Cluster Analysis Based Switching-off Scheme of Base Stations for Energy Saving

Hongzeng He¹, Jingbo Sun¹, Yue Wang¹, Qi Liu² and Jian Yuan¹
¹Department of Electronic Engineering, Tsinghua University, Beijing 100084, China
²China Unicom Research Institute, Beijing 100048, China
Email: {hehz09, sjb06}@mails.tsinghua.edu.cn, {wangyue, liu-qi, jyuan}@mail.tsinghua.edu.cn

Abstract—Recent studies on green communications and networks indicate that most traffic loads could be served by a certain part of base stations during the low load period. In this paper, we present an energy saving approach which is based on the cluster analysis of traffic loads in urban space. Consequently, different traffic patterns are discovered and specific switching-off approaches are applied to the different traffic patterns. Simulation results reveal that a better energy saving performance can be attained compared to the existing approach.

Index Terms—Energy savings, Cellular networks, Cluster analysis

I. INTRODUCTION

Our earth is getting warmer in the last century due to the rising emission of CO₂ and other greenhouse gases. Low-carbon economy and energy saving have now become a hot topic in our daily life. As the rapid development of information and communication technology (ICT) infrastructure, the ICT energy consumption could not be neglected. According to some scientific findings recently, 3% of the world-wide energy consumption could not be neglected. According to some scientific findings recently, 3% of the world-wide energy consumption could not be neglected. According to some scientific findings recently, 3% of the world-wide energy consumption could not be neglected. According to some scientific findings recently, 3% of the world-wide energy consumption could not be neglected. According to some scientific findings recently, 3% of the world-wide energy consumption could not be neglected. According to some scientific findings recently, 3% of the world-wide energy consumption could not be neglected. According to some scientific findings recently, 3% of the world-wide energy consumption could not be neglected. According to some scientific findings recently, 3% of the world-wide energy consumption could not be neglected. According to some scientific findings recently, 3% of the world-wide energy consumption could not be neglected. According to some scientific findings recently, 3% of the world-wide energy consumption could not be neglected. According to some scientific findings recently, 3% of the world-wide energy consumption could not be neglected. According to some scientific findings recently, 3% of the world-wide energy consumption could not be neglected. According to some scientific findings recently, 3% of the world-wide energy consumption could not be neglected. According to some scientific findings recently, 3% of the world-wide energy consumption could not be neglected. According to some scientific findings recently, 3% of the world-wide energy consumption could not be neglected. According to some scientific findings recently, 3% of the world-wide energy consumption could not be neglected. According to some scientific findings recently, 3% of the world-wide energy consumption could not be neglected. According to some scientific findings recently, 3% of the world-wide energy consumption could not be neglected. According to some scientific findings recently, 3% of the world-wide energy consumption could not be neglected. According to some scientific findings recently, 3% of the world-wide energy consumption could not be neglected. According to some scientific findings recently, 3% of the world-wide energy consumption could not be neglected. Accordingly, different traffic patterns are discovered and specific switching-off approaches are applied to the different traffic patterns. Simulation results indicate a better energy saving performance based on cluster analysis, compared to the existing approach.

The outline of this paper is as follows. Section II presents some general green communication network papers and some of the relevant research areas that can be divided into three levels. Section III introduces the implementation of cluster analysis and characteristics of the original data used in our work. In Section IV we describe the details of our energy saving scheme. Section V investigates the simulation results. The benefits and energy saving effects is proposed in this section. In Section VI, we proposed the directions of the future work and end this paper with some concluding discussions.

II. RELATED WORK

Nowadays nearly all the approaches to deal with energy saving issues can be divided into three groups: design level, physical level and algorithm level.

A. Design level

Scientific researches in design level focus on the telecommunication network planning and base station deployment at the design stage. A good network design scheme, such as an optimal base station site selection plan, could bring in a lot of benefits. In this case, not only can we get a higher bandwidth utilization and a lower transmission power, but also guarantee a stipulated signal to noise ratio and service coverage, which leads to a significant energy consumption reduction and cost saving. Some research communities have proposed some algorithms to tackle these issues. It is shown that an optimistic effect can be achieved by the certain usage of genetic algorithm [2] and optimization framework [3]. Accordingly, the base station deployment should put the traffic requirement first in the urban space with high traffic load, in contrast to the maximal service coverage ratio with minimal base stations used in the rural space with low traffic load.

B. Physical level

What we concerned by this level is the energy efficiency of each component in one base station device, including power amplifier (PA) and cooling equipment, etc. According to the statistics, the base stations account for 50% of the energy consumptions in telecommunication networks worldwide. So it matters a lot addition, specific methods are applied to these different traffic patterns. Simulation results indicate a better energy saving performance based on cluster analysis, compared to the existing approach.
whether the telecommunication devices are energy efficient and environment-friendly.

Fig. 1 shows the energy consumption of each component in one BS [4]. As is illustrated in Fig. 1, the power amplifier (PA) device takes 46% of the total energy consumption in one base station, in comparison to the air conditioner 11% and AC/DC devices 11%. It is easy to see that all of these three parts take the greatest proportion of the energy consumptions in one base station. Some techniques have been proposed to improving the energy efficiency of these devices, such as Time-slot Intelligent Shut and Pre-distortion of the Repeater [5], high efficiency Doherty design techniques [6] and so on.

C. Algorithm Level

As we know, the temporal traffic load varies violently by time during one day in each base station. But nowadays most designs of communication networks frequently have not considered for the adaption to the variable traffic load, which means that base stations work in full power mode all day long. There will be a great waste of communication resources and energy resources.

In addition, the spatial traffic load state changes strongly by geographical position, which usually indicates the high traffic in urban city and the low traffic in rural space. The fact leads to a high base station deployment density in urban space, which brings about some idle base stations and service coverage redundancy [7] when the traffic is low.

Since the traffic load is uneven over time and space, it makes it possible to switch off some idle base stations or turn them into sleep mode by making the base station operation mode proportional to the changing traffic load. Some previous work has been published about the quantification of the energy saving via switching off some site during off-peak hours, which can achieve a reduction of energy consumption up to about 40% in a uniform distributed network configuration model [8].

An optimal energy saving switching-off method is proposed to reducing the number of active base stations according to the daily traffic pattern in Ref. [9]. This method achieves the energy savings of about 25%-30%. Besides, if we do some analysis on the original data, a better performance is possible. Some other studies published in [10] point out that we can switch off low utilization base stations via store carry and forward relaying scheme. Considering the service coverage ratio and Quality of Service (QoS) after switching off base stations, they demonstrate some approaches to tackle the traffic load in [11], such as base station antenna beam tilting technology and cellular coverage control.

III. BACKGROUND AND DATA

A. K-means Cluster Analysis

Cluster Analysis is a conventional approach to classifying a set of data into different clusters so that the data in the same cluster could show some similar characteristics in a manner. Cluster Analysis is widely used in many fields, including mathematics, computer science, statistics science, biology science and economics. There are some types of clustering, such as hierarchical algorithms, partitional algorithms, density-based clustering algorithms and subspace clustering methods.

In this paper, we prefer the partitional clustering method which uses a high speed algorithm and is suitable for large datasets. It could determine all the clusters in less time. The k-means clustering algorithm is one of partitional methods. It assigns each element to the cluster whose distance is nearest. Each cluster is characterized by the arithmetic mean for each element over all the data in the cluster. The implement of K-means includes at least the following steps [12]:

- Choose the number of clusters, k.
- Randomly generate k clusters and determine the cluster centers, or directly generate k random points as cluster centers.
- Assign each point to the nearest cluster center, where “nearest” is defined with respect to one of the distance measures discussed above.
- Re-compute the new cluster centers.
- Repeat the two previous steps until some convergence criterion is met (usually that the assignment hasn’t changed).

As a result of the cluster analysis, we can sort all the data sets into k clusters based on the affinity of each element.

B. Original data

The original data of this paper are from a flourishing district of a big city in China, supported by one telecommunication operator. This region covers an area about 470.8 square kilometers. It has a population of nearly 3.1 millions, which means an average density of six thousand and five hundred persons per square kilometer.

The original data contains the traffic load statistics of each base station over 8 days in Erlang, including weekdays and weekends. The data is updated every hour which means that there are traffic data of more than 190 hours. The traffic statistics indicate a violent fluctuation on the temporal scale, which is shown in Fig. 2. We observe variations during 24 hours in one day and differences of amplitude between weekday period and
weekend period. For example, the traffic load of weekend period is always much lower than that of weekday period.

**Fig. 3** is a distribution map of the base station deployments based on the information provided by the original data. Each blue point represents one base station location. The total number is about several hundred. We can see a much higher base station density in the urban space than there in the rural area. As a matter of fact, there will be more redundancy in the cellular coverage that makes rooms for switching-off some of the base stations in the low load period [7].

All of these temporal and spatial characteristics make it possible for us to implement the energy saving algorithm based on cluster analysis.

### IV. ENERGY SAVING SCHEME

We now use these approaches and real datasets to carry out the energy saving algorithm based on cluster analysis. First, we introduce a normalization method to make the original datasets convenient for the calculation afterwards. Then we use a K-Means approach to create segmentations based on the affinity between the traffic loads of different base stations. In this step, the analysis results exhibit certain patterns of the telecommunication networks in urban city. At last, according to the patterns we have found, an energy saving algorithm is applied to the real network. We now describe the details of our works mentioned above.

#### A. Normalization

Data normalization is necessary for comparing the relative change of the traffic loads in different base stations in a more detailed way. Using this approach, we can make original data easier for deeper observation and further calculation.

Our normalization steps are as follows [13]:

1. For each base station $b_i \in B$ ($i=1, 2, \ldots, N$), we do averaging to the value of 8 days (192 hours) that is represented by

$$M = \text{mean}\{\text{erlang}(b_i, \tau)\} \quad i=1, 2, \ldots, N$$  \hspace{1cm} (1)

where $b_i$ is the base station in our datasets and $\tau$ is the time period across 192 hours.

2. We then calculate the variance of traffic load over time in each base station, as in

$$S = \text{std}\{\text{erlang}(b_i, \tau)\} \quad i=1, 2, \ldots, N.$$  \hspace{1cm} (2)

3. We then normalize the data to the uniform margin over time of 192 hours in (3).

$$\text{erlang}_{\text{norm}_i}^{192} = \frac{\text{erlang}(b_i, \tau) - \text{mean}\{\text{erlang}(b_i, \tau)\}}{\text{std}\{\text{erlang}(b_i, \tau)\}}$$  \hspace{1cm} (3)

where $i=1, 2, \ldots, N$

4. The differences between weekday periods and weekend periods are quite visible in Fig. 2. So it is necessary to consider the two cases separately. We then classify the original data into two groups. One is the traffic load profile of weekday periods, and the other is of weekend periods.

5. Finally we normalize the data of the two groups to one day time separately, which can be expressed as

$$\text{erlang}_{\text{norm}_i}^{24} = \text{erlang}_{\text{norm}_i}^{192}(b_i, \tau_j)/3, \quad i=1, 2, \ldots, N$$  \hspace{1cm} (4)

where $\tau_j \in [1+24(j-1), 24+24(j-1)]$ and $j=1,2,8$, and

$$\text{erlang}_{\text{norm}_i}^{24} = \text{erlang}_{\text{norm}_i}^{192}(b_i, \tau_j)/5, \quad i=1, 2 \ldots, N$$  \hspace{1cm} (5)

where $\tau_j \in [1+24(j-1), 24+24(j-1)]$ and $j=3,4\ldots7$.

As a result, we get two groups of normalized data of the real network in this region, including temporal and spatial traffic load status and base station locations information. Fig. 4 shows the traffic load of 4 sample base stations in 24 hours of one day. We observe that the traffic load during the nighttime is much lower than that in the daytime. The heavy load time is from 11am to 20pm, even less than ten hours. Fluctuations in 24 hours are related to the work and living activities of our inhabitants. In addition, differences between the two groups in Fig. 4 are obviously visible. On the weekdays,
we observe that the traffic load is always higher than that on weekends. Instead of working at the workplace, a large majority of the population prefers to stay at home and places of entertainment. And it leads to a migration of traffic load and, of course, a reduction of load in the workplaces. 

B. K-Means Cluster Analysis

So far, we have put our attention largely on the temporal feature of traffic profile in individual base station. We have identified unevenness and wave properties of the traffic load profile on the temporal scale. Now we focus on the analysis of the whole telecommunication networks, especially spatial properties. Because of the disparity between weekdays and weekends, we analysis these two cases separately in this paper.

We adopt the K-means clustering approach to the normalized data we have obtained. After a series of different attempts, we decide to classify all the base stations into 3 clusters by two K-means clustering stages. In other words, the value of K is 3.

First, we randomly choose 3 base stations as the original center of each cluster, designated by $Z_k$, $k=1, 2, 3$.

Second, for each base station $b_i$, we calculate the distance to each $Z_k$ and put it to the cluster $\chi$ with the smallest distance by

$$
X = \min_{i=1, 2, \ldots, N} \left\| \text{erlang}_{\text{norm}}_{24}(b_i, \tau) - Z_k \right\|.
$$

(6)

Third, we re-calculate the center of every cluster, $Z_k$ as in

$$
Z'_k = \frac{1}{N_X} \sum_{b_i \in X} \text{erlang}_{\text{norm}}_{24}(b_i, \tau)
$$

(7)

where $N_X$ represents the number of base stations in cluster $X$.

We then repeat stage 2 and 3 until match the following equation (8):

$$
Z'_k = Z_k.
$$

(8)

As a result of K-means cluster approach, we group all the base stations into 3 types. Each type has its own changing law. The average variation traces can be drawn in Fig. 5 and Fig. 6. The average variation of the traffic load normalized in 24 hours in each cluster is much unlike that in other clusters. Cluster 1, signified by the blue line, appears a flatter changing pattern during all the 24 hours. It implies a matter of fact that base stations in this cluster tackle a medium amount of traffic load with a smaller variation. Cluster 2, which is painted red, exhibits higher values during the working hours from 10 pm to 17 pm except the lunch break time. Additionally, the average load of the base stations in black of cluster 3 concentrate in the after office hours from 19 pm to 23 pm. As a result, the varying patterns of the 3 types show significant differences in operating features and working modes of the different base stations. The traffic load information is related to activity intensity of mobile users under coverage of the base station.

We then conclude the geographical features of base stations with the statistics we have got before. Cluster 1 (blue) shows the evenness in contrast to Cluster 2 and Cluster 3, suggesting that the base stations in Cluster 2 cover areas requiring a normal level communication services all through the day, including transport hubs and
other gathering places. Cluster 2 (red) which shows a high traffic level in the daytime may be connected to the working places, such as companies and banks, where is busy and bustling during business hours. Cluster 3 (black) shows a significant peak at night, suggesting that base stations in Cluster 3 cover the utility areas. For example the residential districts, public houses and restaurants. On the weekdays, we observe higher peaks for the rapid changing of position and intensive activities on work-time. On the weekends, there is a smaller slope and lower peaks by contrast. Naturally, we examine two different traffic load patterns between weekdays and weekends. And the same station appears different patterns between weekdays and weekends.

Next, we investigate the spatial features of these 3 groups of base stations. We are aware of an initial impression by projecting these base stations onto a map in different colors, as in Fig. 7. As is shown in Fig. 7, there are some differences of the locations of each cluster. It can be observed more clearly in Fig. 8. Cluster 1 (blue) shows almost a homogeneous distribution over spatial scale, consisting with its equitability on temporal scale. It is helpful to understand the relations between temporal and spatial scale. Cluster 2 (red) concentrates to the urban space of the city. In addition, there are more base stations in Cluster 2 on weekdays than that on weekends, which means a heavier load in Cluster 2 on weekdays. We examine that Cluster 3 maps to surrounding areas with a sparse density, suggesting the leisure places and uptowns. For the analysis of the 3 clusters, we have seen the individual characteristics of the 3 clusters that reflect the particular geographical locations.

C. Switching-off Approach

We hope to apply the energy saving algorithm to the real networks according to the patterns we have just found. In this paper, we use the constant energy consumption profile [14]

\[ P(b_i) = P_{0c} = const. \]  \hspace{1cm} (9)

The constant energy consumption profile assumes that the power consumption of each base station is a constant quantity that is independent on the variation of the traffic load. The constant profile fits well to the current telecommunication networks. We assume that there are two states of base station: active and sleep, where in the later state the energy consumption is negligible. Therefore, the energy consumption could be modeled with two values: 0 and \( P_{0c} \). An energy saving approach based on the average traffic pattern is proposed in [9] as follows:

- Let \( f(t) \) be the average traffic pattern in one day, which is described as a function of time \( t \), with \( t \in [0, 24] \).
- Normalize \( f(t) \) to the peak of 1, so that \( f(0) \) is the peak hour with the value \( f(0)=1 \).
- Assuming that a fraction \( x<1 \) of the base stations is in working mode while a fraction \( 1-x \) of the base stations is switched off during a period when the traffic load is below certain threshold.
- Let \( \tau_1 \) and \( \tau_2 \) be two time points of the switching-off period, with \( f(\tau_1) = f(\tau_2) \). The optimization is given by (10).

\[
\min C(\tau_1, \tau_2) = P_{0c} \left[ T - (\tau_2 - \tau_1) + f(\tau_1)(\tau_2 - \tau_1) \right].
\]

\[
\text{s.t. } f(\tau_1) = f(\tau_2), \hspace{1cm} (10)
\]

where \( C(\tau_1, \tau_2) \) represents the average energy consumed per base station in a day under this approach and \( T \) reflects the 24 hours in one day.

We make some improvements by introducing cluster analysis to this approach. Firstly, we apply this approach to the different traffic patterns of the three clusters to control the on-off states of base stations separately. It provides us with more precise calculations on the energy savings. We determine the best \( \tau_1 \) and \( \tau_2 \) with the biggest energy saving ratio using the optimal method, as is in Table I.

Then we calculate the total energy savings of the improved approach by (11)

\[
\text{Net}_{\text{saving}} = 1 - \frac{C(\tau_1, \tau_2)}{P_{0c}T}, \hspace{1cm} (11)
\]

We have introduced some simplifying assumption about the network environment. For example, the covering radius of active base stations could be extended.

<table>
<thead>
<tr>
<th>TABLE I.</th>
<th>SWITCHING TIME POINT OF THREE CLUSTERS</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \tau_1 )</td>
<td>( \tau_2 )</td>
</tr>
<tr>
<td>Weekdays scenario</td>
<td></td>
</tr>
<tr>
<td>Cluster1</td>
<td>1:27</td>
</tr>
<tr>
<td>Cluster2</td>
<td>20:45</td>
</tr>
<tr>
<td>Cluster3</td>
<td>00:48</td>
</tr>
<tr>
<td>Weekends scenario</td>
<td></td>
</tr>
<tr>
<td>Cluster1</td>
<td>0:34</td>
</tr>
<tr>
<td>Cluster2</td>
<td>23:08</td>
</tr>
<tr>
<td>Cluster3</td>
<td>0:50</td>
</tr>
</tbody>
</table>

© 2012 ACADEMY PUBLISHER
to tackle the entire service requirement. However, some problems may emerge due to the extending of base station coverage, such as same frequency interference. As a result, certain technologies should be employed to protect against these problems, for example smart antenna is a new technology that can reducing the interferences of cellular networks by adjusting the antenna configurations. As a result, the coverage areas of the base stations are reorganized to avoid interferences in the overlapping coverage areas.

Introducing energy saving approaches to the large spatial scales, we introduce some assumptions to the scenarios, including the average traffic curve applied to the base stations in the same cluster and service maintenance after the switching-off operations of certain base stations.

V. SIMULATION RESULTS

Simulations are performed to evaluate of our energy-efficiency approach. We focus on the different results of the two methods with/without cluster analysis on

<table>
<thead>
<tr>
<th></th>
<th>Energy saving ratio [%]</th>
<th>proportion in number [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Weekdays scenario</td>
<td></td>
</tr>
<tr>
<td>Cluster1</td>
<td>25.24</td>
<td>41.12</td>
</tr>
<tr>
<td>Cluster2</td>
<td>38.37</td>
<td>19.73</td>
</tr>
<tr>
<td>Cluster3</td>
<td>25.08</td>
<td>39.15</td>
</tr>
<tr>
<td>Total</td>
<td>27.77</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>Weekdays scenario</td>
<td></td>
</tr>
<tr>
<td>Cluster1</td>
<td>23.68</td>
<td>38.95</td>
</tr>
<tr>
<td>Cluster2</td>
<td>31.13</td>
<td>15.85</td>
</tr>
<tr>
<td>Cluster3</td>
<td>24.52</td>
<td>45.20</td>
</tr>
<tr>
<td>Total</td>
<td>25.24</td>
<td>100</td>
</tr>
</tbody>
</table>

Figure 9. Comparison of performance of two approach in energy saving.
weekdays and weekends. So we investigate energy savings of the algorithm utilizing the average traffic load of each cluster.

We applied the energy saving method to the scene of real network and access the energy saving ratios of our method on different conditions. Analysis results are summarized in Table II. We observe different energy saving levels of each cluster, and of course the total saving amount of the three on weekday scenarios and weekend scenarios. We then inspect the benefits of the approach based on cluster analysis compared to those without clustering, which is shown in Fig. 9. Here we can see that an additional part of about 3% to 4% is saved if we introduce the clustering analysis to energy saving algorithm.

VI. CONCLUSIONS

An energy saving approach based on cluster analysis was proposed in this paper. Our proposed energy saving approach tries to reduce the energy consumption of the entire network. The cluster analysis on original data from an operating cellular network in a big city exploits different traffic patterns of base stations. In addition, specific methods are applied to these different traffic patterns. Simulation results indicate a better energy saving performance based on cluster analysis, compared to the existing approach.

ACKNOWLEDGMENT

This work is support by the National Basic Research Program of China (973 Program) under grants 2007CB307100 and 2007CB307105.

REFERENCES


Hongzeng He is a graduate student in the Department of Electronic Engineering at Tsinghua University. He received his B.S. degree from the School of Electric and Electronic Engineering at North China Electric Power University in 2009. His research interests include mobility management and green networks.

Jingbo Sun is a Ph.D. candidate in the Department of Electronic Engineering at Tsinghua University. He received his B.S. degree from the School of Electronics and Information Engineering at Harbin Institute of Technology in 2006. His research interests include wireless networks and mobility management.

Yue Wang received his Ph.D. degree from the Electronic Engineering Department of Tsinghua University in 2005. He is now an assistant professor at Tsinghua University. His research interests include computer networks and data fusion.
Qi Liu received her B.S. degree and Ph.D degree in communication engineering from Beijing Jiaotong University, Beijing, P.R. China in 2003 and 2009. She is presently working in China Unicom Research Institute, Beijing. Her research interests are mobile and wireless communications.

Jian Yuan received his Ph.D. degree in electrical engineering from the University of Electronic Science and Technology of China, in 1998. He is currently an associate professor in the Department of Electronic Engineering at Tsinghua University, Beijing, China. His core interest is complex dynamics of networked systems and dependability of mobile networks.