EE. Here, we analyze this problem from two aspects. On one hand, we consider the energy saving of the switching off operation. From (4), we can see that the transmit power is related with the RBs occupied for maintain the QoS requirement, thus we can use l_j as a representation for the energy saving when the BS is switched off. On the other hand, we concentrate on lightening the extra transmit power consumption for neighboring BSs caused by the switching off operation. Inspired by Algorithm 1, we roughly think that when a user k that served by a BS j is near to a neighboring BS i , the extra transmit power consumption for BS i to serve user k which is caused by switching off BS j can be lightened to some extent. Based on the above analysis, we define the energy saving weighted load for a BS j as:

$$\Phi_j = \frac{\sum_{k \in \mathbf{U}} x_{k,j}(n_{k,j} + \kappa/d'_{k,j})}{\sum_{k \in \mathbf{U}} x_{k,j}}, \quad (12)$$

where κ is the impact factor which represents the tradeoff between energy saving and extra consumption lightening. When the value of Φ_j is large, a BS j can be switched off to save considerable energy without causing a serious extra consumption. In other words, the BS with the maximum Φ is the proper one to be switched off. The main steps of ESBS are shown in Algorithm 2.

In addition, Algorithm 2 is still feasible for the case in which accurate user locations are not available. In some scenarios such as some complicated channel conditions or

Algorithm 2 Energy Saving Based Strategy (ESBS)

```

1: Input:  $\mathbf{bs} = \mathbf{b}_1 + \mathbf{b}_2, \mathbf{U}, R_{min}, P_{c,j}, p_{tx,j}, P_{s,j}$ 
2: Initial:  $s_j = 1, x_{k,j} = 0, N_j = N_{max}, \forall j \in \mathbf{bs}, \forall k \in \mathbf{U}$ 
3: For each user  $k$  do
4:   Find BS  $j^* = \arg \max SINR_{k,j}$  that satisfies
      $n_{k,j^*} < N_{j^*}$ ;
5:    $x_{k,j^*} = 1, N_{j^*} = N_{j^*} - n_{k,j^*}$ ;
6: End for
7: For each BS  $j \in \mathbf{b}_1$  do
8:   If no users are associated to  $j$ 
9:      $s_j = 0$ ;
10:  Else
11:    Calculate the energy saving weighted load  $\Phi_j$ ;
12:  End if
13: End for
14: Repeat
15:   Find an active BS  $j \in \mathbf{b}_1$  that has maximum
     energy saving weighted load and switch it off;
16:   Assign the users that are served by BS  $j$  to other
     active BSs with best SINR and enough RBs;
17:   Update  $x_{k,j}, N_j, \Phi_j$ ;
18:   Check the network EE and user QoS;
19: Until user QoS cannot be satisfied or the network
     EE won't increase anymore

```

indoor environments, user location information cannot be accurately obtained by its serving BS and Algorithm 1 is no more workable. However, when we omit the user location related part in (12) i.e., $\kappa/d_{k,j}^{\alpha}$, the Algorithm 2 is effective to some extent.

3) RANDOM SLEEP STRATEGY

Here we consider the simplest strategy without relevant computation inside the BS, in which the user distribution and the traffic load are not cared either. For simplicity, the maximization of the network EE is roughly treated as switching off as many BSs in \mathbf{b}_1 as possible on the premise of satisfying user QoS. In this strategy, BSs in \mathbf{b}_1 are randomly selected to be switched on/off according to different initializations. Specifically, when all the BSs are set to be active in the initialization just like the former two algorithms, the BSs in \mathbf{b}_1 are randomly selected to be switched off until the user QoS cannot be satisfied or no more BS can be switched off, and this strategy is denoted as Random Sleep Strategy 1 (RSS1). In the other case, all of the BSs in \mathbf{b}_1 are set to be in sleep mode in the initialization, then these BSs are randomly selected to be switched on until all users' QoS are satisfied or all BSs have been switched on, and this strategy is denoted as Random Sleep Strategy 2 (RSS2). The main steps of RSS1 and RSS2 are shown in Algorithm 3 and Algorithm 4, respectively. In these two algorithms, the BSs in \mathbf{b}_1 only need to receive the switching on/off signals and check for the existence of users that out of service while no other information is required.

Algorithm 3 Random Sleep Strategy 1 (RSS1)

```

1: Input:  $\mathbf{bs} = \mathbf{b}_1 + \mathbf{b}_2, \mathbf{U}, R_{min}, P_{c,j}, p_{tx,j}, P_{s,j}$ 
2: Initial:  $s_j = 1, x_{k,j} = 0, N_j = N_{max}, \forall j \in \mathbf{bs}, \forall k \in \mathbf{U}$ 
3: For each user  $k$  do
4:   Find BS  $j^* = \arg \max SINR_{k,j}$  that satisfies
      $n_{k,j^*} < N_{j^*}$ ;
5:    $x_{k,j^*} = 1, N_{j^*} = N_{j^*} - n_{k,j^*}$ ;
6: End for
7: For each BS  $j \in \mathbf{b}_1$  that has no users do
8:    $s_j = 0$ ;
9: End for
10: Repeat
11:   Randomly select an active BS  $j \in \mathbf{b}_1$  and
     switch it off;
12:   Assign the users that are served by BS  $j$  to
     neighboring BSs with enough RBs;
13:   Update  $x_{k,j}, N_j$ ;
14:   Check the user QoS;
15: Until user QoS cannot be satisfied or no more BS
     can be switched off

```

Algorithm 4 Random Sleep Strategy 2 (RSS2)

```

1: Input:  $\mathbf{bs} = \mathbf{b}_1 + \mathbf{b}_2, \mathbf{U}, R_{min}, P_{c,j}, p_{tx,j}, P_{s,j}$ 
2: Initial:  $s_j = 0, \forall j \in \mathbf{b}_1; s_j = 1, \forall j \in \mathbf{b}_2; x_{k,j} = 0,$ 
      $N_j = N_{max}, \forall j \in \mathbf{bs}, \forall k \in \mathbf{U}$ 
3: For each user  $k$  do
4:   Find BS  $j^* = \arg \max SINR_{k,j}$  that satisfies
      $n_{k,j^*} < N_{j^*}$ ;
5:    $x_{k,j^*} = 1, N_{j^*} = N_{j^*} - n_{k,j^*}$ ;
6: End for
7: Repeat
8:   Check the user QoS;
9:   If all users' QoS are satisfied
10:    End the algorithm;
11:  Else
12:    Randomly select a sleeping BS  $j \in \mathbf{b}_1$ 
     and switch it on;
13:    Re-assign the users;
14:    Update  $x_{k,j}, N_j$ ;
15:  End if
16: Until all BSs have been switched on

```

B. EEMP ALGORITHM

With the solutions in static-EEMP, we can get the ideal BSs working states without considering the switching cost, which give an important reference for the switching operations. Then, we can judge the extra cost along with the energy saving and EE improvement caused by the switching operations. To deal with the EEMP efficiently, we introduce a QoS First Maximum EE (QFMEE) Strategy. The QFMEE consists of two stages: the first stage is to guarantee the user QoS and the second stage is to maximize the network EE. That is the reason for the strategy name, i.e., QFMEE, and the details of QFMEE are presented in the following part of this subsection.

Generally speaking, dense BSs means more RBs and better channel conditions for users and that is one of the main reasons for the deployment of 5G ultra-dense cellular networks [29]. Obviously, sufficient BSs should keep active to guarantee the user QoS and the first stage is to switch on these necessary BSs. For clarity of description, we denote $\mathbf{S}' = \{s'_j\}$ as the states of the BSs in the last time slot, i.e., the states before switching operations. $\mathbf{S}^{idea11} = \{s_j^{idea11}\}$ is defined as the solution of the static-EEMP in current time slot. Similarly, $\mathbf{S}^{idea12} = \{s_j^{idea12}\}$ is defined as the solution of the static-EEMP in the next time slot based on the predicted user distribution. $\mathbf{S}^{mid} = \{s_j^{mid}\}$ is defined as the state of the BSs after stage 1 and $\mathbf{S}^* = \{s_j^*\}$ indicates the BS state after stage 2, i.e., the final solution for the QFMEE. \mathbf{S}^{idea11} shows the most EE efficient working states without switching cost in current time slot, which also gives out the basic active BSs to guarantee the users QoS. After comparing \mathbf{S}' with \mathbf{S}^{idea11} , we can find the BSs that are switched off in \mathbf{S}' but active in \mathbf{S}^{idea11} , these BSs are necessary to be switched on to satisfy the current user distribution which can be defined as \mathbf{S}^{nec} . Thus, we have $\mathbf{S}^{mid} = \mathbf{S}' \cup \mathbf{S}^{idea11}$ and $\mathbf{S}^{nec} = \mathbf{S}^{mid} - \mathbf{S}'$.

After switching on the target BSs, i.e., \mathbf{S}^{nec} , QoS is guaranteed and we begin to execute stage 2 to maximize the network EE. With the satisfaction of user QoS, the network throughput is roughly constant and the EE maximizing problem can be converted to an energy cost minimizing problem. Similar with stage 1, after the comparison between \mathbf{S}^{mid} and \mathbf{S}^{idea11} , we can find the BSs which are active in \mathbf{S}' but switched off in \mathbf{S}^{idea11} and these unnecessary candidate BSs that could be switched off as needed are concerned in this stage. Here we define these BSs as \mathbf{S}^{unnec} and thus $\mathbf{S}^{unnec} = \mathbf{S}^{mid} - \mathbf{S}^{idea11}$. Besides this, the comparison between \mathbf{S}^{unnec} and \mathbf{S}^{idea12} will help us with lightening the switching cost in the next time slot. In theory, the optimal choices for switching off can be got by travelling through all of the combinations, but it's complicated as the BSs increasing. Here, we propose a low-complexity strategy by jointly analyzing the constant energy saving, extra transmit power saving and switching cost for switching off these BSs. The switching off operations for BSs in \mathbf{S}^{unnec} impact the constant energy saving, transmit power, and switching cost both in the current time slot and in the next time slot. Besides the intrinsic switching cost which is related to the number of serving users in the last time slot, switching off a BS in \mathbf{S}^{unnec} can save a constant part of energy in current time slot. However, the impact on next time slot is different for BSs in $\mathbf{S}^{unnec} - \mathbf{S}^{idea12}$ and $\mathbf{S}^{unnec} \cap \mathbf{S}^{idea12}$, respectively. On one hand, the BSs in $\mathbf{S}^{unnec} - \mathbf{S}^{idea12}$ are unnecessary in the next time slot either, thus the switching off operations in the current time slot can not only avoid an extra switching cost but also save a constant part of energy in the next time slot; on the other hand, the BSs in $\mathbf{S}^{unnec} \cap \mathbf{S}^{idea12}$ are necessary in the next time slot and the switching off operations in the current time slot will bring equivalent switching on operations in the next time slot, additionally. Moreover, the variety of transmit power consumption is bound up with

user SINR. To be more specific, the poorer user SINR is, the more energy is saved for transmitting. Thus, we can define the switching off factor like this:

$$\Gamma_j = \begin{cases} (P_{c,j} - P_{s,j})T - \varphi \sum_{k \in U} x_{k,j'} \\ + \sum_{k \in U} \frac{\gamma}{SINR_{x,j'}}, & j \in \mathbf{S}^{unnec} - \mathbf{S}^{idea12} \\ (P_{c,j} - P_{s,j})T - 2p_c^{sw} - \varphi \sum_{k \in U} x_{k,j'} \\ + \sum_{k \in U} \frac{\gamma}{SINR_{x,j'}}, & j \in \mathbf{S}^{unnec} \cap \mathbf{S}^{idea12}, \end{cases} \quad (13)$$

where $x'_{k,j}$ and $SINR'_{k,j}$ are the user association indicator and corresponding SINR before the user distribution changes, respectively. And γ is the impact factor which represents the influence of the variety of transmit power consumption before and after switching. Based on the above analysis, we choose the BS with maximum Γ as the target to be switched off and check the network EE until $\mathbf{S}^* = \mathbf{S}^{idea11}$ or the network EE no more improves.

We execute the QFMEE at the beginning of each time slot until all of the L time slots are completed and the main steps of QFMEE are shown in Algorithm 5. Steps 3-5 are to realize the QoS first part and the rest steps are to maximize the network EE. Although the BSs are switched off one by one, we think the time consumed in this process can be omitted with respect to the length of a time slot, i.e., T .

Algorithm 5 QoS First Maximum EE (QFMEE) Strategy

- 1: **Input:** $\mathbf{bs} = \mathbf{b}_1 + \mathbf{b}_2$, U^1 , U^2 , R_{min} , $P_{c,j}$, $p_{tx,j}$, $P_{s,j}$, p_c^{sw} , φ , \mathbf{S}' , $\{x'_{k,j}\}$
 - 2: **Output:** \mathbf{S}^* , $\{x_{k,j}\}$
 - 3: Get \mathbf{S}^{idea11} and \mathbf{S}^{idea12} based on the changed user distribution U^1 and the predicted user distribution U^2 by applying proper static-EEMP algorithm.
 - 4: According to \mathbf{S}' and \mathbf{S}^{idea11} , find \mathbf{S}^{nec} .
 - 5: Switch on corresponding BSs in \mathbf{S}^{nec} and get \mathbf{S}^{mid} ;
 - 6: Let $\mathbf{S}^* = \mathbf{S}^{mid}$
 - 7: **If** $\mathbf{S}^{mid} \neq \mathbf{S}^{idea11}$
 - 8: According to \mathbf{S}^{mid} and \mathbf{S}^{idea11} , find \mathbf{S}^{unnec} .
 - 9: Calculate the switching off factors Γ of BSs in \mathbf{S}^{unnec} ;
 - 10: **Repeat**
 - 11: Switch off the BS j with maximal Γ
 - 12: Assign the users that are served by BS j to neighboring active BSs with enough RBs;
 - 13: Check the short term network EE within this two time slots
 - 14: Remove the BS that just be switched off from \mathbf{S}^{unnec} ;
 - 15: Update \mathbf{S}^* ;
 - 16: **Until** the short term network EE no more increases or $\mathbf{S}^* == \mathbf{S}^{idea11}$
 - 17: **Else**
 - 18: The algorithm comes to an end
 - 19: **End if**
-

V. SIMULATION RESULTS

In this section, we use Monte Carlo simulations to evaluate the performance of our proposed strategies for static-EEMP and EEMP. In the following, we firstly introduce the parameter settings, and then corresponding simulation results and analysis are presented one by one.

A. PARAMETER SETTINGS

We consider a simulation scenario like following. $N_1 = 16$ new type BSs and $N_2 = 5$ conventional BSs are regularly deployed in the area of $1 \times 1 \text{ km}^2$ as Fig.2 shows. Users are randomly distributed in the same area and their distribution changes at the beginning of every time slot. For the wireless accessing, the total number of sub-channel is $N_{max} = 100$ and each one is assigned with a bandwidth $B = 180 \text{ kHz}$. For the wireless channel condition, we set the pass loss model as $37.6 \times \log(dist) + 148.1$ similar to [30], the shadowing factor is given by a log-normal function with standard deviation of 8dB and small scale fading model is independently and identically distributed (i.i.d.) Rayleigh fading with zero mean and unit variance. As for the AWGN, we set the noise power as $\sigma^2 = -174 \text{ dBm/Hz}$. Other simulation parameters are summarized in Table 2.

TABLE 2. Partial parameters.

Parameter	Value
$p_{tx,j}$ is the transmit power per RB	0.02 W
$P_{e,j}$ is the constant power consumption to maintain the active mode	5 W
$P_{s,j}$ is the BS power in sleep mode	0.3 W
p_c^{sw} is the constant cost for BS switching	200 J
minimum requirement of data rate R_{min}	1 Mbps
Impact factor κ	0.5
Impact factor γ	150
Number of simulated time slots L	13

B. EVALUATION OF STATIC-EEMP ALGORITHM

In this section, we evaluate the proposed strategies for solving the static-EEMP. First of all, we introduce the minimum and maximum network load to reflect the load condition of the network in a typical time slot. Considering the interference that brought from switching on new type BSs, we simulate the maximum throughput of the network in n -BS-on case (i.e., n new type BSs are randomly switched on, $0 \leq n \leq 16$), separately. Thus, the minimum and maximum network loads correspond to the minimal and maximal simulation results, respectively. Concretely, the maximum throughput of the network is obtained as follows. We gradually increase the amount of users and randomly distribute these users in the network until any user is out of service. This experiment is done 1000 times and a 10% trimmed mean of the experiment results is calculated as the maximum throughput. The result is shown in Fig.3, from which we get the minimum network load as 120 users and the maximum network load as 290 users. Actually, when the user number is 290, all BSs are active no matter which strategy is applied

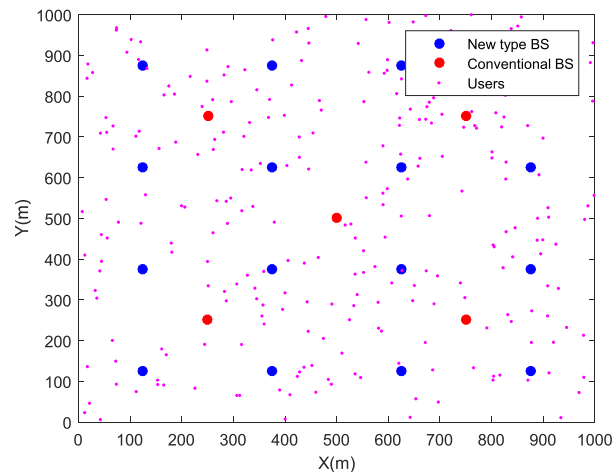


FIGURE 2. Simulation scenario with 280 users.

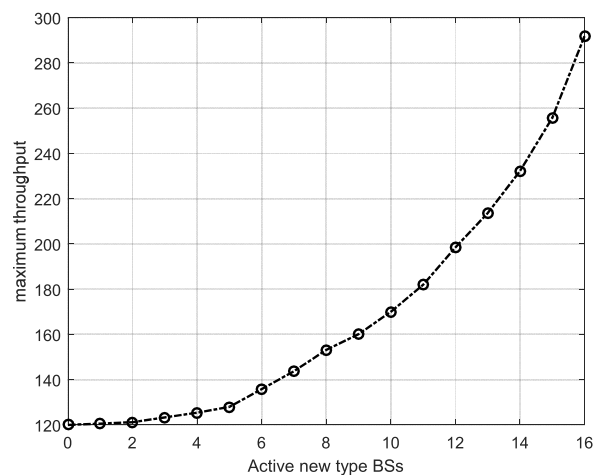


FIGURE 3. The maximum throughput for different number of active new type BSs.

and there is no comparative significance. Thus, we evaluate the performance of proposed TDBS, ESBS, RSS1, and RSS2 by varying the user number from 120 to 280. The case that all BSs keep active is used as the basic case for comparison. 1000 times simulation is made where a corresponding number of users are newly distributed in each time and the average results are adopted.

Fig.4 shows the comparison of the average number of BSs be switched off under different strategies. Particularly, the ESBS under $\kappa = 0$ is simulated to show the effectiveness of the usage of $d_{k,j}$ for lightening the extra consumption. Obviously, ESBS is the most efficient one that can switch off the most BSs and the gap between ESBS and ESBS under $\kappa = 0$ is quite considerable. In other words, ESBS is successfully designed and the tradeoff between energy saving and extra consumption lightening is very significant.

Then, we can see that RSS2 performs worse than ESBS but better than RSS1 or even ESBS under $\kappa = 0$, especially in light network load case. This is because that in RSS2 the

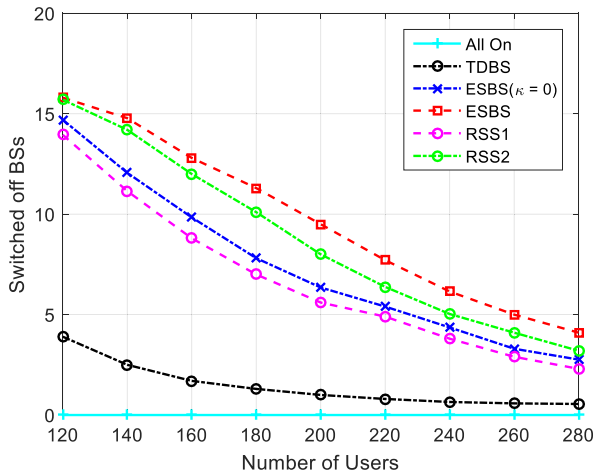


FIGURE 4. Average switched off BSs under different strategies.

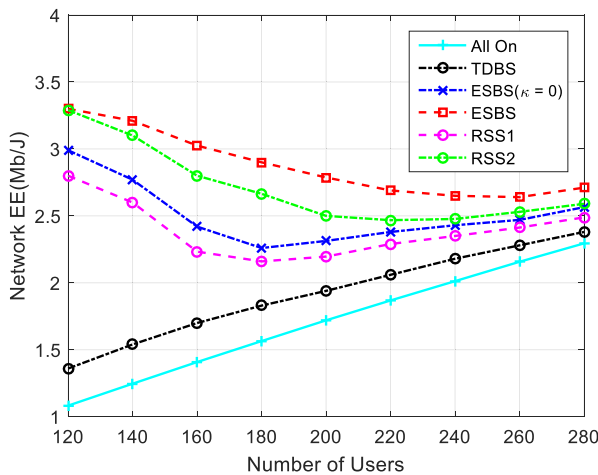


FIGURE 5. Network EE under different strategies.

new type BSs are switched off in the initialization and then switched on one by one. From a statistical point of view, turning off BSs in descending order (as ESBS, TDBS, and RSS1 execute) may lead to more BSs remaining active. Considering the dominating constant power consumption for an active BS, RSS2 may have a relatively good performance. However, this initialization will damage the user QoS at the beginning in practice for the lacking of enough active BSs, although we omit the time cost of switching operations in this paper. In addition, it's surprising that TDBS performs poorly. This is due to its different user association policy with other strategies in which the nearest BS is preferred rather than the one with the best SINR. That means this user association policy is not suitable. As the user number increasing, fewer BSs can be switched off in the premise of guaranteeing the QoS and the gaps between the strategies become narrow. Notably, the performance of RSS2 is still better than that of RSS1 or even ESBS under $\kappa = 0$ in Fig.5-7 which can be ascribed to the similar reason and thus corresponding analyses are omitted in following paragraphs.

Fig.5 illustrates the static network EE under different strategies. Specific to a certain strategy that we proposed (TDBS excepted), the EE gradually decreases as the number of user increasing at first. This is because of the quickly increasing of active BSs which can be seen from Fig.4. However, with the increasing of network load, quite a few BSs have been switched on and the changes of the active BSs number are not obvious. That is to say, the increased network throughput that contributes to the network EE is more significant than the increased cost of switching on BSs. This is why the EE rises slowly later. Then compare to Fig.4, we can find that the strategy who switches off more BSs can always get better EE. That is because a relatively huge constant power consumption to maintain the active mode for a BS and thus the number of BSs that are switched off impact the network EE seriously. Then we can explain the curves of TDBS and All On in which too many BSs keep active from the beginning to the end compared with other strategies.

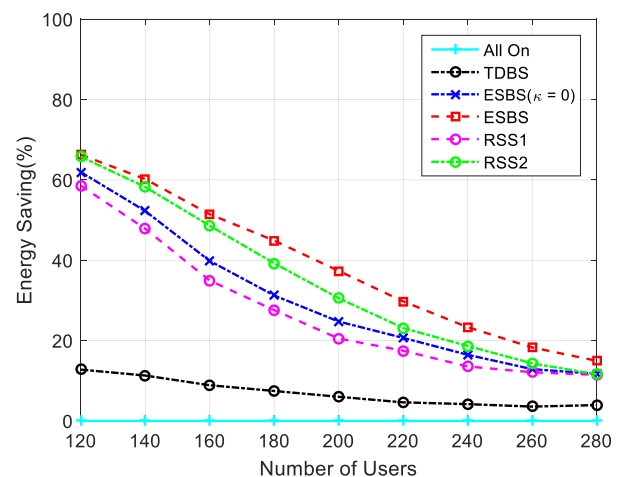


FIGURE 6. Energy saving under different strategies.

Fig.6 represents the percentage of saving in the power consumption under different strategies compared with All-On case. For the dominating consumption of $P_{c,j}$, the energy saving is related to the number of sleeping BSs directly. Thus, the performance of each strategy is quite similar with that in Fig.4. Generally speaking, the energy saving is quite remarkable in a light network load case and as the network load increasing the energy saving decreases gradually. Fig.7 depicts the average energy consumption per active BS under different strategies. Reference to Fig.4, we can see that quite a few BSs are switched off in a light load case, which means that the load for active BS is relatively heavy. That is why the energy cost for the active BS is high at the beginning in Fig.7. As the user number increasing, more and more BSs are switched on rapidly, the average load for active BSs is lightened and the energy cost per active BS decreases naturally. However, when users are more than 220, most of the BSs have been switched on and the BS load no more keep lightening as the user number increasing. Taken Fig.4 together, ESBS performs best and even ESBS under

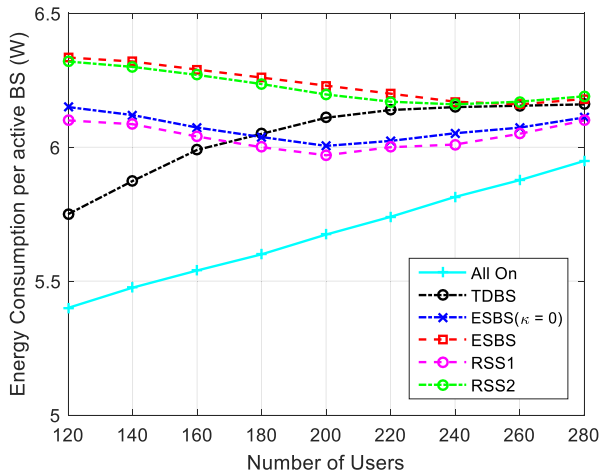


FIGURE 7. Energy consumption per active BS under different strategies.

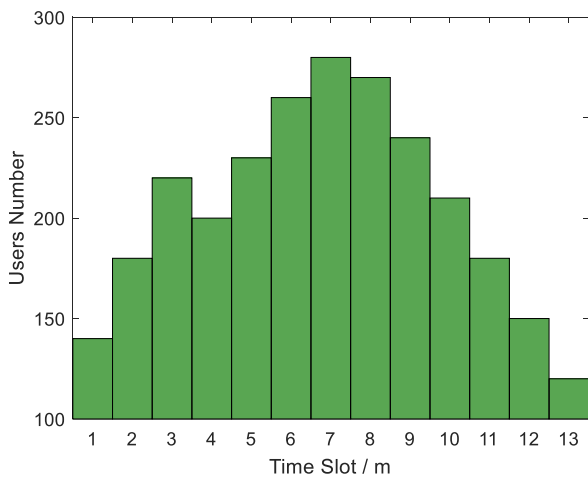


FIGURE 8. Number of user in the network at different time slot.

$\kappa = 0$ performs well because BSs with better SINR are switched on in priority. The curves of TDBS and All On are unique and the reason is similar with that in Fig.5.

From above simulation results, we can see that ESBS always outperforms other strategies and thus we choose ESBS as the best static-EEMP algorithm to maximize the network EE. RSS2 is second only to ESBS, but this strategy may damage the user QoS at the beginning of every time slot in practical situation. RSS1 is simple but mediocre. TDBS performs poorly due to its unsuitable user association policy. In addition, ESBS is still a better choice than RSS1 and RSS2 when the accurate information of users' positions is not available for BSs, i.e., ESBS under $\kappa = 0$. However, TDBS is no longer applicable without user's location.

C. EVALUATION OF EEMP ALGORITHM

In this section, we evaluate the performance of our proposed QFMEE algorithm in several time slots (i.e., $L = 13$), with switching cost considered. Based on the simulation results and analysis in Section IV.B, we choose ESBS as the strategy for solving static-EEMP. The number and the distribution of

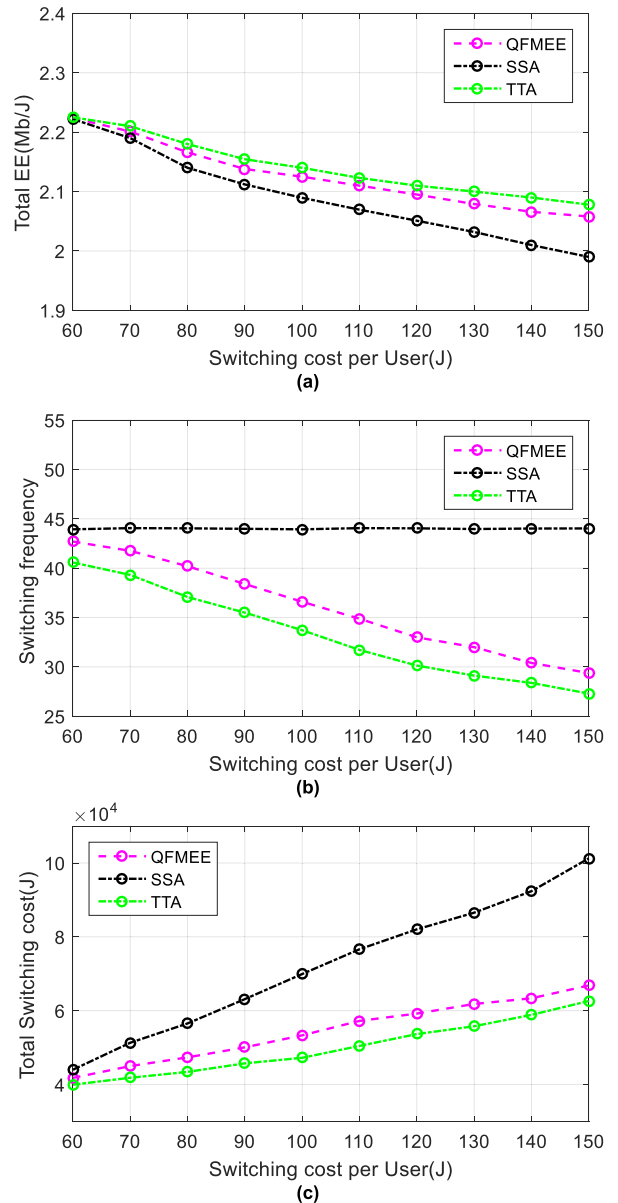


FIGURE 9. Network EE (a), switching frequency (b), and total switching cost (c) of different schemes with varying switching cost per user.

user in the network vary every time slot and the range of users number is set between 120 and 280, which is shown in Fig.8. For the consideration of the switching cost and its impacts on the network EE within the reference period, φ and T are two key parameters. Thus, the influences caused by these two parameters are discussed separately. To verify the effectiveness of our proposed QFMEE, a Static-switching algorithm (SSA) and a Travel-through algorithm (TTA) are chosen as comparison.

SSA: In this scheme, switching operations are executed according to the solution of static-EEMP entirely and the tradeoff between the extra switching off cost and the energy saving is not considered (i.e., let $S^* = S^{ideal1}$).

TTA: This algorithm is changed from the QFMEE. After step 6 in Algorithm 5, all kinds of combinations of practical

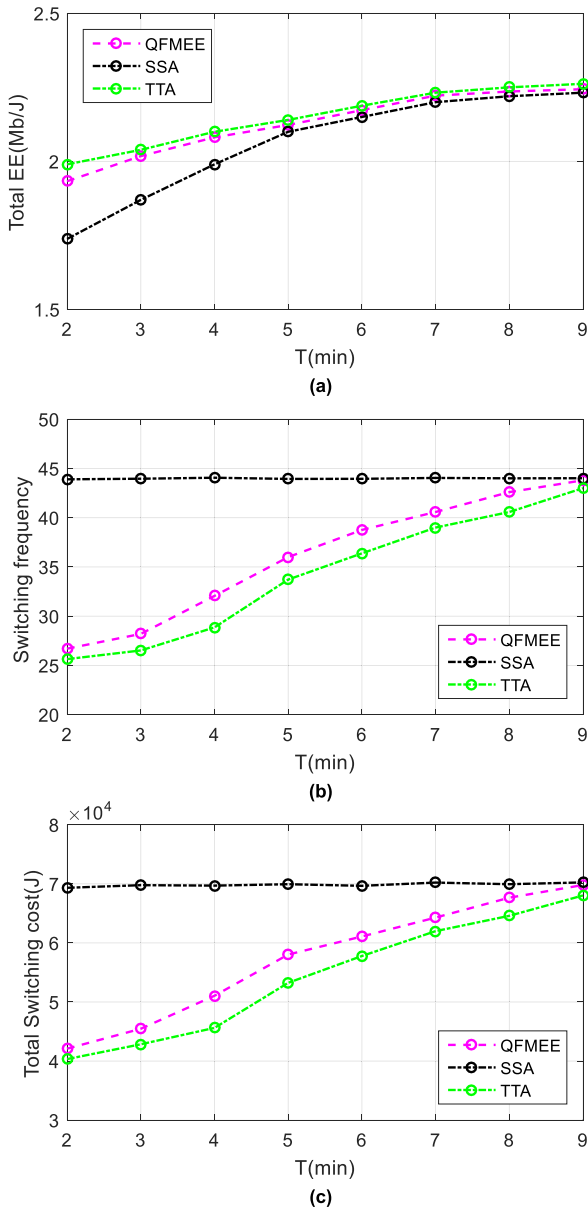


FIGURE 10. Network EE (a), switching frequency (b), and total switching cost (c) of different schemes with varying T .

switching off operations are traveled through, and then the one with best EE is chosen.

As φ increasing, network EE, switching frequency, and total switching cost under $T = 5$ minutes are shown one by one in Fig.9a, Fig.9b, and Fig.9c, respectively. We can see that the network EE decreases with different slope for different schemes in Fig.9a. Concretely, the TTA declines the slowest and the SSA has the maximal dropping slope while the QFMEE performs between them. This can be explained by Fig.9b and Fig.9c. In Fig.9b, we can see that switching frequency in QFMEE and TTA decrease gradually as φ growing while that of SSA remains constant and then Fig.9c gives the corresponding total switching cost. This is due to the fact that some BSs in QFMEE and TTA may keep active rather than be switched off to avoid the over cost in switching off

operation when the user distribution changes. In addition, TTA can find more of these BSs to switch off than that of QFMEE at the cost of increasing computation complexity significantly. Hence, we can conclude that TTA and QFMEE can save energy from controlling the switching off operation and contribute to the network EE and the bigger φ is the more energy can be saved. When φ is small enough (i.e., less than 60 J per user), switching cost is just a bit part for the total energy consumption and the energy saving from controlling the switching off operation is of no significance.

Fig.10 illustrates the impact of T on the corresponding simulation results. From formula (8), the proportion of switching cost in total energy consumption is becoming smaller and smaller with the increasing of T . Thus, the network EE improves either, which can be seen from Fig.10a. Specifically, when T is small (i.e., less than 5 minutes), TTA and QFMEE can help with improving the network EE to varying degrees. Fig.10b and Fig.10c show the switching frequency and total switching cost during the whole L time slots. The size of T does not impact the switching frequency and total switching cost in SSA scheme but the switching frequency and corresponding total switching cost in TTA and QFMEE increase as T becomes longer. This is because the change of user distribution depends on the number of T in the reference period rather than the range of T itself, and so are the switching operations.

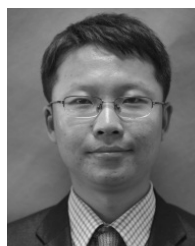
VI. CONCLUSION

In this paper, the EE improvement for dense cellular networks with partial conventional BSs is analyzed, where partial new type BSs can be dynamically switched on/off to save energy and the rest conventional BSs always keep active. Switching cost is remodeled and a short term predictable traffic flow case is considered. We deal with the formulated EEMP which is proved to be NP-hard in two stages. Firstly, the static-EEMP with constant user distribution is considered and several solution strategies are proposed. In the second stage, a QFMEE scheme is proposed to solve the EEMP on the basis of the results from static-EEMP. Simulation results show that the ESBS for static-EEMP outperforms other proposed solution strategies in EE improving as well as energy saving and the proposed QFMEE scheme can significantly improve EE and reduce switching cost when the switching cost occupies a considerable proportion in the total energy consumption.

REFERENCES

- [1] S. Barbarossa, S. Sardellitti, and P. D. Lorenzo, "Communicating while computing: Distributed mobile cloud computing over 5G heterogeneous networks," *IEEE Signal Process. Mag.*, vol. 31, no. 6, pp. 45–55, Nov. 2014.
- [2] J. B. Rao and A. O. Fapojuwo, "A survey of energy efficient resource management techniques for multicell cellular networks," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 1, pp. 154–180, 1st Quart., 2014.
- [3] J. An, K. Yang, J. Wu, N. Ye, S. Guo, and Z. Liao, "Achieving sustainable ultra-dense heterogeneous networks for 5G," *IEEE Commun. Mag.*, vol. 55, no. 12, pp. 84–90, Dec. 2017.
- [4] C. Han, T. Harrold, S. Armour, and I. Krikidis, "Green radio: Radio techniques to enable energy-efficient wireless networks," *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 46–54, Jun. 2011.

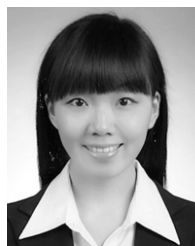
- [5] J. Wu, Y. Zhang, M. Zukerman, and E. K. N. Yung, "Energy-efficient base-stations sleep-mode techniques in green cellular networks: A survey," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 2, pp. 803–826, 2nd Quart., 2015.
- [6] F. Han, S. Zhao, L. Zhang, and J. Wu, "Survey of strategies for switching off base stations in heterogeneous networks for greener 5G systems," *IEEE Access*, vol. 4, pp. 4959–4973, 2016.
- [7] A. Cimmino, T. Pecorella, R. Fantacci, F. Granelli, T. F. Rahman, C. Sacchi, C. Carlini, and P. Harsh, "The role of small cell technology in future smart city applications," *Trans. Emerg. Telecommun. Technol.*, vol. 25, no. 1, pp. 11–20, Jan. 2014.
- [8] A. Ebrahim and E. Alsusa, "Interference and resource management through sleep mode selection in heterogeneous networks," *IEEE Trans. Commun.*, vol. 65, no. 1, pp. 257–269, Jan. 2017.
- [9] P. Chang and G. Miao, "Energy and spectral efficiency of cellular networks with discontinuous transmission," *IEEE Trans. Wireless Commun.*, vol. 16, no. 5, pp. 2991–3002, May 2017.
- [10] J. Kim, H.-W. Lee, and S. Chong, "Traffic-aware energy-saving base station sleeping and clustering in cooperative networks," *IEEE Trans. Wireless Commun.*, vol. 17, no. 2, pp. 1173–1186, Feb. 2018.
- [11] Y. Yang, L. Chen, W. Dong, and W. Wang, "Active base station set optimization for minimal energy consumption in green cellular networks," *IEEE Trans. Veh. Technol.*, vol. 64, no. 11, pp. 5340–5349, Nov. 2015.
- [12] S. Wu, Z. Zeng, and H. Xia, "Coalition-based sleep mode and power allocation for energy efficiency in dense small cell networks," *IET Commun.*, vol. 11, no. 11, pp. 1662–1670, Sep. 2017.
- [13] N. Yu, "Minimizing energy cost by dynamic switching ON/OFF base stations in cellular networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 11, pp. 7457–7469, Nov. 2016.
- [14] P. Chang and G. Miao, "Optimal operation of base stations with deep sleep and discontinuous transmission," *IEEE Trans. Veh. Technol.*, vol. 67, no. 11, pp. 11113–11126, Nov. 2018.
- [15] C. Liu, B. Natarajan, and H. Xia, "Small cell base station sleep strategies for energy efficiency," *IEEE Trans. Veh. Technol.*, vol. 65, no. 3, pp. 1652–1661, Mar. 2016.
- [16] F. H. Panahi, "Green heterogeneous networks via an intelligent sleep/wake-up mechanism and D2D communications," *IEEE Trans. Green Commun. Netw.*, vol. 2, no. 4, pp. 915–931, Dec. 2018.
- [17] J. Wu, E. W. M. Wong, Y. Chan, and M. Zukerman, "Energy efficiency-QoS tradeoff in cellular networks with base-station sleeping," in *Proc. IEEE GLOBECOM*, Singapore, Dec. 2017, pp. 1–7.
- [18] J. Luo, Q. Chen, and L. Tang, "Reducing power consumption by joint sleeping strategy and power control in delay-aware C-RAN," *IEEE Access*, vol. 6, pp. 14655–14667, 2018.
- [19] G. Zhang, Y. Cao, L. Wang, and D. Li, "Operation cost minimization for base stations with heterogenous energy supplies and sleep-awake mode: A two-timescale approach," *IEEE Trans. Cogn. Commun. Netw.*, vol. 4, no. 4, pp. 908–918, Dec. 2018.
- [20] J. Gong, J. S. Thompson, S. Zhou, and Z. Niu, "Base station sleeping and resource allocation in renewable energy powered cellular networks," *IEEE Trans. Commun.*, vol. 62, no. 11, pp. 3801–3813, Nov. 2014.
- [21] W. U. Rehman, A. Hussain, and M. M. Butt, "Joint user association and BS switching scheme for green heterogeneous cellular network," in *Proc. IEEE GLOBECOM WKSHPs*, Abu Dhabi, United Arab Emirates, Dec. 2018, pp. 1–6.
- [22] Z. Li, D. Grace, and P. Mitchell, "Hotspot-oriented green frameworks for ultrasmall cell cloud radio access networks," *IEEE Trans. Veh. Technol.*, vol. 67, no. 1, pp. 703–717, Jan. 2018.
- [23] L. Tang, W. Wang, Y. Wang, and Q. Chen, "An energy-saving algorithm with joint user association, clustering, and ON/OFF strategies in dense heterogeneous networks," *IEEE Access*, vol. 5, pp. 12988–13000, 2017.
- [24] N. Ben Rached, H. Ghazzai, A. Kadri, and M.-S. Alouini, "A time-varied probabilistic on/off switching algorithm for cellular networks," *IEEE Commun. Lett.*, vol. 22, no. 3, pp. 634–637, Mar. 2018.
- [25] Z. Li, D. Grace, and P. Mitchell, "Traffic-aware cell management for green ultradense small-cell networks," *IEEE Trans. Veh. Technol.*, vol. 66, no. 3, pp. 2600–2614, Mar. 2017.
- [26] J. Huang, V. G. Subramanian, R. Agrawal, and R. A. Berry, "Downlink scheduling and resource allocation for OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 8, no. 1, pp. 288–296, Jan. 2009.
- [27] G. Auer, V. Giannini, C. Desset, I. Godor, P. Skillermark, M. Olsson, M. A. Imran, D. Sabella, M. J. Gonzalez, O. Blume, and A. Fehske, "How much energy is needed to run a wireless network?" *IEEE Trans. Wireless Commun.*, vol. 18, no. 5, pp. 40–49, Oct. 2011.
- [28] D. Z. Du, K. I. Ko, and X. Hu, *Design and Analysis of Approximation Algorithms*. New York, NY, USA: Springer-Verlag, 2011.
- [29] X. Ge, S. Tu, G. Mao, and C. X. Wang, "5G ultra-dense cellular networks," *IEEE Trans. Wireless Commun.*, vol. 23, no. 1, pp. 72–79, Feb. 2016.
- [30] *Physical Layer Aspects for Evolved Universal Terrestrial Radio Access (UTRA) (Release 7)*, document 3GPP TR 25.814, V7.1.0, 2006.



ZHANG JIAN was born in Shanxi, China, in 1989. He received the B.S. and M.S. degrees from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2010 and 2013, respectively. He is currently pursuing the Ph.D. degree with the Beijing University of Posts and Telecommunications (BUPT), Beijing, China. His research interests focus on optimization of distributed resource allocation for heterogeneous cellular networks and cloud radio access networks.



WU MUQING was born in 1963. He received the Ph.D. degree. He is currently a Professor with the Beijing University of Posts and Telecommunications and a Senior Member of the China Institute of Communications. His current research interests focus on mobile ad hoc networks, UWB, high-speed network traffic control and performance analysis, and GPS locating and services.



ZHAO MIN received the Ph.D. degree in information and telecommunication systems from the Beijing University of Posts and Telecommunications (BUPT), Beijing, China, in 2014. She is currently a Lecturer with the Laboratory of Network System Architecture and Convergence, BUPT. Her research interest is in the areas of wireless communication systems.

...