Optimal Clustering Technique Using K-means like Algorithm in Wireless Sensor Networks

Yashodha K Bilagi, Vinutha C B, M Z Kurian

1M.Tech Digital Electronics Dept. of ECE SSIT, Tumakuru 572105, Karnataka, India
2Assistant Professor, Dept. of ECE SSIT, Tumakuru 572105, Karnataka, India
3HOD, Dept. of ECE SSIT, Tumakuru 572105, Karnataka, India

Abstract — The advances in wireless communication, digital electronics and microelectromechanical systems (MEMS) technology have given rise to the development of low cost, low power multifunctional sensor nodes which are in smaller size and capable of short distance communication. Most of the times, the sensor nodes have to operate autonomously in an unattended environments where the recharging or replacement of the battery energy is difficult. Therefore, it is very essential to utilize the available sensor node battery energy in an efficient manner.

In this direction, the proposed work considers grouping of sensor nodes into clusters, which effectively increases the scalability & reduces power consumption. Presently, an efficient k-medoids algorithm is used in order to bring about optimal number of clusters which results in improved load balancing & increased lifetime of the network. The proposed work also considers the basic issue of identifying the mobile node location, using samples of RSSI measurement.

Index Terms—Clusters, K-medoids, RSSI, WSN.

I. INTRODUCTION

Wireless Sensor Network (WSN) can be defined as a special class of ad hoc wireless network that can be used to provide a wireless communication infrastructure that allows us to sense, observe and react to the events & phenomena in the natural environment.

The WSN architecture comprises of sensor nodes which are scattered in sensor field & each of the nodes have the capability to collect the data & route the data back to the sink (base station). Data are routed back to the sink (base station) by a multihop infrastructures architecture through the sink. The sink (base station) may communicate with the task manager node via internet or satellite. WSN can be used in many applications[1].

There is no tethered power supply is available in the network. Some form of batteries are necessary to provide energy. Sometimes, some form of recharging by obtaining energy from the environment is available as well (eg: solar cell). But, in some environment where human operating is not possible, in such environment, there is no way to replace the batteries for communication. Hence, power will become crucial part in WSN. Thus, we can reduce the power consumption by some of the important methods like routing, localization and MAC protocol. Among these three protocol, Medium Access Control (MAC) protocol is more helpful as it consumes less power & is more efficient. MAC protocol controls the communication module of WSN node[3].

One of the advantages of wireless sensors networks (WSNs) is their ability to operate in harsh environments in which human monitoring schemes are risky, inefficient and sometimes infeasible. Thus, sensors are expected to be scattered randomly in the area of interest by a relatively uncontrolled means. Grouping of sensor nodes into clusters has been widely pursued by the research community to achieve the network scalability objective [4].

Wireless Sensor Network are tremendously being used in different environment to perform various tasks such as target tracking, search, disaster relief & number of tasks in smart environment. In such applications, node localization is one of system parameter. Hence, it is one of the challenge in WSN. There are different methods used in localization to estimate the distance between two nodes: RSSI, TOA, TDOA & AOA. Received Signal Strength Indicator (RSSI) is more preferred method as it gives accurate distance measure of the nodes[7].

II. RELATED WORK & CONTRIBUTIONS

I.F. Akyildiz, W. Su, Y. Sankarasubramaniam, E. Cayirci (2003) proposed a survey on Wireless sensor networks. Recent advances in wireless communications, digital electronics and microelectromechanical systems (MEMS) technology have enabled the development of low-cost, low-power, multifunctional sensor nodes that are smaller in size and communicate in short distances. These sensor nodes that consist of sensing, data processing, and communicating components, gives the idea of sensor networks based on collaborative effort of a large number of nodes.

Another special feature of sensor networks is the cooperative effort of sensor nodes. Instead of sending the raw data to the sensor nodes which are responsible for the fusion, sensor nodes with their processing abilities will carry out simple computations and transmit only the required and partially processed data.

The above described features ensure a wide range of applications for sensor networks. Thus this paper gives the introduction about the wireless sensor networks in detail1].

B. Mamalis, D. Gavalas, C. Konstantopoulos and G. Pantziou (2009) proposed a paper on Clustering in WSN. The use of WSN has grown by pointing out the crucial need for scalable & energy-efficient routing protocols in...
large-scale environment. The comparison between different clustering algorithms is shown in Table 2.1 [5].

Table 2.1: Comparison between different clustering algorithms

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>LEACH (Low Energy Adaptive Clustering Hierarchy)</th>
<th>EEHC (Energy Efficient Hierarchical Clustering)</th>
<th>HEED (Hybrid Energy Efficient Distributed Clustering)</th>
</tr>
</thead>
<tbody>
<tr>
<td>It is an hierarchical, probabilistic, distributed, one-hop clustering protocol</td>
<td>It is distributed, k-hop hierarchical clustering algorithm</td>
<td>HEED is a distributed, hierarchical, clustering scheme</td>
<td></td>
</tr>
<tr>
<td>All nodes have chance to become CH in order to balance the energy by each node.</td>
<td>Each node is elected as a CH with probability “p” &amp; announces about its election to the neighboring nodes within “k”-hop range.</td>
<td>Only those sensors nodes that have a high residual energy are expected to become CH nodes.</td>
<td></td>
</tr>
<tr>
<td>Disadvantage: As the decision on CH election &amp; rotation is probabilistic, there are chances that a node with a very low power gets selected as CH and the elected CHs will concentrate on one part of network.</td>
<td>Disadvantage: The energy consumption for network operation depends on the parameters p and k of the algorithm.</td>
<td>Disadvantage: A knowledge of the entire network is needed to determine the intracluster communication cost and configuration of those parameters might be difficult in practical world.</td>
<td></td>
</tr>
</tbody>
</table>

A. P. Reynolds, G. Richards, and V. J. Rayward-Smith (2001) proposed a paper on The Application of K-medoids and PAM to the clustering of rules. This paper also describes about K-means algorithm in which each cluster is centered about a point called as centroids, where the centroid’s coordinates are the mean of the coordinates of the objects in the cluster. But the disadvantage of k-means is to calculate the distances of the centroids for each object at each iteration. Hence, K-medoids algorithm was introduced. Rather than calculate the mean of the items in each cluster, a representative item, or medoid, is chosen for each cluster for every iteration.

There are two advantages for using existing rules as the centres of the clusters. Firstly, a medoid rule serves to describe the cluster. Secondly, there is no need for repeated calculation of distances at every iteration, since the k-medoids algorithm can simply look up distances from a distance matrix.

The K-medoids algorithm can be briefed as:

1. Choose k objects at random to be the initial cluster medoids.
2. Assign each object to the cluster associated with the closest medoid.
3. Recalculate the positions of the k medoids.
4. Repeat Steps 2 and 3 until the medoids become fixed. Thus, this algorithm has excellent feature that it requires the distance between every pairs of objects only once and uses this distance at every iterative step and it takes less computation time. The above algorithm runs just like K-means clustering and so this will be called as ‘K-means-like’ algorithm[6].

III. SYSTEM DESIGN

The flowchart of the system design is shown in figure 3.1.

![Flowchart of the System Design](image)

**Network Initialization**: Initially, a wireless sensor network with a size of about 30-50 nodes is placed randomly on the grid by configuring various network parameters. Nodes are given some mobility to implement mobile network scenario.

**Neighbour Node relationship**: By initiating the beacon signals from base station and collecting the reply from the sensor nodes, the Received Signal Strength information is maintained in the base station. The neighbour node identification is accomplished using the collected RSSI information.

**Optimal cluster formation using k-medoids algorithm**: Based on the computed positions of the nodes, the cluster formation is brought about with few cluster members coordinated by cluster heads using an algorithm called k-medoids. The shape of clusters changes every time as the nodes are mobile in nature. Each cluster will be given different colors.
IV. SYSTEM IMPLEMENTATION
MATLAB software has been used for the implementation of the algorithm. MATLAB (matrix laboratory) is a multi-paradigm numerical computing environment. MATLAB allows matrix manipulations, plotting of functions and data, implementation of algorithms, creation of user interfaces. The implementation results are as follows:

4.1 Network initialization

Fig 4.1: Snapshot showing network initialization

The figure 4.1 shows that the nodes are deployed randomly in the network. Each node is configured with various network parameters like:
- Initial energy = 100J
- Communication range = 150m
- Communication model = Energy model
- Area = 100 x 100m

4.2 Cluster formation

Fig 4.2: Snapshot showing cluster formation

The figure 4.2 shows the cluster formation based on k-medoids algorithm & optimal clusters are resulted by executing the above algorithm.

4.3 Identify cluster head & base station

Fig 4.3: Snapshot showing designation of cluster head & base station.

The figure 4.3 shows that the designated leader for each cluster is called as cluster head. The designation of the cluster head is based on the election carried out within each cluster depending on its energy & location.

V. CONCLUSION
As localization and power control are major issues in the WSN, this paper provide an approach to control and optimize the communication power in the mobile sensor network. The position of the node is estimated using RSSI measurement which can be used in a statistical method called k-medoids algorithm. This results in the formation of optimal number of clusters which is beneficial for scalable & load balancing network.

The MATLAB code is used for simulating the sensor network optimal clustering i.e. the k-medoids method of clustering algorithm is used to implement optimal number of clusters.

Further, the energy reduction, latency & packet delivery ratio will be analyzed and compared with the traditional clustering algorithm.

ACKNOWLEDGMENT
My sincere thanks to Mrs. Vinutha C B, Assistant Professor, Department of Electronics & Communication S.S.I.T for her guidance, constant encouragement and whole hearted support I am deeply indebted and i would like to express my sincere thanks to Dr. M.Z.Kurian, HOD, Department of Electronics & Communication, S.S.I.T for providing me an opportunity to do this project and a special gratitude to
Mr. Ningappa Uppal who has assisted me for doing this project.

REFERENCES


