An Improved Modelling of User Clustering for Small Cell Deployment in Heterogeneous Cellular Network

Joyatri Bora1 · Md. Anwar Hussain1

Accepted: 28 May 2022 / Published online: 25 June 2022
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract
Users in practical cellular geographical areas are found to be non-uniformly distributed. Small cell (SC) deployments in heterogeneous user distribution in a cellular geographical area help to meet high data rate user demands for multimedia data communications in hot spots. SCs help to offload traffic burden from the macro cell (MC) base station, and also cater the data traffic need for the edge users where signal strength from the MC base station (BS) is very weak. For deployments of SCs along with the central MC BS (hence called HetNet) in such spatial heterogeneous user distribution, effective user grouping or clustering algorithm is required for appropriate and satisfactory service coverage. We call it service grouping or clustering of users to be put under a SC for data transmission and reception. It does not disturb the spatial positions of users in clustered non-uniform distribution. Efficient grouping or clustering of users and then deploying a SC at optimal location enhances the performance of the HetNet. It is found that the K-means algorithm used for such grouping of users to position SCs is not efficient. A novel and improved user grouping algorithm is proposed in this paper which performs much better compared to the k-means algorithm. The proposed algorithm of modelling of user clustering results in increase in the number of users under SCs, increase in more offloading of data traffic from MC BS thereby increasing data throughput of MC users. The algorithm also increases in the energy efficiencies of the SCs which is considered as one important performance metric. A doubly stochastic poison process (DSPP), also called Cox process, is assumed here for simulation of non-uniform user distributions. We consider Rayleigh distributed small scale fading model, large scale fading factor representing shadow fading, and users’ geographical distances from BSs to evaluate users’ data rates.

Keywords Heterogeneous · Clustering · K-means algorithm · Doubly stochastic poison process · Shadow fading

Joyatri Bora bjoyatri@yahoo.com
Md. Anwar Hussain bubuli_99@yahoo.com

1 Electronics and Communication Engineering Department, North Eastern Regional Institute of Science and Technology, Nirjuli, Arunachal Pradesh 791109, India
1 Introduction

As is prevalent now, for the coverage of mobile phone users a macro cell base station is positioned in the central point of a geographical coverage area (Macro cell). To enhance capacity in the busy areas (hotspots), pico-cell and micro cell base stations which are generally termed as small cells are used at suitable points in the geographical coverage area. Small cells are also used to provide coverage to users at the cell edge where the received signal strength from the macro base station is very weak. Such a cellular communication system is termed as HetNet.

Presently, cellular mobile data communication has far grown than that expected in the last decade of the twenty first century [1]. The extensive growth of data volume generated due to ever increasing number of smart phone users has inspired the use of larger spectrum and more penetration of sophisticated mobile data communication systems in rural areas in addition to the urban usages. Also, the user distribution in a given geographical area is now more non-uniformly distributed than before, with number of hotspots of users where much activity is recorded and data volume generated significantly higher. Also, in areas near the edge of the geographical area, where signal strength from the center BS is very weak, cellular service coverage is needed [2, 3]. Instead of a single BS at the center for coverage of all the users, SCs are also deployed at certain points like the centers of the hotspots or at the cell edge for enhanced throughput or data rate performance of the total cellular data communication system [4–7]. SCs not just offload the traffic from a MC BS in the cell but also boost link quality by reducing the gap of communicating distance, and permit to utilize the spectrum in an efficient manner [6, 7]. In [8], a novel HetNet planning model is proposed that involves deployment of MC and SCs considering least energy consumption and minimum total cost ownership (TCO) that fulfil the traffic demand. While deploying SCs, factors that are given importance are power and bandwidth allocation, traffic requirement, and spectral efficiency. Three heuristic procedures are adopted along with clustering approaches in [9] for joint deployment of SCs and point-to-point fibre links that meets the minimum level of QoS requirement as well as reducing cost of transport and SCs deployment. An observer-based system is used dividing a cell into a series of sectors and identifying the regions of mobile users’ clusters for finding out the optimum location of deployment of SCs [10]. In order to cope with dense deployment scenario for heavy traffic requirements, SCs within a MC are clustered with grouping of 1, 4 or 10 based on 3GPP simulation methodology [11, 12]. Hierarchical clustering algorithm with minimax linkage [13] is used to cluster BSs creating virtual cells based on best channel condition and user affiliation to a BS [14]. A Green Offloading scheme is proposed for offloading traffic from MCs based on Reverse Auction Model [15]. The first price sealed bid mechanism which matches with mobility based offloading process [16] is found to improve system energy efficiency of densely deployed HetNet under the constraints of user QoS requirements, bandwidth and transmission power limitations. Nowadays, a promising technology of 5G is hyper-dense SC deployed HetNet. However, with increased number of large SC BSs the level of power consumption also increases. Therefore, SC BSs is operated according to traffic load variations. During the day when user density is high SC BSs are activated gradually to serve offloaded traffic of MC BS and operated in off mode when traffic is low thus reducing total power consumption of the HetNet [17].

Stochastic Modelling of the user distribution and deployments of SCs for network performance evaluation are now prevalent. Cox process or cluster model is used to model the non-uniform distribution in a cellular area [5, 18–22]. To divide the cellular area in
a number of clusters, \textit{k-means} clustering technique with Voronoi cells may be used [6]. \textit{SC BSs} may now be placed at the centroids of the cells, and the \textit{BS} caters the data service required by the users with a specific circular area of the \textit{SCs}. Similarly, \textit{SC BSs} may be located judiciously near the edge of the \textit{MC}. However, the \textit{k-means} clustering technique has certain defects such as within the circular area of some \textit{SC BSs}, there may be no user at all or the numbers of users are very small. The authors in [23] discuss the impact of \textit{SCs} on the throughput enhancement in \textit{HetNets} and have shown the limitation of \textit{k-means} user clustering approach. It is obvious that the \textit{SC BSs} should be located in such locations where the users’ density is maximum within its \textit{SC} area. This results in offloading of data traffic from the \textit{MC BS} maximum possible which is a requirement in cellular communication design of \textit{HetNets}. Also, by bringing more and more users to the coverage of \textit{SCs} enhances the data throughput of those users [24, 25]. The reduction of power consumption in the \textit{SCs} is a prominent issue. Authors in [26] deal this issue and proposed a novel algorithm based on traffic demands and dynamic switching of redundant \textit{SCs}. \textit{HetNet} based 5G networks are now popular for increased connectivity and capacity. Interference mitigation in such networks is an important requirement. Authors in [27] proposed particle swarm optimization technique for balancing load with interference mitigation. As said above, the interference phenomenon largely affects the capacity and connectivity of 5G networks, authors in [28] proposed a soft frequency reuse scheme for mitigation of interferences and thereby enhancing the throughput. Massive MIMO and mmWave based technologies are suitable candidates for applications in future 5G \textit{HetNet} developments. Authors in [29] discusses potential benefits and challenges of such technologies. Maximization of total system throughput is explored in [30] for 5G \textit{HetNet} multiuser selection in MIMO-OFDMA based system.

As mentioned in the beginning of this section and in the abstract, deployment of small cells along with a central macro base station provide enhanced connectivity and enhanced system throughput, enhanced data offloading and efficient power utilizations. The design of \textit{HetNet} is flexible in the sense that small cells may be positioned based on various requirements. One such requirement is to cover as many numbers of mobile users as possible by a small cell within its coverage area. This provides higher offloading of user traffic from the macro base station which enhances the data rate of macro users under the coverage of the macro base station. Placement of small cells at optimal points in a geographical coverage area is of high demand as it provides highest user data offloading from the macro base station. Also, it caters the need to provide high data rate to the users in hotspots, and services to the edge users.

In this paper, we assume a \textit{HetNet} with a \textit{MC BS} serving the macro users and a few \textit{SCs} deployed in the macro geographical area serving users in cluster locations. \textit{SCs} are placed at the centroid of the clusters of users obtained applying \textit{k-means} algorithm and the proposed algorithm obtained by improving upon the \textit{k-means} algorithm. We obtain the total data throughputs of users under the \textit{MC BS} and under all the \textit{SCs} separately, for both the algorithms. The proposed algorithm increases data throughput of macro users, and also increases total number of users under the \textit{SCs}, which are considered as metrics to show the efficiency of the proposed algorithm. The effective throughput offloading, \textit{OFFLOAD} and energy efficiency of \textit{SC BSs}, \textit{BS\_SC\_W} are two other metrics used for performance evaluation. We consider different cluster scenarios for evaluation of the proposed algorithm. The user spatial heterogeneity is simulated using \textit{Cox} process and the \textit{MC BS} is assumed to be in the center of a circular geographical area.

The rest of the paper is organized as follows: We describe in Sect. 2 the system model and modelling of users resulting in the improved user grouping and obtaining optimum
location to deploy SCs. In Sect. 3, the metrics are formulated for evaluation of the new improved modelling of user clustering. Section 4 describes the simulation studies and discussions. Section 5 concludes the findings of this paper.

2 System Model

We consider a circular geographical area of radius 1500 m which is filled by a non-uniform Cox process that generates users as clusters spatially. A MC BS is positioned at the center and a few SCs of radius 100 m are located in the clusters of users of the geographical area. The users are taken in different numbers in clusters, the Centroid or Centre of which is obtained with k-means clustering algorithm. An example HetNet is shown in Fig. 1, wherein the red outer circular area is the MC and the red pyramid is the MC BS. The small red circular areas are serviced by SC BS which are shown as green pyramids, the solid small circles represent the SC locations obtained by k-means algorithm, and the dashed red circles are the new locations of SCs obtained by applying the proposed algorithm. The black arrow shows the shifting in the location of a SC, and the black dots are the non-uniformly distributed users.

The data traffic of the users under the SCs is serviced by the SC BSs and the data traffic of the remaining users is serviced by the MC BS. The interference model assumed in this paper considers interference between MC BS to different SC BSs, between SC BSs, and SC BSs to MC BS. We propose an improved algorithm for modelling user clustering which performs better than the k-means algorithm. The improvement is shown by metrics that

Fig. 1 An Example HetNet
quantifies the increase in the number of SC users and the increase in the data throughput of the remaining macro users. We also quantify the offloading of data traffic from the MC BS, and also the energy efficiency of the SCs to be used as performance metrics.

It may be noted that the users are spatially clustered because of their non-uniform distribution in the coverage area. In this paper, our grouping or clustering algorithm by modelling of user spatial clustering is to find optimal locations of the SCs for deployment for efficient service coverage in the HetNet. The grouping or clustering of users, while explaining the proposed algorithm is merely to collect users in groups or clusters for downlink data transmission and reception from respective SCs, users’ spatial geographical locations remain unaltered. Hence, the proposed algorithm may be taken to perform service clustering only. Further we considered a static case of user distribution with no mobility of users in the simulation experiment.

We state below the $k$-means clustering algorithm and the proposed improvement in the modelling of users’ clustering from the $k$-means algorithm.

The basic $k$-means algorithm is stated as below [6]:

```
```
Algorithm 1:

a. From the non-uniformly distributed $n$ user points $N_n$, select $k$ random user points as initial centroids.

b. Repeat
   i. Construct $k$ clusters by assigning users closest to its centroid
   ii. Recompute the centroid of each of $k$ clusters.

c. Continue b until centroids do not change.

The improvement of the above algorithm is obtained as stated below.

d. Start with the centroids from the step c above.

e. Draw circles of radius 300 meters around each centroid. Size of the radius is chosen to have no-overlapping or minimal overlapping between the nearby circles around the centroids. It may be a design parameter.

f. Find which user point within 300 meters of radius of each cluster has highest number of neighbors within its 100 meters of radius.

g. The user point having highest number of neighbors around it in 100 meters of radius is taken as the new cluster centroid.

h. Repeat for every cluster from basic k-means algorithm.

The algorithm for the improved service clustering of the users is stated below.
Algorithm 2: Improved Modelling for Service Clustering of users:
BEGIN
1. Draw a circle $A$, circle $(A, R)$, where $A$ is the centre and $R$ is the radius.

2. $Z = \{N1, N2, N3, \ldots, Nn\}$, where $Z$ is the set of $Nn$ users inside the circle $(A, R)$

3. $k\leftarrow$ Random points inside the circle $(A, R)$.

4. For $i=1$: $k$ do

5. Draw a circle $Bi$, circle $(Bi, 300)$ inside the circle $A$, around each random point $i$, of radius $300$ m.

6. EndFor

7. For $i =1$: $k$ do

8 $NUi\leftarrow$ Number of users inside the circle $Bi$.

9. For $j=1$: $NUi$ do

10. Draw a circle $Cj$, circle $(Cj, 100)$ inside the circle $Bi$, of radius $100$ m and center $Cj$ around each user $j$

11. $NUCj\leftarrow$ Number of users inside the circle $Cj$

12 EndFor

13. new random point $k; \leftarrow$ $max (NUC)$, the user within the circle $(Bi, 300)$ having the largest number of users within its 100 m radius.

14. EndFor

END

The benefits of the proposed algorithm are as follows:

i. The SCs which centers are the centroids from the basic k-means algorithm may not contain any user within its 100-m cell radius. Also, the centroids may be located in zones where users are sparsely spread.
The SC centers as obtained from the proposed algorithm contain the maximum number of user points within its 100-meter radius. Hence, the SC centers are located in the densest sites of each cluster in proposed improved algorithm obtained from basic k-means algorithm.

ii. The offloading of traffic from the MC BS is much less as the SCs obtained by k-means algorithm contain no or much smaller number of user points. Since, the SCs as obtained from the proposed algorithm have the maximum possible number of user points in each SC, the offloading of traffic from the MC BS is highest possible.

iii. As offloading of data traffic from MC BS is highest possible, it enhances the data throughput and data rate obtained by each macro users from its MC BS.

To obtain the x number of SC locations (centroids) in the MC BS coverage area, we proceed in two different methods. In the first method, we obtain x number of groups or clusters of users and their centroids applying k-means and the proposed algorithm. We then evaluate the performance of the HetNet designed under the two algorithms. In the second method, applying the two algorithms, we start with y number of SCs in the MC BS coverage area and then switch off some SCs by z numbers, where $z = 1, 2, 3, 4$ and keeping $y-z$ SCs in their original positions as per the algorithm. We then evaluate the performance of the HetNet designed under the two algorithms.

For simulation studies, we consider four different cellular HetNets with total user points in the range of 1000–1400, each one with one MC BS and 4–8 SC BSs. The SCs are distanced from the MC BS by 400 m so that interference from the MC BS to users under the SCs are low. In Fig. 2a and b we show such a cellular HetNet with 4 SCs and 7 SCs respectively, after applying the proposed algorithm. In Fig. 2c we show the network of Fig. 1b with 3 SCs switched off randomly. In Fig. 3a, b and c, show another network with same number of SCs.

The other two networks, considered for our study have been shown in the result and discussion in Fig. 4 and in Fig. 5. For throughput and data rate calculations, we consider Rayleigh distributed small scale fading model, large scale fading factor representing shadow fading, and user distances are considered as user points are at different distances from their servicing BSs.

3 Theoretical Formulation of Metrics

In this Section, we develop four metrics, PI_MUDR, PI_SC, OFFLOAD, PG, and BS_SC_W for evaluating the performance of the proposed modelling of user clustering.

The downlink signal to interference and noise ratio (SINR) for the jth user under the service of the MC BS (only one MC BS), and the kth user under the service of the ith SC BS is as follows.

A. Downlink SINR for the jth macro user, interferences from all SCs:

$$\text{SINR}_j = \frac{\left( P_j^m |h_j|^2 g_{pl,j} \right)}{\left( \sum_{i=1}^{N_s} P_i^s |h_{js}^i|^2 g_{pl,j}^i + N_0 \right)}$$  \hspace{1cm} (1)

The numerator in Eq. 1 stands for jth MC user signal from its serving MC, the first term in the denominator stands for the total interference from all $N_s$ SCs to the jth MC user.
user. $P_j$ is received power, $h_j$ is the small scale fading gain, and $g_{pl,j}$ is the path loss at distance $d$ from the MC BS. $P_j'$ is the power received, $h_j'$ is small scale fading gain, and $g_{pl,j}'$ is the path loss from the $i^{th}$ SC BS to the $j^{th}$ macro user. $N_0$ stands for Gaussian noise power in W Hertz, $N_S$ is total number of SCs. The path loss model is defined as:

$$g_{pl,j} = cd_j^{-a} \quad \text{and} \quad g_{pl,j}' = cd_{ij}^{-a}$$

(2)

where, $\alpha$ is the path loss exponent, $c$ represents large scale fading factor representing shadow fading, $d_j$ and $d_{ij}$ are the distances of the $j^{th}$ user from the MC BS and the $i^{th}$ SC respectively.

The data rate of the $j^{th}$ MC user with bandwidth $W$ Hertz is given by

$$\gamma_j = W \log_2(1 + SINR_j)$$

(3)

The sum data rate of a total $n_m$ users under the MC BS is given by:
The numerator of Eq. (5) stands for user received signal from its serving SC. In the denominator, the first term stands for the interference signal from the MC BS, and the second term stands for the interference signals from the rest $N_s - 1$ SCs. $P_k$ is received power, and $h_k^i$ is the small scale fading gain and $g_{pl,k}^i$ is the path loss, of the $k^{th}$ user in the $i^{th}$ SC with respect to the MC BS. $P_{s,k}^i$ is the power received, $h_{s,k}^i$ is small scale fading gain, and $g_{pl,k}^i$ is the path loss.
An Improved Modelling of User Clustering for Small Cell Deployment…

13

\( g_{pl,k} \) is the path loss, of the \( k \)th user in the \( i \)th SC. And \( P_{s,k} \) is the received power, \( h_{s,k} \) is small scale fading gain, and \( g_{pl,k} \) is the path loss, of the \( k \)th user in the \( i \)th SC with respect to the \( l \)th SC. \((N_s-1)\) is the number of interfering SCs, \( g_{pl,k}, g_{pl,i} \) and \( g_{pl,k} \) follow the definition in Eq. 2.

The data rate of the \( k \)th user in the \( i \)th SC with bandwidth \( W \) Hertz is given by:

\[
\gamma_k^i = W \log_2(1 + SINR_k^i) \tag{6}
\]

The sum data rate of \( n_i \) users under the \( i \)th SCs is given by:

\[
SR_s^i = \sum_{k=1}^{n_i} \gamma_k^i
\]

The sum data rate of all users under total \( N_s \) number of SCs is given by:

Fig. 4 a: HetNet 3, 4 SCs b: HetNet 3, 7 SCs. c: HetNet 3, 4 SCs. 3 SCs removed randomly from a total of 7 SCs.
We can define the percentage increase of total SC users under all SCs, $PI_{SCU}$ as defined below. Let, $n_m^i$, $n_m^*$ be the total number of macro users when clusters are constructed by k-means clustering algorithm and the proposed algorithm, respectively. Let, $\sum_{i=1}^{N_s} n_s^i$ and $\sum_{i=1}^{N_s} n_s^*$ be the total number of SC users under a total $N_s$ number of SCs, following basic (k-means) and proposed clustering algorithms.

$$SR_s = \sum_{i=1}^{N_s} SR_i^s \quad (7)$$

Similarly, the percentage increase of data rates of the macro users, $PI_{MUDR}$ due to the proposed clustering algorithm is defined as:

$$PI_{SCU} = \left( \frac{\sum_{i=1}^{N_s} n_s^i - \sum_{i=1}^{N_s} n_s^*}{\sum_{i=1}^{N_s} n_s^i} \right) \times 100 \quad (8)$$
Both the parameters $PI_{MUDR}$ and $PI_{SCU}$ are used in this paper to evaluate the performance of the proposed clustering algorithm.

The amount of data traffic difference caused by $N_s$ SCs from the MC BS is defined as:

$OFLD = (SM_{DR_B}) - (SM_{DR_A})$  \hspace{2cm} (10)

where, $OFLD$ is the data traffic difference, $SM_{DR_B}$ is sum data rate of total macro users before SCs are deployed, and $SM_{DR_A}$ is sum data rate of total macro users after SCs are added. The total number of users under $N_s$ SCs in the new proposed and basic (k-means) algorithms contribute to $OFLD$, and are multiplied with corresponding $OFLD$ to define a new metric called $OFFLOAD$. Thus, $OFFLOAD$ in the proposed algorithm and in basic (k-means) algorithm is defined as below. The parameter $OFFLOAD$ is used as a metric in this paper to evaluate the performance of the proposed algorithm.

$OFFLOAD_{NEW} = \left( SR_{n_m} - SR_{n_m}^* \right) \times \sum_{i=1}^{N_s} n_s^* \hspace{2cm} (11)$

$OFFLOAD_{OLD} = \left( SR_{n_m} - SR_{n_m}^* \right) \times \sum_{i=1}^{N_s} n_s^* \hspace{2cm} (11)$

We also define the gain in the system data rate because of adding SCs. As is intuitively understood, there would be increase in the sum data rate of the system when service is provided by the MC and the SCs together compared to the sum data rate when service is provided by the MC only. The percentage gain $PG$ in the proposed algorithm, and in the basic (k-means) algorithm are defined as below:

$PG_{New} = \frac{Sys_{Thp_{New}} - Thp_{mc}}{Thp_{mc}} \times 100$ \hspace{2cm} (12)$

$PG_{Old} = \frac{Sys_{Thp_{Old}} - Thp_{mc}}{Thp_{mc}} \times 100$

where, $Sys_{Thp_{New}}$ and $Sys_{Thp_{Old}}$ are the total throughput or sum data rate of the system with macro and SCs as per new proposed algorithm and basic (k-means) algorithm. $Thp_{mc}$ is the sum data rate of the system due to MC only and before addition of SCs.

**Observation** Although we defined percentage gain, $PG$ in Eq. 12, it cannot be used as a metric for evaluating the system performance. This is because the combined data throughput of all SCs in the basic (k-means) clustering algorithm may happen to be several times higher, compared to the case in the improved proposed algorithm, as the number of users under a SC may be very small. The greatest disadvantage of the k-means clustering algorithm is that the centroid of the cluster may be quite separated from its members. Also, the SC which is placed at the centroid has small coverage area. The observed effect on the data throughput of all users under a SC and hence of all SCs is the outcome of round robin user scheduling assumed here for data transmission by a BS. Hence, as shown in Eq. 12, $PG_{New}$ is expected to be smaller than $PG_{Old}$ as $Sys_{Thp_{New}} < Sys_{Thp_{Old}}$. This is further explained in the next section.
As more users are included under a SC in the proposed clustering algorithm compared to the basic (k-means), we define energy efficiency for SCs considering together as below:

\[ BS_{SCW_{New}} = \frac{SC_{STHP_{New}}}{TPs} \]
\[ BS_{SCW_{Old}} = \frac{SC_{STHP_{Old}}}{TPs} \]

where, \( BS_{SCW_{New}} \) and \( BS_{SCW_{Old}} \) are the sum data rate of the SCs per watt as per the proposed and basic (k-means) clustering algorithms, respectively. \( SC_{STHP_{New}} \) and \( SC_{STHP_{Old}} \) are the sum data rate of the SCs together as per the proposed and basic (k-means) clustering algorithm, respectively, \( TP_{s} \) is the total transmission power of the whole SC BSs together. The performance of the proposed clustering algorithm is evaluated using energy efficiency for SCs as defined in Eq. 13.

4 Simulation Results and Discussions

We consider that the MC BS and all SC BSs service data traffic to their users independent of each other, and the users under a BS are scheduled using round robin algorithm. As the users under the macro cell are large in numbers, the round robin time frame length is much longer and results in smaller individual user data throughputs. This is opposite in SCs where the round robin time frame length is much smaller resulting in much higher individual user data throughputs. Further, for simulation studies, we consider the positions of the users and hence their distributions shown in various networks remain static. We consider Rayleigh distributed small scale fading model, large scale fading factor representing shadow fading, and users’ geographical distances from BSs to evaluate users’ data rates.

We consider 4 different HetNets, each in a 1500-m circular geographical area and each with 1 MC BS and 4–8 SC BSs. The MC BS is located at center and SC BSs are located at the centroids of the clusters. We first apply the k-means algorithm, thereby obtaining the locations for the SC BSs. We then apply the proposed algorithm to the cluster points from k-means algorithm, thereby obtaining the new locations of the SC BSs.

In each case, the users within an area of 100 m of the locations of the SC BSs are taken to be the users for service by the SCs. The remaining users in the circular geographical area are serviced by MC BS. Parameters for the HetNets considered in this paper are shown in Table 1.

| Parameters                      | Values            |
|---------------------------------|-------------------|
| Path loss exponent (α)          | 2.5               |
| Large scale fading factor (c)   | \(10^{-3.53}\)    |
| Bandwidth (W)                   | 180 kHz           |
| Time slot for service           | \(0.5 \times 10^{-3}\) s |
| Noise power spectral density    | \(-174 \text{ dBm/Hz}\) |
| Transmit power of MC BS         | 46 dBm            |
| Transmit power of SC BSs        | 23 dBm            |
| Radius of MC                    | 1500 m            |
| Radius of a SC                  | 100 m             |
Of the 4 different HetNets, two are shown in Figs. 2, and Fig. 3 in the section of System Model above. The other two networks are shown in Figs. 4, and 5 below. Figures 4c and 5c shows remaining 4 SCs after 3 SCs are switched off randomly from Figs. 4b and 5b, respectively.

The experimental simulation results of PI_MUDR and PI_SCU for the four HetNets are shown in Figs. 6 and 7, using Eq. 8 and Eq. 9 respectively. It is observed that there are considerable increase in the MC user data rates, and in the number of users in different SCs. The increasing trend is shown in case of all HetNets.

The results of using the proposed algorithm in all 4 HetNets are shown below by simulation in the MATLAB. To evaluate the proposed algorithm, the parameter OFFLOAD, using the proposed and basic (k-means) clustering algorithms as in Eq. 11, experimental simulation results for the four HetNets are shown in Fig. 8 (i-iv) as bar diagrams. The metric shows clear advantage in using the proposed algorithm. The increasing trend is shown in case of all HetNets.

As is observed in the discussion of Eq. 12, the percentage gain, PG in the sum data rate of the system when service is provided by the MC and the SCs together compared to the sum data rate when service is provided by the MC only is worse while using the proposed algorithm. The reason is explained in the observation. The experimental simulation results for the four HetNets are shown in Fig. 9 (i-iv).

The performance metrics BS_SC_WOld and BS_SC_WNew, the sum data rate of the SCs per watt using the basic (k-means) and new proposed algorithm, respectively are evaluated for the four HetNets. The experimental simulation results are shown in Fig. 10 (i-iv) using Eq. 13. As is observed from the bar diagrams, there is increase in the energy efficiency of the combined SCs in all the four HetNets.

![Percentage Increase in macro user Throughput](image)

*Fig. 6* Percentage increase in throughput of macro users applying proposed algorithm for HetNet1, 2, 3, and 4
As stated in Sect. 2, for the second method of HetNet design, following the $k$-means and the proposed algorithm, we start with $y$ number of SCs where $y = 8$ and switch off $z$ number of SCs where $z = 1, 2, 3, 4$ in steps. The remaining SCs $y - z = x$, where $x = 7, 6, 5, 4$ in the HetNet are the results of randomly switching off SCs from the initial scenario of total 8 SCs. The simulation results of switching off $z$ number of SCs are shown in Tables 2 and 3. All four metrics are dependent on SCs that are switched off from initial 8 SCs. We are of the opinion that the decision of which SCs may be switched off depends on certain factors which may be technical, financial as well as administrative. Here we switched off SCs randomly with no obvious reason.

The simulation results of $PI_{MUDR}$, $PI_{SCU}$, as shown in Table 2 are expected to be different from the results shown in Fig. 6 and Fig. 7. Similarly, the simulation results of $BS_{SC}_{W_{New}}$ and $BS_{SC}_{W_{Old}}$ as shown in Table 3 are expected to be different from the results shown in Fig. 10. Hence, the values of all four metrics are also dependent on the user distributions in the HetNets. However, the proposed algorithm always provides positive values of the metrics studied in this paper. As there are a greater number of users in SCs, the centers of which are obtained by using the proposed algorithm, the data traffic offloading from the macro BS by the SCs is always larger than the case of using $k$-means algorithm. The way how we retain or design HetNet for a particular number of SC BSs/SCs may decide the performance of HetNets. The design objectives should satisfy the required service coverage or total transmit power available for the SC BSs.

![Percentage Increase in Small cell user]

Fig. 7 Percentage increase in SC users applying proposed algorithm for HetNet1, 2, 3, and 4
Future Scope

We assumed that all the users in the geographical service area are static while the service clusters are determined by the service provider to place the small cells in the optimum geographical locations. As the users are mobile and changing their locations, we would like to develop the proposed algorithm for such scenarios for dynamic applications. Further, the power optimization for macro cell users by the macro cell base station is also to be addressed. Algorithm for selective withdrawal of SCs to optimize system energy expenditure and its effect on the system throughput shall also be explored in our future works.

Fig. 8 (i): Data offloaded by SCs in HetNet1 applying proposed algorithm and k-means algorithm (ii): Data offloaded by SCs in HetNet2 applying proposed algorithm and k-means algorithm (iii): Data offloaded by SCs in HetNet3 applying proposed algorithm and k-means algorithm (iv): Data offloaded by SCs in HetNet4 applying proposed algorithm and k-means algorithm
Fig. 9 (i): Percentage gain in system data rate for HetNet1 applying proposed Algorithm and k-means Algorithm. (ii): Percentage gain in system data rate for HetNet2 applying proposed Algorithm and k-means Algorithm. (iii): Percentage gain in system data rate for HetNet3 applying proposed Algorithm and k-means Algorithm. (iv): Percentage gain in system data rate for HetNet4 applying proposed Algorithm and k-means Algorithm.
Fig. 10 (i): Energy Efficiency in HetNet1 applying proposed algorithm and k-means algorithm

(ii): Energy Efficiency in HetNet2 applying proposed algorithm and k-means algorithm

(iii): Energy Efficiency in HetNet3 applying proposed algorithm and k-means algorithm

(iv): Energy Efficiency in HetNet4 applying proposed algorithm and k-means algorithm
We proposed an improved grouping or clustering algorithm of user points, non-uniformly distributed (Cox process model) in a circular area, over the $k$-means algorithm. We use the algorithm to obtain the location centers of $SC$s which overlays the $MC$ BS area. In the given circular service area of the $SC$, maximum possible number of user points are obtained. The improved cluster centers (centroids) are obtained from the initial spatial clusters. We first construct users’ groups using the $k$-means algorithm, then an area of radius 300 m is considered around the initial $k$-means centroids. As the $SC$s are deployed to service the hot spots, edge users, or in general to offload the data traffic from the $MC$ BS, location of $SC$s should cover maximum possible users within its service area. By doing so the offloading is maximum possible, and hence increase in the data rate per macro user happens to be maximum possible. We define in this paper two metrics like $PI_{MUDR}$, $PI_{SCU}$, where the first one quantifies the percentage increase in per macro user data rates and the second one quantifies the percentage increase in the total number of users under all $SC$s. We also define two more metrics $OFFLOAD$, and $BS_{SC}_W$, where the first one characterizes the advantages in deploying the $SC$s and the second one characterizes the energy efficiencies of the $SC$s, resulting from application of the proposed algorithm. The simulation results are shown for 4 different $HetNets$ with 4–8 number of $SC$s, with 1000–1400 user points constructed using Cox process. The results show the validity of the claim that the proposed algorithm is an improved one over the basic $k$-means clustering algorithm. We also experimented with random switching off some $SC$s observing that the performances of $HetNets$ remain similar. We intend to explore this further in our future report. We also intend to consider spatial mobility of users within the $MC$ service area.

**Funding** Not Applicable.

**Availability of Data and Material** Standard parameters are used and algorithm developed is of our own.

**Code availability** For simulation Matlab is used.

**Declarations**

**Conflict of interest** No conflicts of Interest.
Table 3  Energy Efficiency (EE)

| Number of SCs | HetNet1 | | | HetNet2 | | | HetNet3 | | | HetNet4 | | |
|---------------|---------|---|---|---------|---|---|---------|---|---|---------|---|
|               | Proposed Algorithm | k-means Algorithm | Proposed Algorithm | k-means Algorithm | Proposed Algorithm | k-means Algorithm | Proposed Algorithm | k-means Algorithm | Proposed Algorithm | k-means Algorithm |
| 4             | 1.2×10^8 | 1.1×10^8 | 2.09×10^8 | 1.77×10^8 | 1.28×10^8 | 1.27×10^8 | 1.09×10^8 | 1.42×10^8 |
| 5             | 1.32×10^8 | 0.94×10^8 | 1.97×10^8 | 1.74×10^8 | 1.49×10^8 | 1.48×10^8 | 1.17×10^8 | 1.33×10^8 |
| 6             | 1.24×10^8 | 0.93×10^8 | 1.89×10^8 | 1.76×10^8 | 1.62×10^8 | 1.61×10^8 | 1.26×10^8 | 1.42×10^8 |
| 7             | 1.31×10^8 | 0.94×10^8 | 1.85×10^8 | 1.68×10^8 | 1.76×10^8 | 1.71×10^8 | 1.28×10^8 | 1.37×10^8 |
| 8             | 1.35×10^8 | 0.99×10^8 | 1.89×10^8 | 1.69×10^8 | 1.92×10^8 | 1.92×10^8 | 1.22×10^8 | 1.31×10^8 |
References

1. Cisco, ‘Cisco visual networking index: Global mobile data traffic forecast update, 2013–2018,” http://www.cisco.com/c/en/us/solutions/service-provider/visual-networking-index-vni/white-paper-listing. html, Feb. 2014, link verified on June. 30th, 2014.

2. Hughes, M. and Jovanovic, V. (2012). “Small cells—effective capacity relief option for heterogeneous networks,” in Proc. IEEE Veh. Technol. Conf. (VTC Fall), pp. 1–6.

3. Claussen, H., Lopez-Perez, D., Ho, L., Razavi, R., & Kucera, S. (2017). Small cell networks: deployment, management, and optimization. Wiley.

4. Chu, N. E. X., & Zhang, J. (2016). Small cell deployment over existing heterogeneous networks. IET Electron Lett., 52(3), 241–243.

5. ElSawy, H., Hossain, E., & Haenggi, M. (2013). Stochastic geometry for modeling, analysis, and design of multi-tier and cognitive cellular wireless networks: A survey. IEEE Commun Surv Tutor., 15(3), 996–1019.

6. Wang, Z., Schoenen, R., Yanikomeroglu, H., & Stillaire, M. (2014) “The impact of user spatial heterogeneity in heterogeneous cellular networks,” In 2014 IEEE Globecom Workshops (GC Wkshps), Austin, TX, 2014, pp. 1278-1283.

7. Qutqut, M. H., Abou-zeid, H., Hassanein, H. S., Rashwan, A. M, Al-Turjman, F. M., “Dynamic small cell placement strategies for LTE heterogeneous networks,” Proc. 2014 IEEE Symposium on Computers and Communications (ISCC), Funchal, 2014, pp. 1-6.

8. Wentao, Z. et.al. (2017). “Approximation Algorithms for Cell Planning in Heterogeneous Networks”, IEEE Trans. On Vehicular Technology, vol.66, no.2.

9. Araujo, W., et al. (2018). Deployment of small cells and a transport infrastructure concurrently for next-generation mobile access networks. PLoS ONE, 13(11), e0207330.

10. ElSawy, H., Sultan-Salem, A., Alouini, M., et al. (2017). Modeling and analysis of cellular networks using stochastic geometry: A tutorial. IEEE J Select Areas Commun, 35(3), 497–508.

11. Heath, S. W., Kountouris, M., & Bai, T. (2013). Modeling heterogeneous network interference using Poisson point processes. IEEE Transactions on Signal Processing, 61(16), 4114–4126.

12. Wang, Y., & Zhu, Q. (2017). Modeling and analysis of small cells based on clustered stochastic geometry. IEEE Communications Letters, 21(3), 576–579.

13. Lee, C. H., Lee, S. H., Go, K. C., et al. (2015). Mobile small cells for further enhanced 5G heterogeneous networks. ETRI Journal, 37(5), 856–866.

14. Wu, Y., Qian, L.P., Huang, J., Shen, X. (2017). Traffic offloading in heterogeneous cellular networks. In: Radio resource management for mobile traffic offloading in heterogeneous cellular networks, SpringerBriefs in Electrical and Computer Engineering. Springer, Cham. https://doi.org/10.1007/978-3-319-51037-8_1.
26. Qais, A., Omar, A. S., Ashraf, A., et al. (2020). Efficient power control framework for small-cell heterogeneous networks. Sensors, 20, 1467. https://doi.org/10.3390/s20051467 (www.mdpi.com/journal/)
27. Tareq, M. S., David, G., Alister, B., & John, S. V. (2019). Load balancing and control with interference mitigation in 5G heterogeneous networks. EURASIP Journal on Wireless Communications and Networking, 2019, 177. https://doi.org/10.1186/s13638-019-1487-0
28. Mayada, O., Salwa, E. R., & Bassant, A. (2021). Electronics 2021, 10, 1723, DOI/https://doi.org/10.3390/electronics10141723 (www.mdpi.com/journal/electronics).
29. Tadio, E. W. & Long, B. L. (2016). “Massive MIMO and mmWave for 5G Wireless Heterogeneous Network,” IEEE Vehicular Technology Magazine, March 2016
30. Ahmad, F., Bernard, C., & Ayman, K. (2018). “User Selection in 5G Heterogeneous Networks based on Millimeter-Wave Beamforming,” IEEE HPCC Conference, June 2018, Exeter UK.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Joyatri Bora received the M.Tech. degree in Electronics Design and Technology from the Tezpur University, Tezpur, India in 2002 and the Ph.D. degree from the Department of Electronics and Communication Engineering, North Eastern Regional Institute of Science and Technology (NERIST), India in 2015. She is a Member of Faculty with the Department of Electronics and Communication Engineering, North Eastern Regional Institute of Science and Technology (NERIST), Itanagar, India, since 2006. She is a member of professional bodies like IEEE, IEI and ISTE. Her current research interests include wireless communication, massive MIMO, Bioelectronics and internet of things.

Md. Anwar Hussain received B.Sc. degree in Physics from Gauhati University, Assam, India, in 1981, B.Tech. and M.Tech. in Radio Physics and Electronics from Calcutta University, West Bengal, India, in 1985 and 1987 respectively. Also he received Ph.D. degree in Electronics and Communication Engineering from Jadavpur University, West Bengal, India, in 2002. He is currently a professor in the department of Electronics and Communication Engineering at North East Regional Institute of Science and Technology (NERIST), Arunachal Pradesh, India. He has more than 31 years of experience in research and teaching field. His research interests includes massive MIMO, high data rate wireless communication and networks, routing and scheduling in multi-hop wireless networks, key distribution in sensor networks, multimedia data encryption and security, mobile computing security, time-series data modeling and prediction, networks-on-chip.