IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 23 October 2024, accepted 13 November 2024, date of publication 25 November 2024, date of current version 4 December 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3506434

RESEARCH ARTICLE

Clustering-Driven Optimization of RRH-BBU Assignment for Green Communication Networks With Big Data Analytics

ASMAA IBRAHIM[1,2], AHMED ELSHEIKH[3], BASSEM MOKHTAR[2,4], (Senior Member, IEEE), AND JOSEP PRAT[1]

[1]Department of Signal Theory and Communications, Universitat Politècnica de Catalunya, 08034 Barcelona, Spain
[2]Computer and Networks Engineering Department, College of Information Technology, United Arab Emirates University, United Arab Emirates
[3]Department of Mathematics and Engineering Physics, Faculty of Engineering, Cairo University, Giza 12613, Egypt
[4]Department of Electrical Engineering, Faculty of Engineering, Alexandria University, Alexandria 21544, Egypt

Corresponding author: Asmaa Ibrahim (asmaa.ibrahim@upc.edu)

This work was supported in part by the Grant CONFLOC funded by MCIN/AEI/10.13039/501100011033/FEDER, UE, under Grant PID2022-137754OB.

ABSTRACT The Centralized Radio Access Networks (CRAN) decentralizes data and control planes by separating the baseband unit (BBU) from the central office, enabling energy-efficient "green networks" through the shutdown of underutilized BBUs. Analyzing extensive Call Detail Records (CDR) as big data, collected by service providers, has gained traction for extracting network features and studying activities. Thus, big data analytics are deemed as potential techniques that various research proposed to analyze the CDR. This paper introduces an energy-efficient CRAN network architecture based on the CRAN framework, focused on an innovative remote radio head (RRH)-BBU assignment. The objective is twofold: minimizing power consumption by deactivating underutilized BBUs and reducing inter-BBU handover rates based on CDR insights. In literature, the problem of assigning RRH to BBU is described as hard nonlinear programming (NLP) problem (bin packing, mixed integer), different suboptimal algorithms have been proposed to offer suboptimal assignment. This study employs clustering techniques to divide the complex NLP problem into simpler optimization tasks, achieving optimal RRH-BBU assignments. The proposed algorithm's effectiveness was assessed using Milan city CDR as a case study, and its performance was validated against Milan's land use map. The results indicated a remarkable 28.8% reduction in power consumption, alongside improvements in inter-BBU handovers.

INDEX TERMS 5G communication systems, CRAN networks architecture, temporal databases, clustering optimization algorithms, spatio-temporal clustering, RRH, BBU, green communications.

I. INTRODUCTION

The 5G network architecture is envisioned to support diverse services with low latency and high reliability. This raises the concept of software-defined networks (SDN), which is based on decentralizing the data plane to support different qualities of service over a common shared infrastructure [1], [2]. A heterogeneous centralized radio access network (HCRAN) presents the fundamentals of decentralizing the data plane by providing a baseband unit (BBU) pool and remote radio heads (RRH) that cover different cell sizes [3], [4]. The core idea of CRAN is to separate the functions of a base station by splitting it into a Base Band Unit (BBU) and a Remote Radio Head (RRH). This approach centralizes the BBUs from multiple locations into a single site (pool), such as a cloud data center, leveraging cloud computing and virtualization. The BBUs handle baseband signal processing, while the RRHs are responsible for signal amplification and modulation. RRHs are installed with antennas at cell sites and connect to the BBU pool via a fronthaul link. By centralizing baseband processing in a virtualized BBU pool, the system can adjust to varying traffic loads and optimize resource use. Multiple

The associate editor coordinating the review of this manuscript and approving it for publication was Yeon-Ho Chung.

operators can share the same BBU pool by utilizing it as a cloud service. Efficiently managing resources in CRAN to meet user demand poses a significant challenge because of user mobility and the dynamic nature of the environment [5]. CRAN architecture poses different challenges as the functionality is split between the RRH and the BBU, the front-haul networks, and the RRH-to-BBU mapping. In CRAN architecture, efficiently allocating radio and computational resources has become a challenge. Radio resources refer to the radio frequency (RF) spectrum, while computational resources include processing power, data storage, and time. The management of these computational resources, centralized at the BBU pool, requires effective oversight of the BBU resources. In this work, we present the RRH-BBU association as a resource management technique that optimizes the resource allocation process by turning off underutilized units over time. The association problem is characterized as a clustering challenge that effectively enhances the allocation of computational resources. In the literature [6], The clustering problem is categorized based on its objective function into location-aware, load-aware, interference-aware, QoS-aware, and throughput-aware RRH clustering. Various studies have tackled each clustering approach. For example, in [7], the author proposes location-aware clustering, which groups RRHs in close geographical proximity to minimize handovers between them. Other studies focus on load-aware clustering; as in [8] dynamic RRH assignment algorithm is designed to offload one or more RRHs from an overloaded BBU to a less loaded one, framing the RRH clustering problem as a generalization of the classical bin packing optimization problem. In this context, BBUs are treated as bins, while RRHs represent the item sets to be packed. In another approach proposed in [9], interference-aware clustering is discussed, where the author first determines RRH reassociation with a BBU when minimum throughput requirements per user are not met, and then seeks to minimize the number of active BBUs by relocating RRHs to those with moderate throughput per user. Lastly, QoS-aware clustering is explored in [10], where authors propose a self-optimizing algorithm for dynamic RRH clustering in CRAN based on load predictions. In the self-optimizing algorithm, traffic load in each cell is forecasted using a Markov model, and the optimal BBU-RRH mapping is determined through genetic algorithms, aiming to maximize QoS. The RRH-to-BBU assignment is proposed as an optimization problem, as it significantly influences the CRAN network's efficiency. Many researchers have proposed ways of association based on maximizing certain objective functions. They used a single objective clustering method, while we implemented a joint load-aware and location-aware RRH clustering approach to tackle the RRH-BBU association issue. We then compared our method with a similar technique described in the second paragraph.

In this paper, we build on the work presented in [11] and propose a joint location-aware and load-aware for time RRH clustering approach to address the RRH-BBU association problem. The proposed clustering approach is a spatiotemporal clustering algorithm that simultaneously does load-aware and location-aware clustering. While [12] introduces a joint optimization problem aimed at further reducing power consumption, it also presents an online optimization framework as a game problem with potential functions among RRHs for associating with active BBUs, which are centrally managed by a central unit. However, their proposed algorithm experiences signaling overload during the re-association process. Alternatively, we propose an offline computational resource management algorithm that employs location-aware and load-aware RRH clustering. This clustering approach is unsupervised learning to predict the switching load of the CRAN networks based on the analysis of Call Detail Records. We are optimizing the association between the computational resources of the BBU and the RRH. While the online process offers certain advantages over the offline method, it comes with increased networks and processing loads. Our model is based on the static locations of the installed RRHs for spatial clustering, as well as the statistical behavior of network load over time. We will discuss network load updates later in the discussion, categorizing them into planned and unplanned events.

We present the problem of RRH-BBU association with the aim of minimizing network power consumption, as a means for creating green communication networks in 5G architecture. Power-saving constraints are reached by switching off underutilized RRHs and BBUs. We introduce an RRH-to-BBU assignment that reduces power and handover rates between the BBUs. The problem is formulated as a hard NLP optimization problem. The proposed algorithm first deploys a time series-based clustering technique on the CDR data to study network utilization over time. Secondly, it provides second-level spatial clustering that divides each temporal cluster into geographical zones to enhance inter-BBU handover rates. Finally, it computes the required BBUs for each temporal-spatial cluster and efficiently optimizes the number of active BBUs using a bin-packing optimization algorithm. The task of assigning RRH to each BBU is framed as a Bin Packing Problem (BPP), which can be addressed using the Knapsack algorithm. In this context, the RRH are treated as objects and the capacity of the BBU as the Knapsack. By clustering the BBU pool, we can transform this single problem into several joint optimization problems. The algorithm allocates one item (object) to each bin (knapsack) to ensure that the total weights of all objects remain within the knapsack's capacity while minimizing the number of knapsacks (BBUs) used.

The proposed algorithm simplifies the problem by dividing it into clusters of joint RRH and BBU. Parallel processing, enabled by dividing the resource allocation for each cluster, divides the optimization problem into multiple simplified optimization problems. The problem is divided into two levels: in the first level, the RRH is clustered based on real collected Call Detail Records (CDR) of the real networks. In the second level, based on the clustered RRH, the BBUs are assigned to each clustered group of RRH. The effectiveness of the proposed algorithm was evaluated using the CDR from Milan as a case study, with its performance validated against the city's land use map. The results showed a significant 28.8% reduction in power consumption, as well as enhancements in inter-BBU handovers. The rest of the paper is organized as follows: The literature review and the contributions are highlighted in section II. Section III proposes the system model and the problem formulation. The simulation results and analysis of Milan City [13] are proposed in Section IV. The paper is concluded in Section V.

II. LITERATURE REVIEW AND CONTRIBUTION

A. RELATED WORK

In this section, we discuss previous work that addresses the RRH-BBU association problem. The problem is divided into two parts. First, we examine research proposing solutions to the BBU-RRH assignment problem, followed by research focused on clustering the BBUs and RRHs.

1) RRH TO BBU ASSIGNMENT

The problem of RRH-BBU mapping has been the focus of much recent research, aiming to maximize objective functions such as energy saving and fulfill certain constraints like quality of service (QoS). In [15], the author proposed a CRAN architecture with separate computational resources from the RRH, as the RRH is deployed as small cells with only transmission functionalities. The allocation problem is described as a two-level scheduling algorithm. In the first level, resources are assigned from cells to each user while satisfying QoS and service continuity. In the second level, the problem of assigning resources from each BBU to RRH is proposed as an optimization problem, maintaining power consumption and minimizing computing resources. The mapping of each BBU to RRH depends on assigning a physical machine (set of BBUs) to all RRH in its coverage area. In [16], the author proposed a two-stage dynamic resource allocation for CRAN, which assigns user equipment (UE) to each RRH with power transmission constraints jointly with RRH-BBU real-time association. The problem of RRH-UE assignment is described as a Mixed Integer Non-Linear Program (MINLP) with signal-to-interference noise ratio constraints. Based on the UE-RRH assignment, the optimal number of BBUs is computed in the second phase, and the RRH-BBU association is described as a Multiple Knapsack Problem (MKP) solved by linear solvers. In the work of [11], the author proposed a BBU assignment that optimizes the efficiency of the BBU pool, with the problem described as bin packing. The assignment of RRH to BBU considers the resource requirements and communication between the RRH by representing the networks as a weighted graph. The algorithm improves power consumption by up to 20% and reduces handovers by 30% by decreasing communication overhead.

In [22], the author proposed a spatial-based clustering technique by grouping neighbor RRHs. The proposed model minimizes the number of active BBUs and reduces the number of handovers. The problem is formulated as bin packing with an NP-hard optimal solution, with the author proposing a heuristic algorithm to drive the optimal solution in large networks. It is illustrated that during periods of high networks load (15 Mbps), spatial clustering is less effective and results in reduced networks QoS. In [23], the author proposed a dynamic BBU virtualization scheme that packs the dynamics of traffic load as bins with finite computing resources in the BBU, targeting minimizing BBU power consumption. Meanwhile, in [24], the author proposed a joint activation and clustering scheme that maximizes networks coverage with QoS constraints. Based on traffic analysis, in [8], the author proposed a traffic-aware RRH-BBU assignment algorithm, dividing the problem into two parts. First, clustering the RRH based on the spatio-temporal variation model, as the author modeled the traffic load of RRH as an exponential function with a time-varying rate parameter. Then, the clustered RRH association with the BBU is described as a bin-packing optimization problem with the BBU as the bin and RRH as an item set. The author proposed a dynamic RRH assignment algorithm that offloads RRH from an overloaded BBU to a less loaded BBU.

2) CLUSTERING ALGORITHMS

In this section, we present the related work on various clustering algorithms recently employed in the literature to address the targeted problem, including spatial, temporal, and spatiotemporal clustering. From the perspective of clustering the RRH problem, predictive data analysis, data mining, and AI for decision-making based on Call Detail Records (CDR) have been proposed in different studies recently. References [18], [19], and [20] propose a supervised mobile traffic signature trained using prior knowledge of ground truth information for specific areas. Meanwhile, in [15], the author suggests unsupervised cell classification based on a mobile signature algorithm that efficiently classifies mobile loads, verified by ground-truth information. This technique has been applied to real mobile data collected from ten cities. Additionally, in [21], the author introduces heat map drawings of significant human activities based on mobile signature characterization, without prior knowledge of ground truth information. As geographical mobile signatures are primarily driven by land use, this produces a common pattern of user traffic in different cities and countries [19], motivating the spatial clustering of CDR. Consequently, various research has proposed dynamic RRH-BBU association based on the longitude and latitude of the RRH using spatial clustering [25]. Residential, entertainment, and work zones exhibit different traffic loads during the week. Residential zones experience the highest traffic during night hours, while work zones peak during the day, and entertainment zones experience peaks on weekends.

On the other side, the temporal clustering of the CDR has been proposed in the literature, as human behavior under stationary and normal circumstances is periodically repeated and this influences the aggregated networks load over a certain period. The real collected CDR shows similar behaviors over a certain period. The CDR supported by the mobile operator describes the traffic load at specific time stamps, usually, the networks traffic load is captured every ten minutes. The CDR is stored as a time series at a specific time and date. Clustering this complex temporal data is a challenge, as the massive data points of CDR represent a single object. The networks load temporal clustering has been proposed recently in research considering certain events as planned [26] and unplanned events [27]. The author of [28] extended this work to include the fine-tuned clustering of snapshots of the traffic demand over multiple periods. The clustering of the time series is utilized to discover frequent and rare patterns of the time series. Time series clustering proposes different tasks as recognizing dynamic changes caused by planned and unplanned events [26], [27], predicting future patterns, and discovering, and classifying different patterns [29]. In literature, the time series clustering methods are classified into three types of whole time series clustering, and the other two categories target clustering single long time series based on either subsequence clustering or time point clustering [30].

Different research extended the clustering to include spatial-temporal clustering. As we described the CDR of each cell over a week as a time series, this data is described as a geo-referenced time series with spatial-temporal data as it records time-changing values at fixed locations. The traditional temporal and spatial clustering methods represent one-way clustering methods, we focus mainly on co-clustering that clusters the data in two dimensions. The concept of data matrix clustering was first proposed, later this concept became used in data analysis of different fields such as bioinformatics data mining and weather temperature records. More researchers extended the co-clustering algorithms to tri-clustering that considers all the values of the recorded data at a certain time and fixed location represented by a 3D data matrix. In [13] the author proposed a 2D clustering algorithm (BBAC_I) that deals with an average of the data recorded over a year at a fixed position in a 2D matrix. And extended the work to a 3D clustering algorithm (BACT_I) that considers the recorded data at each time stamp over the total considered time interval at a fixed position represented by a 3D matrix.

B. CONTRIBUTION

Our contributions are concluded as

1) We propose a novel RRH-BBU assignment technique, as we proved that dividing the BBU into two level clusters, based on the same clustering algorithm of RRH, reduces the system power consumption and decreases the inter BBU handover rate. To our knowledge, this is the first research that extended the RRH clusters to be applied to the BBU in the RRH-BBU assignment phase.

2) We propose RRH clustering based on real collected CDR. The RRH clustering algorithm is described as a time series clustering algorithm that classifies RRH based on its temporal activity. To our knowledge, this is the first research to describe the CDR as time series.

3) We propose space-time series clustering that represents two nested clustering levels based on the CDR and the latitude and longitude of each cell.

4) We compute the optimum number of the BBU to accommodate the maximum real traffic load for each cluster.

5) We enhanced the system power consumption and operating cost minimization by extending the sleep mode to be applied to the BBU based on the assignment technique. We reduce the power consumption by 28.8 %, while in literature the author of [17] proposed 20% power saving based on a weighted graph that depends on user mobilities and neglects the traffic loads.

III. SYSTEM MODEL AND PROBLEM FORMULATION

In this section, we describe the suggested CRAN architecture for large cities and we study the effect of different clustering algorithms for the CRAN architecture on the networks performance. We propose a BBU to RRH mapping scheme based on different clustering algorithms and compare their influences on the power consumption and the backbone handover rate. The proposed clustering can be applied to equipment from any vendor. The vendor does not influence clustering performance since comparisons are made against the vendor's own un-clustered networks.

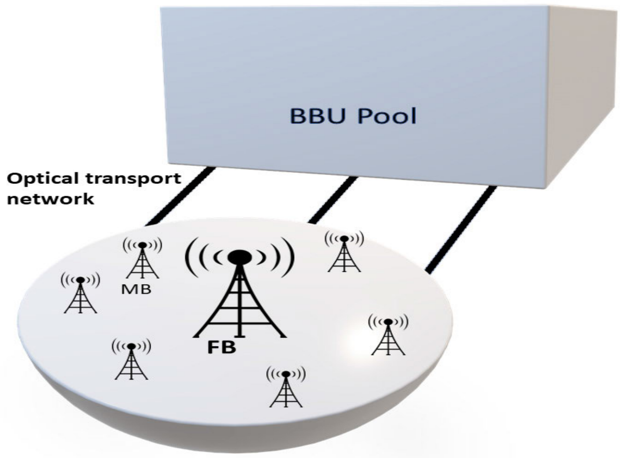


FIGURE 1. HCRAN (femto and micro base stations) architecture connected via an optical transport networks to the BBU pool.

A. CRAN-BASED NETWORKS ARCHITECTURE

The optical transport networks is considered the best candidate for the front-haul networks, as it connects the BBU and the RRH with a reliable and energy-efficient networks that provides low latency, high capacity as it connects a massive number of RRHs, and scalability [3], [4]. On the other side, the functionality split highly influences the system performance and the front-haul networks capacity. As the BBU performs baseband functionalities and is located at a remote office while the RRH, provides the radio functionalities to the end user scheme increases the rate of the front-haul links [14]. Whereas, the flexible functional split scheme, which loads the RRH with more baseband functionalities, enhances the capacity loaded on the front-haul networks and the power consumption. In CRAN architecture with a separated BBU pool and RRH and optical fronthaul transport networks, we consider a decentralization scheme of all baseband operations at the BBU pool, as the RRH performs only the radio operations for the attached user equipment (UE). Based on the traffic analysis and user requests the resources of the BBU

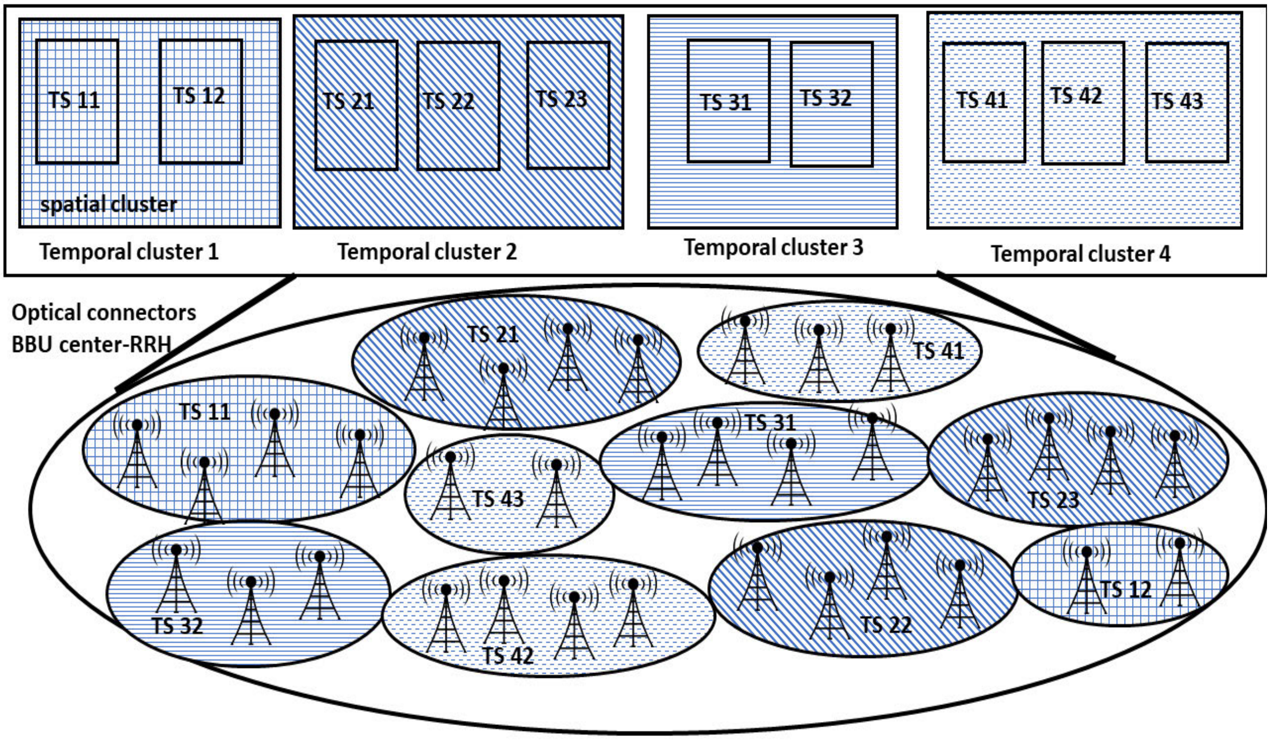


FIGURE 2. The proposed clustering model for temporal and spatial clusters in CRAN (TS11: spatial cluster 1 of temporal cluster 1).

pool must be allocated to each RRH. Mainly the allocation algorithms aim to reduce the networks power consumption and meet the quality of service.

In the proposed HCRAN architecture, the radio operations in the coverage area of the Femtocells are supported by the distributed RRH, while Macro base stations serve the cells with offloaded RRH during sleep modes. As turning off the underutilized units highly improves power consumption, Macro base stations have been introduced to maintain the radio operations in dead zones during low traffic loads as shown in Fig. 1. The proposed clustering algorithm aims to group some RRH based on specific features and assign resources from the connected BBU to this cluster.

In the context of HCRAN with users experiencing high mobility, frequent handovers between small cells can occur due to the small size of these cells. Efficient handover management is therefore crucial and is handled within the BBU pools [38]. In literature handovers between cells have been analyzed for all mobile networks Generations. In 4G mobile networks, handovers are necessary to sustain the Quality of Service (QoS) for ongoing user sessions and to connect users to the most optimal eNodeBs [39]. In LTE/LTE-A networks, inter-eNodeB handovers are managed via the X2 interface when both eNodeBs are connected to the same MME. Conversely, if the serving and target eNodeBs are linked to different MMEs, S1-based inter-eNodeB handovers are required. For a detailed explanation of X2 and S1-based inter-eNodeB handovers, refer to section [40]. The same principles apply to inter-BBU handovers discussed in this paper, with referring to inter-eNodeB as inter-RRH.

B. PROBLEM FORMULATION

In the proposed model, we assume M BBUs in the BBU pool with equal physical resources and computing capabilities, measured by million operations per time slot (MOPTS). The assigned BBU performs all the baseband operations for the attached RRHs. The required resources for baseband operations of RRH I depend on the traffic load and are noted as $R_{i.}$ Assuming each BBU has C MOPTS and up to N RRHs can be attached to each.

$$C_j \geq \sum_{i=1}^{N} \mu_{i,j} R_{i,j} \tag{1}$$

where $\mu_{i,j}$ is the assignment coefficient of RRH i to BBU j

$$\mu_{i,j} = \begin{cases} 1 & RRH\ i\ attached\ to\ BBU\ j \\ 0 & Otherwise \end{cases} \tag{2}$$

The power consumption of each BBU represents the baseband power, $P_{Dynamic}$ represents the dynamic power incurred by the traffic load and can be expressed as a linear function of it.

$$P_{BB} = P_{Dynamic} + P_{Static} \tag{3}$$

$$P_{Dynamic} = P_{filter} + P_{OFDM} + P_{DPD} + P_{FEC} + P_{CPU} + P_{CRPI} \tag{4}$$

$$P_{Dynamic} = \beta R \tag{5}$$

β is the load–power coefficient and R is the traffic load.

Where P_{filter}, P_{OFDM}, P_{DPD}, P_{FEC}, P_{CPU}, P_{CRPI} are the consumed power of the filtering, the OFDM transceivers, digital pre-distortion, CPU, encoder, and serial link to the backbone networks, respectively [41], [42].

The problem of assigning RRH to each BBU is described as a Bin Packing Problem (BPP). That can be solved using the Knapsack algorithm, with the RRH as objects and BBU capacity as the Knapsack. Clustering the BBU pool divides this single problem into multiple joint optimization problems. The algorithm assigns one item (object) to each bin (knapsack) such that the total weights of all objects do not exceed the capacity of the knapsack and minimize the number of the used knapsacks (BBUs).

The proposed networks architecture has N RRHs and M BBUs, each RRH can be only associated with a single BBU. The mathematical formulation of this problem, described by equation (6), and (9), represent an optimization problem that minimizes the number of active BBUs with capacity constraint.

$$minimize \sum_{i=1}^{K} B_i \tag{6}$$

$$\text{Subject to} \sum\nolimits_{i=1}^{N} w_{i,j}\mu_{i,j} \leq C_j B_j \quad j\epsilon\{1, .., K\} \tag{7}$$

$$\sum\nolimits_{j=1}^{K} \mu_{i,j} = 1 \quad \epsilon i\{1, .., N\} \tag{8}$$

where

$$B_i = \begin{cases} 1 & if\ BBU\ i\ is\ used \\ 0 & otherwise \end{cases} \tag{9}$$

$$w_j \leq C \tag{10}$$

where C_i represents the resources of the BBU i and w_j represents the weight of the resources assigned from each BBU to all attached RRHs. The first constraint in (7) ensures that the total assigned resources for all RRH attached to the BBU i are less than the total available resources at this BBU. In comparison, the second constraint verifies that each RRH must be served and attached to a single BBU.

TABLE 1. Parameters of the algorithm.

Cj	The resources of the BBU j
wi,j	The weight of the resources assigned from each BBU j to all attached RRHs i.
µi,j	The assignment factor of RRH I to BBU j

C. THE PROPOSED ALGORITHM

The proposed problem is described as a multi-objective optimization problem, that targets minimizing the networks power consumption and the handover between the BBUs. The algorithm depends on first temporally clustering the BBUs and the RRHs based on the CDR, which enables switching off the BBUs and RRHs of each cluster during the low traffic periods and minimizing the power. Secondly, spatial clustering is performed for each temporal cluster to ensure serving all nearby RRH by certain BBUs to minimize the handovers between the BBUs. Finally, the assignment process is performed for each cluster individually with reduced computational complexity as shown in the algorithm.

Algorithm 1 Clustering Algorithm

RH_i: i^{th} RRH where $i\epsilon\{1, \ldots, n\}$
BU_j: j^{th} RRH where $j\epsilon\{1, \ldots, k\}$
TC_u: u^{th} temporal clustering
SC_t: t^{th} spatial clustering
R_{iw}: The traffic load of i^{th} RRH over week w
NT: Number of temporal clusters
NSP: Number of temporal clusters
$No_{BBU_{ij}}$: Number of BBUs of temporal cluster i and spatial cluster j
BBU_RHH_{ij}: RRH to BBU assignment of temporal cluster i and spatial cluster j
Begin:
data processing R_{iw} to define:
  Cell id_i
  Time $interval_i$
  R_{iw} = Total activity
For $3 <=$ **n_clusters** < 8
    Apply DTW time series clustering (R_{iw}).
    Compute the distortion of each cluster.
End
NSP=min (distortion)
For $1 <$ **i** $<$ **NT**
NSP= DBscan(cluster(i))
End
For $1 <$ **i** $<$ **NT**
For $1 <$ **j** $<$ **NSP**
$No_{BBU}\ BR_{ij}$ = max (sum (R_{iw}))
$BBU\_RRH_{i\,j}$ = Bin_packing (R_{iw})
    End
End

1) CLUSTERING THE BBU & RRH

In literature the tri-clustering algorithms have been proposed to deal with geo-referenced time series, these algorithms analyze the CDR at a certain instant over the time interval at each RRH position. Although for the proposed application the RRH needs to be clustered based on its total activity. Based on this we proposed a two-level clustering that first clusters the CDR of the RRH as a temporal clustering then clusters a temporal cluster based on the location of the attached RRH.

Time Series-Based Clustering involves organizing sequences of data points collected at consecutive time intervals. For instance, clustering EEG signals can help identify patterns associated with various brain states or conditions. Most time-series clustering algorithms can be categorized into three types whole time-series clustering, subsequence clustering, and time point clustering [29]. The Euclidean distance metric has two key limitations that make Dynamic Time Warping (DTW) more effective for time series classification. First, it requires the time series to be of equal length, which poses a challenge when there are missing readings, leading to unequal lengths. Second, it compares the values of the two-time series at each time point independently, only considering the minimum Euclidean distances between the time series [43].

We describe the CDR as a multi-variant time series and refer to the networks temporal clustering as time series clustering. In which, the whole series clustering is considered as a single object and classified based on its similarities. These similarities are measured based on different distance measurements such as discrete time Wrapping (DTW) and Euclidean distance.

An agglomerative time series clustering algorithm, an up-bottom approach, is proposed for clustering the CDR of RRH as a temporal clustering. Temporal clustering is applied first on the CDR of each RRH, we propose K-means discrete time wrapping (DTW) time series clustering. First, the dataset is divided into K clusters and randomly k centroids, then DTW to assign each time series to the nearest cluster centroid and update the centroid based on the newly assigned time series. Where the DTW clustering finds all possible paths between two-time series to provide a distance matrix with a cumulative minimum distance of the three neighbors. Then it selects the minimum distance between the two series [11]. In the proposed algorithm the number of clusters is determined by measuring the distortion factors of each clusters number to find the optimum number of clusters.

Spatial clustering can be categorized into five main types: Partition clustering, Hierarchical clustering, Fuzzy clustering, Density-based clustering, and Model-based clustering. Given that CDR represents a vast geographical dataset containing latitude and longitude information for numerous RRH locations, spatial clustering must identify arbitrary cluster shapes, efficiently manage large quantities of points (ensuring scalability) and detect and eliminate noise and outliers. These requirements drive the adoption of density-based clustering algorithms [44].

The proposed spatial clustering algorithm divides the RRH of the CRAN architecture into clusters based on the longitude and latitude of each RRH. DBSCAN, K-means, and many different clustering algorithms are defined to cluster geographical data based on longitude and latitude. DBSCAN (Density-Based Spatial Clustering of Applications with Noise) is a dense-based clustering technique that forms a cluster based on dense connectivity analysis. BDSCAN is based on identifying the radius of the connected area with a minimum number of objects for each object in a cluster. There are two main parameters for the DBSCAN, as for each point of a cluster the neighborhood of a distance (R) must contain at least the number of points equals (Minpts) [11].

2) RRH TO BBU ASSIGNMENT

As mentioned, various RRH to BBU assignment techniques with different objectives have been proposed in the literature. The objective of power reduction in the communication networks arises as an application for green networks with reduced power consumption. We introduce an RRH to BBU assignment with power reduction and reducing handovers between the BBUs. The problem is formulated as a bin packing optimization problem that is considered an NP-hard problem. The proposed algorithm simplifies the problem by dividing it into clusters of joint RRH and BBU. Parallel processing, enabled by dividing the resource allocation for each cluster, divides the optimization problem into multiple simplified problems. The best-fit bin packing optimization is proposed, in which the item is packed into a bin by leaving the smallest residual space.

The complexity order of the optimization problem is $O(nlogn)$, where n is the number of the available bins (BBUs). The complexity order of the optimization problem based on a clustered model with M temporal clusters, Pj spatial clusters in temporal cluster j, and njk BBU in temporal cluster j and spatial cluster k is given by (11).

$$O(\sum_{j=1}^{M}\sum_{k=1}^{P_j} n_{jk}\log n_{jk}) \tag{11}$$

IV. RESULTS AND ANALYSIS USING REAL TRAFFIC: MILAN CITY

In this section, we present simulation results of the proposed RRH-BBU assignment based on the clustering algorithm.

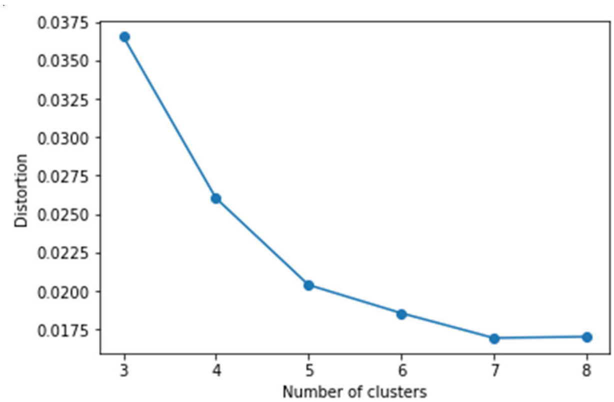


FIGURE 3. Distortion of different clustering numbers (With 7 clusters optimal due to constant distortion level at 7 clusters).

A. DATA

The CDR of Milan City is used as a case study area to simulate the proposed algorithm using big data analytics of collected real mobile traffic. The Italian telecom operator shared a CDR record for 10,000 cells with 235 × 235 m spatial resolution covering the area of Milan city [31]. The data records mobile activities sample every 10 minutes over 2 months from the first of November till the end of December 2013. The recorded activities are divided into call activities, SMS activities, and internet activities. The data has been processed and stored as data frames with multiple columns representing the square ID, time interval, and total activity (aggregated call, SMS, and internet activity). Each RRH is

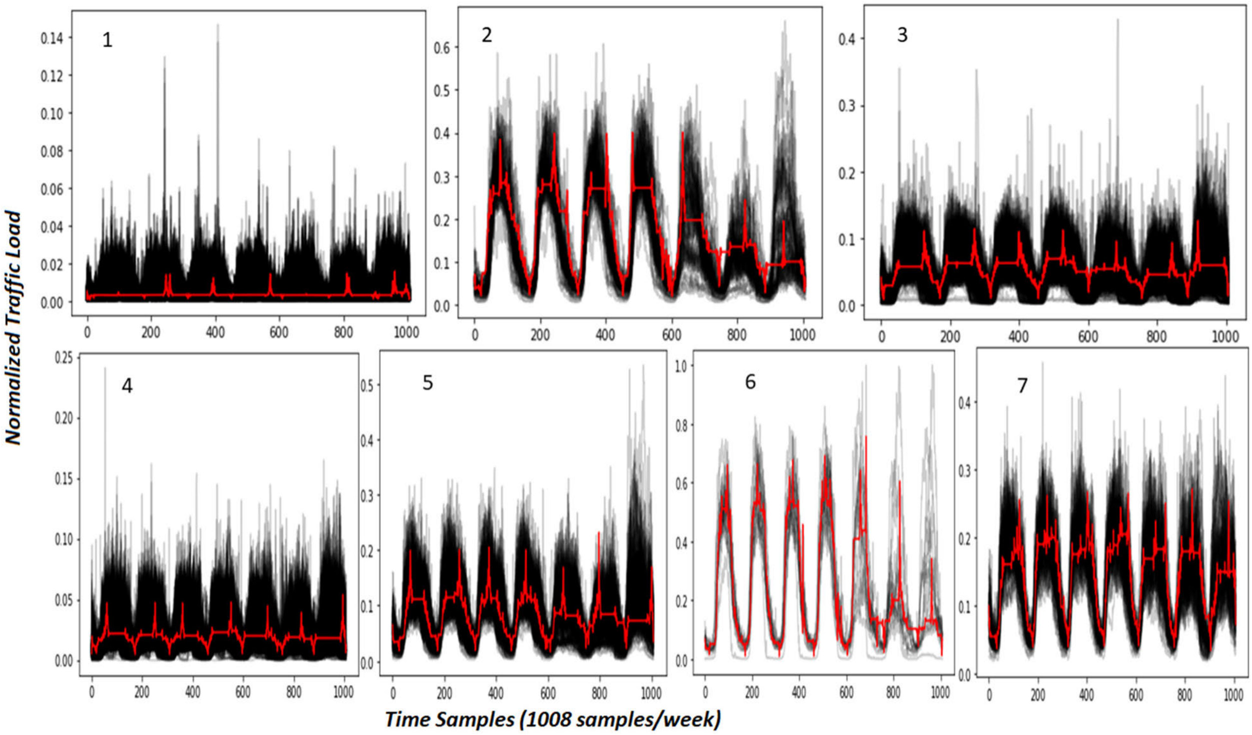


FIGURE 4. The Normalized traffic load of different clusters and cluster centers over week (Monday - Sunday).

served with only BBU and multiple RRHs can share the same BBU based on the aggregated traffic load. Each cell is served by multiple RRHs, and the proposed CDR is considered as a load of a single RRH.

B. SIMULATION RESULTS

The procedure of the applied algorithm on the CDR is shown in Algorithm 1, the data is processed using big data analytics as Dask data frames [32], [33], [34], [35], [36]. The total activity of the cellular mobile user is studied to classify and study the nature of the geographical area.

First, the data is grouped by the square ID and time intervals to analyze the traffic of each cell over the total period. We propose the observations of a week from Monday (4th November) to Sunday (10th November) as a first training set, the time samples representing this period are 1008 samples. The samples of each cell are treated as a time series producing 10,000 time series to be clustered. The values of the traffic to ensure accurate measurements, as the cells show records with enormous variations. The time series clustering is based on the K-Means algorithm that uses DTW as a similarity metric between different time series. As stated in the algorithm, the number of clusters is determined in an iterative manner, and the best number of clusters is chosen based on the first minimal distortion level among different numbers of clusters.

For the proposed data the cells are clustered into 7 clusters, and at 7 clusters the distortion level begins to be constant as shown in Fig.3. In Fig. 3, we plotted the distortion level in relation to the number of temporal clusters. The results show that as the number of clusters increases, the distortion level of the data assigned to each cluster decreases up to an optimal point. Beyond this point, further increases in the number of clusters lead to a rise in the distortion level. For the proposed dataset, this optimal number of clusters is found to be 7.

Fig. 4 shows the clusters center of the seven clusters and the activity of the cells belonging to each one. Each cluster center can indicate the nature of the geographical area, as clusters 2, and 6 show pure workspace areas with low traffic during the weekends and high traffic during the weekdays. Whereas cluster 3 represents the residential area with high traffic load during night hours. Some clusters show different behaviours composed of the basic classifications (residential, entertainment, transportation, and work area). These different combinations produce different clusters as shown in clusters 5, and 7. Cluster 5 shows moderate activities during all the hours of day and night, which shows the mixed area of workspace and entertainment area. Moreover, it is noticed that some cells show low traffic records compared with other cells as in clusters 1, and 4. The low traffic load cells are distributed over the total area of Milan and represent areas with low population ratio and transportation.

The results of the clustering system have been verified by comparing the results of the clustering algorithm and the land use of Milan City. Fig. 5. A show the published land use map of Milan city, while Fig. 5. B shows the proposed simulated temporally clustered cells of Milan using the cell location given in Fig. 6. As shown in Fig. 5, the agriculture area and the green areas are represented by cluster 1 with low normalized traffic at the city edges. Whereas the city center has the highest traffic load represented by Cluster 3 as a residential

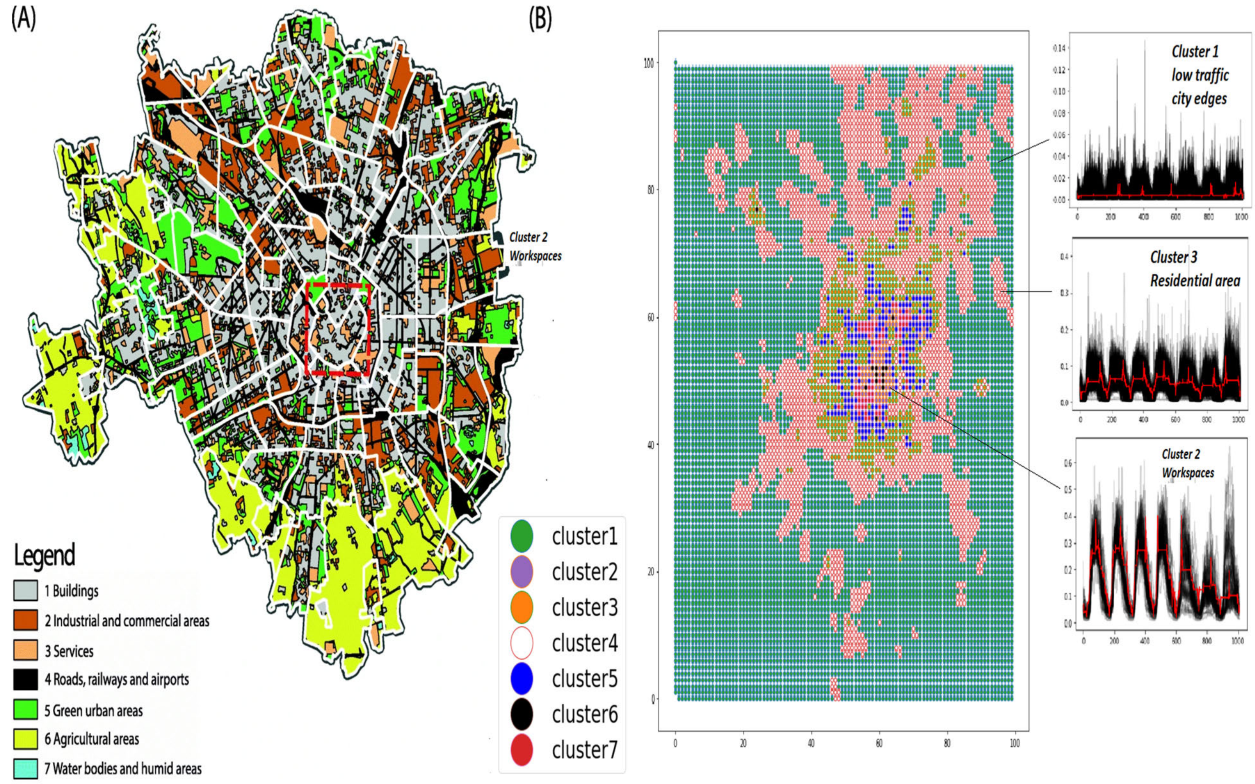


FIGURE 5. A. Published Milan land use map/ B. The land use map based on the time series clustering.

area and Cluster 2 as a workspace. Moreover, various clusters are shown at the city center representing entertainment and other activities as clusters 4, 5, and 7.

The second phase of clustering represents a spatial clustering for each temporal cluster, this cluster phase reduces the handover between different BBUs by assigning near RRHs to the same BBU. The accurate handover rate can be computed using the mobility of each user between different RRHs, in this work, we proposed the algorithm without computing the handover rate due to lack of user mobility data. The DBSCAN proposes a spatial clustering based on connectivity with minimum distance between different objects [33]. In Milan CDR data the square id represents the cell position as telecom Italy shared a map for the distribution of the cell over the covered area as in Fig. 6. The DBSCAN algorithm is applied on each temporal cluster individually to produce inner spatial clusters. Changing the radius highly affects the number of the produced spatial clusters, so this radius must be carefully studied and picked according to the case.

The cell id is translated into cartesian coordinates based on the published cell map given in Fig. 6, as shown the distance between two adjacent cells in the vertical axis is 100 and 1 in the horizontal axis. According to the Euclidian distance, the radius of the connected area, for the used data, is adjusted to the $\sqrt{2}$ after normalizing the vertical distance to 1. A minimum number of points of each cluster (Minpts) parameter of DBSCAN needs to be computed carefully, as it highly influences the number of formulated clusters and causes some random points to represent the un-clustered points.

9901	9902	....	9999	10000
9801	....	....	9899	9900
....	....	....	....	....
101	102	....	....	200
1	2	3	......	100

FIGURE 6. Milan cell grid.

As shown in Fig. 7, increasing the Minpts decreases the number of the used BBUs. As the Minpts increase the random points, which will be ignored in the BBU assignment process. This tradeoff, between neglecting some points and increasing the installed BBUs, is resolved by measuring the required BBUs for the traffic of the neglected point and comparing it with the corresponding number of saved BBUs from the neglect.

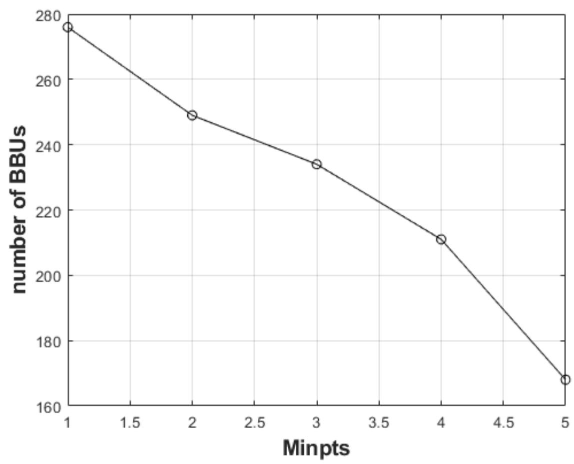


FIGURE 7. **The effect of the Minpts of the DBSCAN algorithm on the number of the spatial clusters.**

As shown in Fig. 8, for 3 Minpts the neglected random points save 42 BBUs, while it can be served by assigning 21 BBUs. The optimum point is at 3 as it verifies the tradeoff between saving the number of BBU while serving all the cells. On the other side, at 3 Minpts the additional BBUs cannot be included in the turning off BBU algorithm, as they serve random points with different temporal behavior. Consequently, this scheme decreases the power efficiency of the networks, so the performance metrics have been studied at 1 Minpts. Fig. 9 shows the spatial clustering for the single temporal cluster (cluster 1), all the near-connected cells are grouped in a single spatial cluster.

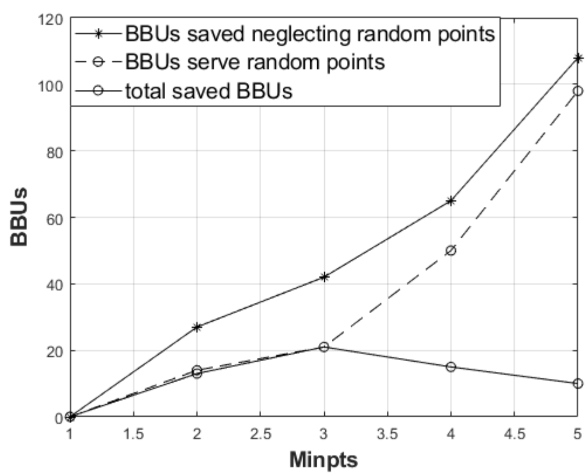


FIGURE 8. **The effect of the Minpts of the DBSCAN algorithm on the number of the added and reduced BBUs to serve the neglected.**

Table 2 shows the number of the inner spatial cluster, number of RRHs, and number of required BBUs of each temporal cluster. The number of the required BBUs is computed based on the normalized traffic required by computing and the ceiling the traffic of each spatial cluster TRk, according to (12).

$$BBU = £\sum\nolimits_{j=1}^{M}\left\lceil \sum\nolimits_{k=1}^{P_j} TR_k \right\rceil \tag{12}$$

where £ is the mapping ratio between the traffic l]oad and the number of the required BBUs.

TABLE 2. **Number of spatial clusters and RRHs in each temporal cluster.**

Temporal Clusters	No. of BBU		Spatial Cluster		No. of RRHs	
	1	3	1	3	1	3
Cluster 1	51	31	24	16	6657	
Cluster 2	15	9	13	5	36	
Cluster 3	56	49	66	27	613	
Cluster 4	75	71	64	32	2319	
Cluster 5	40	31	43	14	249	
Cluster 6	7	2	5	2	10	
Cluster 7	32	22	26	10	116	
Random Cells	-	21	-	-	-	168

Table 3 shows the total number of BBUs and the power consumption of 1, 3 Minpts clustered system and the un-clustered system. The required BBUs for an un-clustered system use the traffic load of all the RRHs by adding the traffic load and ceiling it according to (13).

$$\mathrm{BBU} = £\left\lceil \sum\nolimits_{u=1}^{V} TR_u \right\rceil \tag{13}$$

where V is the number of RRHs in the system.

As expected, the number of the clustered system is higher than the un-clustered system due to the multiple ceiling function used in the clustered system, as the assigned number of each temporal cluster must be integer.

TABLE 3. **Number of BBU in each cluster.**

	Clustered system		UN- Clustered system
	1	3	
No. of BBU	276	255	226
Power savings	28.8%	26.89%	0%

Finally, the RRH-BBU assignment is a bin packing problem assignment. The assignment problem is divided into multiple problems based on the number of temporal and spatial clusters. The proposed algorithm computes the required number of BBUs for each cluster based on the maximum value of the RRH's total traffic. Then, as mentioned, the algorithm performs nested loops over the temporal and the inner spatial clusters to assign the RRHs to the proposed BBUs. For the study case of Milan, we propose the number of active BBUs over time.

The proposed bin packing assignment is performed every hour based on the maximum traffic load over this time interval, as the maximum traffic load is computed and the required number of BBUs to serve the required traffic load using the bin packing algorithm. Fig. 10 shows the number of required

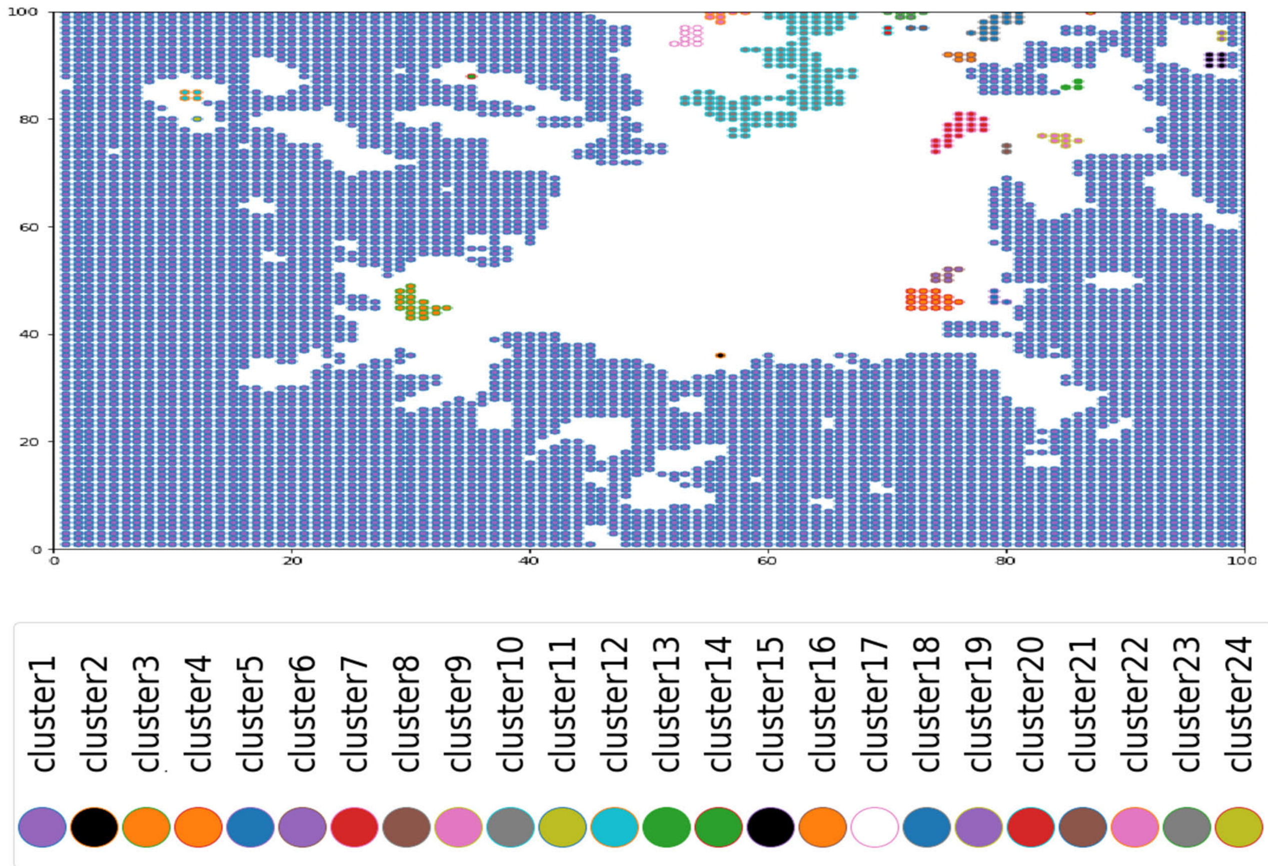


FIGURE 9. Spatial clusters distribution of temporal cluster 1.

active BBUs to serve the load demand based on DBSCAN with 1 Minpts, the required objectives are achieved in two phases.

First, each spatial cluster is treated as a single entity to ensure the objective of minimizing inter-BBU handover. The accurate handover enhancement rate is computed with the user movement scheme between different cells. As this data is not shared from the operator, we depend on decreasing the total handover rate with grouping all the nearby cells. Then, the algorithm is performed separately on each temporal cluster to decrease the power consumption by turning off the under-utilized BBUs over the monitoring time interval (1 week). The power saving of the DBSCAN algorithm with 1 and 3 Minpts are compared using the average power saving in each scheme according to (14).

$$\text{Power\_saving} = \frac{BBUun - \sum_{t=1}^{time\_points(168)} BBU\_CLU_t}{BBUun} \tag{14}$$

where BBU_{UN} represents the total BBUs of the un-clustered system, and BBU_CLUt represents the total active BBU of the clustered system at time t. As shown in Table 2, DBSCAN with 1 Minpts saves 28.8% of the total power consumption of the un-clustered system, while the 3 Minpts scheme saves 26.89%. taking into consideration, the higher number of installed BBUS in the 1 Minpts scheme. Turning off underutilized BBUs reduces active power consumption over time. As indicated in Equation (14), increasing the number of active BBUs at any given moment leads to higher power consumption and lower power savings.

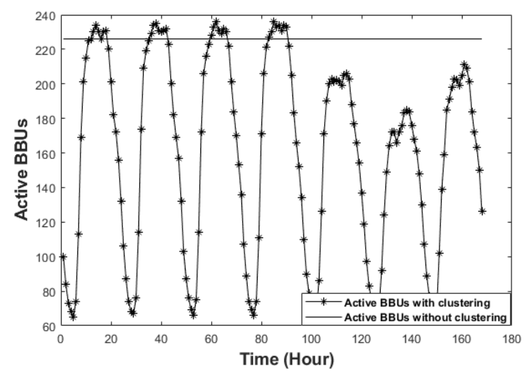


FIGURE 10. The number of active BBUs over one hour using clustering and traditional schemes.

Fig. 10 illustrates that power savings are calculated by subtracting the number of active BBUs in a standard un-clustered networks from the number of active BBUs using the proposed models, thereby determining the reduced power consumption achieved by switching off BBUs.

V. CONCLUSION

In this paper, we proposed an RRH-BBU assignment based on a clustering algorithm that targets minimizing the power consumption and the inter-BBU handover. The proposed algorithm computes the required number of installed BBUs to accommodate the maximum traffic load, deploys time series clustering as a temporal clustering method, and applies the DBSCAN algorithm to divide each temporal cluster into several spatial clusters based on the cell location. Then the problem of assigning RRH of each cluster is described as a bin packing optimization problem to find the optimum number of BBUs for each cluster [45].

The proposed algorithm has been validated using real-world CDR of Milan city and it is verified by the published Milan land use map. The inter-BBU handover signals have been enhanced by assigning near RRHs to the same spatial cluster and same BBU, to avoid inter-BBU handover. The accurate handover rate can be computed using the user mobility between different RRHs, as stated we provide the algorithm without computing the handover rate due to a lake of user mobility data. It is shown that the algorithm reduces the total power consumption of the current deployed networks by 28.8 %, by assigning all the active RRHs to certain BBUs based on the traffic load and switching off the unassigned BBUs. This research could be expanded to include diverse CDR datasets, as each city exhibits unique temporal and spatial traffic load patterns influenced by its cultural behaviors. We encourage service providers to share CDR datasets of different cities, which would enable researchers to apply and extend this work to various cities, taking into account different traffic load distributions and cultural behaviors.

REFERENCES

[1] J. Ordonez-Lucena, P. Ameigeiras, D. Lopez, J. J. Ramos-Munoz, J. Lorca, and J. Folgueira, "Network slicing for 5G with SDN/NFV: Concepts, architectures, and challenges," *IEEE Commun. Mag.*, vol. 55, no. 5, pp. 80–87, May 2017, doi: 10.1109/MCOM.2017.1600935.

[2] D. Camps-Mur, J. Gutierrez, E. Grass, A. Tzanakaki, P. Flegkas, K. Choumas, D. Giatsios, A. F. Beldachi, T. Diallo, J. Zou, P. Legg, J. Bartelt, J. K. Chaudhary, A. Betzler, J. J. Aleixendri, R. Gonzalez, and D. Simeonidou, "5G-XHaul: A novel wireless-optical SDN transport network to support joint 5G backhaul and fronthaul services," *IEEE Commun. Mag.*, vol. 57, no. 7, pp. 99–105, Jul. 2019, doi: 10.1109/MCOM.2019.1800836.

[3] C.-Y. Chang, R. Schiavi, N. Nikaein, T. Spyropoulos, and C. Bonnet, "Impact of packetization and functional split on C-RAN fronthaul performance," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2016, pp. 1–7.

[4] (2013). *C-RAN: The Road Towards Green RAN*. [Online]. Available: http://labs.chinamobile.com/cran/

[5] W. Ejaz, S. K. Sharma, S. Saadat, M. Naeem, A. Anpalagan, and N. A. Chughtai, "A comprehensive survey on resource allocation for CRAN in 5G and beyond networks," *J. Netw. Comput. Appl.*, vol. 160, Jun. 2020, Art. no. 102638.

[6] R. T. Rodoshi, T. Kim, and W. Choi, "Resource management in cloud radio access network: Conventional and new approaches," *Sensors*, vol. 20, no. 9, p. 2708, May 2020.

[7] U. Karneyenka, K. Mohta, and M. Moh, "Location and mobility aware resource management for 5G cloud radio access networks," in *Proc. Int. Conf. High Perform. Comput. Simul. (HPCS)*, Genoa, Italy, Jul. 2017, pp. 168–175, doi: 10.1109/HPCS.2017.35.

[8] D. Mishra, P. C. Amogh, A. Ramamurthy, A. A. Franklin, and B. R. Tamma, "Load-aware dynamic RRH assignment in cloud radio access networks," in *Proc. IEEE Wireless Commun. Netw. Conf.*, Apr. 2016, pp. 1–6.

[9] K. Boulos, M. E. Helou, K. Khawam, M. Ibrahim, S. Martin, and H. Sawaya, "RRH clustering in cloud radio access networks with re-association consideration," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Barcelona, Spain, Apr. 2018, pp. 1–6.

[10] M. Mouawad, Z. Dziong, and M. Khan, "Quality of service aware dynamic BBU-RRH mapping based on load prediction using Markov model in C-RAN," in *Proc. IEEE Int. Conf. Internet Things (iThings) IEEE Green Comput. Commun. (GreenCom) IEEE Cyber, Phys. Social Comput. (CPSCom) IEEE Smart Data (SmartData)*, Jul. 2018, pp. 1907–1912.

[11] A. I. M. Badr. (2023). *Machine Learning Algorithms for 5G Optical Networks*. [Online]. Available: http://hdl.handle.net/2117/403554

[12] K. Boulos, K. Khawam, M. El Helou, M. Ibrahim, H. Sawaya, and S. Martin, "An efficient scheme for BBU-RRH association in C-RAN architecture for joint power saving and re-association optimization," in *Proc. IEEE 7th Int. Conf. Cloud Netw. (CloudNet)*, Oct. 2018, pp. 1–6.

[13] 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, no. 1, pp. 1–15, Oct. 2015.

[14] C. Ranaweera, E. Wong, A. Nirmalathas, C. Jayasundara, and C. Lim, "5G C-RAN with optical fronthaul: An analysis from a deployment perspective," *J. Lightw. Technol.*, vol. 36, no. 11, pp. 2059–2068, Jun. 3, 2018, doi: 10.1109/JLT.2017.2782822.

[15] E. Aqeeli, A. Moubayed, and A. Shami, "Power-aware optimized RRH to BBU allocation in C-RAN," *IEEE Trans. Wireless Commun.*, vol. 17, no. 2, pp. 1311–1322, Feb. 2018.

[16] M. Y. Lyazidi, N. Aitsaadi, and R. Langar, "Dynamic resource allocation for cloud-RAN in LTE with real-time BBU/RRH assignment," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2016, pp. 1–6.

[17] B. J. R. Sahu, S. Dash, N. Saxena, and A. Roy, "Energy-efficient BBU allocation for green C-RAN," *IEEE Commun. Lett.*, vol. 21, no. 7, pp. 1637–1640, Jul. 2017.

[18] F. Girardin, F. Calabrese, F. D. Fiore, C. Ratti, and J. Blat, "Digital footprinting: Uncovering tourists with user-generated content," *IEEE Pervasive Comput.*, vol. 7, no. 4, pp. 36–43, Oct. 2008.

[19] R. A. Becker, R. Caceres, K. Hanson, J. M. Loh, S. Urbanek, A. Varshavsky, and C. Volinsky, "A tale of one city: Using cellular network data for urban planning," *IEEE Pervasive Comput.*, vol. 10, no. 4, pp. 18–26, Apr. 2011.

[20] J. L. Toole, M. Ulm, M. C. Gonzalez, D. Bauer. (2012). *Inferring Land Use From Mobile Phone Activity*.

[21] P. Secchi, S. Vantini, and V. Vitelli, "Analysis of spatio-temporal mobile phone data: A case study in the metropolitan area of Milan," *Stat. Methods Appl.*, vol. 24, no. 2, pp. 279–300, Jul. 2015.

[22] K. Boulos, M. E. Helou, and S. Lahoud, "RRH clustering in cloud radio access networks," in *Proc. Int. Conf. Appl. Res. Comput. Sci. Eng. (ICAR)*, Beirut, Lebanon, Oct. 2015, pp. 1–6, doi: 10.1109/ARCSE.2015.7338135.

[23] M. Qian, W. Hardjawana, J. Shi, and B. Vucetic, "Baseband processing units virtualization for cloud radio access networks," *IEEE Wireless Commun. Lett.*, vol. 4, no. 2, pp. 189–192, Apr. 2015.

[24] H. M. Soliman and A. Leon-Garcia, "QoS-aware joint RRH activation and clustering in cloud-RANs," in *Proc. IEEE Wireless Commun. Netw. Conf.*, Apr. 2016, pp. 1–6.

[25] Z. H. Fakhri, M. Khan, F. Sabir, and H. S. Al-Raweshidy, "A resource allocation mechanism for cloud radio access network based on cell differentiation and integration concept," *IEEE Trans. Netw. Sci. Eng.*, vol. 5, no. 4, pp. 261–275, Oct. 2018.

[26] F. Calabrese, F. C. Pereira, G. D. Lorenzo, L. Liu, C. Ratti. (2010). *The Geography of Taste: Analyzing Cell-Phone Mobility and Social Events*.

[27] J. P. Bagrow, D. Wang, and A.-L. Barabási, "Collective response of human populations to large-scale emergencies," *PLoS One*, vol. 6, no. 3, Mar. 2011, Art. no. e17680.

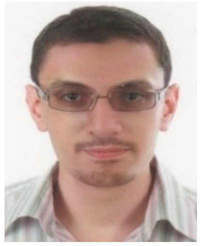
[28] D. Naboulsi, R. Stanica, and M. Fiore, "Classifying call profiles in large-scale mobile traffic datasets," in *Proc. IEEE INFOCOM*, 2014, pp. 1–20.

[29] S. Aghabozorgi, A. Seyed Shirkhorshidi, and T. Ying Wah, "Time-series clustering—A decade review," *Inf. Syst.*, vol. 53, pp. 16–38, Oct. 2015.

[30] E. Keogh and J. Lin, "Clustering of time-series subsequences is meaningless: Implications for previous and future research," *Knowl. Inf. Syst.*, vol. 8, no. 2, pp. 154–177, Aug. 2005.

[31] (2015). *Telecommunications—SMS, Call, Internet—MI*.
[32] K. Khan, S. U. Rehman, K. Aziz, S. Fong, and S. Sarasvady, ''DBSCAN: Past, present and future,'' in *Proc. 5th Int. Conf. Appl. Digit. Inf. Web Technol. (ICADIWT)*, Feb. 2014, pp. 232–238.
[33] M. Abadi. (2015). *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. [Online]. Available: https://www.tensorflow.org/learn
[34] F. Chollet. (2015). *Keras*. [Online]. Available: https://github.com/fchollet/keras
[35] W. McKinney, ''Data structures for statistical computing in Python,'' in *Proc. 9th Python Sci. Conf.*, 2010, pp. 51–56.
[36] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay, ''Scikit-learn: Machine learning in Python,'' *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Nov. 2011.
[37] L. Buitinck, ''API design for machine learning software: Experiences from the scikit-learn project,'' in *Proc. ECML PKDD Workshop Lang. Data Mining Mach. Learn.*, 2013, pp. 108–122.
[38] H. Zhang, C. Jiang, J. Cheng, and V. C. M. Leung, ''Cooperative interference mitigation and handover management for heterogeneous cloud small cell networks,'' *IEEE Wireless Commun.*, vol. 22, no. 3, pp. 92–99, Jun. 2015.
[39] M. Khan, R. S. Alhumaima, and H. S. Al-Raweshidy, ''QoS-aware dynamic RRH allocation in a self-optimized cloud radio access network with RRH proximity constraint,'' *IEEE Trans. Netw. Service Manage.*, vol. 14, no. 3, pp. 730–744, Sep. 2017.
[40] *General Packet Radio Service (GPRS) Enhancements for Evolved Universal Terrestrial Radio Access Networks (E-UTRAN) Access*, Standard TS23.401, 3GPP, 2016.
[41] C.-Y. Chang, N. Nikaein, R. Knopp, T. Spyropoulos, and S. S. Kumar, ''FlexCRAN: A flexible functional split framework over Ethernet fronthaul in cloud-RAN,'' in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2017, pp. 1–7.
[42] Y. Zhou, J. Li, Y. Shi, and V. W. S. Wong, ''Flexible functional split design for downlink C-RAN with capacity-constrained fronthaul,'' *IEEE Trans. Veh. Technol.*, vol. 68, no. 6, pp. 6050–6063, Jun. 2019.
[43] J. Faouzi, ''Time series classification: A review of algorithms and implementations,'' *Mach. Learn. Emerg. Trends Appl.*, vol. 2, pp. 1–24, Jun. 2022.
[44] C. Aggarwal and C. Reddy. (2013). *Data Clustering Algorithms and Applications*.
[45] C. A. Mandal, P. P. Chakrabarti, and S. Ghose, ''Complexity of fragmentable object bin packing and an application,'' *Comput. Math. Appl.*, vol. 35, no. 11, pp. 91–97, Jun. 1998.

ASMAA IBRAHIM received the master's and bachelor's degrees (Hons.) from Cairo University, Egypt. She is currently pursuing the Ph.D. degree with Universitat Politècnica de Catalunya (UPC), Catalonia, Spain. She is a Postdoctoral Fellow with UAEU, was an ESR with CTTC, Barcelona, Spain, in the framework of the Marie-Curie ITN 5G STEP-FWD project, and as a Teaching Assistant with the AUC, Zewail, and the International Academy for Engineering and Media Science (IAEMS), Egypt. Her M.Sc. thesis focused on optimized resource allocation and interference management techniques in indoor optical communications. She has published number of papers during her M.Sc. studies in international journals and conferences. Her research interests include 5G wireless communication networks, wireless optical communications, and visible light communications.



AHMED ELSHEIKH received the B.Sc. degree in electronics and electrical communications engineering and the M.Sc. degree in mathematics and engineering physics from Cairo University, Giza, Egypt, in 2011 and 2014, respectively, and the Ph.D. degree in mathematics and industrial engineering from the University of Montreal, Montreal, QC, Canada, in 2018. He is currently an Assistant Professor with the Mathematics and Engineering Physics Department, Cairo University. His research interests include applied machine learning in various domains, such as engineering physics, industrial engineering, and communications.

BASSEM MOKHTAR (Senior Member, IEEE) received the B.Sc. and M.Sc. degrees in electrical engineering from Alexandria University, Egypt, in 2004 and 2008, respectively, and the Ph.D. degree in computer engineering from Virginia Tech, USA, in 2014. He is currently an Assistant Professor of computer engineering with the College of Information Technology, United Arab Emirates University, Al Ain, United Arab Emirates. He has experience in teaching bachelor's and graduate courses related to computer and electrical engineering and computer science. He has been involved in a set of funded multidisciplinary research projects integrating advances in fields of machine learning-based applications in sensor networks, nanotechnology, and flexible electronics. Also, he participated as PI, Co-PI, and a Researcher in a set of funded research projects that cross multi-disciplinary research fields, including the Internet of Things, the Internet of Vehicles, and software-defined networking and flexible electronics. Additionally, he has been involved as a Reviewer for various conferences and journals, including IEEE INTERNET OF THINGS JOURNAL, *IEEE Vehicular Technology Magazine*, and *Journal of Networks and Computer Applications* (Elsevier). His research interests include embedding intelligence in networking operations, autonomic resilient networking, network semantics reasoning, and machine learning applications in computer networks.

JOSEP PRAT received the M.S. degree in telecommunications engineering, in 1987, and the Ph.D. degree from Universitat Politècnica de Catalunya (UPC), Barcelona, in 1995. He is currently a Full Professor with the Optical Communications Group, Signal Theory and Communications Department, UPC, and coordinates the Optical Access Networks Laboratory. He has mainly investigated broadband optical communications with emphasis on FTTH access networks and high bit-rate WDM transmission systems. He led the FP7 European project SARDANA (''Scalable Advanced Ring-Based Passive Dense Access Networks Architecture'') on next-generation FTTH networks, winning the 2011 Global Telecommunications Business Innovation Award in the Fixed Networks Infrastructure Category, and has participated in the international projects COCONUT, ACCORDANCE, Euro-Fos, BONE, ePhoton/One, LION, MEPHISTO, MOON, SONATA, and RACE1027, on optical transport and access networks. He was a Guest Scientist with the University College London, in 1998, and Stanford University, in 2016. He has been the Subdirector of the ETSETB Telecom School and a member of the Government Counsel of UPC. He has published more than 200 international works and edited the books *Fiber-to-the-Home Technologies* and *Next-Generation FTTH Passive Optical Networks* (Springer Ed.), and has supervised 16 Ph.D. Thesis. He was an Associate Editor of the IEEE-PTL and a TPC member of OFC and ECOC.