Enhanced Clustering Techniques for Hyper Network Planning using Minimum Spanning Trees and Ant-Colony Algorithm

1,2 Lamiaa Fattouh Ibrahim
1Department of Computer Sciences and Information, Institute of Statistical Studies and Research, Cairo University, Cairo, Egypt
2Department of Information Technology, Faculty of Computing and Information Technology, King AbdulAziz University, B.P. 42808 Zip Code 21551- Girl Section, Jeddah, Saudi Arabia

Abstract: Problem statement: The process of network planning is divided into two sub steps. The first step is determining the location of the Multi Service Access Node (MSAN). The second step is the construction of subscriber network lines from MSAN to subscribers to satisfy optimization criteria and design constraints. Due to the complexity of this process artificial intelligence and clustering techniques have been successfully deployed to solve many problems. The problems of the locations of MSAN, the cabling layout and the computation of optimum cable network layouts have been addressed in this study. The proposed algorithm, Clustering density-Based Spatial of Applications with Noise original, minimal Spanning tree and modified Ant-Colony-Based algorithm (CBSCAN-SP-ANT), used two clustering algorithms which are density-based and agglomerative clustering algorithm using distances which are shortest paths distance and satisfying the network constraints. This algorithm used wire and wireless technology to serve the subscribers demand and place the switches in a real optimal place. Approach: The density-based Spatial Clustering of Applications with Noise original (DBSCAN) algorithm has been modified and a new algorithm (NetPlan algorithm) has been proposed by the author in a recent work to solve the first step in the problem of network planning. In the present study, the NetPlan algorithm is modified by introduce the modified Ant-Colony-Based algorithm to find the optimal path between any node and the corresponding MSAN node in the first step of network planning process to determine nodes belonging to each cluster. The second step, in the process of network planning, is also introduced in the present study. For each cluster, the optimal cabling layout from each MSAN to the subscriber premises is determining by introduce the Prime algorithm which construct minimal spanning tree. Results: Experimental results and analysis indicate that the CBSCAN-SP-ANT algorithm was effective, leads to minimum costs for network construction and make the best grade of service. Conclusion: Using mobile network to serve the area with low density is decreasing the cost of design the fixed wire network. Also, using the modified ANT algorithm and minimum spanning tree, are helping to construct the cable layout from each MSAN to subscribers when the network is complicated and the number of intersections and streets are large.

Key words: Multi Service Access Node (MSAN), agglomerative algorithm, clustering techniques, network planning, cable layout network, short path, Minimum Spanning Tree (MST), ant-colony algorithm, meta-heuristic, NetPlan algorithm

INTRODUCTION

The goal of a clustering algorithm is to partition a given data set into clusters or groups, which are not predefined, such that the data points in a cluster are similar to each other more than points in different clusters. These groups are formed according to some measures of goodness that differ according to application.

The field of “ant algorithms” studies models derived from the observation of real ant's behavior and uses these models as a source of inspiration for the design of novel algorithms for solution of optimization and distributed control problems (Dorigo and Stulzle, 2004; Gunes et al., 2002; Othman et al., 2007; AL-Salami, 2009).

Ant colony algorithms are a subset of swarm intelligence and consider the ability of simple ants to solve complex problems by cooperation. The interesting point is, that the ants do not need any direct communication for the solution process, instead they
communicate by stigmergy. The notion of stigmergy means the indirect communication of individuals through modifying their environment. Several algorithms which are based on ant colony problems were introduced in recent years to solve different problems, e.g., optimization problems, image segmentation (Ouadfel and Batouche, 2007; Ibrahim et al., 2005; Hashim and Abd, 2010).

Multi Service Access Node MSAN is one of the dominant access delivery methods and comes in modular units which may be equipped with line cards supporting a number of services, differing capacities, ranging from a few dozen lines up to thousands of lines.

In many countries, there is tremendous demand for new business and residential telephone service. The continuous increasing number of subscribers make congestion in MSAN and cause degradation in grade of service and in sometimes impossible to add new subscribers which lead to using the mobile tower. Also, the existing of the natural obstacle is affecting on distribution the MSAN on the regions.

In this study we determine place of each MSAN using NetPlan algorithm (Ibrahim et al., 2009) after modified by introduce the modified Ant-Colony-Based algorithm to find the optimal path between any node and the corresponding MSAN node in the first step of network planning process to determine nodes belong to each cluster. When the congestion in MSAN is occur mobile tower is playing as auxiliary tool with maximum 100 subscribers.

The second step is also introduced in the present paper. For each cluster, the optimal cabling layout from each MSAN to the subscriber premises must be determined. The problem is to find the least cost network by minimizing the distances from the MSAN to the subscribers belonging to it. Thus the problem of connecting n nodes with a minimum distance network is the problem of finding a minimal spanning tree in a connected weighted graph of n vertices. The Prime algorithm is used to construct the optimal cabling layout for each cluster. The final step is the calculation of total cost.

Due to the complexity of this process, Artificial Intelligence (AI) (Fahmy and Douligeris, 1997; El-Dessouki et al., 1999) and partitioning clustering techniques (Khaled et al., 2003; Ibrahim and Al Harbi, 2008a; 2008b; Harby and Ibrahim, 2008; Ibrahim, 2005; 2006; Han et al., 2001) successfully deployed in a number of areas.

The NetPlan algorithm: Network Planning package (NetPlan) (Ibrahim et al., 2009) is divided into two steps:

Step 1: Applying the modified DBSCAN algorithm.
Step 2: Applying Agglomerate algorithm to the resulted clusters.

Modified DBSCAN algorithm: Two parameters must me determine before we starts applying the DBSCAN. These parameters are MinPts and Eps. In network planning the cable length must be at maximum 2.5 km for 0.4 cm diameters to achieve an acceptable grade of service. So, we make the value of EPS take the value of shortest path from core (MSAN) to the most remote point (subscribers) which is 2.5 km. The original DBSCAN Algorithm uses Euclidian distance (that means the direct distance between the MSAN and nodes); The Direct Euclidean distance ignores the presence of streets and paths that must be taken into consideration during clustering. In NetPlan, a clustering based solution is presented depending on using the physical shortest available routes. To apply shortest path we selected Dijkstra algorithm.

When the congestion in MSAN is occur or the number of subscribers is less than 100 we use mobile tower as auxiliary tool to serve this small number of subscribers. Therefore the value of MinPts is set to 101. The DBSCAN classify nodes to:

- Core point which is a subset of candidate MSAN location
- Noise point: In real planning all subscribers must be served so noise point is served using the mobile tower which can serve at maximum 100 subscribers because that number is the maximum number of subscriber who can be served by mobile tower
- Border point that belong to ascertain cluster

NetPlan use Dijkstra algorithm to calculate shortest path from one node to all MSANs (the reason is to determine the nearest suitable MSAN that will serve this node).

Agglomerative clustering technique: The agglomerative clustering technique is hierarchal clustering technique. It starts with the points as individual clusters and at each step merge the closest pair of clusters depends on a notion of cluster proximity and the faraway node from MSAN which should be at maximum 2.5 km to achieve the required grade of service. In NetPlan, after we distribute the nodes into different clusters (each cluster has served by one MSAN) and maybe there are two cores (MSANs) are close in distance (less than 1.25 Km). So, we need to decrease the cost of constructing a new MSAN if one of them can hold the loads (subscribers) of the two MSANs. Figure 1 shows Pseudo code of agglomerative algorithm used in NetPlan.
Fig. 1: Pseudo code of agglomerative algorithm

```
For (i=1 to candidate No.)
    For (j=1 to number of node)
        Calculate the shortest path from MSAN(i) to node(j)
        If (shortest path < 2.5 km)
            Then current load(i) = current load(i) + load of node
                End For
        If (Current load(i) := 101)
            Then add MSAN to cores
            End for
    For each node in the city select the best switch for it by calculating the shortest path between the nodes and each MSAN (path < 2.5)
    Calculate the load of each core
    For each two clusters  // Agglomerative
        If the shortest path between the cores < 2.5
            Then
                If (sum of two clusters' load < 1536)
                    Then
                        First assume that the first core is core of the two clusters and apply the previous algorithm for each node in the two clusters.
                        Second assume that the second core is core of the two clusters and apply the previous algorithm for each node in the two clusters.
                        If (the two cores are suitable to be a core for all nodes of two clusters)
                            Then
                                Calculate the Cost Function = \sum (load \times short path distance from node to core)
                                Choose the appropriate core depend on minimum Cost Function value.
                                Else let the suitable one the core for two clusters.
```

Fig. 2: NetPlan algorithm

**Implemented of NetPlan algorithm:** The user first inserts the location of candidate MSAN. NetPlan uses these candidate locations as candidate core points and chooses the one which satisfies the condition to be a core node. After this step the NetPlan determine the boundaries of cluster by calculate the short path from node to each core and allocate the node to the minimum short path core. The following step is the agglomerative step with mix two clusters if the overall load is less than the max load of MSAN which is 1536 and if also the maximum short path for all node to core is less or equal 2.5 km and the third condition if the distance between the two core is less the half of 2.5 km. NetPlan introduces a cost function which is the sum of multiplication of load of node and short path between node and core. This cost function is used to choose the best MSAN when we merge two clusters. Figure 2 shows pseudo code of NetPlan algorithm used.

**MATERIALS AND METHODS**

The proposed algorithm CBSCAN-SP-ANT modified NetPlan algorithm by introduce the modified Ant-Colony-Based algorithm to find the optimal path between any node and the corresponding MSAN node in the first step of network planning process to determine nodes belong to each cluster. The second step, in the process of network planning, is also introduced in the present study. For each cluster, the optimal cabling layout from each MSAN to the subscriber premises is determining by introduce the Prime algorithm which construct minimal spanning tree.

The problem statement:

- **Input:** A set P data points \{p_1, p_2, ..., p_n\} in 2-D map which represent intersection nodes, coordinates of each node, a map of streets, distribution of the subscribers’ loads within the city and the location of base station in mobile network in this city
- **The available cable sizes, the cost per unit for each size and the maximum distance of wire that satisfied the allowed grade of service**
- **Objective:** Partition the city into k clusters \{C_1, C_2, ..., C_k\} that satisfy clustering constraints, such that the cost function is minimized with high grade of services
- **Output:** k clusters, the location of MSAN, the cable layout from each MSAN to subscriber, boundaries of each cluster and finally the total cost to construct the network

**Computing the shortest distance from node to MSAN:** The ANT colony optimization meta-heuristic is a particular class of ant algorithms. Ant algorithms are multi-agent systems, which consist of agents with the behavior of individual ant (Bonabeau et al., 1999; Wittner and Helvik, 2004).

The basic idea of the ant-colony optimization meta-heuristic is taken from the food searching behavior of real ants. When ants are on the way to search for food, they start from their nest and walk toward the food. When an ant reaches an intersection, it has to decide which branch to take next. While walking, ants deposit pheromone, which marks the route taken. The concentration of pheromone on a certain path is an
indication of its usage. With time the concentration of pheromone decreases due to diffusion effects.

All ants take the shortest path after an initial searching time.

The Ant colony algorithm depends mainly on two steps as follows:

- Tour construction
- Pheromone update

Tour construction: At each construction step, the probability with which ant \( k \), currently at node \( i \), chooses to go to node \( j \) at the \( t \)th iteration of the algorithm is:

\[
P^k_{ij}(t) = \frac{[T^k(i)(t)]^\alpha \cdot [\eta^k_{ij}]^\beta}{\sum_{i \in N_j^k} [T^k(i)(t)]^\alpha \cdot [\eta^k_{ij}]^\beta}
\]

If \( j \in N_k \) and 0 otherwise

Where:
- \( \eta_{ij} = 1/d_{ij} \) is an a priori available heuristic value. It is known as visibility of node \( j \) from node \( i \)
- \( T_{ij} = \) Intensity of pheromone trail between node \( i \) and node \( j \)
- \( N_j^k = \) The set of nodes which ant \( k \) has not yet visited
- \( \alpha = \) Parameter to regulate the influence of \( T_{ij} \)
- \( \beta = \) Parameter to regulate the influence of \( \eta_{ij} \)
- \( d_{ij} = \) Distance between node \( i \) and node \( j \) (in case of shortest path problem)

Pheromone update: After all ants have constructed their tours, the pheromone trails are updated:

\[
T^k_{ij}(t+1) = (1-p) \cdot T^k_{ij}(t) + \sum_{k=1}^{N} \Delta T^k_{ij}(t)
\]

where, \( 0 < p \leq 1 \) is the pheromone trail evaporation.

\( \Delta T^k_{ij}(t) \) is the amount of pheromone which ant \( k \) puts on the arcs it has visited; it is defined as follows:

\[
\Delta T^k_{ij}(t) = \begin{cases} 
1/L^k(t) & \text{if arc}(i,j)\text{is used by ant } k \\
0 & \text{otherwise}
\end{cases}
\]

where, \( L^k(t) \) is the total length of the tour, from source to destination, made by the ant \( k \) at iteration \( t \). By the last equation, the better the ant's tour is, the more pheromone is received by arcs belonging to the tour.

Figure 3 shows the fundamental Algorithm of Ant Algorithm.

The ANP Ant-Colony-Based Network Planning algorithm (Ibrahim et al., 2005) modified the previous describe algorithm in Fig. 3 due to a number of problems are raised when we try to apply this algorithm such as:

- Loss of ants due to some node is connected with the network by terminates edge and the ant is not permitted to visit node twice
- Ant is blocked when it reaches terminates node and has no way to reach the location of the switch
- Loss of ants due to internal loop in a graph

Figure 4 shows ANP Ant-Colony-Based Network Planning Algorithm. In This algorithm, for each node we put a number of ants equal the number of edges (no ants at switch location) to substitute the losses of ants and at the first step we move each ant at different route edge.

Optimal cabling layout: For each cluster, the optimal cabling layout from each exchange to the subscriber premises must be determined. The problem is to find the least cost network by minimizing the distances from the exchange to the subscribers belonging to it. Thus the problem of connecting \( n \) nodes with a minimum distance network is the problem of finding a minimal spanning tree in a connected weighted graph of \( n \) vertices. The Prime algorithm is used to construct the optimal cabling layout for each cluster.

The Prime algorithm (O'Rourke, 1994) is used to grow the minimum spanning tree from a root (core) to find the least expensive network that connects all nodes together. The Minimum Spanning Tree (MST) is a tree that connects all nodes in the cluster with minimum cost and achieves the maximum grade of service. Figure 5 shows Prime algorithm.

Figure 6 shows pseudo code of CBSCAN-SP-ANT Algorithm used.
RESULTS AND DISCUSSION

Table 1 compares related study. In Gravity Center algorithm (Dessouki et al., 1999), the city is divided into four quadrants (clusters) at the center of gravity if the capacity is more than the capacity of max switch capacity allow or distance from node to switch is more than the max distance allow. The switches will be located at the center of gravity of each cluster. If the constraints are not satisfied in any of the four quadrants the same partitioning method is applied to the quadrant which does not satisfy the constraints. This yields that the number of clusters equal seven partitions. This method will be iterated until the network constraints are 7, 10. This study doesn’t reflect the real nature of the clusters, or the number of the suitable clusters, it is always incrementing the number of clusters by three.

COD-CLARANS (Tung et al., 2001) and CSPwCLARANS (El-Dessouki et al., 1999) algorithms depend mainly on CLARANS which is design to deal with large database by using multiple different samples. These two algorithms is very powerfully when we plan a large city, but not acute when we plan small city due the sampling use.
Table 1: A comparison between related works

<table>
<thead>
<tr>
<th>Algorithm name</th>
<th>Algorithm type</th>
<th>Input Parameters</th>
<th>Results</th>
<th>Constraint</th>
<th>Location of exchange</th>
<th>Type of distance and Type of network</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gravity center</td>
<td>Gravity center algorithm</td>
<td>Data Points</td>
<td>Divide a block in 4,7,10...block</td>
<td>Yes, network constraints</td>
<td>At the gravity center</td>
<td>Shortest path distance</td>
</tr>
<tr>
<td>COD-CLARANS</td>
<td>partitioning method</td>
<td>Data points Number of clusters (k) Maximum number of neighbors</td>
<td>Medoids of clusters</td>
<td>Yes, obstacles constraints</td>
<td>At the medoids</td>
<td>Obstructed distance</td>
</tr>
<tr>
<td>CSPw-CLARANS</td>
<td>Partitioning method</td>
<td>Medoids of clusters</td>
<td>Medoids of clusters</td>
<td>Yes, network</td>
<td>At medoids with min Constraints</td>
<td>Shortest Path distance</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CSPw-CLARANS</td>
<td>Partitioning method</td>
<td>Data points</td>
<td>Medoids of clusters</td>
<td>Yes, network</td>
<td>At medoids with min Constraints</td>
<td>Shortest Path distance</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ANP Ant-Colony-based network planning algorithm</td>
<td>Gravity center algorithm</td>
<td>Data points</td>
<td>Divide a block in 4,7,10...block</td>
<td>Yes, network constraints</td>
<td>At the gravity center</td>
<td>Shortest path distance</td>
</tr>
<tr>
<td>CWSP-PAM-ANT</td>
<td>Partitioning method</td>
<td>Data points</td>
<td>Medoids of clusters</td>
<td>Yes, network</td>
<td>At medoids with min Constraints</td>
<td>Shortest path distance</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NetPlan</td>
<td>Density-based and agglomerative clustering algorithms</td>
<td>Data points Candidate switch location</td>
<td>Core of the cluster</td>
<td>Yes, network constraints</td>
<td>near node to core of the cluster</td>
<td>Shortest path distance</td>
</tr>
<tr>
<td>CBSCAN-SP-ANT</td>
<td>Density-based and agglomerative clustering algorithms</td>
<td>Data points Candidate switch location</td>
<td>Core of the cluster</td>
<td>Yes, network constraints</td>
<td>near node At core of the cluster</td>
<td>Shortest path distance</td>
</tr>
</tbody>
</table>

satisfied. The resulting number of clusters may be 1, 4, ant-colony-based network planning algorithm (Ibrahim et al., 2005) used Gravity center to find the location of switch and applied a modified version of Ant-colony algorithm to find the shortest path. The algorithm is very powerful when the network is complicated and we have a large number of intersection and streets but has the disadvantages of Gravity center algorithm mentioned above.

CWSP-PAM (Ibrahim, 2005) algorithm depends mainly on PAM clustering algorithm. This algorithm use Floyd-Warshall algorithm to find short path.

The CWSP-PAM-ANT (Ibrahim, 2006) (Clustering with Shortest Path-PAM and ANT-Colony algorithms) modified the CWSP-PAM algorithm by introduce the Ant-Colony-Based algorithm to find the optimal path between any node and the corresponding switch node.

The NetPlan algorithm used two clustering algorithms which are density-based and agglomerative clustering algorithm using distances which are shortest paths distance calculated by Dijkstra algorithm and satisfying the network constraints. This algorithm uses wire and wireless technology to serve the subscribers demand and place the switches in a real place which are determined by the candidate locations entered by the user of the package. This algorithm has two disadvantages. First, Dijkstra algorithm fail when the network is complicated and we have a large number of intersection and streets. The second disadvantage, the algorithm does not determine the cable layout from MSAN to subscriber and does not calculate the total cost.
Fig. 7: Cost comparisons between gravity center and CBSCAN-SP-ANT algorithms

Figure 7 Compare between the total cost calculated by Gravity Center algorithm, CWSP-PAM-ANT and CBSCAN-SP-ANT algorithm when the database size is increased gradually to show the better result of CBSCAN-SP-ANT algorithm against Gravity Center algorithm and CWSP-PAM-ANT algorithm. The figure showed the decrease of cost when used CBSCAN-SP-ANT where the Gravity Center algorithm has the largest number of cluster which mean the largest number of switch and largest cost. In CBSCAN-SP-ANT algorithm the number of MSAN is smallest due to the use of mobile network when a small number of subscribers are existed in non density areas and the use of agglomerative clustering algorithms to merge small clusters where the CWSP-PAM-ANT algorithm introduce MSAN in any constructed cluster even if the number of subscribers is small (less than 100 subscribers) which increase the total cost.

CONCLUSION

Clustering analysis is one of the major tasks in various research areas. The clustering aims at identifying and extracting significant groups in underlying data. Based on certain clustering criteria the data are grouped so that the data points in a cluster are more similar to each other than points in different clusters. Ant colony algorithms are a subset of swarm intelligence and consider the ability of simple ants to solve complex problems by cooperation. In this study, we have studied the problem of network planning and the construction of optimal cable network layouts. The algorithm CBSCAN-SP-ANT algorithm is introduced to solve this problem. This algorithm used two clustering algorithms which are density-based and agglomerative clustering algorithm using shortest paths calculated by ANP algorithm, which is a modified of ant colony algorithm and also used Minimum spanning algorithm, Prime algorithm, to construct the cable layout and satisfying the network constraints. This algorithm uses wire and wireless technology to serve the subscribers demand and place the switches in a real place which are determined by the candidate locations entered by the user of the package. Experimental results and analysis indicate that CBSCAN-SP-ANT algorithm leads to minimum costs for network construction in an area where accuracy is needed and the network is complicated and we have a large number of intersections and streets.

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.

Another objective, which is currently under consideration, is to extend the present work to design a mobile network. Such networks have others complex constraints and varies of equipments.

REFERENCES

Gunes, M., U. Sorges and I. Bouazizi, 2002. ARA-the ant-colony based routing algorithm for MANETs. Proceeding of the International Conference on Parallel Processing Workshops, (ICPPW’02), Aachen University of Technology, Germany, pp: 79-85. DOI: 10.1109/ICPPW.2002.1039715