Received August 22, 2016, accepted September 28, 2016, date of publication March 16, 2017, date of current version April 24, 2017.

Digital Object Identifier 10.1109/ACCESS.2016.2627527

Hybrid Clustering Scheme for Relaying in Multi-Cell LTE High User Density Networks

MARYAM HAJJAR[1], GHADAH ALDABBAGH[1,2], NIKOS DIMITRIOU[3], (Senior Member, IEEE), AND MOE Z. WIN[2], (Fellow, IEEE)

[1]Computer Science Department, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah 23218, Saudi Arabia
[2]Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[3]Institute of Informatics and Telecommunications, NCSR Demokritos, Athens 153 10, Greece

Corresponding author: G. Aldabbagh (galdabbagh@kau.edu.sa)

This work was supported by the National Plan for Science, Technology and Innovation (MAARIFAH), King Abdulaziz City for Science and Technology, Saudi Arabia, under Award 12-INF 2743-03.

ABSTRACT LTE usage is rapidly increasing with the increased demand for high data rate services. LTE cells with high load would be insufficient to handle all users' traffic. This may cause blocking for some users or degrade the quality for others. Several techniques are adopted to increase the capacity of LTE cells. This paper will address specifically the use of some users as relays to others in order to increase the capacity of a specific cell. The result is a two-hop topology network with some users connected directly to the base station (BS) and others using some of the already connected users as relays to access the BS. Different techniques could be used to configure the users in such a topology. The paper proposes a new algorithm for relay selection in a multi-cell scenario based on K-means and selection strategy.

INDEX TERMS Clustering, frequency reuse, K-means, LTE.

I. INTRODUCTION

In a conventional LTE network, a User Equipment (UE) communicates directly with a Base Station (BS) within a specific coverage area. Nevertheless, as the number of LTE users is continuously increasing with the emergence of smart devices and 4G/5G service requirements, relay schemes are being considered to increase the capacity of the cell. Relaying is one of the promising schemes that enable reusing the available frequency within a cell area without the need for additional spectrum. In a relay-enhanced LTE network, a user may communicate directly to the BS or use a near relay station to relay its data to the BS. In such topology, the communication will require less power. Therefore, interference is minimized, and frequency reuse could be applied, which in turn enhances the capacity of the system.

Relay devices are mostly fixed stations, being a part of the network infrastructure. However, since density occurs only occasionally and in many cases it is unpredictable, it would be feasible instead to use some users' devices as relays. This will require low deployment and maintenance cost compared to the fixed station solution [1]. Moreover, in dense situations with a high number of UEs, it would be more likely to find closer relay links in the case of relaying through other users.

The allocation of users in such two-layer topology creates clusters of adjacent users, where each cluster uses a single relay to send all its users' data to the BS. This can be viewed as a clustering problem. Clustering is an unsupervised learning problem, defined as the process of grouping more similar data objects in the same cluster and less similar data objects in different clusters [2]. Clustering has been adopted in wireless networks to group nodes that are geographically adjacent together so they can communicate more easily. Clustering approaches could be found in mobile ad-hoc networks (MANETs), Wireless Sensor Networks (WSNs), and LTE networks. In MANETS, clustering facilitates communications, as there is no central operator. Moreover, it enables frequency reuse which in turn increases the system's capacity [3]. Clustering in WSNs solves problems of lifetime and scalability of the network by preserving energy of the sensor nodes [4], while in LTE clustering is used to enable reuse and increase capacity of the cell [5], [6]. The two most widely used clustering algorithms in wireless networks are K-means and Hierarchical Agglomerative Clustering (HAC).

K-means works by clustering the data points (nodes) into K clusters. It starts by randomly determining K centroids that the clusters will be allocated around. Then, a similarity value is calculated for each node with each centroid. Each node then joins the cluster of the closest centroid. After that, the algorithm assigns new K centroids by computing the mean of all points inside each cluster. The algorithm will iterate re-assigning clusters and computing new centroids until no change in the K centroids is acquired. Mostly distance is used as a similarity metric in K-means. K-means has $O(Nki)$ complexity, where N is the number of nodes, k is the number of clusters and i is the number of convergence iterations [7].

Wide ranges of studies address the use of K-means clustering algorithm in WSNs [4], [8]–[14]. Both centralized and distributed approaches of K-means in WSNs can be found [4], [14]. Cluster Heads (CHs) are chosen from each cluster based on two factors, distance from the centroid and residual energy [4], [11]. The study in [11] uses K-means with the adjustment of the chosen centroids; in each iteration, every centroid is assigned as the actual node that is nearest to the center of the cluster. The study in [13] proposes a modified K-means algorithm that includes for each cluster three CHs to take turns as active CHs. This will enable load balancing and increase network life time.

K-means has also been applied in cellular networks and in LTE networks. For instance, the study in [15] uses K-means to specify the locations of relay stations. Moreover, in [5] a two-hop relay system is adopted by applying K-means to generate low power usage clusters. The study suggests that direct communications with the BS are done through the LTE band, while communications inside the clusters are done through another frequency band (white-space). On the other hand, the study in [6] addresses a similar approach using only the LTE band.

HAC algorithm works in a bottom-up approach. It stars with every data point (node) as a distinct cluster. Then, in every successive step, the most similar pair of clusters are merged until all nodes are contained within one cluster [16]. This kind of approach creates a dendrogram tree topology. The algorithm can also create disjoint clusters by cutting the dendrogram tree at a specific point. This cutting point could either be a predefined number of clusters or a predefined similarity value. HAC has a complexity of $O(N^3)$, where N represents the number of nodes [17].

HAC has been commonly applied in WSNs [10], [18], [19]. The bottom up approach of the algorithm helps in implementing self-organizing WSNs. The study in [18] uses HAC to group nodes into disjoint clusters by cutting the dendrogram tree at a predefined threshold. This threshold could either be the transmission radius or the number of required clusters. Also, another threshold is defined in this study for the minimum cluster size. Clusters that are smaller than this threshold will merge with their closest cluster. CHs are then chosen based on two parameters, lower level in the hierarchy tree and energy level. Moreover, the study in [19] compares between two different methods for calculating distances between clusters in HAC, namely the single link and the complete link. In the single link method, distances between clusters are measured based on the minimum distance between a node in one cluster and a node in another, while in the complete link method, distances are based on the maximum distance. Results have shown that the single link is less efficient compared to the complete link. However, the complete link has higher complexity. HAC has also been implemented to cluster nodes in LTE networks as proposed in [20].

Although HAC and K-means are both used for creating disjoint clusters, they perform differently. In K-means, the number of clusters K has to be defined in advance, while in HAC this information is not required. However, K-means has an iterative improving procedure in contrast to HAC that iterates to build up the clusters. K-means is also computationally efficient with much lower complexity compared to HAC [7], [17]. Moreover, the results from the study in [20] show that HAC and K-means have similar performance in terms of capacity when clustering is done on an LTE network.

Some other studies in LTE and cellular networks [1], [6], [21]–[23] use selection strategies to enable the relaying system rather than clustering. In this manner, each node chooses its best relay (i.e. CH) from a predefined pool of candidate relays. The selection can be done based on different criteria. Some are based on choosing the relay with least distance while others are based on the minimum path loss [6], [21], [22]. Using path loss is better than distance because it gives better estimation of the link quality. Some other factors for selecting relay nodes include SIR [1], [21], transmission power [21], throughput [24], and SNR [24]. These techniques optimize performance only from the UE point of view without considering the overall system performance and capacity.

In [6] a comparison between clustering techniques and selection strategies in LTE is investigated. The study shows that clustering techniques are better than selection strategies in terms of network capacity. However, studies concerned with clustering in LTE [5], [6] have considered a single cell scenario only. In high user density situations, interference will be a major issue when multiple cells are deployed, and it will force a limitation on the location of CHs. Therefore, different considerations should be adopted.

Motivated by the above discussion, this paper proposes a way to create small cells called *clusters* in an LTE macro cell within a multi-cell scenario. These clusters will be created based on promoting a specific user as a Cluster Head (CH) to act as a temporary (ad-hoc) relay station. This CH will relay all other users' data that are in the same cluster to the BS. These cluster members are called *slaves* (Fig. 1). The clustering approach is based on proper allocation of CH and slave nodes to meet the constraint of low power clusters. Transmitting and receiving in such low power clusters will enable reusing the frequency in other clusters without causing significant interference and hence capacity is increased. This increase in capacity is achieved without the need of additional infrastructure.

The rest of the paper is organized as follows. Section II explains the system model and frequency planning. The proposed algorithm is presented in Section III. Simulation setup along with results are given in Section IV. Finally, the paper is concluded in Section V.

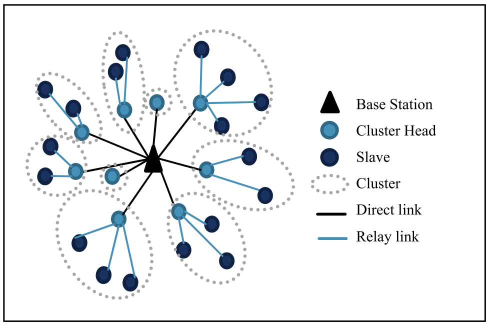

FIGURE 1. Proposed topology.

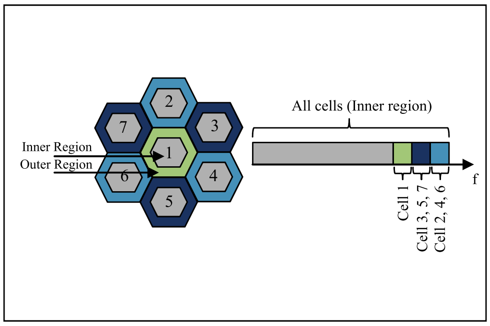

FIGURE 2. Applied frequency planning in proposed scheme.

II. FREQUENCY PLANNING AND SYSTEM MODEL

The system considered in this paper is composed of an LTE cell in the center and its first tier interfering cells (Fig. 2). The topology of the center cell consists of two layers of communication. The first layer has users that communicate directly to the BS while the second layer has users that communicate with first layer users as relays to access the BS. This kind of association forms clusters. Each cluster has only one CH, which is a first layer user and a number of slaves that connect to this CH (Fig. 1). Each connected user will be allocated a single Resource Block (RB). Each user has a specific SINR requirement. A slave user will have an SINR requirement α and a throughput requirement Th_{min}, while a CH will have an aggregated SINR and throughput requirements of all its slaves' requirements plus its own, such as:

$$\alpha_{ch} = cluster\_size \times \alpha \tag{1}$$

$$Th_{min,ch} = cluster\_size \times Th_{min} \tag{2}$$

where α and Th_{min} are the SINR and throughput requirements of a slave, while α_{ch} and $Th_{min,ch}$ are the SINR and throughput requirements of a CH. $cluster\_size$ is defined as the total number of slaves belonging to the cluster plus the CH itself:

$$cluster\_size = total\_slaves + 1 \tag{3}$$

A. INTER-CELL INTERFERENCE

Soft Frequency Reuse (SFR) [25] is assumed in this paper. Even though full frequency reuse can provide more spectrum resources than SFR, it would increase interference in a multi cell network. Interference would be even much more significant when the network is heavily loaded. In such case, interference will extremely degrade the performance and hence users will not be able to have reliable connectivity especially at the cell edge region. Therefore, frequency planning is crucial to minimize inference and increase capacity while providing acceptable service.

SFR is applied as shown in Fig. 2. 1/3rd of available RBs are used in the outer cell region and full frequency reuse is assumed in the inner cell region. Users at the inner region of cell 1 will be affected by inter-cell interference from all surrounding cells' inner regions.

B. POWER CONTROL

This study focuses on the uplink power control. Power control will be used for all co-channel users. This includes the first tier interferes at the inner cell region of all surrounding cells along with other co-channel users inside the cell (due to clustering). As for the outer cell region, co-channel users from the same cell are only considered which possibly include one macrocell user and a few slaves communicating with their CHs.

Power allocation depends on the SINR requirement of each user (α) and is detailed as follows:

The SINR equation for a single user i,

$$SINR_i = \frac{\frac{P_{tx,i}}{L_{ij}}}{I_i + N} = \alpha_i \tag{4}$$

where,
N is the noise. L_{ij} is the path-loss between i and j (having j as the receiver of i) and it is calculated based on the log-distance path-loss model as:

$$L_{ij} = r_{ij}^{e} 10^{\frac{\xi_{ij}}{10}} \tag{5}$$

r is the distance between i and j. e is the path-loss exponent, and ξ is the shadowing value between i and j.

I_i is the interference power received at node i. It is assumed to be equal to the summation of powers of all other n signals that use the same RB as node i as:

$$I_i = \sum_{\substack{k=1 \\ k\neq i}}^{n} \frac{P_{tx,k}}{L_{ki}} \tag{6}$$

From (4), the required transmission power P_{tx} for a single user i based on its SINR requirement α_i can be computed as:

$$P_{tx,i} = L_{ij} \times \alpha_i \left(I_i + N\right) \tag{7}$$

Powers for all co-channel users (users that are allocated the same RB), are computed using a linear equation system as:

$$\begin{bmatrix} \frac{1}{a_1 L_{11}} & -\frac{1}{L_{21}} & \cdots & -\frac{1}{L_{n1}} \\ -\frac{1}{L_{12}} & +\frac{1}{a_2 L_{22}} & \cdots & -\frac{1}{L_{n2}} \\ \vdots & \vdots & \vdots & \vdots \\ -\frac{1}{L_{1n}} & -\frac{1}{L_{2n}} & \cdots & +\frac{1}{a_n L_{nn}} \end{bmatrix} \begin{bmatrix} P_{tx,1} \\ P_{tx,2} \\ \vdots \\ P_{tx,n} \end{bmatrix} = \begin{bmatrix} N \\ N \\ \vdots \\ N \end{bmatrix} \tag{8}$$

Assuming that the matrices are **A**, **P**, and **B** respectively, this can be written as:

$$\mathbf{A} \cdot \mathbf{P} = \mathbf{B} \Longleftrightarrow \mathbf{P} = \mathbf{A}^{-1}\mathbf{B} \tag{9}$$

The allocation of powers for all co-channel users will be based on the results of the power vector **P**. The resulting power vector is considered achievable if:

$$P_{tx,min} < P_{tx,i} < P_{tx,max}, \quad \forall P_{tx,i} \in \mathbf{P} \tag{10}$$

III. PROPOSED ALGORITHM

The proposed relay scheme is based on applying K-means algorithm at the outer cell region where no co-channel interference from other cells is considered. K will be assumed to be equal to 33 as this is the maximum allocation for direct users in this region which can use 33 RBs (33% of all RBs). At the inner cell region, K-means will have low performance since we cannot assign any K nodes as CHs. As a result, some of the obtained clusters wouldn't have any connected users. This happens because of the high inter-cell interference in this area where not all locations are suitable for CH connectivity. Therefore, K-means will not be applied at the inner region. Instead, a selection strategy scheme will be used based on first allocating a number of nodes with best link quality as CHs and then connecting slaves to their nearest CH. Each user will be allocated a single RB. There should be room for 67 CHs at the inner region which means 67 clusters. However, since not all users can successfully connect, a smaller number of clusters will be obtained. A summary of used notations in the algorithm is given in Table 1.

TABLE 1. Summary of notations.

Parameter/function	Description
K	Number of clusters
$\mathbb{N}_{outer}=\{n_1,n_2,n_3,....n_{\lvert Nouter \rvert}\}$	Set of nodes in outer region.
$\mathbb{N}_{inner}=\{n_1,n_2,n_3,....n_{\lvert Ninner \rvert}\}$	Set of nodes in inner region.
$S=\{S_1,S_2,\ldots,S_i,\ldots.,S_k\}$	Set of clusters. S_i represents a set of all nodes in cluster i.
$V=\{v_1,v_2,\ldots\ldots.,v_k\}$	Set of clusters' centroids.
$C=\{ch_1, ch_{2,..} ch_{i,..} ch_{\lvert C \rvert}\}$	Set of CHs.
$RB_{inner}=\{r_1,r_2,\ldots.,r_{100}\}$	Set of RBs allowed for cell inner region
$RB_{outer}=\{r_{68},r_{77},\ldots.,r_{100}\}$	Set of RBs allowed for cell outer region
Resource_Allocation (a,b)	Function that allocates resources for user *a* transmitting to receiver *b*.

A. OUTER CELL REGION CLUSTERING

At the outer cell region, K-means will be used to cluster nodes based on their path-loss similarity. The objective is to decrease the path loss of the overall outer cell region:

$$\arg min \sum_{i=1}^{k} \sum_{n \in S_i} L_{n,v_i} \tag{11}$$

Where k is the total number of clusters, S_i is the set of nodes belonging to the i^{th} cluster, n is a node in the cluster S_i, and v_i is the centroid of the i^{th} cluster. Decreasing path loss will result in low power values, which maximizes the ability to reuse RBs. The proposed scheme for outer cell region clustering is detailed in Algorithm 1.

Algorithm 1 Outer Cell Region Clustering

Phase 1: K-means at outer cell region

1. Take K number of centroids initially at random places in the outer cell region.
2. Calculate the path loss from each node to all centroids.
3. Assign each node to the cluster whose path loss from cluster center is minimum of all cluster centers.
4. Recalculate the new centroid for each cluster using the mean of all points inside the cluster:

$$v_i = \frac{1}{|S_i|} \sum_{j=1}^{|S_i|} n_j \tag{12}$$

5. If there is change in position of any centroid then go to STEP 2, otherwise stop.

Phase 2: Allocating CHs and slaves

For each cluster S_i,

a. Choose the node n_j with least path loss to the centroid v_i, and assign it as the CH of this cluster (ch):

$Resource\_Allocation\ (n_j, BS)$

$ch \leftarrow n_i$

b. Sort remaining nodes in the cluster based on their path loss to ch, and store them in the set S_i.

c. For every node n in S_i:

Connect n to the CH (ch) :

$Resource\_Allocation\ (n, ch)$

In Phase 1, clustering is done for the outer cell region using K-means. K-means algorithm was chosen based on its lower complexity compared to the HAC algorithm. Phase 2, consists of allocating CHs and slaves. The overhead of choosing a CH is low since it only depends on choosing the node with least path loss to the BS.

B. INNER CELL REGION CLUSTERING

As mentioned before, selection strategy will be used in the inner cell region for maximizing direct link associations. After allocating all remaining resources for the direct connections to the BS as CHs, each unconnected user n will make a decision for which CH to use as a relay. The decision will be based on finding the minimum path loss value $L_{n,min}$:

$$L_{n,min} = \arg \underset{all\ ch \in C}{min} \{L_{n,ch}\} \tag{13}$$

Based on this, a user will be connected directly to the BS as a CH as long as there are available free RBs. When all RBs have been used once, a user may be only connected via a CH as a slave. The allocation is confirmed if there exist a suitable CH and an RB that could be reused successfully. Otherwise, the user will be blocked. It should be noticed that at the end, all blocked users from the outer cell region will be

Algorithm 2 Selection Strategy at Inner Cell

Phase 1: Assigning CHs

For each node n_i in $\mathbb{N}_{\text{inner}}$,

Try to connect n_i to the BS: *Resource_Allocation* (n_i, *BS*)

If successfully connected,

1. Assign n_i as a CH.
2. Update C.

If $|C| = 100$ ***or passed through all nodes,***

End Phase 1.

Phase 2: Associating slaves

For each unallocated node n_i in $\mathbb{N}_{\text{inner}}$,

1. Compute path loss between n_i and every ch in C.
2. Sort C ascendingly.
3. For each ch_j in sorted C,

If $|S_j| < max\_cluster\_size$,

Try to connect n_i to ch_j

Resource_Allocation (n_i, ch_j)

reconsidered for allocation as slaves to CHs in the inner cell region. The proposed scheme for inner cell region clustering is detailed in Algorithm 2.

C. RESOURCE ALLOCATION

This function works for both (UE-CH) and (CH-BS) links. In general, the algorithm inputs are annotated as transmitter a and receiver b. There are two groups of resources, RB_{outer}, which consists of 33 RBs to be used in outer cell region, and RB_{inner}, which consists of all 100 RBs to be used in inner cell region. Because resource allocation is done for the outer region first, CHs in that region will consume 33 RBs. Therefore, there will only be 67 RBs available for CHs in the inner region out of the 100. However, reusing is still permitted for 100 RBs. The function is described in Algorithm 3.

IV. SIMULATION SETUP AND RESULTS

In this section, an evaluation of the performance of the algorithm is done based on a MATLAB simulation model. The proposed scheme will be compared against the use of standalone K-means clustering and the selection strategy scheme. A high user density environment is assumed with low mobility. A total of 100 RBs are used at a single sub-frame, and the scheduling requirement will be fixed to 1 RB per user. The BS is centered at the middle of the cell with coverage of 100 meters. This will give us a total area of 31,400 m^2. A 20 MHz band is assumed with a total of 18 MHz effective bandwidth. Different densifications of users are tested assuming a maximum density of 1 user/m^2, which yields 31,400 people. Having 10% of them as actual users, gives a total of 3140 maximum users at a given time frame. The simulation will test a single sub-frame duration which consists of a maximum of approximately 300 users currently accessing resources in a round-robin fashion. Interference

Algorithm 3 Resource_Allocation (a, b)

If a ***is in outer cell region,***

Use RB_{outer} as available RBs.

Else if a ***is in inner cell region,***

Use RB_{inner} as available RBs.

1. Sort available RBs based on least reused.
2. Take out RBs that have been already acquired by other users sending to b.
3. for every RB,
   a. Collect all co-channel users (cell/first tier).
   b. ***If no any co-user,***
      Compute power based on (7)
      Else,
      Do mathematical system (8).
   c. Based on (9), if power vector is achievable, then connect the user and exit. Otherwise, do not connect and continue trying other RBs.

TABLE 2. Parameters used in simulation.

Parameter	Value
Cell radius	100 m
Cell area	31,400 m^2
Effective Bandwidth	18 MHz
No. of Resource Blocks	100
Number of nodes |N| per sub-frame	100, 150, 200, 250, 300
Number of iterations	100
Max cluster size	5
Path-loss exponent	4
Shadowing	6 dB
Position of BS (x,y)	(0,0)
Noise	-121 dBm [26]
α	10 dB
Th_{min}	15 Kbps (VoLTE [27])
$P_{\text{tx, ,max}}$	26 dBm [26]
$P_{\text{tx, min}}$	-50 dBm [26]
K (no. of clusters in K-means)	33
Duplexing	FDD
Modulation scheme	64 QAM

from the first tier cells is assumed with a total of 100 active users in each. All provided results are the averages of 100 iterations. Table 2 lists the main parameters of the simulation. Moreover, below are some assumptions and considerations:

- The simulation environment is composed of a single cell with applied relays, and 6 conventional LTE cells at the first tier.
- The BSs stores users' information.
- All BSs are fixed and located at the center of each cell.
- User nodes are randomly distributed in each cell area.
- Users have limited mobility due to high user density (e.g. densely crowded events).
- Simulation is done for 1 ms duration (single sub-frame) acquiring 10% of all active users.
- 1 RB in a single sub-frame duration is allocated for each user per frame.
- VoLTE (Voice Over LTE) service is assumed with a throughput requirement of 15 Kbps [27].

- Based on LTE maximum throughput [28], the maximum throughput for a single RB in a single sub-frame is estimated as 75 Kbps.
- The maximum cluster size is set to be 5 based on having 75 Kbps as the maximum throughput and 15 Kbps as per connection throughput requirement.

The simulation is tested for four different schemes:

A. Path Loss Relay Selection (Whole Cell)

In this scheme, the LTE center cell uses relaying; all direct users are assigned first as CHs. Then, unallocated users will try connecting to CHs sequentially based on minimum path loss (similar to what was described in Algorithm 2 but throughout the whole cell). This is denoted in the graphs as Selection Scheme.

B. K-Means Clustering (Whole Cell)

The LTE center cell applies relaying using the same described K-means but throughout the whole cell area and with the assumption of $K = 100$. This is denoted in graphs as K-means.

C. Proposed (Hybrid)

The paper proposed scheme using both K-means and selection strategy in the center cell. Denoted in graphs as Proposed Scheme.

D. Conventional LTE (No Relays)

LTE center cell has conventional direct connectivity with no relaying.

All four schemes are applied in multi-cell scenarios as the model given in Fig. 2 with the surrounding 6 cells having conventional LTE with no relays.

The simulation is done for different numbers of users ranging between 100 and 300 with a total of 100 different iterations for each. The results show an average blocking probability of 8% for 300 users in the proposed scheme compared to 10% and 13% in K-means and selection strategy schemes respectively as shown in Fig. 4. The proposed scheme enhances blocking probability by 5% compared to the selection strategy. This enhancement is due to the fact that adding k-means at outer cell region improves the cluster formation in that area to have less power and therefore allocate more users. Also, the proposed scheme enhances the blocking probability by 2% compared to the K-means scheme. The K-means has less performance because of its limitation at the inner-cell region where inter-cell interference is dominant. As a consequence, it will fail to have valid cluster formation where some of the resulting clusters have no suitable CH with sufficient SINR to carry any of its cluster members' data. Therefore, it is better to exploit all the possible CH locations first as done in the proposed scheme. Moreover, as a comparison to conventional LTE without relays as shown in Fig. 3, the proposed scheme enhances the blocking probability by 65%. Fig. 5 shows the average user power for the different schemes. It can be shown that the proposed scheme and K-means scheme both require lower power than the selection strategy scheme since the applied K-means algorithm is based on minimizing path loss which results in lower power. It can be seen from Fig. 4 and Fig. 5 that the selection strategy has the worst performance in terms of both power and served capacity. Total throughput of the center cell reaches 4.1 Mbps in the proposed scheme compared to 3.9 Mbps in Selection Scheme and 1.5 Mbps only in conventional LTE.

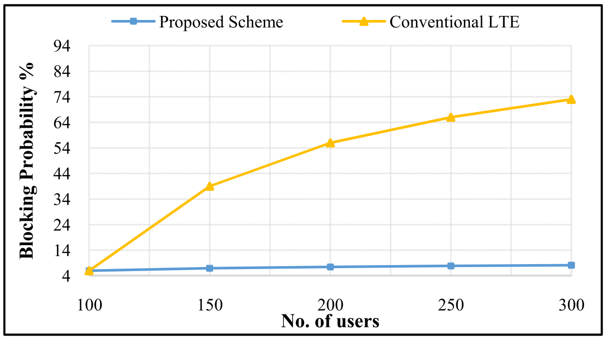

FIGURE 3. Blocking probability of proposed scheme compared to Conventional LTE.

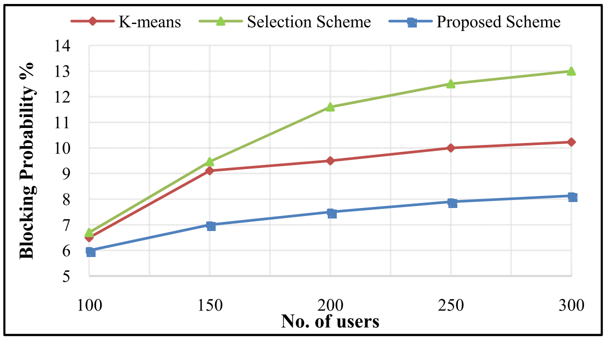

FIGURE 4. Blocking probability of proposed scheme compared to Selection and K-means schemes.

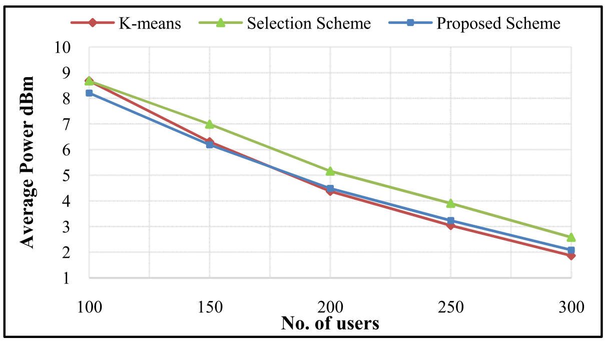

FIGURE 5. Average power of proposed scheme compared to Selection and K-means schemes.

V. CONCLUSIONS

A hybrid clustering technique was proposed for multi-cell high user density scenarios in LTE networks. The proposed scheme is based on using K-means at the outer cell region and selection strategy in the inner cell region. The scheme shows improved performance in terms of power and capacity compared to the use of each method alone in the entire cell area. Also it proves that capacity enhancement through relaying is also achievable in multi-cell scenarios with inter-cell interference. The proposed scheme enhances blocking probability 5% compared to the selection strategy scheme, 2% compared to the K-means scheme, and 65% compared to the conventional LTE.

ACKNOWLEDGMENT

The authors acknowledge with gratitude the Science and Technology Unit in King Abdulaziz University for their technical support.

REFERENCES

[1] H. Nourizadeh, S. Nourizadeh, and R. Tafazolli, "Performance evaluation of cellular networks with mobile and fixed relay station," in *Proc. IEEE 64th Veh. Technol. Conf. (VTC)*, Sep. 2006, pp. 1–5.

[2] T. S. Madhulatha. (2012). "An overview on clustering methods." [Online]. Available: https://arxiv.org/abs/1205.1117.

[3] J. Y. Yu and P. H. J. Chong, "A survey of clustering schemes for mobile ad hoc networks," *IEEE Commun. Surveys Tut.*, vol. 7, no. 1, pp. 32–48, 1st Quart., 2005.

[4] P. Sasikumar and S. Khara, "k-means clustering in wireless sensor networks," in *Proc. 4th Comput. Intell. Commun. Netw. (CICN)*, 2012, pp. 140–144.

[5] H. Tabrizi, G. Farhadi, and J. M. Cioffi, "CaSRA: An algorithm for cognitive tethering in dense wireless areas," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2013, pp. 3855–3860.

[6] M. Hajjar, N. Dimitriou, and G. Aldabagh, "Using clustering techniques to improve capacity of LTE networks," in *Proc. Asia–Pacific Conf. Commun. (APCC)*, Oct. 2015, pp. 68–73.

[7] R. Evans, B. Pfahringer, and G. Holmes, "Clustering for classification," in *Proc. 7th Int. Conf. Inf. Technol. Asia*, Jul. 2011, pp. 1–8.

[8] P. G. Yong, K. Heeseong, J. H. Woon, and Y. H. Yong, "A novel cluster head selection method based on k-means algorithm for energy efficient wireless sensor network," in *Proc. 27th Int. Conf. Adv. Inf. Netw. Appl. Workshops (WAINA)*, 2013, pp. 910–915.

[9] H. Harb, A. Makhoul, D. Laiymani, A. Jaber, and R. Tawil, "K-means based clustering approach for data aggregation in periodic sensor networks," in *Proc. IEEE 10th Int. Conf. Wireless Mobile Comput. Netw. Commun. (WiMob)*, Oct. 2014, pp. 434–441.

[10] G. Kumar, H. Mehra, A. R. Seth, P. Radhakrishnan, N. Hemavathi, and S. Sudha, "An hybrid clustering algorithm for optimal clusters in wireless sensor networks," in *Proc. IEEE Students' Conf. Elect. Electron. Comput. Sci. (SCEECS)*, Mar. 2014, pp. 1–6.

[11] T. Liansheng, G. Yanlin, and C. Gong, "A balanced parallel clustering protocol for wireless sensor networks using k-means techniques," in *Proc. 2nd Int. Conf. Sensor Technol. Appl. (SENSORCOMM)*, 2008, pp. 300–305.

[12] A. Talgini, V. Shakarami, F. Sheikholeslam, and A. Chatraei, "Aerial node placement in wireless sensor networks using fuzzy k-means clustering," in *Proc. 8th Int. Conf. E-Commerce Developing Countries, With Focus E-Trust (ECDC)*, Apr. 2014, pp. 1–7.

[13] S. Periyasamy, S. Khara, and S. Thangavelu, "Balanced cluster head selection based on modified k-means in a distributed wireless sensor network," *Int. J. Distrib. Sensor Netw.*, vol. 12, no. 3, p. 11, Feb. 2016.

[14] J. Zhou, Y. Zhang, Y. Jiang, C. L. P. Chen, and L. Chen, "A distributed k-means clustering algorithm in wireless sensor networks," in *Proc. Int. Conf. Informat. Cybern. Comput. Social Syst. (ICCSS)*, 2015, pp. 26–30.

[15] J. Chang and Y. Chen, "A relay station deployment with particular k-means strategy for multi-hop relay network systems," presented at the TANET, 2013.

[16] M. N. Murty and V. S. Devi, *Pattern Recognition: An Algorithmic Approach*. Berlin, Germany: Springer, 2011.

[17] D. Davcev and J. M. Gómez, *ICT Innovations*. Berlin, Germany: Springer, 2010.

[18] C. H. Lung and C. Zhou, "Using hierarchical agglomerative clustering in wireless sensor networks: An energy-efficient and flexible approach," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, Nov. 2008, pp. 1–5.

[19] T. K. Jain, D. S. Saini, and S. V. Bhooshan, "Performance analysis of hierarchical agglomerative clustering in a wireless sensor network using quantitative data," in *Proc. Int. Conf. Inf. Syst. Comput. Netw. (ISCON)*, 2014, pp. 99–104.

[20] M. Hajjar, "Using clustering techniques to improve capacity of LTE networks," M.S. thesis, Master Comput. Sci., King Abdulaziz Univ., Jeddah, Saudi Arabia, 2016.

[21] H. Nourizadeh and R. Tafazolli, "Capacity improvement of WCDMA cellular system through different relaying strategies," in *Proc. Int. Conf. Wireless Mobile Commun. Netw.*, 2005, pp. 3–6.

[22] V. Sreng, H. Yanikomeroglu, and D. D. Falconer, "Relayer selection strategies in cellular networks with peer-to-peer relaying," in *Proc. IEEE 58th Veh. Technol. Conf. (VTC)*, Oct. 2003, pp. 1949–1953.

[23] H. Wei, H. Jing, and W. Haiming, "UL performance evaluation of relay enhanced FDD LTE-advanced networks," in *Proc. 7th Int. Conf. Wireless Commun. Netw. Mobile Comput. (WiCOM)*, Sep. 2011, pp. 1–5.

[24] W. Li-Chun, S. Wen-Shan, H. Jane-Hwa, A. Chen, and C. Chung-Ju, "Optimal relay location in multi-hop cellular systems," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Apr. 2008, pp. 1306–1310.

[25] Y. Yiwei, E. Dutkiewicz, H. Xiaojing, M. Mueck, and F. Gengfa, "Performance analysis of soft frequency reuse for inter-cell interference coordination in LTE networks," in *Proc. Int. Symp. Commun. Inf. Technol. (ISCIT)*, Oct. 2010, pp. 504–509.

[26] H. Holma and A. Toskala, *LTE for UMTS: Evolution to LTE-Advanced*. Hoboken, NJ, USA: Wiley, 2011.

[27] L. Norell, "Telephony services over LTE end-to-end," Ericsson, Stockholm, Sweden, Tech. Rep.34, 2010.

[28] F. Rezaei, M. Hempel, and H. Sharif, "LTE PHY performance analysis under 3GPP standards parameters," in *Proc. IEEE 16th Int. Workshop Comput. Aided Modeling Design Commun. Links Netw. (CAMAD)*, Jun. 2011, pp. 102–106.

MARYAM HAJJAR was born in Riyadh, Saudi Arabia, in 1985. She received the B.S. and M.S. degrees in computer science from King Abdulaziz University (KAU), Jeddah, Saudi Arabia, in 2008 and 2016, respectively.

She is currently a Research Assistant with the Faculty of Computing and Information Technology, KAU. She is also a Teacher Assistant with the Computer Skills Unit, KAU. Her current research interests include relay deployment and selection in wireless networks, high density networks, and pattern recognition.

GHADAH ALDABBAGH received the B.Sc. degree in computer engineering from the College of Engineering, University of Illinois at Urbana-Champaign, Champaign, IL, USA, in 1991, the M.Sc. degree in data communication networks and distributed system from the Computer Science Department, University College London (UCL), U.K. She received the Ph.D. degree from the Department of Electronic and Electrical Engineering, UCL, in 2010. She worked on several European Union projects in Grid Computing. From 2013 to 2014, she joined Professor John M. Cioffi's Dynamic Spectrum Management Research Group as a Visiting Professor, Stanford University. In 2016, she joined the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, as a Research Associate where she was involved on the topic of location-aware ad hoc sensor networks. She holds an associate professor position with the Computer Science Department, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia. Her research interests are mainly in the area of communication systems, wireless communications, inter-domain QoS, network measurement, complexity issues, performance evaluation, communication and information theory protocols and algorithms, and networking theory. In collaboration with Professor Moe Win's research Group, her current focus is to develop scalable, distributed, and energy efficient techniques for scheduling and routing in ad hoc sensor networks with localization-awareness.

NIKOS DIMITRIOU (SM'11) received the Diploma degree in electrical and computer engineering from the National Technical University of Athens, Greece, in 1996, and the M.Sc. degree (Hons.) in mobile and satellite communications and the Ph.D. degree in mobile communications from the University of Surrey, U.K., in 1997 and 2001, respectively. Since 2001, he has been actively involved in numerous research projects in the area of Wireless Communications via his affiliations with the Institute of Accelerating Systems and Applications, the National and Kapodistrian University of Athens, the National Center of Scientific Research Demokritos, the Institute of Informatics and Telecommunications, Greece, and the Computer Science Department, King Abdulaziz University, Saudi Arabia. His current research interests include radio resource management for multi-tier HetNets, energy efficient mobility management in composite wireless networks, and robust routing for mobile-airborne ad-hoc networks.

MOE Z. WIN (S'85–M'87–SM'97–F'04) received both the Ph.D. in Electrical Engineering and the M.S. in Applied Mathematics as a Presidential Fellow at the University of Southern California (USC) in 1998. He received the M.S. in Electrical Engineering from USC in 1989 and the B.S. (*magna cum laude*) in Electrical Engineering from Texas A&M University in 1987.

He is a Professor at the Massachusetts Institute of Technology (MIT) and the founding director of the Wireless Information and Network Sciences Laboratory. Prior to joining MIT, he was with AT&T Research Laboratories for five years and with the Jet Propulsion Laboratory for seven years. His research encompasses fundamental theories, algorithm design, and experimentation for a broad range of real-world problems. His current research topics include network localization and navigation, network interference exploitation, intrinsic wireless secrecy, adaptive diversity techniques, and ultra-wideband systems.

Professor Win is an elected Fellow of the AAAS, the IEEE, and the IET, and was an IEEE Distinguished Lecturer. He was honored with two IEEE Technical Field Awards: the IEEE Kiyo Tomiyasu Award (2011) and the IEEE Eric E. Sumner Award (2006, jointly with R. A. Scholtz). Together with students and colleagues, his papers have received numerous awards, including the IEEE Communications Society's Stephen O. Rice Prize (2012), the IEEE Aerospace and Electronic Systems Society's M. Barry Carlton Award (2011), the IEEE Communications Society's Guglielmo Marconi Prize Paper Award (2008), and the IEEE Antennas and Propagation Society's Sergei A. Schelkunoff Transactions Prize Paper Award (2003). Highlights of his international scholarly initiatives are the Copernicus Fellowship (2011), the Royal Academy of Engineering Distinguished Visiting Fellowship (2009), and the Fulbright Fellowship (2004). Other recognitions include the International Prize for Communications Cristoforo Colombo (2013), the *Laurea Honoris Causa* from the University of Ferrara (2008), the Technical Recognition Award of the IEEE ComSoc Radio Communications Committee (2008), and the U.S. Presidential Early Career Award for Scientists and Engineers (2004).

Dr. Win was an elected Member-at-Large on the IEEE Communications Society Board of Governors (2011–2013). He was the Chair (2005–2006) and Secretary (2003–2004) for the Radio Communications Committee of the IEEE Communications Society. Over the last decade, he has organized and chaired numerous international conferences. He is currently serving on the advisory board of the IEEE Communication Letters. He served as Editor-at-Large (2012–2015) for the IEEE Wireless Communications Letters, as Editor (2006–2012) for the IEEE Transactions on Wireless Communications, and as Area Editor (2003–2006) and Editor (1998–2006) for the IEEE Transactions on Communications. He was Guest-Editor for the Proceedings of the IEEE (2009) and for the IEEE Journal on Selected Areas in Communications (2002).