MJRET

Open Access

Multidisciplinary Journal of Research in Engineering and Technology

Journal homepage: www.mjret.in

ISSN: 2348 - 6953

Suhas K. Pawar
Dept. Computer Engineering,
KJCOEMR,Pune, India.

Abhishek R. Tawde
Dept. Computer Engineering,
KJCOEMR,Pune,India.

Archana Pokharkar
Dept. Computer Engineering,
KJCOEMR,Pune, India.

Priya Panjwani
Dept. Computer Engineering,
KJCOEMR,Pune, India

Prof. Suhas Patil
Dept. Computer Engineering,
KJCOEMR,Pune, India

A Survey of Cluster formation Protocols in Wireless Sensor Networks

Abstract

In modern world use of WSNs is increased in many applications and many people are doing research in WSNs for enhancing the use of WSNs. Routing protocols are divided into flat, hierarchical routing based on network structure. In this paper, we discussed cluster formation and cluster head election protocols and their advantages and disadvantages. By considering the advantages of clustering we developed taxonomy on clustering routing method. At last we conclude and summarized the paper along with future direction.

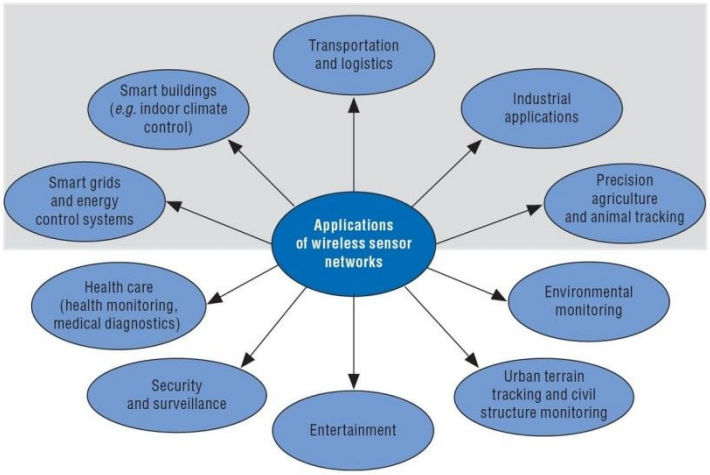
Keywords: *Wireless sensor networks (WSNs), Cluster head (CH), Base Station (BS), Member Nodes (MNs).*

1. INTRODUCTION

Wireless sensor networks are new a class of distributed systems that are an integral part of the physical space they inhabit. Unlike most computers, which work primarily with data created by humans, sensor networks reason about the state of the world that embodies them.

This bridge to the physical world has captured the attention of many researchers, encompassing a broad spectrum of ideas, from environmental protection to military applications [1]. Collection of many detection stations called sensor nodes is known as a sensor network, each node is small, lightweight and portable. Many disaster management applications require networks of sensors that can be easily deployed. In such applications wires or cabling is not practically possible. To overcome these drawbacks wireless sensor networks are used. Wireless sensor networks are fast, easy to install and maintain.

One of the major advantages of WSNs is that they can operate in dangerous environments in which human monitoring schemes is not possible or fails. Figure 1 involves various applications of WSNs like security, industrial, environmental and health care systems.



Applications of WSNs

2. RELATED WORK:

Xuxun Liu et al.[1] presents a survey on clustering algorithms. Classifications of clustering algorithms based on variable convergence time protocols and constants convergence were presented by the author. Concepts related to the process of clustering like structure of cluster, types of cluster, advantages of clustering and some proactive and reactive algorithm were surveyed by author. Key design challenges and the performance issues with clustering protocols based on clustering algorithms based on the identity, clustering algorithms on neighborhood information were discussed by author.

Author also classified a clustering algorithms based on the parameters of cluster formation and CH election criteria and also focuses on clustering routing protocols on the basis of cluster, cluster-head, clustering characteristics and entire proceeding of the algorithm. Then comparison between different type clustering routing protocols is made by author with respect to energy efficiency, cluster stability, scalability, delivery rate, load balancing and algorithm complexity.

3. GOALS OF CLUSTERING:

Less energy consumption, less load, scalability and more robustness are some of the advantages of clustering routing protocols which are discussed in this section as follows:-

1. **Less energy Consumption**

In scheme of routing of clustering, Aggregation of data helps in dramatically reduces the transmission of data and saves the energy. Using clustering , number of sensor nodes can be reduce which are performing long distance communication tasks in inter-clustering or intra-clustering. Thus energy is consumed. Data transmission is done by the CHs in clustering routing scheme by which saving of energy consumption is done.

2. **Data Aggregation**

The purpose of data aggregation is for aggregating of data from multiple nodes to eliminate the redundant transmission and provide aggregated data to the BS. It is the best way to save energy. The CH first collects the aggregated data and then transmits to the BS.

3. **Scalability**

Responsibilities of CHs are aggregation of data, data processing, dissemination of information and management of network and the MNs for events information sensing in their surroundings. At each sensor nodes the amount of size of the routing table stored can be reduced by localizing the route set up with the help of clustering topology. Compared with other types of topology, this is more scalable to tasks in the environments.

4. **Network lifetime maximization**

Network lifetime is the term which cannot be neglected in WSNs because sensor nodes are impel in supply of power, bandwidth of transmission.

5. **Connectivity Guarantee**

In WSNs, transmission of data to one or more base stations through the single hop or multi-hop routing. To check whether a data is delivered successfully or not to BS is determined by the connectivity of node to next hop of its own. If the nodes cannot communicate with each other than the data cannot be transfer to the BSs.

6. **Avoidance of energy holes**

The data is delivered to sink node or BS by using the multi-hop routing. The sensor nodes which are closer to the BS transmit more number of packets than the sensor nodes which are away from BS. Due to this, the nodes nearer to BS, first it decrease their energy, hole is leaving near to BS, whole network partitioning and the nodes which are outside preventing them for sending of data to BS, nodes which are remaining still have some amount of energy. This concept is known as energy hole.

7. **Load balancing**

Load balancing is helpful for extending the lifetime of network in WSNs. Distribution of sensor nodes are done even among the clusters for cluster construction where the data processing and intra-cluster management performed by the CHs. Due to the equal size of clusters, it extends the lifetime of networks and prevents exhaustion of energy of CHs.

8. Latency Reduction

Transmission of data is done hop by hop, by using the flat routing scheme, but the data is transmitted by the CHs in scheme of clustering routing. Because of this reduction in hops is occurred from source to BS, so it decreases latency.

4. CLUSTERING ROUTING PROTOCOLS FOR CLUSTER FORMATION:

1. LEACH

Heinzelman et al. [2] proposed a Low Energy Adaptive Clustering Hierarchy (LEACH). LEACH selects the sensor nodes as CHs randomly. This is the main objective of the LEACH. Due to selection of CHs randomly, the high energy is dissipated in the communication to the BS is distributed in network of sensor nodes. The LEACH has two main phases, first is Set-up phase and second is Steady-state phase. In set-up phase all the nodes in cluster decides to become CH or not for that round. The decision of CHs is determine by percentage of CHs in the network and how many times the node become a CH. If the value of that node is less than the threshold then the node become a CH. The threshold is determined by following formula:

$$T(n) = P/1-P(r \bmod 1/P), \text{ if } n \in G$$

$$T(n) = 0, \text{ otherwise}$$

Where: desired percentage of CHs is denoted by P, Current round is denoted by r, The nodes which are not elected as CHs for last 1/P rounds is denoted by the set G.

Advantages of LEACH:

1) If the node becomes CH than for later round it cannot be CH. i.e. node that attended as CH in some round then that node cannot be CH again for next rounds.
2) By the utilization of schedule of TDMA avoidance of needless collisions for CHs are done.
3) Members of cluster are open or close communication. Due to this, a prevention of extreme energy dissipation is takes place.

Disadvantages of LEACH:

1) It is not suitable for long-range networks.
2) Dynamic clustering brings extra overhead.

2. HEED

Younis and Fahmy introduced a Hybrid Energy-Efficient Distributed clustering (HEED) [3]. It is a multi-hop clustering algorithm. It gets an energy-efficient clustering routing with unambiguous concern of energy. HEED is different from LEACH because it does not selects CH randomly. The way of construction of cluster is achieved on the fusion combination of two factors. One factor is residual energy and second factor is communication cost of intra-cluster. In HEED, chosen CH have comparatively high average residual energy

related to MNs. Getting an evenly distributed CHs all over the networks. In HEED, two parameters are used to select the CHs, one is residual energy and other is cost of intra-cluster communication of applicant nodes.

Advantages of HEED:

1) It contains fully distributed clustering method by using two parameters for the CH selection.
2) Provides CH distribution and load balancing through the network.
3) It provides multi-hop communication between CHs and BS to support more energy conservation and scalability difference from single-hop.

Disadvantages of HEED:

1) HEED undergoes a consequent overhead meanwhile it needs some iteration to form cluster.
2) The CHs which are near to the sink node dies earlier because of it has extra work load.
3) HEED causes visible energy dissipation which causes to decrease the lifetime of network.

3. DWEHC

Ding et al.[4] projected a Distributed Weight-based Energy-Efficient Hierarchical Clustering protocol (DWEHC). Improvement of HEED by constructing cluster sizes and improved the intra-cluster topology with awareness of the nodes. In process of CH election both DWEHC and HEED has similarities like no supposition about size of network and density. Residual energy is comes into focus for CH election in DWEHC. All nodes implements DWEHC separately and after some iterations algorithm ends. It defines a weight factor for CH election process. Weight of the each node is calculated by:

$$\boldsymbol{Wweight(S) = \frac{Eresidual(S)}{Einitial(s)} \sum_u R - d/6R}$$

Where, Eresidual(s) is residual energy for node S, Eintial is the opening energy of the node s, R is range of cluster, the distance between node s and adjacent node u is denoted by d. Node having highest weight is elected as a CH and remaining node becomes a MNs. Initially MNs are considered as first-level nodes and talk with CH directly. On the basis of distance knowledge its neighbors it decides whether node is stay within first level or become in h-level. Where, h-level gives the number of hops from CH to itself. If the energy is saved by the MN at the time of reaching to destination then MN became h-level member. Creation of multi-level structure takes place in DWEHC for intra-cluster communication and a cluster range R is for the limit the number of levels.

Advantages of DWEHC:

1) DWEHC is completely distributed clustering technique founded on the nodes energy reservation and the nearness to the neighbors for CH election.
2) Lower energy consumption in intra-cluster and inter-cluster routing because of stable CHs distribution.
3) Clustering process in DWEHC ends in few repetitions.

Disadvantages of DWEHC:

1) Single-hop inter-communication takes place in DWEHC. Therefore it is not appropriate to networks of large-region.
2) Produces a great control message overhead compared to other protocols.

4. PEGASIS

Lindsey proposed Power-Efficient Gathering in Sensor Information Systems (PEGASIS) [5]. It is an improvement of LEACH. Each node communicates with its neighbors and become a leader for the transmission to the sink node. In PEGASIS, nodes are located randomly. Wireless communication, data fusion, positioning and data detection are the abilities of sensor nodes. Node forms the chain, which is assigned by the sink and transmission to all nodes by greedy algorithm. During the procedure of chain establishment it is expected that all nodes have the universal information of network and greedy method is employed. The chain building is initiated from furthermost node commencing the sink node and the closest node to this node is the next node for the chain. Dead node is eliminated by reconstruction of chain bypassing it. Each node receives data from its neighbor node and fuses this data with its own data and then node passes this fused data to the leader node and leader node passes this data to sink node in the chain network.

Advantages of PEGASIS:

1) PEGASIS has ability to overtake LEACH and for dissimilar network sizes and topologies cause of it decreases the upstairs of dynamic cluster creation in LEACH and decreases the data transmission.
2) The energy load is isolated equally in network.

Disadvantages of PEGASIS:

1) It is the requirement of having a whole understanding of the network topology at every node for chain building and all nodes must be capable to communicate directly to the sink. So, this arrangement is incompatible for networks with a time fluctuating topology.
2) Long-term transportations straight from the node to the sink consume lots of energy.
3) The communication undergoes from extreme delays caused by the single chain for aloof nodes and a great possibility for any node converted to a blockage.
4) It is a tough mission for all nodes to preserve a whole database about the location of each node in the network. Also the network is not scalable because all nodes must have worldwide information of the network and service the greedy algorithm.

5. PANEL

Position-based Aggregator Node Election Protocol (PANEL) was introduced by Buttyan and Schaffer [6,7]. The main objective of PANEL is to select aggregators. It means that CHs for regular and tenacious data storage applications. The nodes are arranged in a limited area and it is separated into geographical cluster. The clustering is determined before the positioning of network and each node having the geographical information of the cluster within same cluster. At starting of each era, a reference point Rj is calculated in each cluster j by the nodes in dispersed manner in terms of the epoch number by using:

$$\overrightarrow{R_j} = \overrightarrow{Q_j} + \overrightarrow{Q}$$

Where, $\overrightarrow{Q_j}$ is the position of lower-left corner of cluster j. Also, the present epoch number e is known by each node and the calculation contains in calling a pseudo-random function H(e) that maps e to relational position Q inside the cluster. i.e. {H(e)=Q}. After reference point is calculated, the node which is closest to the reference point will be chosen as CH for given epoch. Re-computation for the reference points of clusters is done and re-execution of procedure of CH election takes place for next epochs. Intra-cluster communication helps in CH election. Establishment of routing tables takes place due to this communication. The intra-cluster routing is used to track a message to the aggregator point of cluster if the message is intimate the cluster.

Advantages of PANEL:-

1) PANEL is an energy-efficient protocol that gives load balancing due to elected aggregator.
2) It supports synchronous and asynchronous applications.

Disadvantages of PANEL:-

1) It cannot be useful for WSN dynamics because assumption which are determined by cluster before positioning.
2) Geographical location info of the nodes is used to decide which node should be the aggregators. This is a limitation in WSNs, since the geographical location is not always presented without singular situation. For example, GPS-like hardware and software.

Table 1.Taxonomy on Different Clustering Protocols in WSNs.

Protocol	Cluster features			Cluster head abilities			Clustering Procedure		
	Cluster computation	Intra cluster topology	Inter cluster connection	Mobility	Node category	Function	Approach	Aim of node combination	Cluster head selection
LEACH	Fluctuating	Static	Direct link	Stationary	Sensor	Relaying	Distributed	Save energy	Random
ACE	Fluctuating	Adaptable	Direct link	Relocatable	Sensor	Aggregation and relaying	Distributed	Scalability and load balancing	Random
HEED	Fluctuating	Fixed 1-Hop	Direct link/multi-hop	Stationary	Sensor	Aggregation and relaying	Distributed	Save energy	Random
DWEHC	Fluctuating	Adaptable [Multilevel]	Direct link	Stationary	Sensor	Aggregation and relaying	Distributed	Save energy	Random
EECH	Fluctuating	Adaptable	Direct link/multi-hop [Hierarchical]	Stationary	Sensor	Aggregation and relaying	Distributed	Save energy	Random
LCA	Fluctuating	Fixed1-Hop	Direct link/multi-hop	Mobile	Sensor	Aggregation and relaying	Distributed	Connectivity	Random
Adaptive clustering	Fluctuating	Fixed 1-Hop	Direct link/multi-hop	Mobile	Sensor	Relaying	Distributed	Bandwidth gain and quality of service	Random
PANEL	Constant	Single-hop	Multi-hop	Stationary	Sensor	Aggregation and Relaying	Distributed	Load Balancing	Random
PEGASIS	Fluctuating	Multi-hop	Single-hop	N/A	Sensor	N/A	Distributed	Load Balancing	Random

6. K-MEANS

K-mean is the most famous position based clustering algorithm. Each k-cluster are represented by the mean (weighted average) of its points, hence the name K-mean algorithm. The advantage of partitioning algorithms is that they determine clusters at once. The k-means algorithm is an algorithm to cluster *n* objects based on attributes into *k* partitions, where k<n (i.e. Algorithm divides the region into k clusters).An algorithm for partitioning (or clustering) N data points into K disjoint subsets S_j containing data points so as to minimize the sum-of-squares criterion

$$J = \sum_{j=1}^{K} \sum_{n \in Sj} |x_n - \mu_j|^2$$

Where $x_n$is a vector representing the n^{th} data point and u_j is the geometric centroid of the data points in S_j.

K-Means Advantages:

1) It is relatively efficient and fast.

2) If variables are large in number, then K-Means most of the epochs computationally faster than hierarchical clustering, by keeping k smalls.
3) K-Means creates close-fitting clusters than hierarchical clustering, particularly if the clusters are globular.

5. FUTURE AND RESEARCH WORK

WSN requires specific energy efficient protocols. A number of clustering protocol sometimes forms clusters and assign CHs statically and so the wastage of energy is occurred. So developing an energy efficient algorithm is research work for less energy consummation. Finding energy efficient algorithm for cluster formation is research work of clustering in WSNs. Developing a real time traffic clustering protocol is also a research work. Developing or expanding the existence cluster formation algorithm to overcome the drawbacks of its own is a future work in WSNs.

6. CONCLUSION

The crucial objective of the protocol design is to retain the sensors working for a long time for increase network's lifetime. In this paper, we have charted and summarized recent research works focused on the cluster-based routing protocols for WSNs. As this is a wide area, this paper has enclosed only few routing protocols. The protocols discussed in this paper have separate advantages and drawbacks. The issues affecting cluster formation and CH communication are open subjects for future research. With the help of above study we developed taxonomy for clustering routing methods based on the cluster features, cluster head abilities and clustering procedure. By studying the different clustering routing protocols we conclude that K-mean is relatively efficient and fast as compared to other clustering

routing protocols. K-mean is faster than hierarchical clustering by keeping k small and it creates close fitting clusters than hierarchical clustering.

ACKNOWLEGEMENT

We would like to thank our guide Prof. Soumitra.Das, Prof.Suhas Patil for the guidance and support. We will forever remain grateful for the constant support and guidance extended by guide, for the completion of paper. We also thank to editor and referees for their comments.

REFERENCES

[1]. A Survey on Clustering Routing Protocols in Wireless Sensor Networks by Xuxun Liu School of Electronic and Information Engineering, South China University of Technology, Guangzhou 510641.

[2]. Heinzelman, W.R.; Chandrakasan, A.; Balakrishnan, H. Energy-Efficient Communication Protocol for Wireless Microsensor Networks. In Proceedings of the 33rd Annual Hawaii International Conference on System Sciences, Maui, HI, USA, 4±7 January 2000; pp. 10±19.

[3]. Younis, O.; Fahmy, S. HEED: A hybrid, energy-efficient, distributed clustering approach for adhoc sensor networks. IEEE Trans. Mobile Comput.2004, 3, 366±379.

[4]. Ding, P.; Holliday, J.; Celik, A. Distributed Energy Efficient Hierarchical Clustering for Wireless Sensor Networks. In Proceedings of the 8th IEEE International Conference on Distributed Computing in Sensor Systems (DCOSS), Marina Del Rey, CA, USA, 8±10 June 2005; pp. 322±339.

[5]. Lindsey, S.Raghavendra, C.Sivalingam, K.M. Data gathering algorithms in sensor networks using energy metrics. IEEE Trans. Parallel Distrib.Syst. 2002, 13, 924±935.

[6]. Buttyan, L.; Schaffer, P. PANEL: Position-Based Aggregator Node Election in Wireless Sensor Networks. In Proceedings of the 4th IEEE International Conference on Mobile Ad-hoc and Sensor Systems Conference (MASS), Pisa, Italy, 8±11 October 2007; pp. 1±9.

[7]. Buttyan, L.; Schaffer, P. PANEL: Position-based aggregator node election in wireless sensor networks .Int. Distrib.Sens. Netw.2010, 2010, 1±16.