Enhancing Clustering Algorithm to Plan Efficient Mobile Network

Lamiaa Fattouh Ibrahim*  
Department of Computer Sciences and Information System, Institute of Statistical Studies and Research, Cairo University Giza, Egypt

Manal El Harby  
College of Education, UMM AL-QURA University, Macca, Saudi Arabia

*Corresponding Author: Department of Information Technology, Faculty of Computer and Information Technology, King Abdulaziz University B.P. 42808 Zip Code 21551- Girl Section, Jeddah, Saudi Arabia

ABSTRACT

With the rapid development in mobile network effective network planning tool is needed to satisfy the need of customers. However, deciding upon the optimum placement for the base stations (BS) to achieve best services while reducing the cost is a complex task requiring vast computational resource. This paper addresses antenna placement problem or the cell planning problem, involves locating and configuring infrastructure for mobile networks. The Cluster Partitioning Around Medoids (PAM) original algorithm has been modified and a new algorithm M-PAM (Modified-Partitioning Around Medoids) has been proposed by the authors in a recent work. In the present paper, the M-PAM algorithm is modified and a new algorithm CWN-PAM (Clustering with Weighted Node-Partitioning Around Medoids) has been proposed to satisfy the requirements and constraints. Implementation of this algorithm to a real case study is presented. Results demonstrate the effectiveness and flexibility of the modifying algorithm in tackling the important problem of mobile network planning.

General Terms

Data mining, network planning.

Keywords

Clustering techniques, network planning, cell planning and mobile network.

1. INTRODUCTION

The design objective of early mobile radio systems was to achieve a large coverage area by using a single, high powered transmitter with an antenna mounted on a tall tower. While this approach achieved very good coverage, it also meant that it was impossible to reuse those same frequencies throughout the system, since any attempts to achieve frequency reuse would result in interference. Faced with the fact that government regulatory agencies could not make spectrum allocations in proportion to the increasing demand for mobile services, it became imperative to restructure the radio telephone system to achieve high capacity with limited radio spectrum while at the same time covering very large areas [1].

Cellular telephony is designed to provide communications between two moving units, called mobile stations (MSs), or between one mobile unit and one stationary unit, often called a land unit [2]. A service provider must be able to locate and track a caller, assign a channel to the call, and transfer the channel from base station to base station as the caller moves out of range. Each cellular service area is divided into regions called cells. Each cell contains an antenna and is controlled by a solar or AC power network station, called the base station (BS). Each base station, in turn, is controlled by a switching office, called a mobile switching center (MSC). The MSC coordinates communication between all the base stations and telephone central office.

Cell planning is challenging due to inherent complexity, which stems from requirements concerning radio modeling and optimization. Manual human design alone is of limited use in creating highly optimized networks, and it is imperative that intelligent computerized technology [3], Tuba Search TS [4], [5] Genetic Algorithm GA [6] and clustering algorithm [7]-[9] have been successfully deployed in mobile network designs. Clustering analysis is a sub-field in data mining that specializes in techniques for finding similar groups in large database [10]. Its objective is to assign to the same cluster data that are more close (similar) to each other than they are to data of different clusters. The application of clustering in spatial databases presents important characteristics. Spatial databases usually contain very large numbers of points. Thus, algorithms for clustering in spatial databases do not assume that the entire database can be held in main memory. Therefore, additionally to the good quality of clustering, their scalability to the size of the database is of the same importance [11]. In spatial databases, objects are characterized by their position in the Euclidean space and, naturally, dissimilarity between two objects is defined by their Euclidean distance [12]. Clustering techniques has been successfully deployed in wire [13],[14],[15] and in Wireless Local Loop [16] network planning.

This paper introduces the spatial clustering to solve the Mobile Networking Planning problem. This paper is an extension version of papers [17] and [18]. Section 2 discusses main phases used in radio network planning. In sections 3 The Cluster Partitioning Around Medoids (PAM)
2. MAIN PHASES USED IN RADIO NETWORK PLANNING

The radio network planning process can be divided into different phases [19]. At the beginning is the Preplanning phase. In this phase, the basic infrastructural properties of the future network are investigated, for example, what kind of mobile services will be offered by the network, what kind of requirements the different services impose on the network, the basic network configuration parameters and so on. The second phase is the main phase. A site survey is done about the to-be-covered area, and the possible sites to set up the base stations are investigated. All the data related to the geographical properties and the estimated traffic volumes at different points of the area will be incorporated into a digital map, which consists of different pixels, each of which records all the information about this point. Based on the propagation model, the link budget is calculated, which will help to define the cell range and coverage threshold. There are some important parameters which greatly influence the link budget, for example, the sensitivity and antenna gain of the mobile equipment and the base station, the cable loss, the fade margin etc. Based on the digital map and the link budget, computer simulations will evaluate the different possibilities to build up the radio network part by using some optimization algorithms. The goal is to achieve as much coverage as possible with the optimal capacity, while reducing the costs also as much as possible. The coverage and the capacity planning are of essential importance in the whole radio network planning. The coverage planning determines the service range, and the capacity planning determines the number of to-be-used base stations and their respective capacities.

In the third phase, constant adjustment will be made to improve the network planning. Through driving tests the simulated results will be examined and refined until the best compromise between all of the facts is achieved. Then the final radio plan is ready to be deployed in the area to be covered and served.

3. THE CLUSTER PARTITIONING AROUND MEOIDS (PAM) AND M-PAM ALGORITHMS

The PAM (Partitioning Around Medoids) algorithm, also called the K-medoids algorithm, represents a cluster by a medoid [20]. Initially, the number of desired clusters is input and a random set of k items is taken to be the set of medoids. Then at each step, all items from the input dataset that are not currently medoids are examined one by one to see if they should be medoids. That is, the algorithm determines whether there is an item that should replace one of the existing medoids. By looking at all pairs of medoids, non-medoids objects, the algorithm chooses the pair that improves the overall quality of the clustering the best and exchanges them. Quality here is measured by the sum of all distances from a non-medoid object to the medoid for the cluster it is in. A item is assigned to the cluster represented by the medoid to which it is closest (minimum distance or direct Euclidean distance between the customers and the center of the cluster they belong to).

The PAM algorithm [20] is shown in Figure 1. By assuming that $K_i$ is the cluster represented by medoid $t_i$. Suppose $t_i$ is a current medoid and it is wished to determine whether it should be exchanged with a non-medoid $t_j$. Do this swap only if the overall impact to the cost (sum of the distances to cluster medoids) represents an improvement.

The total impact to quality by a medoid change is given by $TC$:

$$TC = \sum_{i=1}^{k} \sum_{j \in C_i} \text{dis}(n_i, n_j)$$

**Algorithm PAM**

**Input:**

- $D = \{t_1, t_2, t_3, \ldots, t_n\}$ // set of elements
- $A$ // adjacency matrix showing distance between elements.
- $k$ // Number of desired clusters.
- $K$ // set of clusters.

**PAM Algorithm:**

1. Arbitrarily select $k$ medoids from $D$;
2. repeat
   1. For each $t_i$ not a medoid do
      1. Compute square error function $TC_{ih}$;
      2. Find $i, h$ where $TC_{ih}$ is the smallest;
      3. If $TC_{ih} < \text{Current } TC$ then
         1. Replace medoid $t_i$ with $t_j$;
         2. Until $TC_{ih} \geq \text{Current } TC$;
      4. For each $t_i \in D$ do
         1. Assign $t_i$ to $K_j$ where $\text{dis}(t_i, t_j)$ is the smallest over all medoids;
3. **Figure 1: PAM algorithm**

**The problem statement:**

- A set $P$ data points $\{p_1, p_2, \ldots, p_n\}$ in 2-D map, subscribers loads and communication constraints.
- **Objective**: Partition the city into $k$ clusters $\{C_1, C_2, \ldots, C_k\}$ that satisfy clustering constraints, such that the cost function is minimized with high grade of services.
- **Input**: Set of $n$ objects (map), set of streets associate with their load.
The proposed algorithm contains three phases. The following sections describe these three phases.

4.1 Phase I: Pre-Planning
This phase is divided into two steps. Step 1, convert map from raster form to digital form. Step 2, determine the initial number of clusters.

4.1.1 Map and their Data Entries
The maps used for planning are scanned images obtained by the user. It needs some preprocessing operations before it is used as a digital map. The streets and intersection nodes on the raster maps, the beginning and ending of each street are transformed into data nodes, defined by their coordinates. The streets themselves are transformed into links between data nodes. The subscriber’s loads are considered to be the weights for each node. Figure 2 show map transformation. For each intersection node and street the user can right click to input the characteristics of intersection node (no, name, capacity) or street (street number, street name, street load).

![Figure 2: map transforming](image)

(a) raster map  
(b) draw squares and streets on raster map  
(c) map after transforming to digital map

All above data is saved in database. Visual basic dot Net as work environment with oracle is used to create tables and data input.

4.1.2 Determine Initial Number of Clusters
In cell planning, planned area is divided to number of cells, each cell served by BS which guarantees the quality of service for all subscribers. In this paper, GSM technology of radio network planning is used to calculate number of cell need by coverage planning and calculate number of cell need by capacity planning for planned area.

- Number of cell needed by coverage planning: \( \text{cell needed} = \frac{\text{Total area}}{\text{area of the cell}} \)
- Number of cell needed by capacity planning: \( \text{cell needed} = \frac{\text{Total number of subscribers}}{\text{Total subscribers per cell}} \)
- Initial number of clusters: \( K = \text{the maximum of the two values} \)

4.2 Phase II: Main-Planning stage
In this phase, the goal is to split the entire database into clusters.

4.2.1 Partition Database
After initial \( k \) is known, CWN-PAM algorithms is used to determine the optimal location of base station and its boundary of the served area for each cluster. The CWN-PAM algorithm is based mainly on the idea of the Modified Partitioning Around Medoids (M-PAM). The cost function is modified in the M-PAM.

4.2.2 Modified Cost Function to Handle Node Load
In database contains \( n \) points \( \{n_1, n_2, ..., n_k\} \), Where \( n_k \) is the medoid (the real data point that satisfies minimum cost) of cluster \( C_k \). \( n_i \) denote to non-medoid points and \( k \) is the number of cluster. The Direct Euclidean distance from a point \( n_i \) to \( n_j \) is the dist(\( n_i, n_j \)). The cost function in M-PAM (equation 1) is modified. The TC is modified to WTC where:

\[
\text{WTC} = \sum_{b=1}^{k} \sum_{n_i \in C_b} L_{hi} \text{dis}(n_{hi}, n_i) \quad (2)
\]

\( L_{hi} \) is the subscriber load cost of this distance. According to equation (2), medoids, location of the base stations, move toward the heavy loaded (weighted) nodes. Figure 3 shown map of a city and number in map represent the subscriber loads. Block 1 and block 3 has homogeneous distribution loads therefore the medoids are not changed when applying M-PAM and CWN-PAM algorithms. Block 2 and 4 has non-homogeneous distribution load therefore medoids are changed locations when applying M-PAM and CWN-PAM algorithms (medoids move toward the heavy loaded nodes).

![Figure 3 Map after apply M-PAM and CWN-PAM algorithms](image)

4.3 Phase III: Adjustment Stage
The base station satisfy the constraints if and only if \( k = 1 \) for each cluster. If any of clusters are not satisfy the requirements, more clusters is added and redistribute nodes to close base station.

For each cluster, Coverage and capacity plan are applied to calculate the number of needed base station BS. If any cluster needs more than one base station, one of the following methods is applied. Method I: number of clusters is increased on the whole data (area), and then goes to the Phase II. Method II: number of clusters, on just the cluster that had a problem on its mobile constraints, is increased. Figure 4 describes the new algorithm CWN-PAM.

5. CASE STUDY
By applying CWN-PAM algorithm method I and method II to different datasets Table 1 and different cell range the results shown in figure 5 – figure 7 are obtained.
The experiments show that method II does not have any better solution with small datasets (like dataset I and dataset II) or in the homogeneous density maps. But when large number of subscribers is deployed in different area with different densities, it produces a better solution and cost minimization. Because the clustering algorithm is applied only on the clusters that had a problem on its mobile requirements whether coverage or capacity or even both instead of clustering a whole planned area by adding more clusters and redistribute nodes by using clustering algorithms which sometimes repeat the problem that depending on the redistributing process.

To compare the different densities, two maps with (30,101) nodes are considered. For each map the load of their nodes is changed. The data base is shown in table 2.

Table 2. Data base entries used for comparison (same maps with different densities)

| Data set ID | Number of Nodes | Coverage Area | Number of Subscribers |
|-------------|-----------------|---------------|-----------------------|
| Data set I  | 30              | 327600 [m²]   | 3000                  |
| Data set II | 30              | 327600 [m²]   | 7158                  |
| Data set III| 101             | 337800 [m²]   | 4000                  |
| Data set IV | 101             | 337800 [m²]   | 18595                 |

Figure 8 to 11 show the results of applying CWN-PAM algorithm method I and method II to numbers of nodes 30 and 101 but when increase density (numbers of subscribers).
Dataset I with 30 nodes and 3000 subscribers and dataset III with 101 nodes and 4000 subscribers, since the number of subscribers is small; the number of base stations had different values with different cells range. Dataset II with 30 nodes and 7158 subscribers and dataset IV with 101 nodes and 18595 subscribers, since the number of subscribers is big, therefore the problem in these maps is in their capacity. The number of base stations had not based effected with different cells ranges. But also method II obtain the minimum numbers of Base station.

### Table: Number of BS's

| Number of BS's in CWN-PAM by method I for map 30 node | Number of BS's in CWN-PAM by method II for map 30 node |
|-------------------------------------------------------|-------------------------------------------------------|
| ![Figure 8](image8.png)                                | ![Figure 9](image9.png) |
6. COMPARISON BETWEEN PROPOSED METHOD AND OTHER METHODS

Table 3 described the different comparison between the proposed method and other methods using in mobile network planning. There are two methods that are frequently used here: Tabu search [4], [5] and Genetic Algorithms [6]. Table 3 shows the comparison between relative works. Tabu Search and Genetic algorithm, needed a huge numbers of estimated input parameters which can be affected to the results.

7. CONCLUSION AND FUTURE WORK

In this paper the proposed algorithm CWN-PAM which use modifies clustering technique M-PAM to the mobile network planning problem has been presented. This algorithm is a medoid clustering algorithm using weighted Direct Euclidean distance that satisfy the network constraints and where the weights used are the subscriber loads. Due to the modification in cost function, the location of the base stations (medoid), move toward the heavy loaded (weighted) nodes and increase there grade of service. Experimental results and analysis indicate that the CWN-PAM algorithm when applying method two is effective, and leads to minimum costs (minimum number of base stations) in network have a large number of intersection and streets where number of subscribers are deployed with different densities. It is expected that by applying this system to a number of areas belonging to different countries with different sizes, one can verify its capabilities more universally.

8. REFERENCES

[1] Theodore S. Rappaport, “Wireless Communications: Principles and Practice”, Second Edition, 2002 Pearson Education.
[2] Behrouz A.Forouzan, “Data Communications and Network” fourth edition, Mc Graw- Hill 2007.
[3] S.M. Allen, B. Belloul, S. Hurley, S. Saunders, R.M. Whitaker, Smart Cell Planning and Optimisation for UMTS, 3G2004, pp 34–38, October 18, 2004, IEE press, ISBN 0 86341 388 9.
[4] M. St-Hilaire, S. Chamberland, S. Pierre," A Tabu Search Heuristic for the Global Planning of UMTS Networks", IEEE International Conference on Wireless and Mobile Computing, Networking and Communications, pp. 148-151, June 2006.
[5] YufeiWu, Samuel Pierre, “Mobile Radio Network Planning Aspects”, Third IEEE International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob 2007) p. 26.
[6] Qiang Guan, Yu Liu, Yiping Yang, Wensheng Yu, “Genetic Approach for Network Planning in the RFID Systems”, Proceedings of the Sixth International Conference on Intelligent Systems Design and Applications (ISDA’06), Jinan, Shandong, China, 2006.
[7] Lamiaa Fattouh Ibrahim, Manal Hamed Al Harbi, “Using Modified Partitioning Around Medoids Clustering Technique in Mobile Network Planning”, International Journal of Computer Science Issues Volume 9, Issue 6, November 2012.
[8] Ibrahim, L.F. and H.A. Salman, 2011. Using hyper clustering algorithms in mobile network planning. Am. J. Applied Sci., 8: 1004-1013.
[9] Salman, H.A.; Ibrahim, L.F.; , ”Efficient Mobile Network Planning Algorithm in the Presence of Obstacles,” Wireless and Mobile Communications (ICWMC), 2010
[10] Jean F. Hou, Clustering with obstacle entities, a thesis submitted in partial fulfillment of the requirements for the degree of master, Simon Fraser University November 1999.
[11] Alexandros Nanopoulos , Yannis Theodoridis , Yannis Manolopoulos, C2P: Clustering based on Closest Pairs, Proceedings of the 27th International Conference on Very Large Data Bases, p.331-340, September 11-14, 2001.
[12] Man Lung Yiu, Nikos Mamoulis: Clustering Objects on a Spatial Network. SIGMOD Conference 2004: 443-454.
[13] Lamia Fattouh Ibrahim, " Using of Clustering and Ant-Colony Algorithms CWS-PAM-ANT in Network Planning", International Conference on Digital Telecommunications(ICDT 2006), Cap Esterel, French Riviera, France, 26-31 August 2006.
[14] Lamia Fattouh Ibrahim, "Using of Clustering Algorithm CWS-PAM for Rural Network Planning", 3rd International Conference on Information Technology and Applications, ICITA'2005, Sydney, Australia, July 4-7, 2005.
[15] Lamiaa Fattouh, Omar Karam, Mohamed A. El Sharkawy, Walaa Khaled, "Clustering For Network Planning", WSEAS Transactions on Computers, Issue 1, Volume 2, ISSN 1109-2750, January 2003.
[16] Omar H. Karam, Lamia Fattouh, Nourhan Youssif, Ahmad E. Abdelazim “Employing Clustering Techniques in Planning Wireless Local Loop Communication Systems: PLANAIR” 11th International Conference On Artificial Intelligence Applications Cairo, Egypt, February 23-26, 2005.
[17] Ibrahim, L.F., Al Harbi, M.H., "Employing Clustering Techniques in Mobile Network Planning," New Technologies, Mobility and Security, 2008. NTMS '08. , vol., no., pp.1-9, 5-7 Nov. 2008
[18] El Harby, M., Ibrahim, L.F., "Employing of Clustering Algorithm CWN-PAM in Mobile Network Planning," Systems and Networks Communications, 2008. ICSNC
Table 3. Relative Works

| Algorithm          | Input Parameters | Results | Location of BS’S | Constraints | Type Of Distance | Complexity     |
|--------------------|------------------|---------|------------------|-------------|------------------|----------------|
| Genetic Algorithm  | - Data points    | # of clusters | Optimal placement | Fitness Function | Euclidean distance | $O(NK\ln K)$ |
|                    | - Population     |          |                  |             |                  |                |
|                    | - Initial        |          |                  |             |                  |                |
|                    | probability      |          |                  |             |                  |                |
|                    | - Mutation       |          |                  |             |                  |                |
|                    | probability      |          |                  |             |                  |                |
|                    | - Crossover      |          |                  |             |                  |                |
|                    | probability      |          |                  |             |                  |                |
|                    | - Number of      |          |                  |             |                  |                |
|                    | iteration        |          |                  |             |                  |                |
|                    | - Selection      |          |                  |             |                  |                |
|                    | pressure         |          |                  |             |                  |                |
| Tabu Search        | - Data points    | # of clusters | Optimal placement | Tabu List  | Euclidean distance | $O(NK\log K)$ |
|                    | - Init probability|          |                  |             |                  |                |
|                    | - Generation     |          |                  |             |                  |                |
|                    | probability      |          |                  |             |                  |                |
|                    | - Recency factor |          |                  |             |                  |                |
|                    | - Frequency      |          |                  |             |                  |                |
|                    | factor           |          |                  |             |                  |                |
|                    | - Number of      |          |                  |             |                  |                |
|                    | iteration        |          |                  |             |                  |                |
|                    | - Number of      |          |                  |             |                  |                |
|                    | neighbors        |          |                  |             |                  |                |
| PAM                | - Data points    |Clusters medoid | Mediods | TC_{ih} = \sum C_{ih} | Euclidean distance | $O(K(n-K)^{2})$ |
|                    | - K              | # of clusters |          |             |                  |                |
| M-PAM              | - Data points    |Clusters medoid | Mediods | TC_{ih} = \sum C_{ih} | Euclidean distance | $O(K(n-K)^{2})$ |
|                    | - K              | # of clusters |          |             |                  |                |
| CWN-PAM            | - Data points    |Clusters medoid | Mediods | WTC_{ih} = \sum L_{ih} | Euclidean distance | $O(K(n-K)^{2})$ |
|                    | - K              | # of clusters |          |             |                  |                |