Complementary Base Station Clustering for Cost-Effective and Energy-Efficient Cloud-RAN

Longbiao Chen, Thi-Mai-Trang Nguyen, Gang Pan, Jérémie Jakubowicz,
Linjin Liu, Xiaoliang Fan, Johnthan Li, Cheng Wang

Complementary Base Station Clustering for Cost-Effective and Energy-Efficient Cloud-RAN

Longbiao Chen*†, Thi-Mai-Trang Nguyen†, Gang Pan‡, Jérémie Jakubowicz§,
Linjin Liu*, Xiaoliang Fan*, Johnthan Li*, Cheng Wang*

*Fujian Key Laboratory of Sensing and Computing for Smart City, Xiamen University, Xiamen, China
†Sorbonne Universities, University of Paris VI, UMR 7606, LIP6, 4 Place Jussieu, 75005 Paris, France
‡College of Computer Science, Zhejiang University, Hangzhou, China
§Institut Mines-Télécom; Télécom SudParis; UMR CNRS SAMOVAR, Evry 91000, France
Email: longbiao.chen@xmu.edu.cn

Abstract—The fast growing mobile network data traffic poses great challenges for operators to increase their data processing capacity in base stations in an efficient manner. With the emergence of Cloud Radio Access Network (Cloud-RAN), the data processing units can now be centralized in a data center and shared among several base stations. By clustering base stations with complementary traffic patterns to the same data center, the deployment cost and energy consumption can be reduced. In this paper, we propose a two-phase framework to find optimal base station clustering schemes in a city-wide Cloud-RAN. First, we design a traffic profile for each base station, and propose an entropy-based metric to characterize the complementarity among base stations. Second, we build a graph model to represent the complementarity as link weight, and propose a distance-constrained clustering algorithm to find optimal base station clustering schemes. We evaluate the performance of our framework using two months of real-world mobile network traffic data in Milan, Italy. Results show that our framework effectively reduces 12.88% of deployment cost and 9.45% of energy consumption compared with traditional architectures, and outperforms the baseline method.

Index Terms—Cloud-RAN; base station clustering; mobile network data; network optimization

I. INTRODUCTION

Today, mobile network data traffic is growing explosively as Internet-enabled smartphones and tablets become increasingly popular [1]. According to Cisco [2], global mobile network data traffic has grown 18-fold over the past five years. In order to satisfy the fast growing data traffic demand, mobile network operators need to increase their *data processing capacity*, such as deploying more base stations, and adding more data processing units to the base stations. However, the deployment cost of these network infrastructures are becoming increasingly high, and may exceed operator's revenue as network scale grows [3]. Moreover, the energy consumption of mobile network infrastructures are substantially increasing, taking up more than 3% of the worldwide electric energy consumption nowadays [4]. Therefore, optimizing the *deployment cost and energy consumption* has become a necessity for mobile network operators [5].

Interestingly, for a mobile network, although the overall data traffic demand at the city scale is growing, the demands in
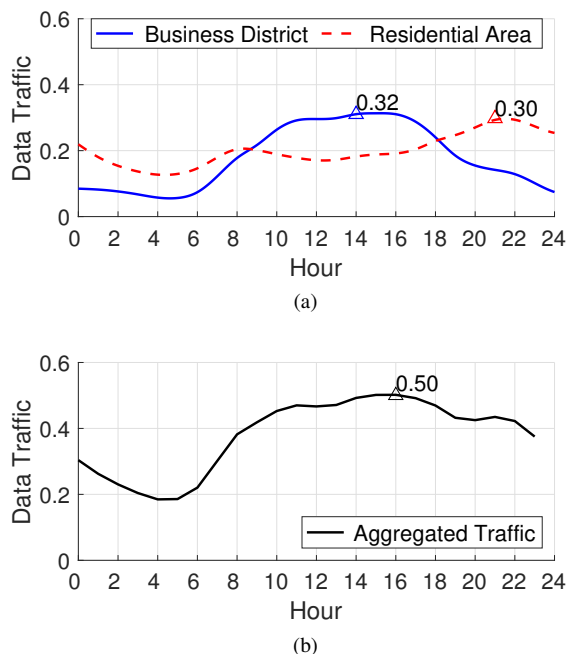

Fig. 1. (a) Data traffic patterns in different areas of Milan during a typical weekday. The blue solid line denotes the data traffic in a business district (Centro Sacra Mall), and the red dashed line corresponds to the data traffic in a residential area (Quintosolo district). (b) The aggregated data traffic pattern of the two areas. Triangles denote the peak traffic hour and intensity.

different areas and different periods of time are not evenly distributed. For example, as shown in Fig.1a, the data traffic in the business district (denoted in blue solid line) observes several peaks during working hours, while the traffic in the residential area (denoted red dashed line) is relatively higher in morning and evening hours than in working hours. Such a *spatio-temporal non-uniform property* of traffic demand poses great challenges in optimizing deployment cost and energy consumption. On one hand, the data processing capacity of each base station needs to cover its peak traffic intensity, leading to high deployment cost. On the other hand, the data processing capacity in individual base station is wasted during off-peak hours, resulting in low energy efficiency.

Fortunately, with the rapid evolution of mobile network architecture, the emergence of *Cloud Radio Access Network (Cloud-RAN)* [6] has presented new opportunities to answer the above mentioned challenges. In Cloud-RAN, the *data processing units* (i.e., Baseband Units, BBUs) are removed from base stations, and are centralized in a virtual *BBU pool* and shared among several base stations [5]. In this way, the data traffic from these base stations can be aggregated and processed in the BBU pool. By clustering base stations with *complementary traffic patterns* to a BBU pool, the total number of BBUs needed to be deployed can be reduced, and the utilization rate of the BBU pools can be increased [7]. For example, in Fig.1, if we cluster the base stations in the business district and in the residential area to a BBU pool, the aggregated traffic pattern will become relatively stable (Fig.1b), leading to potentially higher utilization rate of the pool. Moreover, the maximum BBU pool capacity can be reduced from the sum of the peak capacities of the two base stations (0.32+0.30) to a lower value (0.50), and thus reducing the deployment cost of BBUs.

Therefore, in order to unlock the power of the Cloud-RAN architecture, one of the key problem is to design an optimal *clustering scheme* between base stations and BBU pools [7], [8], so as to minimize the deployment cost and energy consumption of the entire network infrastructure. More specifically, given a set of base stations in a city, our objective is to cluster base stations with complementary traffic patterns to the same BBU pools. To achieve these goals, we need to address the following issues:

1) **How to characterize the data traffic pattern of a base station?** The data traffic generated in each base station is highly dynamic in different time of day, and in different day of week. For example, the traffic volume in business districts may be higher during the working hours than at midnight, and be lower in weekends than in weekdays. Therefore, we need to build appropriate data traffic profiles to characterize the data traffic patterns of base stations under different *temporal contexts*.

2) **How to measure the complementarity among base stations?** In order to reduce deployment cost, the peak traffic volumes of the base stations clustered to the same BBU pool should be scattered in different temporal contexts, so that the capacity of the BBU pool can be shared among these base stations. Moreover, the aggregated traffic volume in different temporal contexts should be close to the capacity of the BBU pool to increase the pool utilization. Therefore, we need to take into account both aspects, i.e., *the peak time distribution and the capacity utilization*, to design an effective metric to measure the *complementarity* of base stations.

3) **How to cluster the complementary base stations to BBU pools?** Given the traffic patterns and the complementarity measurements of base stations, there are potentially large numbers of schemes to cluster base stations to BBU pools. However, in order to reduce
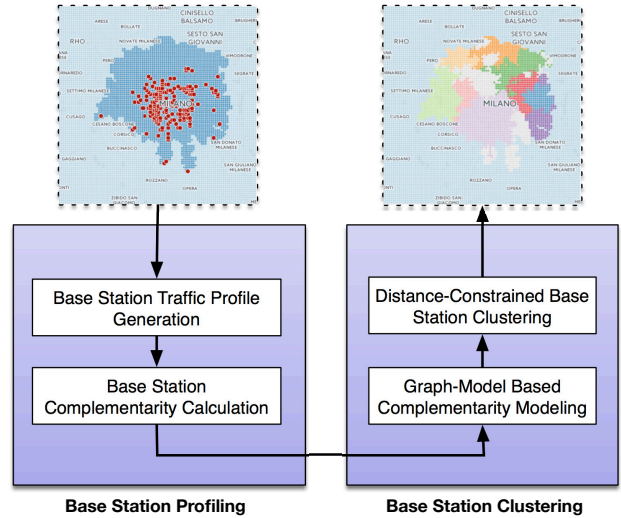

Fig. 2. Framework overview.

the signal propagation delay and the cost of optical fiber between base stations and BBU pools [5], the *distance* between a BBU pool and its affiliated base stations should be constrained within a reasonable range. Therefore, we need to design an effective algorithm to find the optimal base station clustering scheme under *distance constraint*.

With the above-mentioned research objectives and issues, the **main contributions** of this paper are:

- We propose a two-phase framework to find optimal base station clustering schemes for cost-effective and energy-efficient Cloud-RAN. In the first phase, we design a *temporal-context-based data traffic profile* to characterize the data traffic pattern of each base station. Based on the data traffic profiles, we propose an *entropy-based metric* to characterize the *complementarity* among base stations, taking into consideration both the peak time distribution and the capacity utilization of the aggregated traffic of a cluster of base stations. In the second phase, we first build a *weighted graph model* to represent the complementarity of base stations as link weight, and then propose a *distance-constrained clustering algorithm* to find optimal clustering schemes between base stations and BBU pools.

- We evaluate the performance of our framework using two months of real-world mobile network data from Milan, Italy. Results show that our framework effectively reduces 12.88% of deployment cost and 9.45% of energy consumption compared with non-Cloud-RAN architectures, and outperforms the other clustering method.

II. FRAMEWORK OVERVIEW

We propose a two-phase framework to cluster base stations to a number of BBU pools, so that the deployment cost and the energy consumption of the entire network infrastructure can be reduced to a maximal extent. As shown in Fig.2, in the

base station profiling phase, given a set of base stations, we first generate a traffic profile for each base station according to its historical data traffic record, and then calculate the complementarity of base stations using a proposed entropy-based metric. In the base station clustering phase, we first build a graph model to represent the complementarity connections among base stations, and then propose a distance-constrained clustering algorithm to cluster base stations with complementary traffic patterns, and aggregate their traffic to BBU pools.

III. BASE STATION PROFILING

In order to cluster base stations with complementary traffic patterns to the same BBU pool, we need to characterize the traffic pattern of each base station, and then measure the complementarity of base stations. Since the base station data traffic is varying significantly across different hours and different days, we design a temporal-context-based traffic profile for each base station using their historical data traffic record. In order to characterize the complementarity of base stations, we focus on the peak distribution and capacity utilization of their aggregated traffic, and design an entropy-based metric to characterize the complementarity of base stations.

A. Base Station Traffic Profile Generation

In the mobile network architecture, a set of base stations are deployed over geographical areas called cells [9]. This base station provides the cell with the network coverage which can be used for transmission of voice and data. With the emergence of smartphones and tablets, the data traffic generated from users connected to the base stations is increasing rapidly. In order to benchmark the data processing capacity of base stations, many operators have collected large scales of base station traffic statistics data and make them publicly available [1]. In this paper, we exploit the dataset released by Telecom Italia for the Big Data Challenge initiative [10], which contains two months of network traffic data from 11/01/2013 to 12/31/2013 in Milan, Italy. We also collect the locations of active base stations in Milan from CellMapper.com[1], and derive the traffic intensity of each base station during the two months on an hourly basis. The traffic data pre-processing steps will be detailed in the evaluation section.

Based on the historical traffic data, we observe that the traffic pattern of a base station is highly dynamic under different temporal contexts. For example, Fig.3 shows the traffic intensity of a base station located in a business district in Milan from 11/04/2013 to 11/10/2013. We can see clear weekday-weekend patterns as a result of regular working schedules. We also observe significant traffic peaks during the working hours of weekdays. Based upon the observations, we propose to characterize the traffic pattern of each base station using a temporal-context-based profile. More specifically, given a base station s_i and its historical traffic intensity vector measured in hours, we aggregate and average the traffic from Monday to
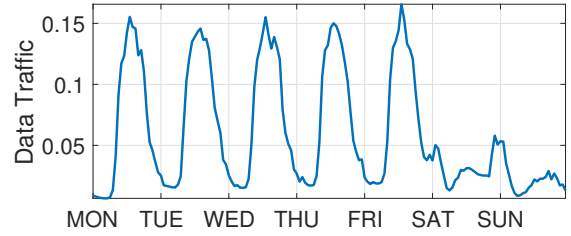
Fig. 3. An illustrative example of the traffic pattern of a base station location in a business district (Wagner district, Milan) from 11/04/2013 to 11/10/2013.

Friday in each week to build a *typical weekday traffic profile*, i.e.,

$$\mathbf{f}_w(s_i) = [u_1, u_2, \ldots, u_{24}] \tag{1}$$

Similarly, we build a *typical weekend traffic profile* by aggregating and averaging the traffic in Saturday and Sunday of each week, i.e.,

$$\mathbf{f}_n(s_i) = [v_1, v_2, \ldots, v_{24}] \tag{2}$$

Finally, we concatenate the weekday and weekend traffic profiles to obtain the temporal-context-based traffic profile:

$$\mathbf{f}(s_i) = [\mathbf{f}_w(s_i), \mathbf{f}_n(s_i)] \tag{3}$$

B. Base Station Complementarity Calculation

We consider the following two aspects to design an effective complementarity metric of base stations.

1) Peak Distribution: To reduce deployment cost, the peak traffic intensity of a set of base stations clustered to the same BBU pool should be scattered in different temporal contexts, so that the capacity of the BBU pool can be shared among these base stations. To this end, we design an entropy-based metric to measure the peak distribution of a set of base station. Specifically, given a set of clustered base stations $S = \{s_1, \ldots, s_N\}$, we first find the peak hours in their traffic profiles, respectively, i.e.,

$$T(s_i) = \{t_{i_1}, t_{i_2}, \ldots, t_{i_m}\}, \quad 1 \leq i_m \leq 24 \tag{4}$$

Then, we calculate the *Shannon entropy* [11] of the peak hours of the set of clustered base stations $T(S) = \cup T(s_i)$ as follows:

$$H(S) = -\sum_{k=1}^{K} p_k \log p_k \tag{5}$$

where $K = \sum i_m$ corresponds to the total number of peaks, and p_k is the probability of observing the corresponding peak hour in the set $T(S)$. A larger entropy measurement of a base station cluster indicates that the base stations are more complementary to each other w.r.t. traffic patterns.

2) Capacity Utilization: In order to increase the utilization rate of a BBU pool, the aggregated base station traffic in different temporal contexts should be close to the capacity of the pool. To this end, we design the following metric to quantitatively measure the capacity utilization of a BBU pool:

$$U(S) = \frac{\text{mean } \mathbf{f}(S)}{\text{max } \mathbf{f}(S)} \qquad (6)$$

where $\mathbf{f}(S) = \sum_{i=1}^{N} \mathbf{f}(s_i)$ denotes the aggregated traffic profile of the base stations cluster. We assume the pool capacity to be the maximum of the aggregated traffic intensity, and derive the average capacity utilization rate correspondingly.

Finally, we calculate the complementarity of the base station cluster as follows:

$$M(S) = U(S) * H(S) = -\frac{\text{mean } \mathbf{f}(S)}{\text{max } \mathbf{f}(S)} \sum_{k=1}^{K} p_k \log p_k \qquad (7)$$

IV. BASE STATION CLUSTERING

In this phase, our objective is to cluster base stations with complementary traffic patterns to a set of BBU pools. One intuitive method is to exhaustively search for base stations with complementary traffic patterns and iteratively cluster them. However, since there are tremendous number of clustering schemes, such a method can be computationally impossible as the network scale increases. Moreover, the distance between base stations and BBU pools should also be constrained within a range, since the propagation delay between base stations and BBU pools may exceed quality-of-service requirements as distance increases, and the cost of laying optical fiber for communication between base stations and BBU pools should be limited within budget.

Therefore, we propose a graph-model-based, distance-constrained algorithm to effectively cluster neighboring base stations to BBU pools. First, we construct a weighted graph model to represent the relationship of base stations, exploiting graph links to express the base station distance constraints, and link weights to characterize the base station complementarity measurement. Then, we propose a community-detection-based algorithm to iteratively cluster base stations into clusters, so that the complementarity of base stations is maximal within each cluster and minimal across different clusters.

A. Graph-Model Based Complementarity Modeling

We model the complementarity among base stations as an undirected, weighted graph $G = (V,E)$, where $V = \{s_1, \ldots, s_N\}$ denotes the set of N base stations, and E denotes the set of links between two base stations.

We then define the adjacency matrix A of graph G, which is an $N \times N$ symmetric matrix with entries $a_{i,j} = 1$ when there is a link between base station s_i and base station s_j, and $a_{i,j} = 0$ otherwise $(i, j = 1, \ldots, N)$. We use the geographic distance of two base stations to determine whether they are adjacent or not. More specifically, for base station s_i and base station s_j, we define:

$$a_{i,j} = \begin{cases} 1, & \text{if } dist(s_i, s_j) \leq \tau \\ 0, & \text{otherwise} \end{cases} \qquad (8)$$

where $dist(s_i, s_j)$ is the geographic distance between the two base stations, and τ is a *neighborhood threshold* controlling the geographic distance of neighboring base stations.

Given two neighboring base stations, we use their complementarity measurement to determine their link weight, i.e.,

$$w(s_i, s_j) = M(\{s_i, s_j\}) * a_{i,j} \qquad (9)$$

We consider the case of normalized symmetric positive weights ($w(s_i, s_j) \in [0,1]$) with no loops ($w(s_i, s_i) = 0$). We note that $w(s_i, s_j) = 0$ when there is no link between s_i and s_j ($a_{i,j} = 0$).

B. Distance-Constrained Base Station Clustering

In this step, we need to cluster base stations to BBU pools, so that each cluster consists of neighboring base stations with complementary traffic patterns. As the link weight of graph G encodes the complementary of base stations, we need to cluster base stations with high link weights together, which can be identified as a community detection problem [12].

Problem: Given graph $G = (V,E)$, we first define a set of clusters $\mathbb{P} = \{C_1, \ldots, C_K\}$, where

$$\cup_{\forall C_k \in \mathbb{P}} = V \quad \text{and} \quad \cap_{\forall C_k \in \mathbb{P}} = \emptyset \qquad (10)$$

Then, given a base station v, we define the *connectivity* of v to a cluster C as the sum of link weights between v and the base stations in the cluster C:

$$con(v,C) = \sum_{v' \in C} w_{v,v'} \qquad (11)$$

Finally, we define the *adjacent clusters* $\mathbb{C}(v)$ of v as

$$\mathbb{C}(v) = \{C | con(v,C) > 0, C \in \mathbb{P}\} \qquad (12)$$

With the above definition, our objective is to find an optimal set of clusters $\mathbb{P}$, so that the internal connectivity within a cluster is higher than the inter-cluster connectivity, i.e.,

$$\forall v \in C_k, \ con(v,C_k) \geq max\{con(v,C_l), C_l \in \mathbb{P}\} \qquad (13)$$

We also need to bound the distance span of a cluster within the neighborhood threshold, i.e.,

$$\forall v, v' \in C_k, \ dist(v,v') \leq \tau \qquad (14)$$

Solution: Based on [13], [14], we propose a *Distance-Constrained Complementarity-Aware (DCCA)* algorithm to cluster base stations. The basic idea of DCCA is iteratively assigning base stations to the adjacent clusters, where the *gain* of assigning base station v to cluster C is iteratively evaluated by a *value function* as follows:

$$value(v,C) = con(v,C) \times log\left(\frac{\tau}{max\{dist(v,v')\}}\right) \qquad (15)$$

The DCCA algorithm greedily assigns the base stations to the adjacent cluster with highest value[2] until none of the base stations are moved among clusters [14]. As the convergence of such a greedy approach is difficult to prove, we set a maximum iteration number *max_iter* to ensure the algorithm will stop.

---

[2]If two clusters yield the same value, we randomly choose one.

TABLE I
DATASET DESCRIPTION

Dataset	Item	Value
Network Traffic	# Grids	10,000
	Grid size	55,225 m^2
	Average traffic intensity	0.19
Base Station	# Base stations	182
	Average coverage	885,724 m^2
Data collection period		11/01/2013–12/31/2013

Algorithm: The DCCA algorithm is initialized by assigning each base station in the graph to a unique cluster label. In each iteration, we randomly populate a list of base station *L*, and traverse the list to update the cluster label of each base station. The label update process is as follows. First, we remove the base station from its current cluster, and find the set of adjacent clusters to the current base station. Then, we compute the value function for all the adjacent clusters, and assign the base station to the cluster with the highest value. We mark the the base station as *moved* among clusters if its new cluster label is different from the old one. After we finish iterating over the base station list, we decide whether to perform another iteration or finish the algorithm based on the following stop criteria: (1) the user specified maximum iteration number *max_iter* is reached, or (2) none of the base station are moved among clusters.

V. EVALUATION

In this section, based on a real-world mobile network traffic dataset, we evaluate the performance of our framework by assessing its ability to reduce deployment cost and energy consumption. We first describe the experiment settings, and then present the evaluation results. We also conduct a parameter study to help select optimal parameters for our framework.

A. Experiment Settings

Datasets: The Telecom Italia Big Data Challenge dataset [10] contains two months of network traffic data from 11/01/2013 to 12/31/2013 in Milan, Italy. The area of Milan is partitioned into 100×100 grids with grid size of about 235×235 square meters. In each grid, a normalized traffic intensity is recorded on an hourly basis. We also compile a base station dataset from CellMapper.com, which consists of the locations and coverage areas of active base stations observed in the two months. The details of these two datasets are listed in TABLE I.

Evaluation Plan: Based on the above-mentioned datasets, we map the grids to the coverage areas of base stations, and aggregate the traffic data to the corresponding base stations. We then select 8 weeks of data from 11/04/2013 to 12/29/2013 to generate 182 traffic profile for the base stations. Finally, we construct a graph of 182 base stations and their corresponding link structure based on the complementarity metrics, and perform the DCCA algorithm to cluster base stations to a set of BBU pools.
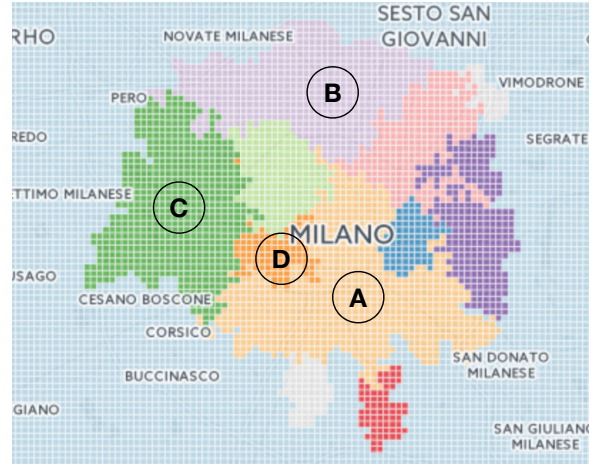

Fig. 4. Base station clustering results in Milan using the proposed method. Each colored area denotes a cluster and its corresponding coverage range.

Evaluation Metrics: We focus on the reduction of deployment cost and capacity utilization to evaluate the effectiveness of the proposed framework. For the deployment cost metric, we compare the overall deployment capacity of all the clusters (i.e., BBU pools) with the overall deployment capacity of all the base stations without clustering, and derive the following metric to measure the percentage of deployment cost reduction rate of a proposed clustering scheme:

$$Cost(\mathbb{P}) = \frac{\sum_{C_k} \max \mathbf{f}(S_i)}{\sum_{s_i} \max \mathbf{f}(s_i)} \quad (16)$$

where $\mathbb{P} = \{C_k\}$ is the clustering scheme. For the capacity utilization metric, we compare the mean capacity utilization of all the clusters with the mean capacity utilization of all the base stations, i.e.,

$$Util(\mathbb{P}) = \frac{\text{mean}_{C_k} U(C_k)}{\text{mean}_{s_i} U(s_i)} \quad (17)$$

Baseline Method: We compare our method with the *distance-constrained (DC)* baseline method, which simply clusters neighboring base stations to clusters without considering their traffic complementarity. The graph modeling and clustering algorithm are similar to the propose method.

B. Evaluation Results

Fig.4 shows the base station clustering results in Milan using the optimal parameter settings (which will be discussed later). In general, we obtain 12 clusters to setup BBU pools. We can see that many clusters (e.g., Cluster *A*, *B*, and *C*) are composed of an urban part and a suburban part, indicating that the traffic patterns in these areas are potentially complementary. We also note that cluster *D* is concentrated in a relatively small area, indicating the diverse traffic patterns within this area. The reason is probably due to the hybrid functions of this area, which consists of a large residential district (the *Washington neighborhood*), a national theater (*Teatro Nazionale CheBanca*), and a transportation hub consisting of 6 metro stations (*e.g., Wagner Station*).

TABLE II
EVALUATION RESULTS

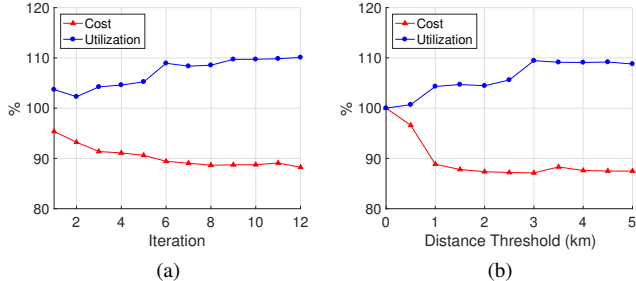|	Cost	Utilization
Proposed (**DCCA**)	87.12%	109.45%
Baseline (**DC**)	96.68%	102.95%

Fig. 5. Parameter studies. (a) Maximum iteration number parameter impact. (b) Distance threshold parameter impact.

TABLE II shows the evaluation results of the proposed method and the baseline method under the optimal parameter settings. We can see that both methods reduce the deployment cost and increase the capacity utilization of the network, indicating the advantages of the Cloud-RAN architecture. The proposed *DCCA* method saves 12.88% of deployment cost while the baseline *DC* method only achieves a 3.32% cost reduction, indicating the importance of involving base station traffic complementarity property in the clustering process. Similarly, the increase rate of capacity utilization in the two methods are 9.45% and 2.95%, respectively. The relatively low value of the baseline method implies that simply clustering neighboring base stations to BBU pools may result in sub-optimal solutions w.r.t. energy consumption.

C. Parameter Impact Study

We also study the impacts of the two important parameters, i.e., the maximum iteration number *max_iter* and the distance threshold τ. In Fig.5a, we can see that the deployment cost and capacity utilization ratio converge to stable values respectively as the *max_iter* values increases. In Fig.5b, we observe that as the cluster size increases, the deployment cost continues to decrease and the capacity utilization continues to increase. However, due to propagation delay and optical fiber costs, a very large cluster is not realistic in deployment. Based on the above study, we finally choose *max_iter* = 12 and $\tau = 3km$.

VI. RELATED WORK

A. Cloud Radio Access Network

Cloud Radio Access Network (Cloud-RAN) is a novel mobile network architecture to address the challenges faced by operator while trying to meet the fast-growing traffic demand. The details of the Cloud-RAN concept can be found [6]. The basic idea of Cloud-RAN is to pool the data processing units from multiple base stations into centralized BBU pools, so that

the pool capacity can be shared among these base stations. Since fewer BBUs are needed and higher BBU utilization can be achieved, the Cloud-RAN architecture can reduce the network deployment cost and energy consumption [5]. Therefore, Cloud-RAN is seen as a typical realization of the fifth generation (5G) network in the year 2020 horizon [15].

One of the key problem in the Cloud-RAN architecture is to design optimal base station clustering scheme to build BBU pools. An optimal scheme should facilitate the pool capacity utilization, reduce the BBU deployment cost, and also prevent the propagation delay between base stations and BBU pools [5]. To this end, Bhaumik et al. [7] proposed CloudIQ, a framework for partitioning a set of base stations into groups and process the signals in a shared data center. However, the one data center topology may result in potential delay between distant base stations and the data center [5]. Lee et al. [16] proposed a base station cooperation scheme with dynamic clustering in Cloud-RAN, however the objective of the cooperation is to derive the signal-to-interference for base station evaluation. One of the very relevant ideas to our work is illustrated in [1], which explores various means of integrating big data analytics with network optimization in 5G, especially by exploiting historical data to optimize resource allocation in centralized baseband units in Cloud-RAN.

B. Mobile Data Analytics

With the emergence of ubiquitous sensing and computing diagrams [17], a massive number of mobile data can now be collected either by mobile crowdsensing paradigms [18]–[20] or from operators' infrastructures. These heterogeneous mobile big data are being extensively analyzed in the literature to retrieve interesting and informative information [21]–[24]. For example, Barlacchi et al. [10] released a large-scale Call Detail Records (CDR) dataset from Telecom Italia, containing two-months of calls, SMSs and network traffic data from the city of Milan and the province of Trentino, Italy. Based on the dataset, Furno et al. [25] proposed a data analytics framework to builds profiles of the city-wide traffic demand, and identifies unusual situations in network usages, aiming at facilitating the design and implementation of cellular cognitive networking. Cici et al. [26] studied the decomposition of cell phone activity series, and connect the decomposed series to socio-economic activities such as regular working patterns and opportunistic social events [27].

VII. CONCLUSION

In this work, we identify two of the most important goals in Cloud-RAN optimization, i.e., cost-effectiveness and energy-efficiency, and propose a base station clustering framework to achieve these goals. More specially, we first exploit the historical data of base stations to generate the corresponding traffic profiles, and propose an entropy-based metric to characterize the complementary of base stations. Upon the basis, we build a graph model to represent the complementarity of base stations, and propose a distance-constrained clustering algorithm to find optimal base station clustering scheme. Evaluation results

using two months of real-world data show that our framework effectively reduces 12.88% of deployment cost and 9.45% of energy consumption compared with traditional architectures. Our work can facilitate green and soft Cloud-RAN architecture in the next generation 5G network era.

In the future, we plan to improve this work in the following three directions. First, we plan to involve more contextual factors in building the base station traffic profiles. Second, we will explore the methods for dynamic capacity adjustment in BBU pools. Third, we plan to evaluate our framework in broader areas, such as Trentino Italy, and to study the traffic complementarity property under different contextual settings.

REFERENCES

[1] K. Zheng, Z. Yang, K. Zhang, P. Chatzimisios, K. Yang, and W. Xiang, "Big data-driven optimization for mobile networks toward 5G," *IEEE Network*, vol. 30, no. 1, pp. 44–51, Jan. 2016.

[2] Cisco, "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016–2021 White Paper," Cisco, San Jose, CA, USA, Tech. Rep., 2016.

[3] J. Research, "Mobile Operator Business Models: Challenges, Opportunities & Adaptive Strategies 2011-2016," Juniper Research, New York, Tech. Rep., 2011.

[4] G. Y. Li, Z. Xu, C. Xiong, C. Yang, S. Zhang, Y. Chen, and S. Xu, "Energy-efficient wireless communications: Tutorial, survey, and open issues," *IEEE Wireless Communications*, vol. 18, no. 6, pp. 28–35, Dec. 2011.

[5] A. Checko, H. L. Christiansen, Y. Yan, L. Scolari, G. Kardaras, M. S. Berger, and L. Dittmann, "Cloud RAN for Mobile Networks - A Technology Overview," *IEEE Communications Surveys Tutorials*, vol. 17, no. 1, pp. 405–426, Firstquarter 2015.

[6] C. M. R. Institute, "C-RAN: The road toward Green RAN," China Mobile Research Institute, Beijing, China, Tech. Rep., 2011.

[7] S. Bhaumik, S. P. Chandrabose, M. K. Jataprolu, G. Kumar, A. Muralidhar, P. Polakos, V. Srinivasan, and T. Woo, "CloudIQ: A Framework for Processing Base Stations in a Data Center," in *Proceedings of the 18th Annual International Conference on Mobile Computing and Networking*, ser. Mobicom '12, pp. 125–136.

[8] Y. S. Chen, W. L. Chiang, and M. C. Shih, "A Dynamic BBU #x2013;RRH Mapping Scheme Using Borrow-and-Lend Approach in Cloud Radio Access Networks," *IEEE Systems Journal*, vol. PP, no. 99, pp. 1–12, 2017.

[9] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, May 2005, google-Books-ID: GdsLAQAAQBAJ.

[10] G. Barlacchi, M. De Nadai, R. Larcher, A. Casella, C. Chitic, G. Torrisi, F. Antonelli, A. Vespignani, A. Pentland, and B. Lepri, "A multi-source dataset of urban life in the city of Milan and the Province of Trentino," *Scientific data*, vol. 2, 2015.

[11] J. Lin, "Divergence measures based on the Shannon entropy," *IEEE Transactions on Information Theory*, vol. 37, no. 1, pp. 145–151, Jan. 1991.

[12] M. E. J. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Physical Review E*, vol. 69, no. 2, p. 026113, 2004.

[13] L. Chen, D. Zhang, L. Wang, D. Yang, X. Ma, S. Li, Z. Wu, G. Pan, T.-M.-T. Nguyen, and J. Jakubowicz, "Dynamic Cluster-based Over-demand Prediction in Bike Sharing Systems," in *Proc. ACM UbiComp*, ser. UbiComp '16, pp. 841–852.

[14] U. N. Raghavan, R. Albert, and S. Kumara, "Near linear time algorithm to detect community structures in large-scale networks," *Physical Review E*, vol. 76, no. 3, p. 036106, 2007.

[15] C. L. I, C. Rowell, S. Han, Z. Xu, G. Li, and Z. Pan, "Toward green and soft: A 5G perspective," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 66–73, Feb. 2014.

[16] N. Lee, R. W. Heath, D. Morales-Jimenez, and A. Lozano, "Base station cooperation with dynamic clustering in super-dense cloud-RAN," in *2013 IEEE Globecom Workshops (GC Wkshps)*, pp. 784–788.

[17] D. Zhang, B. Guo, and Z. Yu, "The emergence of social and community intelligence," *Computer*, vol. 44, no. 7, pp. 21–28, 2011.

[18] L. Wang, D. Zhang, Y. Wang, C. Chen, X. Han, and A. M'hamed, "Sparse mobile crowdsensing: Challenges and opportunities," *IEEE Communications Magazine*, vol. 54, no. 7, pp. 161–167, Jul. 2016.

[19] B. Guo, Z. Wang, Z. Yu, Y. Wang, N. Y. Yen, R. Huang, and X. Zhou, "Mobile Crowd Sensing and Computing: The Review of an Emerging Human-Powered Sensing Paradigm," *ACM Computer Survey*, vol. 48, no. 1, pp. 1–31, 2015.

[20] J. Wang, Y. Wang, D. Zhang, L. Wang, C. Chen, J. W. Lee, and Y. He, "Real-time and generic queue time estimation based on mobile crowdsensing," *Frontiers of Computer Science*, vol. 11, no. 1, pp. 49–60, Feb. 2017.

[21] C. Chen, D. Zhang, N. Li, and Z.-H. Zhou, "B-Planner: Planning Bidirectional Night Bus Routes Using Large-scale Taxi GPS Traces," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 4, pp. 1451–1465, 2014.

[22] L. Chen, D. Zhang, X. Ma, L. Wang, S. Li, Z. Wu, and G. Pan, "Container Port Performance Measurement and Comparison Leveraging Ship GPS Traces and Maritime Open Data," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 5, pp. 1227–1242, May 2016.

[23] D. Yang, D. Zhang, L. Chen, and B. Qu, "NationTelescope: Monitoring and visualizing large-scale collective behavior in LBSNs," *Journal of Network and Computer Applications*, vol. 55, pp. 170–180, 2015.

[24] M. Tan, B. Wang, Z. Wu, J. Wang, and G. Pan, "Weakly Supervised Metric Learning for Traffic Sign Recognition in a LIDAR-Equipped Vehicle," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 5, pp. 1415–1427, May 2016.

[25] A. Furno, D. Naboulsi, R. Stanica, and M. Fiore, "Mobile Demand Profiling for Cellular Cognitive Networking," *IEEE Transactions on Mobile Computing*, vol. PP, no. 99, pp. 1–1, 2016.

[26] B. Cici, M. Gjoka, A. Markopoulou, and C. T. Butts, "On the Decomposition of Cell Phone Activity Patterns and Their Connection with Urban Ecology," in *Proceedings of the 16th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, ser. MobiHoc '15, pp. 317–326.

[27] L. Chen, J. Jakubowicz, D. Yang, D. Zhang, and G. Pan, "Fine-Grained Urban Event Detection and Characterization Based on Tensor Cofactorization," *IEEE Transactions on Human-Machine Systems*, vol. PP, no. 99, pp. 1–12, 2016.