Enhancement of RDMAR Protocol by k-means Distance-Based Nodes-Clustering in a MANET

Omar F Hamad, Mi-Young Kang, Jin-Han Jeon, Ji-Seung Nam

Multimedia Data Communications Lab., Department of Computer Engineering
Chonnam National University, Gwangju - 500 757, South Korea

omarfh@gmail.com
kmy2221@yahoo.co.kr
jhjeon23@naver.com
jsnam@chonnam.ac.kr

***Abstract*— k-means distance-based nodes clustering technique proposed enhance the performance of RDMAR protocol in a Mobile Ad-hoc NETwork (MANET). To limit the flood search to just a circular local area around the source, the Relative Distance Micro-discovery Ad Hoc Routing (RDMAR) protocol uses the Relative Distance (RD). If the distance of flood discovery is further limited by clustering the nodes with similar characters in to one group, different from the dissimilar characters' group, the performance of the RDMAR implementation can be elevated. The k-means algorithm, similar to the one in unsupervised learning in pattern classification, can be recursively applied to re-classify the clusters as the MANET environment, resource availability, and node demands change. This technique can be more effective in a MANET with comparatively moderate change of the dynamicity and slow change in nodes' demands plus highly accumulated groups of nodes at given sub-areas.**

***Keywords*:** k- means algorithm, distance-based clustering, RDMAR, MANET.

I. INTRODUCTION

Nodes clustering in a MANET can be considered as an efficient and the most important unsupervised Mobile Ad Hoc Network learning problem that deals with finding a character in a collection of unlabeled information nodes. Loosely, nodes clustering can be said to be the process of organizing nodes into groups (physically or logically related) whose members are similar in some way. It is to collect nodes which are physically or logically "similar" between them and are "dissimilar" to the nodes belonging to the other nodal clusters.

If we consider a simple graphical example of mobile nodes' distribution at a certain time, t_o, as shown in Figure 1, and if we consider distance being the main character or feature or criterion for nodes clustering, we can clearly observe that two or more nodes belong to the same Nodes-Cluster if they are "close" according to a given geometrical distance, as shown in Figure 2.

This geometrical distance based nodes clustering is what we, in this paper, refer as *Distance-Based Nodes Clustering*. In clustering the nodes, there is no absolute "best" criterion and, therefore, the nodes must supply this criterion so that the nodes' results suit the majority nodes' needs. In nodal clustering, the main idea is to cluster MANET's nodes discover groups of nodes of similar access or communication patterns.

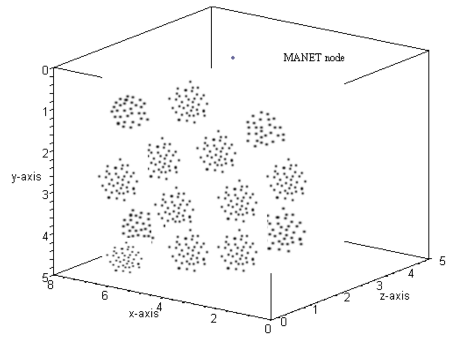

Fig. 1: MANET's un-clustered nodes

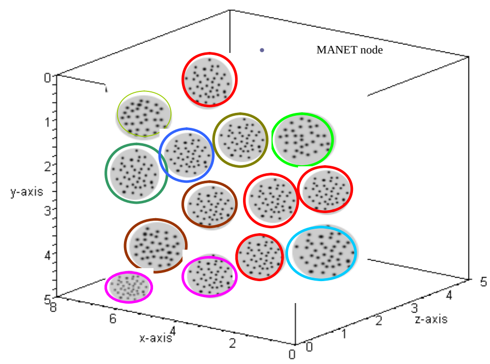

Figure 2: MANET's distance-based clustered nodes

II. MANET'S NODES CLUSTERING ALGORITHM

The MANET's nodes clustering algorithm should satisfy the scalability requirements as well as be able to deal with different types of criterion and characters. Besides, the algorithm must be able to discover nodes clusters with arbitrary shape and configuration as diverse as possible. The algorithm must possess minimal requirements for domain knowledge to determine input parameters for nodes clustering, it must have ability to deal with noise and outliers nodes, and it must be insensitive to order of input records. In addition to that, this algorithm must resist high dimensionality, interpretability, and usability to the large amount of candidate nodes.

In this proposal, the exclusive nodes clustering algorithm (k-means nodes clustering) will be discussed. The classification algorithms that can be thought of include the hierarchical nodes clustering, overlapping (Fuzzy C-means) nodes clustering, and probabilistic (Mixture of Gaussians) nodes clustering.

In the case of k-means distance based nodes clustering, if an obvious distance measure doesn't exist, we must define it. It is always not easy especially in multi-dimensional spaces. However, in our case, we will be defining the Relative Distance (RD), the one used in RDMAR protocol, as our distance measure.

III. K-MEANS NODES CLUSTERING

The procedure of k-means nodes clustering follows a simple and easy way to classify a given MANET's nodes set through a certain number of clusters (assume k clusters) fixed a priori. The Procedure to classify the MANET's nodes into k clusters can be simplified as below: -

- According to the geometric distances, define k central nodes, one for each nodes cluster.
- The central nodes should be placed in a cunning way because different locations have different result.
- So, the better choice is to place them as much as possible far away from each other.
- Make each node belonging to a given cluster and associate it to the nearest central node.
- When no node is pending, the first step is completed and an early groupage is done.
- At this point we need to re-calculate k new central nodes as barycenters of the clusters resulting from the previous step. This includes the new joined nodes and excludes the nodes that have left the MANET.
- After we have these k new central nodes, a new binding has to be done between the same MANET nodes set and the nearest new central nodes. A loop will be generated recursively.
- As a result of this loop we may notice that the k central nodes change their local location step by step until no more changes are done. In other words central nodes do not move any more.
- Finally, this algorithm aims at minimizing an *objective function*, in this case a squared error function. The objective function,

$$J = \sum_{j=1}^{n} \left(\sum_{i=1}^{k} \left| -x_i^j + C_j \right|^2 \right)$$

where,

$$\left| -x_i^j + C_j \right|^2$$

is a chosen distance measure between a MANET node

$$x_i^j$$

and the nodes cluster centre

$$c_j$$

is an indicator of the distance of the *n* nodes from their respective nodes cluster centers.

Finally, when no more logical moving of the MANET nodes outside the localized area (nodes cluster), then the RDMAR protocol is applied.

This enhance the performance of the RDMAR protocol in the sense that the RDMAR protocol does not have to apply RDM function for the whole MANET possible area, but rather only on the nodes cluster area or region. In case it happens that a new node or two old nodes from different nodes clusters want to communicate, then the k-means algorithm is recursively run and the new central nodes re-created.

IV. A SIMPLEST EXAMPLE SCENARIO

Let us look at common used scenario as depicted in Figure 3 and let us assume that n distance vectors $\mathbf{x}_1$, $\mathbf{x}_2$, ..., $\mathbf{x}_n$ all from the same MANET nodes region and that they fall into k compact nodes clusters such that $k \le n$ and that $\mathbf{m}_i$ is the mean of the vectors in nodes cluster i.

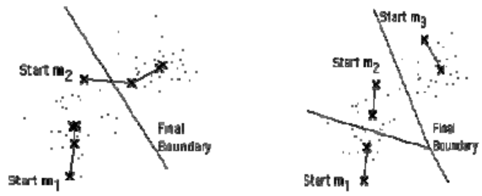

Figure 3: A simple example scenario

If the nodes clusters are well separated, we can use a minimum-distance clustering to separate them. We can say that $\mathbf{x}$ is in nodes cluster i if $\| \mathbf{x} - \mathbf{m}_i \|$ is the minimum of all the k distances. This suggests that, for finding the k-means of nodes clusters, we make initial guesses for the means $\mathbf{m}_1$, $\mathbf{m}_2$, ..., $\mathbf{m}_k$. Until there are no changes in any nodes clusters' mean, we use the estimated means to cluster the nodes into nodes clusters.

Simply, for i from 1 to k, we replace $\mathbf{m}_i$ with the mean of all of the odes for nodes cluster I, and then we end the loop which is repeated recursively. Figure 3 shows a simple version of the k-means procedure. It can be viewed as a greedy algorithm for partitioning the n nodes into k nodes clusters so as to minimize the sum of the squared distances to the cluster central nodes.

In general, there is no general theoretical solution to find the optimal number of nodes clusters for any given MANET's set of nodes. A simple approach is to compare the results of multiple runs with different k nodes clusters and choose the best one according to a given criterion and the performance required against the trade-off. Care need to be taken because increasing k results in smaller error function values.

V. A SIMPLE EXAMPLE SCENARIO

From Figure 4, we can observe that there are four similar and dissimilar groups of nodes around a MANET region. The similar groups are, respectively, marked with letters A, B, C, and D. With conventional RDMAR, the RDM process involves the whole region regardless of the fact that the nodes in one group, say C, never communicate with the one in B.

This may increase the resources required for route establishment, maintenance, and destroying. Conversely, in Figure 5, the MANET nodes are grouped into four similar nodes clusters and hence there are four separately done RDMAR operations when needed.

If the node marked A intends to communicate with any node marked A, there is a simple use of RDMAR protocol at the local region and hence the route discovery, maintenance and destroying. Hence there is improvement in performance of the original RDMAR protocol.

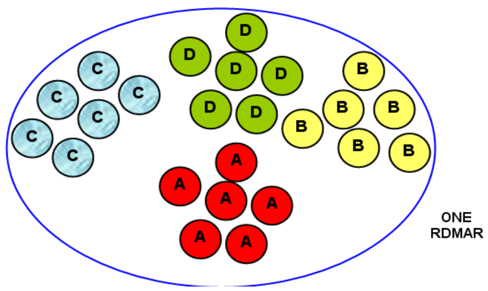

Figure 4: RDMAR on the whole region – Inefficient!

In MANET, the criterion for nodes clustering may differ from service to service. However, in this particular case, we use Euclidean distance as the criterion for nodes clustering and hence the nodes with closer geometrical resemblance belong to one nodes cluster.

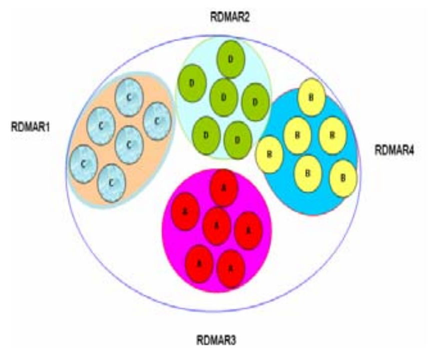

Figure 5: RDMAR on the nodes clusters regions – Efficient!

To further refine the nodes clustering process, we increase the number of k. In the above case (Figure 5), the 4 nodes clusters correspond to k=1 which is always larger or equal to the number of all the expected nodes on the MANET regions (24 in this case).

VI. SIMULATION FOR K-MEANS NODES CLUSTERING

An assumption is made, here, that the MANET nodes around the region are not changing their characters once they have joined the region. Furthermore, the mobility is around the pre-defined region. The best examples where this k-means distance-based nodes clustering can be applied include the university campuses, sports complex during the tournament, airport complexes, and such types of regions.

TABLE I

THE GEOMETRICAL POSITIONS OF 150 NODES AROUND MANET'S REGION

x	y	z	x	y	z	x	y	z	x	y	z	x	y	z	x	y	z
5.1	3.5	1.4	5.0	3.0	1.6	7.0	3.2	4.7	6.6	3.0	4.4	6.3	3.3	6.0	7.2	3.2	6.0
4.9	3.0	1.4	5.0	3.4	1.6	6.4	3.2	4.5	6.8	2.8	4.8	5.8	2.7	5.1	6.2	2.8	4.8
4.7	3.2	1.5	5.2	3.5	1.5	6.9	3.1	4.9	6.7	3.0	5.0	7.1	3.0	5.9	6.1	3.0	4.9
4.6	3.1	1.5	5.2	3.4	1.4	5.5	2.3	4.0	6.0	2.9	4.5	6.3	2.9	5.6	6.4	2.8	5.6
5.0	3.6	1.4	4.7	3.2	1.6	6.5	2.8	4.6	5.7	2.6	3.5	6.5	3.0	5.8	7.2	3.0	5.8
5.4	3.9	1.7	4.8	3.1	1.6	5.7	2.8	4.5	5.5	2.4	3.8	7.6	3.0	6.6	7.4	2.8	6.1
4.6	3.4	1.4	5.4	3.4	1.5	6.3	3.3	4.7	5.5	2.4	3.7	4.9	2.5	4.5	7.9	3.8	6.4
5.0	3.4	1.5	5.2	4.1	1.5	4.9	2.4	3.3	5.8	2.7	3.9	7.3	2.9	6.3	6.4	2.8	5.6
4.4	2.9	1.4	5.5	4.2	1.4	6.6	2.9	4.6	6.0	2.7	5.1	6.7	2.5	5.8	6.3	2.8	5.1
4.9	3.1	1.5	4.9	3.1	1.5	5.2	2.7	3.9	5.4	3.0	4.5	7.2	3.6	6.1	6.1	2.6	5.6
5.4	3.7	1.5	5.0	3.2	1.2	5.0	2.0	3.5	6.0	3.4	4.5	6.5	3.2	5.1	7.7	3.0	6.1
4.8	3.4	1.6	5.5	3.5	1.3	5.9	3.0	4.2	6.7	3.1	4.7	6.4	2.7	5.3	6.3	3.4	5.6
4.8	3.0	1.4	4.9	3.6	1.4	6.0	2.2	4.0	6.3	2.3	4.4	6.8	3.0	5.5	6.4	3.1	5.5
4.3	3.0	1.1	4.4	3.0	1.3	6.1	2.9	4.7	5.6	3.0	4.1	5.7	2.5	5.0	6.0	3.0	4.8
5.8	4.0	1.2	5.1	3.4	1.5	5.6	2.9	3.6	5.5	2.5	4.0	5.8	2.8	5.1	6.9	3.1	5.4
5.7	4.4	1.5	5.0	3.5	1.3	6.7	3.1	4.4	5.5	2.6	4.4	6.4	3.2	5.3	6.7	3.1	5.6
5.4	3.9	1.3	4.5	2.3	1.3	5.6	3.0	4.5	6.1	3.0	4.6	6.5	3.0	5.5	6.9	3.1	5.1
5.1	3.5	1.4	4.4	3.2	1.3	5.8	2.7	4.1	5.8	2.6	4.0	7.7	3.8	6.7	5.8	2.7	5.1
5.7	3.8	1.7	5.0	3.5	1.6	6.2	2.2	4.5	5.0	2.3	3.3	7.7	2.6	6.9	6.8	3.2	5.9
5.1	3.8	1.5	5.1	3.8	1.9	5.6	2.5	3.9	5.6	2.7	4.2	6.0	2.2	5.0	6.7	3.3	5.7
5.4	3.4	1.7	4.8	3.0	1.4	5.9	3.2	4.8	5.7	3.0	4.2	6.9	3.2	5.7	6.7	3.0	5.2
5.1	3.7	1.5	5.1	3.8	1.6	6.1	2.8	4.0	5.7	2.9	4.2	5.6	2.8	4.9	6.3	2.5	5.0
4.6	3.6	1.0	4.6	3.2	1.4	6.3	2.5	4.9	6.2	2.9	4.3	7.7	2.8	6.7	6.5	3.0	5.2
5.1	3.3	1.7	5.3	3.7	1.5	6.1	2.8	4.7	5.1	2.5	3.0	6.3	2.7	4.9	6.2	3.4	5.4
4.8	3.4	1.9	5.0	3.3	1.4	6.4	2.9	4.3	5.7	2.8	4.1	6.7	3.3	5.7	5.9	3.0	5.1

The main purpose of this simulation is to cluster the nodes in the data file, here in this paper, referred as "kmean_nodes_clusters data" using K-Means Distance-Based Nodes Clustering.

The specific objective it to perform K-Means nodes clustering using an iterative algorithm that assigns nodes in a MANET's region to nodes clusters so that the sum of distances from each node to its nodes cluster central node, over all nodes clusters, is as minimum as possible.

The simulation, on the "kmean_nodes_clusters.dat", will find the natural groupings among the MANET's nodes which can later be used to apply the RDMAR protocol in a sub-region localized among the regions with similar characters.

With k-means Distance-Based nodes clustering, we firstly must specify the number of nodes clusters that we wish to create. We load the data file "kmean_nodes_clusters.dat"and call k-means function with the desired number of nodes clusters set to k, and using squared Euclidean distance to get an idea of how well-separated the resulting nodes clusters are, we can make a silhouette plot. The silhouette plot displays a measure of how close each node in one cluster is to nodes in the neighboring nodes clusters.

VII. APPLICABILITY OF THE SCHEME IN DEVELOPING COUNTRIES

Considering that the nodes in Figure 4 belong to a local school with students grouped into four similar and dissimilar groups of nodes around a MANET region. With conventional RDMAR, the RDM process involves the whole region regardless of the fact that classes making the nodes in each group never communicate with the others in other groups. This may increase the resources required for route establishment, maintenance, and destroying.

In Figure 5, the MANET nodes are grouped into four similar nodes clusters of the classes and hence there are four separately done RDMAR operations when needed, each belonging to one classroom.

If the nodes marked A, belonging to one class, intends to communicate with any node marked A, the same class, there is a simple use of RDMAR protocol at the local region and hence the route discovery, maintenance and destroying. Hence there is improvement in performance of the original RDMAR protocol by applying at individual classes rather than applying a single protocol for the dissimilar classes.

The criterion for nodes clustering classes may differ from service to service. However, in the case of the different class, an equivalent of Euclidean distance is used as the criterion for nodes clustering and hence the nodes with closer geometrical resemblance belong to one nodes cluster which means a separate class.

The classification process is further increased by increasing the number of k. In the above case of the classes (as exemplified in Figure 5), the 4 nodes clustered classes correspond to k=1 which is always larger or equal to the number of all the expected nodes on the MANET regions.

VIII. SIMULATION RESULTS

The silhouette plot displays a measure of how close each node in one nodes cluster is to nodes in the neighboring clusters. For k=2, the nodes clustering results were as shown here in Figure 6.

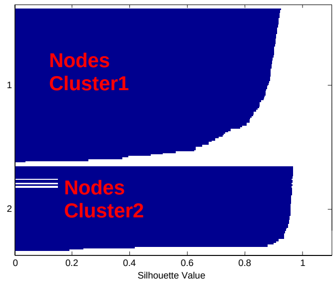

Figure 6: How close nodes from one cluster are to the other clusters; k=2

For k=3, we have three node clusters and the way the nodes are separated is in Figure 7.

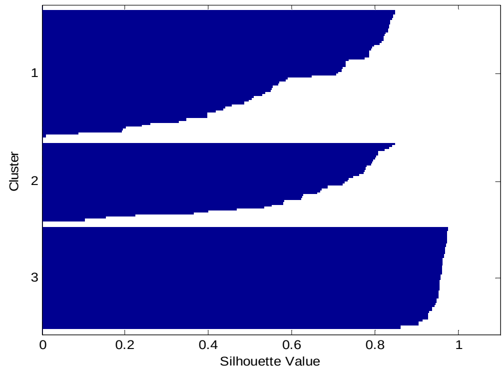

Figure 7: How close nodes from one cluster are to the other clusters; k=3

For k=4 through 10, the nodes separation relationship is, in Figure 8 through 14.

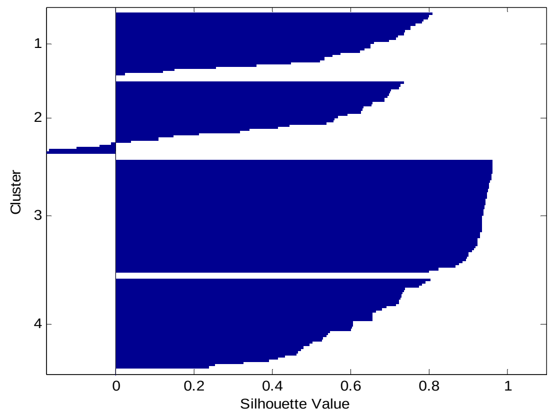

Figure 8: How close nodes from one cluster are to the other clusters; k=4

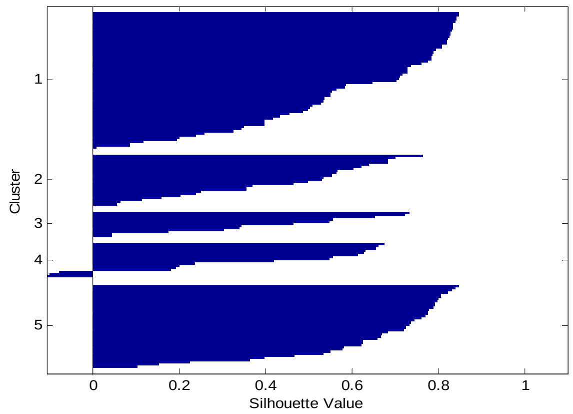

Figure 9: How close nodes from one nodes cluster are to the other nodes clusters; k=5

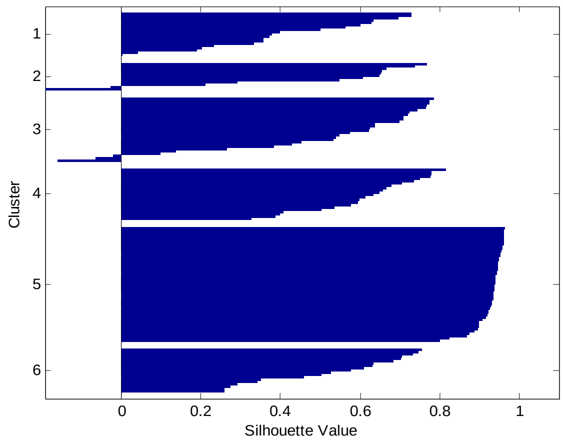

Figure 10: How close nodes from one nodes cluster are to the other nodes clusters; k=6

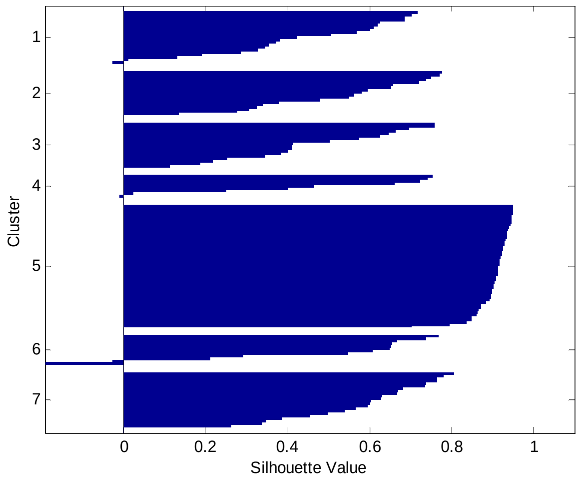

Figure 11: How close nodes from one nodes cluster are to the other nodes clusters; k=7

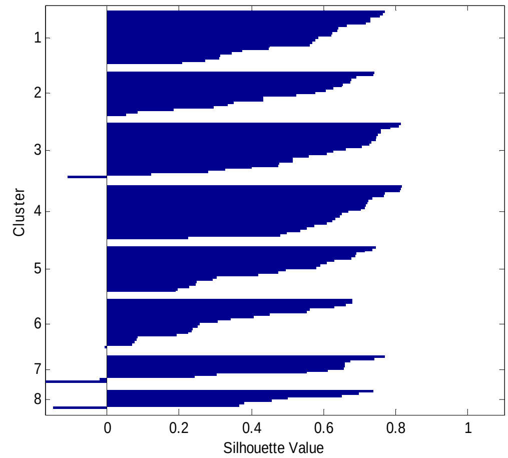

Figure 12: How close nodes from one nodes cluster are to the other nodes clusters; k=8

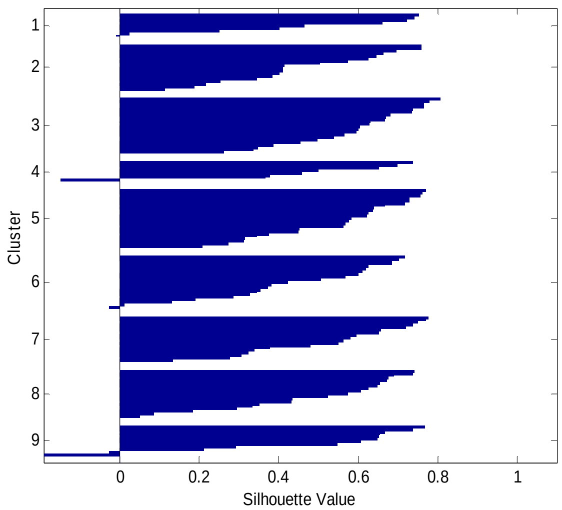

Figure 13: How close nodes from one nodes cluster are to the other nodes clusters; k=9

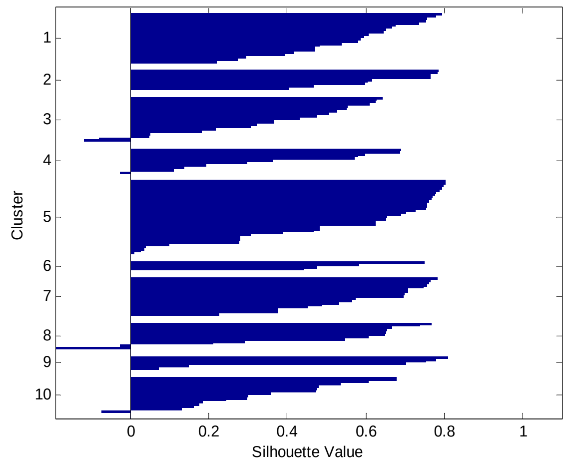

Figure 14: How close nodes from one nodes cluster are to the other nodes clusters; k=10

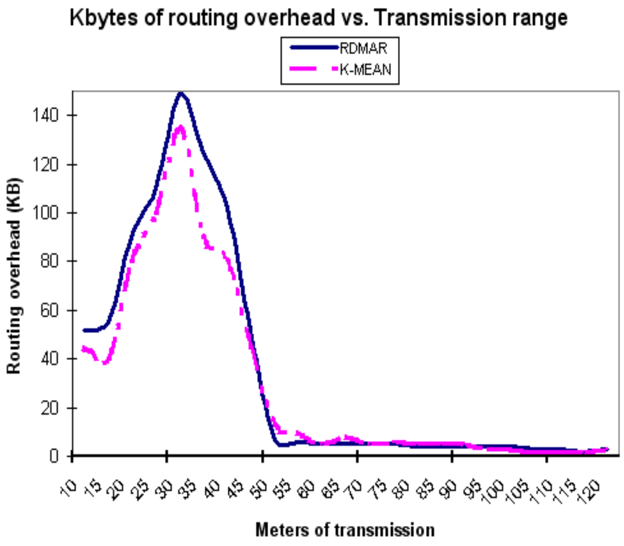

Figure 15: Kilobytes of routing overhead vs. Meters of transmission distance

In terms of the routing overhead, Fig. 15 shows that there is a little improvement especially when the number of ah-hoc nodes is less, while the overhead is quite similar to the original RDMAR protocol for the node beyond 48.

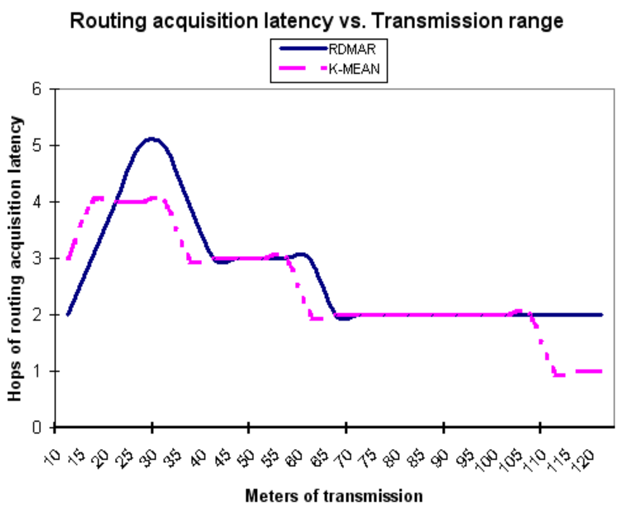

Figure 16: Hops of routing acquisition latency vs. Meters of transmission distance

Fig. 16 shows that the routing acquisition latency is either equal to or better than the traditional RDMAR. For the nodes beyond 110, the average latency seems to favorably drop.

IX. CONCLUSION

In this paper, a possibility of enhancing the RDMAR protocol's performance has been suggested. The suggestion introduces the neural network based unsupervised learning method of k-means distance-based nodes clustering. In this mechanism, the nodes are categorized as per the characteristics keeping them together.

Euclidean distance has been used as the criterion to cluster the MANET's nodes and the resultant nodes clusters can be used to implement RDMAR locally. Instead of the RDMAR being applied over the whole MANET region, the region is divided into sub-local regions and sub-RDMARs are applied (RDMAR1, RDMAR2, RDMAR3, RDMAR4, etc).

Among the disadvantages of this scheme are: the fact that the k-means Distance-Based Nodes Clustering might not be a very effective scheme as the learning time for the system might be a constraint while nodes are waiting for the path to communicate; the fact that the mobility of the nodes in a give MANET region is not certain and hence the node can not guarantee to remain at a give nodes cluster for a certain expected time; and the recursion in operating the k-means algorithm might take longer when the nodes are very unstable.

The technique here presented can be more effective in a MANET with comparatively moderate change of a dynamicity and slow change in nodes' demands plus highly accumulated groups of nodes at given sub-areas. The higher the number of the groupings, the higher the complexity and the better the service classification.

REFERENCES

[1] Aamir Ansaari and Dileep Perchani, Enhanced MANET Routing and Service Discovery In Konark.

[2] A Tutorial on Clustering Algorithms – k-means clustering.

[3] G. Agelou and R. Tafazolli, RDMAR: A bandwidth-efficient Routing Protocol for Mobile Ad Hoc Netwroks.

[4] Murthy and Manoj, Ad Hoc Wireless Netwroks: architecture and Protocols.

[5] Lecture materials – Internet Engineering Course, Chonnam National University, Gwangju, Korea - Spring 2006 – Prof. Park Jae Hyung.

[6] Rui Xu and Donald Wunsch: Survey of Clustering Algorithms: In IEEE TRANSACTIONS ON NEURAL NETWORKS, VOL. 16, NO. 3, MAY 2005. pp 645-679.

[7] Suman Banerjee, Samir Khuller: A Clustering Scheme for Hierarchical Control in Multi-hopWireless Networks. In WINS 2005.

[8] Wendi B. Heinzelman, Chandrakasan, and Hari Balakrishnan: An Application-Specific Protocol Architecture For Wireless Microsensor Networks: In IEEE Transactions On Wireless Communications, Vol. 1, No. 4, October 2002.