Research Journal of Applied Sciences, Engineering and Technology 4(8): 973-977, 2012
ISSN: 2040-7467
© Maxwell Scientific Organization, 2012
Submitted: November 25, 2011 Accepted: December 27, 2011 Published: April 15, 2012

Clustering in Ad Hoc Networks using Nearest Neighbor

[1]V. Anand, [1]N. Sairam and [2]M. Thiyagarajan
[1]School of Computing, SASTRA University, Thanjavur, Tamil Nadu, 613401, India
[2]School of Engineering, Amrita Vishwa Vidya Peetham, Coimbatore, Tamil Nadu, 641105, India

Abstract: In this study, we propose an algorithm to route the data from a source to destination through a path with minimum distance in an ad hoc network. To achieve this, sets of nodes based on their neighborhood are grouped to form clusters. The algorithm forms clusters and the cluster leaders are chosen based on the minimum distance. The source and destination nodes choose their cluster leader and commute through their cluster leaders. The proposed work is attempted on different set of nodes and attained the optimal path.

Key words: Clustering, cluster leader, neighbors, optimal path

INTRODUCTION

An ad hoc network is a multihop communication network, which supports users without any infrastructure. The main problem to be addressed in ad hoc networks is scalability problem (Jiangchuan *et al*., 2004). Clustering helps in solving this problem. Partitional clustering is nonhierarchical. Since a set of clusters is output, the clusters are generated based on the number of nodes. In this paper, based on the distance between nodes, clusters are formed by grouping nodes in a network. Each node is informed about its own as well as the position of its neighbors. The source of a data knows the position of the destination. Then, the cluster leaders are identified by the source and destination. Through the optimal path, they send/receive the data through their cluster leaders.

LITERATURE REVIEW

Clustering: Clustering is to group the given set of nodes. The grouping is done by computing the distance between them. The groups identified are referred as clusters. There are many types of clustering algorithms. Some of them are discussed here.

Hierarchical algorithms: In these algorithms, successive clusters are identified using the established clusters. These algorithms are either agglomerative or divisive. Agglomerative algorithms begin with each element as a separate cluster and join them into consecutively larger clusters. Divisive algorithms begin with the entire set and proceed to split it into consecutively small clusters.

Partitional algorithms: Here, all clusters are determined at once. There are various types of partitional clustering algorithms such as minimum spanning tree, k-means clustering (Tapas *et al*., 2002), nearest neighbor algorithm. In Minimum Spanning Tree (MST) algorithm, inconsistent edges are identified and removed. Then, a set of ordered pairs are formed as clusters. In k-means the set of clusters are selected at random initially. Then the items are moved between clusters till desired set is formed. High degree of similarity among items in a cluster is obtained. In nearest neighbor algorithm, items are iteratively merged into the existing clusters that are close to each other.

Clustering in Ad hoc networks: A transient association of nodes, which do not depend upon any fixed support infrastructure, is an ad hoc network. It is a wireless network used for communication. In mobile ad hoc networks, the location of the network nodes may change the topology swiftly. In topology maintenance, a quick change in topology leads to a rise in the overhead message; the clustering schemes for mobile ad hoc networks (Dali and Chan, 2007). In a multihop packet-switched network, nodes are required to route packets between the source and destination even if they are not directly connected. Ad hoc networking technology is one, which can be formed everywhere, regardless of the location. In an ad hoc network, users and computing devices will be able to connect conveniently and even transparently. In mobile ad hoc networks, nodes communicate without stationary infrastructure. Intermediate nodes act as routers, when messages are exchanged between nodes. As proposed by (Frank *et al*., 2010), node clustering is an important operation performed in graph analysis. As stated by Frank *et al*. (2010), the spectral techniques works better for larger graphs and that the sparse network model works well with smaller number of clusters. The algorithm (Alneu and Nils, 2011) improves the k-nn classification accuracy.

Independent dominating sets as clusters: Insisting that the dominating set is also an independent set can produce a relatively small number of clusters of a given graph.

Corresponding Author: V. Anand, School of Computing, SASTRA University, Thanjavur, Tamil Nadu, 613401, India

(Baker and Ephremides, 1981) has developed an algorithm, the linked cluster algorithm for ad hoc networks. In this algorithm, each node has a dedicated TDMA time slot to avoid collisions. A vertex v is chosen as a cluster-head by a neighbor u if v has the highest vertex ID within N(u). The chosen vertices form an independent dominating set (Gerla and Tsai, 1995), which proposed two clustering algorithms, one based on vertex ID and vertex degree. In the lowest-ID algorithm, each vertex with the lowest ID is chosen as a cluster-head. In the highest degree algorithm, each vertex with the highest degree is the cluster head. The cluster-heads chosen by these two algorithms form an independent set. But, these algorithms do not work on all graphs.

Dominating sets as clusters: When the network topology changes, choosing independent dominating sets as clusters creates problems. A distributed greedy algorithm (Liang and Haas, 2000) for dominating Sets, which is a centralized greedy algorithm. (Wu and Li, 1999; 2001) developed a localized distributed algorithm for finding small connected dominating sets in which each node only needs to know its distance two neighborhood.

PROPOSED APPROACH

The objective of the proposed stuy is to find the optimal path between source and destination in an ad hoc network, so that data can be routed in efficient way. In the proposed research, the objective is achieved by implementing the three phases: cluster generation, cluster leader identification and optimal path identification. Grouping nodes based on the nearest neighbor algorithm forms clusters. Then, each node in a cluster identifies its leader, which may be itself. The cluster leader forwards the data to the leader of the designated cluster. The optimal path between the source and destination node is determined by computing the minimum distance.

Cluster generation: In this phase, the cluster is generated using nearest neighbor algorithm. A node ‘i’ is randomly selected and added to first cluster. Subsequently, another node ‘j’ is chosen and the distance between the i^{th} node and j^{th} node is compared with the threshold of the first cluster and decides whether to add the j^{th} node to the first cluster or not. The process is repeated for each node in the network, until the first cluster is formed with a set of nodes. This is done iteratively until all the clusters are generated with the remaining nodes (nodes not yet added to the clusters) in the network, based on the threshold of the clusters.

Cluster leader identification: This part of the proposed approach aims at identifying the cluster leaders. The node that has data to transmit (source) and the node to receive the data (destination) identify its leader node, based on the minimum distance between the source/destination and the other nodes in the cluster. The cluster leader changes from time to time. There is a chance for each node to be a cluster leader.

Optimal path identification: Here, identifying an optimal path is possible when the source and destination are in the same cluster or in different clusters. When the source and destination are in the same cluster, irrespective of the direct path between them, the data are routed through the identified optimal path (minimum distance). When the source and destination are in different clusters, the optimal path through which the data are routed through the cluster leaders or through the direct path, if exists, whichever is optimal.

IMPLEMENTATION DETAILS

Cluster generation: Clusters are generated based on the nearest neighbor algorithm. Based on the threshold of the clusters, each node in the network is added to the clusters.

Pseudocode for cluster generation:

```
Get the distance matrix d.
FOR  k=1 to n

t = {set of nodes yet to be include in cluster}
kc= t(1);

FOR EACH node j in kc
FOR EACH node i in t do the following

IF d(i, j) <= k^th threshold
add node i in kc

END IF
END FOR

END FOR
Add the elements in kc in to the k^th cluster
ENDFOR
```

Cluster leader identification: The source/destination in a cluster identifies its leader, which may be itself. The cluster leader forwards the data to the leader of the designated cluster.

Pseudocode for cluster leader identification:

```
sclus= cluster containing source node
FOR EACH node i in   sclus

IF d(source,i)<=thresh(sclus)
 source_lead=i

END IF  ENDF OR
Similary find desti_lead
```

d =

0	2	4	10	3	6	4	7	10	8	7	2	2	1	4	8	10	1
2	0	1	2	1	2	1	2	2	2	2	1	1	1	1	2	2	1
4	1	0	4	2	3	2	3	4	3	3	1	1	1	2	3	4	1
10	2	4	0	3	6	4	8	11	9	8	2	2	1	4	9	11	1
3	1	2	3	0	2	2	2	3	3	2	1	1	1	2	3	3	1
6	2	3	6	2	0	3	5	6	5	5	2	2	1	3	5	6	1
4	1	2	4	2	3	0	3	4	3	3	1	1	1	2	3	4	1
7	2	3	8	2	5	3	0	8	7	6	2	2	1	3	7	8	1
10	2	4	11	3	6	4	8	0	9	8	2	2	1	4	9	11	1
8	2	3	9	3	5	3	7	9	0	7	2	2	1	3	7	9	1
7	2	3	8	2	5	3	6	8	7	0	2	2	1	3	7	8	1
2	1	1	2	1	2	1	2	2	2	2	0	1	1	1	2	2	1
2	1	1	2	1	2	1	2	2	2	2	1	0	1	1	2	2	1
1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1
4	1	2	4	2	3	2	3	4	3	3	1	1	1	0	3	4	1
8	2	3	9	3	5	3	7	9	7	7	2	2	1	3	0	9	1
10	2	4	11	3	6	4	8	11	9	8	2	2	1	4	9	0	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0

clus =

1	14	15	16	17	18	0	0	0	0	0	0
2	3	4	5	6	7	8	9	10	11	12	13

Enter the source 17
Enter the destination 10
y = 4
x = 9
path = 17   14   2   10
distance = 4

Fig. 1: Source and destination in different clusters, with optimal path containing cluster leaders.

d =

0	2	2	3	2	2	3	2	3	3	3	1	3	3	1	1	3	2
2	0	2	4	2	3	5	3	5	4	5	1	4	5	2	1	4	4
2	2	0	3	2	3	4	3	4	3	4	1	3	4	2	1	3	3
3	4	3	0	3	6	9	5	8	7	8	2	7	8	3	2	7	7
2	2	2	3	0	3	4	3	4	3	4	1	3	4	2	1	3	3
2	3	3	6	3	0	7	4	6	6	6	2	6	6	2	2	6	5
3	5	4	9	4	7	0	6	10	9	10	2	9	10	3	2	9	8
2	3	3	5	3	4	6	0	6	5	6	2	5	6	2	2	5	5
3	5	4	8	4	6	10	6	0	8	9	2	8	9	3	2	8	7
3	4	3	7	3	6	9	5	8	0	8	2	7	8	3	2	7	7
3	5	4	8	4	6	10	6	9	8	0	2	8	9	3	2	8	7
1	1	1	2	1	2	2	2	2	2	2	0	2	2	1	1	2	2
3	4	3	7	3	6	9	5	8	7	8	2	0	8	3	2	7	7
3	5	4	8	4	6	10	6	9	8	9	2	8	0	3	2	8	7
1	2	2	3	2	2	3	2	3	3	3	1	3	3	0	1	3	2
1	1	1	2	1	2	2	2	2	2	2	1	2	2	1	0	2	2
3	4	3	7	3	6	9	5	8	7	8	2	7	8	3	2	0	7
2	4	3	7	3	5	8	5	7	7	7	2	7	7	2	2	7	0

clus									
	1	12	15	16	0	0	0	0	0
	2	3	50	0	0	0	0	0	0
	4	0	00	0	0	0	0	0	0
	6	7	89	10	11	13	14	17	18

Enter the source 7
Enter the destination 16
y = 4
x = 2
path = 7   16
distance = 2

Fig. 2: Source and destination in different clusters, with direct link as optimal path.

```
d =

     0     7     6     7     3     4     2     1     2     7     3     5     7     2     7     6     7     5
     7     0     7     8     3     4     3     1     2     8     3     5     8     3     8     7     8     5
     6     7     0     7     3     4     2     1     2     7     3     5     7     2     7     6     7     5
     7     8     7     0     4     5     3     1     2     9     4     6     9     3     9     7     9     6
     3     3     3     4     0     2     2     1     1     4     2     3     4     2     4     3     4     3
     4     4     4     5     2     0     2     1     1     5     2     3     5     2     5     4     5     3
     2     3     2     3     2     2     0     1     1     3     2     2     3     1     3     2     3     2
     1     1     1     1     1     1     1     0     1     1     1     1     1     1     1     1     1     1
     2     2     2     2     1     1     1     1     0     2     1     2     2     1     2     2     2     2
     7     8     7     9     4     5     3     1     2     0     4     6     9     3     9     7     9     6
     3     3     3     4     2     2     2     1     1     4     0     3     4     2     4     3     4     3
     5     5     5     6     3     3     2     1     2     6     3     0     6     2     6     5     6     4
     7     8     7     9     4     5     3     1     2     9     4     6     0     3     9     7     9     6
     2     3     2     3     2     2     1     1     1     3     2     2     3     0     3     2     3     2
     7     8     7     9     4     5     3     1     2     9     4     6     9     3     0     7     9     6
     6     7     6     7     3     4     2     1     2     7     3     5     7     2     7     0     7     5
     7     8     7     9     4     5     3     1     2     9     4     6     9     3     9     7     0     6
     5     5     5     6     3     3     2     1     2     6     3     4     6     2     6     5     6     0


clus =

     1     8     9    10    11    12    13    14    15    16    17    18
     2     0     0     0     0     0     0     0     0     0     0     0
     3     5     6     7     0     0     0     0     0     0     0     0
     4     0     0     0     0     0     0     0     0     0     0     0
```

Enter the source 17
Enter the destination 10
y = 2
x = 9
path = 17   8   10
distance   = 2
Enter the 1to continuel

Fig. 3: Source and destination in same cluster through intermediate node.

Optimal path identification: The optimal path between the source and destination node is determined by computing the minimum distance between them.

Pseudocode for optimal path identification:

f1 = d (source,source_lead); f2 = d(source_ lead, desti_ lead)
f3 = d(desti_lead, destination)

y = f1+f2+f3
x = d(source, destination)

IF x $\leq$ y then
distance = x

path = [source, destination]
ELSE

distance = y
path = [source, source_lead, desti_lead, destination]
END IF

Advantages and limitations of the proposed approach: The advantages of the proposed approach are:

- Scalability of the network is improved by clustering
- Works fast, since the data are transmitted through the optimal path from source to destination, regardless of the cluster.
- Effective for large number of nodes.

The limitation of the proposed approach is, grouping of nodes to form cluster depends more on threshold.

EXPERIMENTAL RESULTS

The experiments were done with large amount of data. The Fig. 1 shows the results of the experiments done with the source and destination in different clusters, with optimal path containing cluster leaders. The results of the experiments done with source and destination in different clusters, with direct link as optimal path are shown in Fig. 2. Initially, the results are shown by having the source and destination in different clusters. Then, the

Fig. 3 shows the results with source and destination in same cluster through intermediate nodes.

CONCLUSION

The proposed approach takes as input the source and destination and determines the optimal path, based on whether they are in the same cluster or in different clusters. The proposed approach is effective for large number of nodes, since the nearest neighbor algorithm is used. The results prove that it would work fast. But, the computation cost is to be considered for large number of nodes. Besides cluster generation, cluster leader identification, and optimal path identification, there are a lot of research areas relevant to clustering are in the limelight.

REFERENCES

Alneu de Andrade Lopes and Nils Murrugarra-Llerena, 2011. An Adaptive Graph-Based K-Nearest Neighbor, ICMC, USP.

Baker, D. J. and A., Ephremides, 1981, The architectural organization of a mobile radio net work via a distributed algorithm. IEEE Trans. Commun. COM, 29: 1694-1701.

Dali, W. and H.A. Chan, 2007. Clustering Ad Hoc Networks: Schemes and Classifications, Annual IEEE Communications Society on Sensor and Ad Hoc Communications and Networks, pp: 920-926.

Frank, L., B. Ramnath and W.C. Willia, 2010. Node Clustering in Graphs: An Empirical Study, NIPS, Vancouver, BC, Canada.

Gerla, M. and J.T.C. Tsai, 1995. Multicluster, mobile, multimedia radio network. Wireless Networks, 1: 255-265.

Jiangchuan, L., P.C. Yuanzhu and L.L. Arthur, 2004. Clustering Algorithms for Ad Hoc Wireless Networks. In: Pan, Y. And Y. Xiao (Eds.), Ad Hoc and Sensor Networks. Nova Science Publishers.

Liang, B. and Z.J. Haas, 2000. Virtual backbone generation and maintenance in ad hoc network mobility management, INFOCOM, pp: 1293-1302.

Tapas, K., M.M. David, S.N. Nathan, D.P. Christine, S. Ruth and Y.W. Angela, 2002. An efficient k-Means clustering algorithm: Analysis and implementation. IEEE Trans. Pattern Analy. Mach. Intell., 24(7): 881-892.

Wu, J. and H. Li, 1999. On calculating connected dominating set for efficient routing inad hoc wireless networks, DIAL M'99, Seattle, pp: 7-14.

Wu, J. and H. Li, 2001. A dominating-set-based routing scheme in ad hoc wireless networks, Special Issue on Wireless Networks. Telecommunication Syst. J., 3: 63-84.