Abstract—Small cells are widely used to offer flexible deployment and improve signal quality. To serve an increasing number of users with large traffic demands, they are densely deployed in many hotspot regions to share the load of a macrocell. This network scenario is critical in the upcoming 5G epoch. However, the issue of managing small cells to provide load balance and save energy is rarely discussed. Therefore, the paper proposes a load-aware small-cell management (LSM) mechanism for 5G networks, which organizes small cells into groups. In each group, a chief base station (BS) is elected to manage mobile devices (MDs) in all small cells. When a cell contains many MDs, they are transferred to other cells to alleviate congestion in its BS. Beside, a BS serving only few MDs will sleep to save energy. Then, each BS allocates resource to its MDs with the objective of reducing packet discard of real-time flows. Through simulations, we show that LSM can better support green communications by saving more energy of BSs and keeping high network throughput.

I. INTRODUCTION

Owing to low cost and flexibility, telecom-operators prefer deploying many small cells to consolidate their networks [1]. Small-cell BSs (s-BSs) not only improve signal quality in hotspot areas like shopping malls but also share the load of a macrocell BS (m-BS) whose signal coverage includes them [2]. We can foresee that the coexistence of macrocells and small cells will be the common scenario in 5G networks [3].

On the other hand, the OFDMA technique is widely used in wireless broadband communication devices such as WiMAX and LTE-A to support high-speed downlink transmissions [4], [5]. These protocols organize the spectrum resource into a 2D array of resource blocks (RBs) for a BS to allocate resource to MDs [6]. For example, LTE-A defines that one RB has 0.5 ms duration and 180 kHz bandwidth. Depending on the channel condition of an MD, each RB can transmit different number of data bits due to the usage of different modulations.

How to efficiently allocate RBs to MDs will decide system performance of a wireless network [7], [8]. Various resource management methods are thus developed to improve network throughput, maintain system fairness, or support quality of service (QoS) for real-time applications [9]. However, many methods are dedicated to resource allocation in one single cell. How to manage resource in a macrocell along with many small cells is rarely discussed in the literature.

Besides, it attracts considerable attention to support green communications by reducing energy consumption of BSs, especially in the network scenario where small cells are densely deployed [10]. Specifically, the density of MDs in small cells may change drastically over different times [11], where one representative is the region with business office buildings in a city. Generally speaking, there are usually a lot of MDs in that region on working days, but the same region will contain very few MDs on holidays. Because turning off idle s-BSs can greatly save their energy [12], it is economic to find out those s-BSs serving just few MDs and make them sleep, so as to avoid unnecessary energy consumption.

Motivated by the above observations, this paper proposes a load-aware small-cell management (LSM) mechanism for 5G networks where small cells are densely deployed. Our idea is to organize s-BSs into groups according to their geographic locations. In each group, a chief s-BS is selected to manage the MDs for its member s-BSs. In particular, when a member s-BS is busy in serving many MDs, the chief s-BS can transfer some of these MDs to nearby cells in order to alleviate its traffic load. On the other hand, when a member s-BS has just few MDs, the chief s-BS can transfer all of its serving MDs and make the s-BS sleep accordingly. In this way, most s-BSs could have sufficient resource to serve their MDs, and we are able to turn off more idle s-BSs to save their energy.

The paper contributes in developing an efficient management mechanism for 5G small cells by adaptively arranging their MDs, so as to balance traffic loads of s-BSs while making more s-BSs sleep to reduce energy consumption. Simulation results show that our proposed LSM mechanism not only keeps high network throughput but also saves more energy of BSs, which provides better support of green communications.

The rest of this paper is organized as follows. The next section discusses related work. In Section III, we propose the LSM mechanism. Then, performance evaluation is presented in Section IV. Finally, Section V concludes this paper and gives some future research directions.

II. RELATED WORK

A. Resource Management Schemes

The work [9] surveys a few classical schemes to manage the spectral resource for MDs. In particular, the maximum through-
put (MT) scheme always finds the MD with best data rate \( r_i \) to get resource. To support fairness [13], the proportional fair (PF) scheme serves MDs based on their \( r_i/\tau_i \) values, where \( \tau_i \) is an MD’s average data rate. The modified largest weighted delay first (M-LWDF) scheme uses \( (w_i d_i r_i)/\tau_i \) to select MDs, where \( w_i \) is a weight and \( d_i \) is the MD’s packet delay. Then, the exponential proportional fair (EXP/PF) scheme alters M-LWDF by \( \exp[(w_i d_i - d_A)/(1 + \sqrt{d_A})] \), where \( \exp[\cdot] \) is the exponential function and \( d_A \) is average packet delay.

Some work focuses on the issue of user fairness on data transmissions [14]. For example, [15] uses a bankruptcy game to model the resource management problem and solves it based on some rules in the game theory. The study [16] computes a credit for each MD by the gap between its amount of received data and the amount of resource that the system expects to give. MDs then compete for resource based on credit values. In [17], MDs are clustered into good-channel, average-channel, and bad-channel groups. A number of MDs in each group are picked to get resource to keep fairness. The work [18] adopts the Nash’s solution to allocate resource, which makes the result of resource allocation satisfy Pareto-optimal [19].

Several studies target at reducing packet delays of real-time flows. Liu et al. [20] tailor PF by giving a high priority to packets with the most urgent deadline. The work [21] allocates resource to MDs following the MT policy, and asks non-urgent flows to give back a portion of resource. Such resource is given to real-time flows whose packets are about to be dropped. Samia et al. [22] model the resource allocation problem as a cooperative game, and try to decrease dissatisfaction of real-time flows in terms of packet delays. The study [23] estimates the amount of resource given to real-time flows to support QoS, and uses MT and M-LWDF to allocate resource.

However, the above studies allocate resource merely in one single cell. None of them considers resource management in a macrocell together with multiple small cells. Consequently, it motivates us to develop the LSM mechanism to deal with resource management in this network scenario.

B. Energy Management Schemes

There are different schemes to reduce energy consumption of BSs. With the help of MIMO beam-forming and multiplexing gain, [24] develops a MAC protocol for BSs to transmit data in an energy-efficient manner. The work [25] selects suitable subchannels for each BS to send data, so as to reduce its transmitted power. The study [26] proposes an energy-saving strategy in cognitive-radio networks by allowing BSs to switch to a power-saving mode when there is no packet transmission. Wang and Lee [27] adaptively add s-BSs to reduce the load and transmitted power of each m-BS, under the limitation of budget.

A number of studies decide when to make a BS sleep to save its energy. Abdallah et al. [28] shorten the overlapped duration when neighboring BSs wake up for data transmissions, so as to avoid signal interference between adjacent cells. The work [29] uses the game theory to predict the traffic profile of MDs, and turns off BSs when no packets wait for transmissions. The study [30] models user traffics by a Markov process, and lets BSs sleep when the traffic load is light to save energy. Unlike these studies, our work proactively transfers MDs among small cells, so as to make the s-BSs that serve only few MDs sleep to reduce their energy consumption.

III. THE LSM MECHANISM

We consider a macrocell inside which there are multiple s-BSs to provide service, as illustrated in Fig. 1. Then, our LSM mechanism consists of four stages to manage small cells:

- **BS grouping:** As shown in Fig. 1, we first divide s-BSs into groups based on their geographic locations.
- **Load estimation:** Given the MDs in each small cell, we then calculate the traffic load of its s-BS.
- **MD arrangement:** The chief s-BS transfers MDs among its member cells for load balancing and energy saving.
- **RB allocation:** Each BS finally allots RBs to its MDs.

Below, we detail the design of each stage.

A. Stage of BS Grouping

In the first stage, our goal is to group together those s-BSs which are close to each other, so the chief s-BS can arrange MDs in its member cells. To do so, we amend the enhanced agglomerative hierarchical clustering (eAHC) approach [31], which recursively groups s-BSs by referring to their locations. In particular, eAHC contains three steps:

1) Initially, every s-BS is treated as one single group.
2) We then merge two groups \( \hat{G}_i \) and \( \hat{G}_j \) such that they have the shortest inter-group distance \( D(\hat{G}_i, \hat{G}_j) \), which is defined by the Euclidean distance between two farthest s-BSs \( b_x \) and \( b_y \), where \( b_x \in \hat{G}_i \) and \( b_y \in \hat{G}_j \). Note that \( \hat{G}_i \) and \( \hat{G}_j \) are allowed to be merged only when \( D(\hat{G}_i, \hat{G}_j) \leq d_D \), where \( d_D \) is a threshold to limit the diameter of a group.
3) Step 2 is repeated until no groups can be further merged. After dividing s-BSs into groups, we then arbitrarily select one s-BS to be the chief s-BS in each group.
B. Stage of Load Estimation

To estimate the load of an s-BS, we need to obtain SINR of each of its serving MDs. Supposing that an MD \( m_i \) is located in an s-BS \( b_j \)'s cell, its SINR value is calculated as follows:

\[
s_{i,j} = \frac{p(m_i, b_j)}{\zeta + \sum \{p(m_i, b_k) \mid \forall b_k \in \hat{B} \setminus \{b_j\}\}}.
\]  

(1)

In Eq. (1), \( p(m_i, b_j) \) gives the strength of power received by \( m_i \) which is emitted from \( b_j \), \( \zeta \) indicates noise interference from the environment, and \( \hat{B} \) denotes the set of all BSs. Then, according to the Shannon theorem [32], we can estimate the capacity of a small cell with BS \( b_j \) by

\[
C_j = \tau \cdot \log \left(1 + \frac{\sum_{m_i \in \hat{M}_j} s_{i,j}}{|\hat{M}_j|}\right),
\]  

(2)

where \( \tau \) is \( b_j \)'s channel bandwidth and \( \hat{M}_j \) is the set of MDs located in \( b_j \)'s cell. Let us denote by \( r_i \) the traffic demand of \( m_i \). Then, the load of \( b_j \) can be estimated by

\[
\lambda_j = \frac{\sum_{m_i \in \hat{M}_j} r_i}{C_j}.
\]  

(3)

C. Stage of MD Arrangement

For each group of s-BSs, we further divide it into three sub-groups: busy, normal, and idle. In particular, when an s-BS \( b_j \) has load of \( \lambda_j > \delta_H \), it is added to the busy sub-group. On the other hand, if it has load of \( \lambda_j < \delta_L \), \( b_j \) is added to the idle sub-group. Otherwise, we add \( b_j \) to the normal sub-group. Here, \( \delta_H \) and \( \delta_L \) denote the upper-bound and low-bound thresholds on the traffic load, respectively, and we suggest setting \( \delta_H \geq 1 \) and \( \delta_L \leq 1/5 \).

For each s-BS \( b_j \) in the busy sub-group, the chief s-BS iteratively transfers the farthest MD in \( b_j \)'s cell to the nearest small cell whose s-BS belongs to the normal sub-group and its load is kept below \( \delta_H \) after the transfer of that MD, until \( \lambda_j \leq \delta_H \). In case that no such small cells can be found, the chief s-BS will ask the m-BS \( b_m \) to take the MD (if \( \lambda_m \leq \delta_H \)). Afterwards, the chief s-BS seeks to transfer all MDs of each s-BS in the idle sub-group according to the above rules. When an s-BS in the idle sub-group has no MDs to serve (due to the operation of MD transfer), it can be turned off to save energy.

D. Stage of RB Allocation

In the last stage, each BS allocates RBs to its MDs. Specifically, when a BS \( b_j \) has load of \( \lambda_j \leq 1 \), it means that \( b_j \) has sufficient resource to satisfy the traffic demands of all serving MDs. Otherwise, \( b_j \) first gives RBs to those MDs whose real-time packets are about to be dropped. Then, each MD will compete for \( b_j \)'s residual RBs according to the MT policy discussed in Section II-A.

IV. PERFORMANCE EVALUATION

In this section, we evaluate LSM’s performance by MATLAB, where Fig. 2 presents the network topology. In particular, the m-BS has cell range of 1.5 km, transmitted power of 46 dBm, and channel bandwidth of 20 MHz. On the other
hand, each s-BS has cell range of 0.25 km, transmitted power of 30 dBm, and channel bandwidth of 5 MHz. When an s-BS is in the sleeping state, it spends just one half of its energy. In addition, we follow the LTE-A standard [33] to model wireless transmissions. Specifically, the effect of path loss between an MD $m_i$ and a BS $b_j$ is evaluated by a log-distance model:

$$128.1 + 37.6 \log \left( \frac{D_{m_i, b_j}}{m_i} \right) \quad \text{if } b_j \text{ is the m-BS}, \tag{4}$$

$$38 + 30 \log \left( \frac{10^{3} D_{m_i, b_j}}{b_j} \right) \quad \text{otherwise}, \tag{5}$$

where the distance $D(m_i, b_j)$ is measured in km. On the other hand, the shadowing effect (for fading) is modeled by a log-normal distribution whose mean is 0. If $b_i$ is the m-BS, the standard deviation is set to 10 dB; otherwise, it is set to 6 dB.

We also increase the number of MDs from 60 to 660 to observe its effect. Each MD has a 242 kbps video flow and a 12 kbps constant-bit-rate flow. About 2/9 of MDs congregate in the lower left corner of the network to simulate the hotspot region, as illustrated in Fig. 2. Other MDs are then uniformly distributed over the whole network. For comparison, we adopt the MT scheme discussed in Section II-A for each BS to manage its resource individually, which attempts to increase throughput by giving each RB to the MD with the best channel quality. Besides, in the MT scheme, when an s-BS has no MDs to serve, it will go to the sleeping state to save energy. For our LSM mechanism, we set $\delta_H = 1$ and $\delta_L = 1/5$.

Fig. 3 shows the experimental result of network throughput. It is obvious that network throughput increases when the number of MDs grows, because we consider the aggregate throughput of MDs. Thanks to the arrangement of MDs by the chief s-BS in each group, our LSM mechanism can balance the traffic load of each s-BS, especially in the hotspot region. In this case, we can make sure that most s-BSs have sufficient resource to serve their MDs. Consequently, our LSM mechanism always has higher network throughput than the MT scheme. In particular, it can averagely improve 20% of network throughput as comparing with the MT scheme.

Fig. 4 presents the amount of energy consumed by BSs. When there are only 60 MDs in the network, the MT scheme has an opportunity to search out idle s-BSs, since the density of MDs is pretty low. However, when the number of MDs reaches to 180, the MT scheme can no longer find out any s-BS which serves no MDs. In this case, every s-BS has to become active to provide service, even though some small cells actually contain very few MDs. On the contrary, our LSM mechanism can adaptively arrange MDs for each group of s-BSs. In this way, there is a higher opportunity for the chief s-BS to ask some of its member s-BSs sleeping to save energy. Only when there are more than 420 MDs in the network, will the LSM mechanism perform the same with the MT scheme, because the density of MDs becomes much higher.

Through the results in both Fig. 3 and 4, we demonstrate that our LSM mechanism is capable of keeping high network throughput and saving more energy of BSs when there are fewer MDs in the network, which verifies that the proposed mechanism can better support green communications.

V. CONCLUSION

For the upcoming 5G epoch, it is an inevitable trend to densely deploy small cells to improve service quality. How to manage small cells to improve performance while reducing energy consumption is thus a critical issue. Therefore, this paper develops the LSM mechanism by organizing small cells into groups. In each group, the chief s-BS adaptively arranges MDs for its member s-BSs to not only balance their traffic loads but also make more s-BSs sleep to save energy. Through MATLAB simulations, we show that our LSM mechanism can provide better support of green communications.

We finally discuss several research directions. First of all, because this paper considers managing the downlink resource, we will address how to manage the uplink resource for MDs, especially with the existence of relay stations [34], in the future work. Second, the DVB-H (digital video broadcasting-handheld) service becomes more popular to provide many multimedia applications such as mobile television [35]. Thus, we will study how to deal with DVB-H traffics in 5G networks with dense small cells, which should consider how to fast recover the lost DVB-H packets due to the broadcast nature [36]. Third, since Wi-Fi networks are also densely deployed to provide broadband Internet access in hotspot regions [37], it deserves investigation to integrate these small cells with Wi-Fi cells to balance traffic loads of macrocells [38]. Fourth, it is interesting to address device-to-device (D2D) communications in 5G networks, where MDs can directly communicate with each other in an ad hoc fashion [39], [40].

ACKNOWLEDGEMENT

You-Chiun Wang’s research is co-sponsored by the Ministry of Science and Technology under Grant No. MOST 106-2221-E-110-022-MY2, Taiwan.

REFERENCES


