



Cluster Based Routing Algorithm for Wireless Sensor Networks

S. Gopikrishnan¹, R.Janani², D.Shyam Prakash³

Assistant Professor, Department of CSE, Karpagam College of Engineering, Coimbatore, Tamilnadu, India ¹

Assistant Professor, Department of IT, Karpagam College of Engineering, Coimbatore, Tamilnadu, India ^{2,3}

ABSTRACT: Nowadays sensors are very essential for day today life to monitor environment where human cannot get involved very often. Wireless Sensor Networks (WSN) is employed in a wide range of applications such as environmental monitoring, traffic control, trajectory monitoring etc. Sensors sense and collect a large amount of data which are continuous over time, which in turn need to be forwarded to sink for further decision making process. This paper presents k-means cluster node clustering (K-CRA) algorithm for grouping sensor nodes there by reducing number of nodes transmitting data to sink node, it reduces the communication overhead and in this manner increase the network performance. A clustering in WSN involves selecting cluster heads and assigning cluster members to it for efficient data relay. Clustering sensor data act as a core task of mining sensor data. These algorithms face major new challenges such as limited communication bandwidth constraints in power supply, and storage resources. Furthermore preserve and detect modes (PDM) algorithm reduces the number of nodes within the cluster there by without compromising the coverage region. K-CRA contributes to reduced power consumption, time efficiency of the algorithm is evaluated through simulation experiments and the results are presented.

KEYWORDS: WSN, clustering, PDM, data mining

I. INTRODUCTION

Advances in wireless communications made to cultivate tiny hardware components with low-cost, low-power, multifunctional and intelligent sensor nodes. These devices are small in size and communicate in short distances over an RF (radio frequency) channel. These tiny nodes, which consist of components for sensing, processing, and communicating data, realize the objectives of sensor networks [13].

Wireless sensor networks promise significant improvements over traditional sensors. A Wireless Sensor Network is composed of a large number of integrated sensor nodes that are densely deployed either inside the phenomenon or very close to it, and collaborate through a wireless network in collecting environmental information or reacting to specific events [1, 2]. Some typical applications of sensor networks include medical monitoring, natural event monitoring, object tracking, monitoring product quality, and combat field reconnaissance [3]. The majority of sensor networks applications fall into the querying class of applications in which it is required to continuously collect and integrate data for future analysis and mining. Special characteristics of WSNs lead us to new research challenges in mining data

sensed in sensor network. Sensors have serious resource constraints including power supply, communication bandwidth, processor capacity and Storage. Moreover, data which is sensed in WSNs is generated and streamed continuously. These challenges make traditional mining techniques inapplicable in sensor networks [4], as the traditional mining process is usually centralized, computationally expensive, and focuses on disk stored data.

Clustering is a basic task grouping the data in data mining. It is usually used as a preparatory step in data mining for future data analysis. In this paper, a distributed version of K-means cluster clustering algorithm is proposed for the purpose of clustering sensors' node in a wireless sensor network. The nodes within the clusters will forward the data to sink through cluster head where aggregation of data takes place.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 1, January 2015

II. PROBLEM DEFINITION

In this section, we describe our problem in detail. Initially we have to recall some of the traditional clustering concepts for better understanding of our proposal.

2.1. Cluster Analysis in Data Mining

Data Clustering is defined as the process of grouping the data objects into cluster based on the similarities and their functionalities. This process is much easier to collect and process the data sensed from the environment. The dissimilar sensors are placed in the nearby clusters. The difference between the clusters are dependent on the characteristics of each cluster and they are measured by functions such as Euclidean distance, Manhattan distance, etc [5, 6]. There are number of traditional data clustering algorithms developed in recent past. Many of these algorithms are designed to deal with data which is stored in a traditional database management system. But in a distributed environment we intend to analyze data which is distributed across multiple sites. Because of limited bandwidth, privacy, data analysis in a distributed environment faces many challenges [7]. There are some algorithms proposed for the problems of cluster analysis in distributed environments. In [7], a number of distributed clustering algorithms are well discussed and grouped based on rounds of message passing between data sites and a central site. A wireless sensor network is made up of energy efficient sensors in a distributed environment. So, data clustering algorithms for sensor networks should address challenges mentioned above as well as sensor constraints in communication and computation.

2.2. Cluster-Based Sensor Networks

Due to energy limitations in sensor networks, managing the sensors energy and safeguarding of the power is a crucial factor to attain prolonged network lifetime. Among many studies on considering sensors' energy, clustering sensors into groups has proven to be a more efficient and adaptive approach [8]. Cluster-based mechanism [9] is adopted for node communication and routing. In a clustered sensor network, sensors communicate data only to cluster-heads and then the cluster-heads transfer the aggregated data to processing center or a base station usually called sink. The base station can be a specialized device or just one of the sensors [10]. The essential operation in sensor node clustering is to select a set of cluster heads among the nodes in the network, and cluster the rest of the nodes with these heads. Cluster-heads are selected according to some negotiated rules. Usually, more powerful nodes in the topology play the role of cluster heads and other nodes are responsible for sensing data and forwarding them to cluster nodes. Communication is usually the main factor of energy consumption in sensors [9]. The amount of energy used is totally dependent on the distance the data has to travel between the sender and the receiver. Since in a clustered sensor network, sensors only communicate data to cluster-heads over smaller distances, the total in the network will be much lower than the situation in which every sensor communicates directly to the base station. There are already a lot of works related to clustering in sensor networks [11]. Several heuristic algorithms have been proposed to choose cluster-heads. Moreover, some approaches attempt to estimate the optimal number of cluster-heads in a sensor network [8]. The main goal of sensor network clustering is to balance energy consumption over the whole network. LEACH, ASCENT, SPAN, GAF, ACE and HEED all try to preserve and balance the energy dissipation of the network using cluster-based architectures, thus prolonging the network lifetime [11].

III. NODE CLUSTERING IN WIRELESS SENSOR NETWORKS

Suppose that N sensor nodes are dispersed uniformly in a field to detect D attributes. Moreover, assume that these sensors are clustered into groups resulting in a clustered sensor network with pre determined cluster-heads. Our goal is to develop an algorithm that clusters the sensor nodes for a data relay without compromising coverage area. The majority of the algorithms are not concentrated on the convergence time and computation time

3.1 Proposed work

Initial assumptions made for following work

- The network is homogeneous (sensors with the same behavior and they sense similar data and they contain same amount of initial energy)
- The cluster-heads are dominant for performing computations and long range transmissions to the sink
- Each sensor node can transfer data directly only with other nodes in its cluster.
- Number iteration in k-means cluster controlled by sum of mean square error

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 1, January 2015

Generally K-means cluster algorithm is used to cluster the sensors for efficient data relay. Initially, k random points c_1, \dots, c_k are chosen as the cluster representatives by the sink node. The main drawbacks of the above mentioned algorithm are that the chosen cluster heads may not be able to obey its assigned function as a representative. Moreover in case of smaller networks it would result in a high cost for clustering the predefined number of k , whereas the clusters could be limited. The proposed K-CRA algorithm conserves the energy by delimiting the clustering factors in a number of ways and thus improving efficiency.

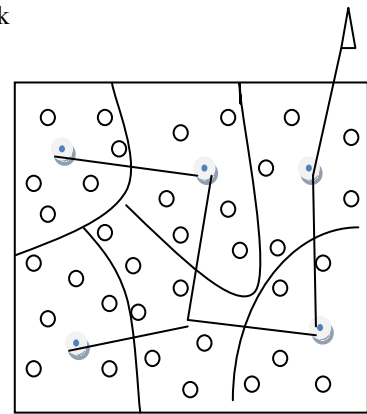
Algorithm K-CRA

```

1: Choose  $k$  arbitrary points as  $c_j$  ( $j = 1, \dots, k$ )
2: repeat
3: for all  $s_i \in S$  do /*assign each sensor to clusters*/
4:  $Q_i \leftarrow C$  /*candidate clusters*/
5: to prune candidate representatives from  $C = \{c_1, \dots, c_k\}$ .
6: for all  $c_j \in C$  do
7: if  $MDR_i \subseteq V(c_j)$  then
8:  $Q_k \leftarrow \{c_j\}$  /*The one and only one candidate*/
9: if  $|Q_i| = 1$  then /*only one candidate remains*/
10:  $h(i) \leftarrow j$  where  $c_j \in Q_i$ 
11: else
12: for all  $c_j \in Q_i$  do /*remaining candidates*/
13: Compute  $ED(s_i, c_j)$ 
14:  $h(i) \leftarrow \arg \min_j: c_j \in Q_i \{ED(s_i, c_j)\}$ 
15: for all  $j = 1, \dots, k$  do /*readjust cluster representatives*/
16:  $c_j \leftarrow$  centroid of  $\{s_i \in S \mid h(i) = j\}$ 
17: until  $C$  and  $h$  become stable  $h(i) \leftarrow j$ 

```

sink



To improve the performance of K-means cluster, we need to reduce the time spent on ED calculations, because it dominates the execution time of the algorithm. Since we know that computation cost is much more than transmission cost. One way to achieve this is to avoid ED computations whenever possible. For a given sensor i , the set Q_i stores the set of candidate cluster representatives that are potentially the closest to s_i . Initially, $Q_i = C$, the set of all cluster representatives. A pruning algorithm is applied to prune candidate representatives from Q_i that are guaranteed to be not the closest to sensor s_i . If all but one candidate cluster remains in Q_i , sensor s_i is assigned to that cluster. Otherwise, we compute The expected distances between s_i and each remaining cluster in Q_i . Sensors are then assigned to the cluster that gives the smallest expected distance. We describe a few pruning algorithms in the following sections. A good pruning algorithm should desirably reduce the set Q_i to a very small cardinality, so that the number of ED calculations that need to be performed in line 13 is as few as possible. Thus we can reduce ED calculation by K-CRA. Above method would swallow a considerable amount of energy of the individual sensor in the network. Additionally, another algorithm (PDM) would render its function to make the lifetime of the entire network to be enhanced without comprising coverage area within its cluster by minimizing number of nodes contributing to sense and forward data.

3.2 PDM Algorithm for Sensor Activation

After the first iterative step, the further selected subset of nodes enters into the monitoring state while the remaining sensor nodes go to conserve mode. The selected observe nodes provide full coverage over the monitored field during this iteration. In PDM, each sensor node assigns itself an activation delay that is proportional to its current function cost $C(s) = 1/E(s)$. In this way, sensor nodes with smaller cost have a better chance of becoming active sensors in the upcoming communication round. Each sensor node then waits for this period of time before deciding whether it will stay observe mode during the next communication round. While waiting for observe mode delay to expire, the sensor node can receive the observe mode messages from its neighboring nodes, which have smaller activation delays (smaller cost), if they decide to become active during the upcoming iteration. If, after its observe mode delay time expires, the sensor node determines that its sensing area is completely covered by its neighboring nodes, it turns observe mode off for the upcoming iteration. Otherwise, the sensor node broadcasts an observe mode message to inform its neighbors about its decision to remain active. In this way, the lower cost nodes have priority to decide whether they should be active. All sensor nodes in the network jointly take part in the activation phase, regardless of the cluster to which they belong.

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 1, January 2015

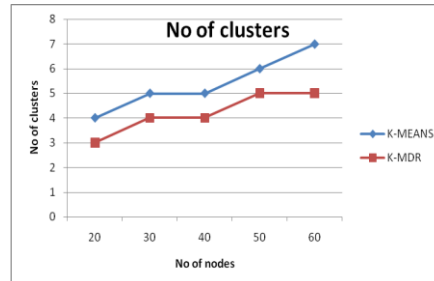


Fig 3.2.1 Comparison with K-Means

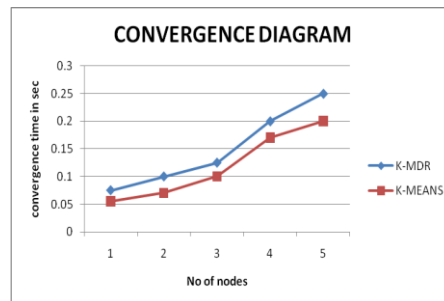


Fig 3.2.2 Convergence diagram

IV. CONCLUSION

The simulation results of the proposed work are carryout with MATLAB. And comparison between K-CRA and K-MEANS CLUSTER algorithm are given above. From the results obtained, we find that K-CRA with PDM was proved to be effective and efficient algorithm for node clustering. Further more optimal number of nodes was participated per iteration in clusters without compromising network coverage.

REFERENCES

- [1] F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E.Cayirci, "A Survey on Sensor Networks," IEEECommunications Magazine, pp. 102, 2002.
- [2] D. Culler, D. Estrin, and M. Srivastava, "Overview of Sensor Networks," IEEE Computer magazine, pp. 41-49,2004.
- [3] M. Ilyas, and I. Mahgoub, HANDBOOK OF SENSOR NETWORKS: COMPACT WIRELESS AND WIRED SENSING SYSTEMS.CRC Press, 2004.
- [4] X. Ma, D. Yang, S. Tang, Q. Luo, D. Zhang, et al., "Online Mining in Sensor Networks," Lecture Notes inComputer Science, pp. 544-550, 2005.
- [5] J. Han, and M. Kamber, DATA MINING – CONCEPTS AND TECHNIQUES.Morgan Kaufmann Publishers, 2004.
- [6] G. Forman, and B. Zhang, "Distributed data clustering can be efficient and exact," SIGKDD Explorations, vol. 2(2),pp. 34-38, 2000.
- [7] J.C.d. Silva, C. Giannella, P. Bhargava, H. Kargupta and M. Klusch, "Distributed data mining and agents,"Engineering Applications of Artificial Intelligence, vol. 8,pp. 791-807, 2005.
- [8] H. Kim, S. W. Kim, S. Lee, and B. Son, "Estimation of the Optimal Number of Cluster-Heads in Sensor Network," LNAI, pp. 87-94, 2005.
- [9] S. Ghiasi, A. Srivastava, X. Yang, and M. Sarrafzadeh, "Optimal Energy Aware Clustering in Sensor Networks,"Sensors, vol. 2, pp. 258-269, 2002.
- [10] S. Bandyopadhyay, and E.J. Coyle, "An Energy Efficient Hierarchical Clustering Algorithm for Wireless Sensor Networks," IEEE INFOCOM, 2003.
- [11] S. Lee, J. Yoo, and T. Chung. "Distance-based Energy Efficient Clustering for Wireless Sensor Networks," in Proceedings of the 29th Annual IEEE International Conference on Local Computer Networks, 2004.
- [12] G. Cheng, and A. Zell, "CL2: A Multi-dimensional Clustering Approach in Sensor Networks," Lecture Notes in Computer Science, vol. 3289, pp. 234-245, 2004.638
- [13] Amirhosein ,Taherkordi "Communication-Efficient Distributed Clustering Algorithm for Sensor Networks",2009i.