Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 49 (2015) 147 – 154

Procedia Computer Science

ICAC3'15

Comparison of Clustering Algorithms to Design New Clustering Approach

Sumedha Sirsikar[a], Kalyani Wankhede[b]

[a and b]*Maharashtra Institute of Technology (MIT), Paud Road, Pune 411038, India*

Abstract

Wireless Sensor Network (WSN) is a multi-hop sensor network system in which sensor nodes are deployed in monitoring area to sense some environmental parameters. WSN is used to bridge the gap between physical world of humans and virtual world of electronics. WSNs have number of applications in daily life such as remote monitoring of environment, habitat, agriculture, health care, automobiles, and disaster prone zones. WSNs have characteristics such as limited resources, multi-hop routing and dynamic network topology. Clustering is introduced in WSN because it has proven to be an effective approach to provide better data aggregation and scalability for large WSNs. Clustering also conserves the limited energy resources of the sensors. This paper does comparative study of existing clustering algorithms in WSNs based on centralized, distributed or hybrid method and highlights the challenges in clustering. In addition we proposed a new clustering system which uses at most two-hop for intra-cluster communication. Reduction in number of clusters and CHs prolongs network lifetime.

Peer-review under responsibility of organizing committee of the 4th International Conference on Advances in Computing, Communication and Control (ICAC3'15)

Keywords:Wireless Sensor Networks, Clustering, Energy Efficiency, Network Lifetime.

1. Introduction

In wireless sensor network (WSN) the sensor nodes are often grouped into individual disjoint sets called a cluster. Each cluster comprises of Cluster Head (CH) and its members [1]. CHs are nodes that consume more energy than cluster members when they involve in aggregating, processing and routing data. CHs collect data from the sensors. Clustering is especially important for sensor network applications where a large number of ad-hoc sensors are deployed for sensing purpose. In the network if each and every sensor node starts to communicate then all sensors will be engaged in data transmission hence network will enter into enormous congestion and data collisions. This

Peer-review under responsibility of organizing committee of the 4th International Conference on Advances in Computing, Communication and Control (ICAC3'15)

doi:10.1016/j.procs.2015.04.238

situation will lead to drain limited energy from the network. Clustering of node will address these issues. Scalability feature of the Wireless Sensor Networks (WSN) is used to meet load balancing and efficient resource utilization constraints. Use of clustering in a hierarchical WSN facilitates efficient utilization of limited energy of sensor nodes and hence extends network lifetime. Although sensor nodes in clusters transmit messages over a short distance (within clusters), more energy is drained from CHs due to message transmission over long distances (CHs to the base Station) compared to other sensor nodes in the cluster. Periodic re-election of CHs within clusters based on their residual energy is carried out to balance the power consumption of each cluster. In addition, clustering increases the efficiency of data transmission by reducing number of sensors that are attempting to transmit data in the WSN, aggregating data at CHs via intra-cluster communication and reducing total data packet loses.

1.1. Components of a clustering in WSN

The following are the components of a clustered WSN [5] as shown in figure 1:
Sensor node: A sensor node is the main component of a WSN. Sensor nodes perform functions such as sensing; data storage; routing; and data processing.
Clusters: Clusters are the hierarchical units for WSNs. Large sensor networks need to be broken down into clusters to simplify tasks such as communication between the base station and the cluster heads.
Cluster heads: Cluster heads (CHs) are the leader of a cluster. CHs are often required to organize activities in the cluster. These tasks include data aggregation, organizing and relaying the communication schedule of a cluster.
Base Station: The base station (BS) provides the communication link between the sensor network and the end-user. It is normally the sink in a WSN.

In this paper we present some existing clustering algorithms for WSNs. We also compare these algorithms based on metrics such as residual energy, uniformity of CH distribution, cluster size, number of hops and cluster formation methodology. The rest of the paper is divided as follows:
Section II presents the challenges for clustering algorithms and the working process of clustering. In Section III, we present the comparison, advantages and limitations of some clustering algorithms. Section IV concludes the paper.

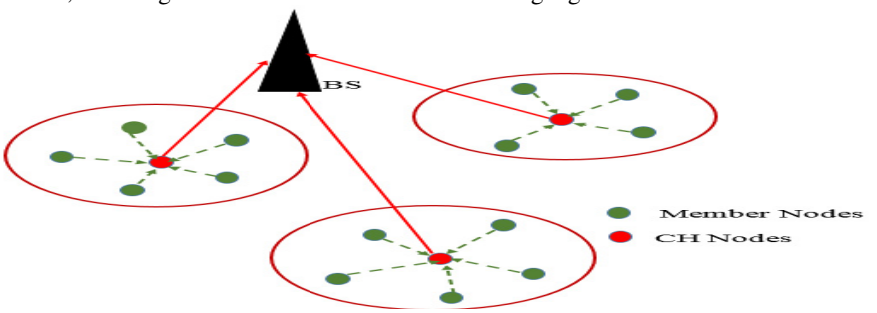


Fig. 1.Clustering in wireless sensor network

2.Related work

2.1. Challenges for Clustering Algorithms

Clustering schemes play an important role in WSN;these can effectively improve the network performance. Following are key limitations of clustering algorithms [7]:
• Limited Energy: Wireless sensor nodes are small size battery operated sensors, so they have limited energy storage. It is not practicable to recharge or replace their batteries after exhaustion. The clustering algorithms are more energy efficient compared to the direct routing algorithms. This can be achieved by balancing the energy consumption in sensor nodes by optimizing the cluster formation, periodically re-electing CHs based on their residual energy, and efficient intra-cluster and inter-cluster communications.
• Network Lifetime*:* The energy limitation on nodes results in a limited network lifetime for nodes in a network. Clustering schemes help to prolong the network lifetime of WSNs by reducing the energy usage in the communication within and outside clusters

• Limited Abilities*:* The small physical size and small amount of stored energy in a sensor node limits many of the abilities of nodes in terms of processing, memory, storage, and communication.
• Cluster formation and CH selection*:* Cluster formation and CHs selection are two of the important operations in clustering algorithms. Energy wastage in sensors in WSN due to direct transmission between sensors and a base station can be avoided by clustering the WSN. Clustering further enhances scalability of WSN in real world applications. Selecting optimum cluster size, election and re-election of CHs, and cluster maintenance are the main issues to be addressed in designing of clustering algorithms. The selection criteria to isolate clusters and to choose the CHs should maximize energy utilization.
• Data Aggregation*:* Data aggregation eradicates duplication of data. In a large network there are often multiple nodes sensing similar information. Many clustering schemes providing data aggregation capabilities [2] must carefully select a suitable clustering approach.
• Repair Mechanisms*:* Due to the nature of Wireless Sensor Networks, they are often prone to node mobility, node death, delay and interference. All of these situations can result in link failure.
• Quality of Service (QoS)*:* From an overall network standpoint, we can look at QoS requirements in WSNs. Many of these requirements are application dependent such as acceptable delay and packet loss tolerance.

2.2. Clustering Methodology

The three methodology are used for clustering process are centralized, distributed and hybrid [1]. In centralized clustering, the clusters and cluster heads are made by an authority (centralized authority). In distributed(decentralized) clustering, all the nodes in the clusters can took the decision of becoming cluster head for the current round. Hybrid clustering is the mixture of both of above. Figure 2 shows the taxonomy of clustering methods with respect to its features.

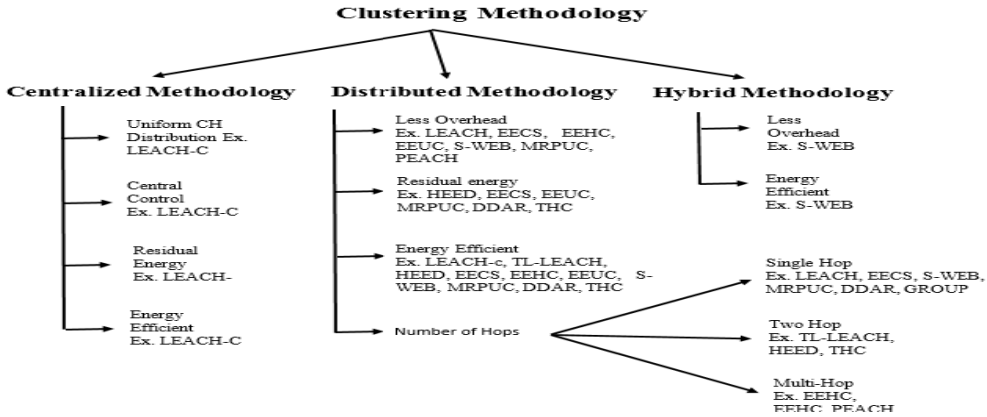


Fig.2. Taxonomy of Clustering Methodology

2.2.1. Centralized Clustering Methodology

In centralized clustering, the clusters and cluster heads are made by an authority (centralized authority. Following are the key features of the centralized methodology:

- *Central Control:*Central control is used to form fixed number of clusters. Also decides which sensor node will become CH by considering energy at each node. Due to central control it gives guarantee about number of cluster head.

  LEACH-centralized (LEACH-C), uses central control to produce better clusters by dispersing the cluster head nodes throughout the network. It gives better performance than LEACH because clusters are adaptive in LEACH [3].

- *Uniform CH Distribution:* Uniform CH distribution means CHs are not cluttered. Cluttered CHs can cause long distance between member nodes and their corresponding CH. In other hand uniform distribution gives high energy consumption for intra-cluster communication.

  To determine good clusters in LEACH-C, the BS needs to ensure that the energy load is evenly distributed among all the nodes. So, it gives uniform distribution of CH which determines optimal number of clusters.
- *Residual energy:* Residual Energy is the remaining energy of the sensor node. Sometimes it is considered when sensor node to be elected as a CH.

  CH selection in LEACH-C is based on sensor node having more residual energy. It gives equal chance of becoming CH for every sensor nodes.
- *Energy Efficiency:*Distributed methodology decides whether the algorithm is energy efficient or not.

  LEACH-C [2] is an energy efficient protocols, in which each node calculates the minimum transmission energy to communicate with its cluster head and only transmits with that power level changing the CH is probabilistic in LEACH; there is a good chance that a node with very low energy gets selected as a CH [3]. All nodes get equal chance to become a CH so energy is uniformly distributed.

2.2.2 Distributed Clustering Methodology

In distributed (decentralized) clustering, all the nodes in the clusters can took the decision of becoming cluster head for the current round. Following are the key features of the distributed methodology:

- *Less Overhead in CH selection:*In distributed clustering methodology there is no central control. All nodes are involved in the process of cluster formation; hence work is distributed among all nodes results in less overhead.

  LEACH [2] (Low Energy Adaptive Clustering Hierarchy) forms clusters based on the received signal strength and uses the CH nodes as routers to the base-station [6, 7]. All the data processing such as data fusion and aggregation are local to the cluster, hence no overhead is wasted making the decision of which node becomes cluster head as each node decides independent of other nodes.

  EECS, EEHC [4], EEUC, S-WEB, MRPUC, PEACH gave low overhead in CH selection.
- *Residual Energy:*Residual Energy is the remaining energy of the sensor node. Sometimes it is considered when sensor node to be elected as a CH.

  Energy Efficient Clustering Scheme (EECS) is a clustering algorithm in which cluster head candidates compete for the ability to elevate to cluster head for a given round. This competition involves candidates broadcasting their residual energy to neighbouring candidates.

  In Hybrid Energy Efficient Distributed Clustering (HEED) [8] CHs are chosen based on two important parameters: residual energy and intra-cluster communication cost. Residual energy of each node is used to probabilistically choose the initial set of CHs, as commonly done in other clustering schemes.

  (EEUC) proposed to balance the energy consumption among clusters, in which the cluster sizes near the sink node are much smaller than the clusters far away from the sink node in order to save more energy in intra-cluster communications and inter-cluster communications [9].

  MRPUC [9] use each node to gather the correlative information of its neighbour nodes and elects a node with maximum residual energy as the CH.

  THC algorithm is proposed to increase the energy saving in sensor nodes. CH selection is based on remaining energy and node degree. THC working is divided into three phases initial, distributed node clustering and actual data transmit.

  In DDAR, The BS computes both the average node energy (E_{avg}) and the average node distance (D_{avg-BS}). If the node energy (E_{node}) is greater than or equal to E_{avg} and if the distance of the node from the BS ($D_{node-BS}$) is less than or equal toD_{avg-BS}, then the nodes are eligible to become CH for this round. Therefore, the nodes are selected as CHs only when the following two conditions are satisfied [12].

$$E_{node} \geq E_{avg}$$

$$D_{node-BS} \leq D_{avg}$$

- *Energy Efficiency:*Distributed methodology decides whether the algorithm is energy efficient or not.

  In TL-LEACH (Two Level LEACH), Secondary nodes collect data from nodes in their respective clusters. The two-level structure of TL-LEACH reduces the number of nodes that need to transmit to the base station, effectively reducing the total energy usage hence energy efficient [10].

  In HEED, each sensor sets its probability of becoming a cluster headCH_{prob}, as follows:

$$CH_{prob} = C_{prob} * E_{residual}/E_{max},$$

Where$E_{residual}$ is the current energy in the sensor, and $E_{max}$is the maximum energy, which corresponds to a fully charged battery. It provides uniform distribution of CH nodes and better load balancing.So reduces energy consumption [8].

EECS extends LEACH algorithm by dynamic sizing of clusters based on cluster distance from the base station. The result is an algorithm that addresses the problem that clusters at a greater range from the base station requires more energy for transmission than those that are closer. Ultimately, this improves the distribution ofenergy throughout the network, resulting in better resource usage.

EEHC [8] algorithm is divided into two phases; initial and extended. In the first stage, also called single-level clustering, each sensor node announces itself as a CH with probability p to the neighbouring nodes within its communication range. These CHs are named as the volunteer CHs. All nodes that are within k hops range of a CH receive this announcement either by direct communication or by forwarding [11].

The second phase, called multi-level clustering builds h levels of cluster hierarchy. The algorithm ensures h-hop connectivity between CHs and the base station. In inter cluster communication this algorithm ensures that the energy dissipated by CHs far from the base station is reduced because these CHs don't need to transmit to base station.

EEUC [9] is a distance based scheme similar to EECS and it also requires every node to have global knowledge such as its locations and distances to the sink node. It tries to prolong the network lifetime and to balance the load among the nodes. It solves the hot spot problems; cluster size is proportional to the distance to base Station resulting into energy efficient.

MRPUC is a distributed clustering scheme which operates in rounds, and each round is separated into three phases: cluster setup, inter-cluster multi-hop routing formation and data transmission. The regular nodes join clusters where the cluster heads have more residual energy and are closer to them. An inter-cluster routing tree is constructed as network backbone, and data is transmitted to BS via multiple [9]. This algorithm prevents early CHs death because the inter cluster communication also depends on the residual energy.Hence provides energy efficient approach.

Dynamic clustering and distance aware routing protocol (DDAR) for WSN which takes into account of node distance for selecting CH, dynamic approach of selecting the CH nodes, and two level hierarchy of clustering with super cluster head (SCH) node near to the BS. It also introduces the SCH in the network to save the energy of the CHs which are far from the BS [12]. The introduction of SCH nodes in the network divides the clusters into two clustering hierarchy: child and parent clusters. It consumes about 15.5% less energy than conventional LEACH protocol.

Two Hop Clustering (THC) protocol employs 2-hop clustering method, which reduces the amount of power required by the member node to communicate with the CH nodes. All the 2-hop away member nodes forward their data to the CH nodes via 1-hop nodes; which significantly reduces the amount of energy consumption. THC uses the uses next head concept, which reduces the number of negotiation messages required after every round to select the CH [13].

- *Number of Hops (Single/Multi-hop):* In cluster formation hop count is considered. There are single hop, two hop and multi hop clusters present in various clustering algorithms. Single hop means all member nodes are 1-hop away from CH (similarly 2-hop away from CH in two hop clustering).

  LEACH, LEACH-C, EECS, S-WEB, MRPUC, DDAR, and GROUP uses single hop cluster formation.

  TL-LEACH [10] is extension of LEACH algorithm utilizes two levels of cluster heads (primary and secondary) in addition to the other simple sensing nodes. In this, the primary cluster head in each cluster communicates with the secondary, and the corresponding secondary communicate with the nodes in their sub-cluster. Hence TL-LEACH forms cluster by two hops.

  There are two algorithms HEED, THC [13] which uses two hop cluster formation. EEHC, EEUC, and PEACH [6] use multi-hop communication for cluster formation.

2.2.3. Hybrid Clustering Methodology

Hybrid clustering is the combination of both centralized and distributed methodology. Following are the key features of the centralized methodology:

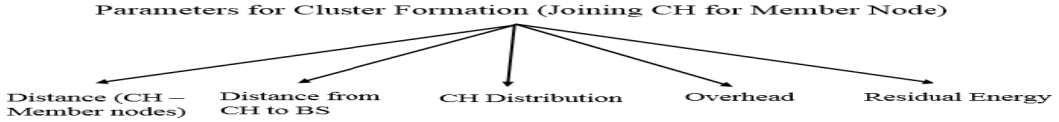
- *Less Overhead in CH Selection:*Hybrid methodology is combination of centralized and distributed methodologies so as location of sensor nodes takes from the BS and for further work is carried out from member nodes. So it gives less overhead in CH selection.

  Sensor Web(S-WEB) algorithm is a hybrid methodology in which BS generates the beacons and remaining tasks are performed by member nodes. Here there is no single central control, hence less overhead in CH selection [14].
- Energy Efficiency:Hybrid methodology decides whether the algorithm is energy efficient or not.
  S-Web is energy efficient because beacons are generated by BS and remaining tasks are given to member nodes.

2.3. Parameters considered in Clustering Process

There are two main steps in clustering, which are CH selection and cluster formation. CH selection could be classified into three types, centralized by the BS, distributed by the sensor nodes or hybrid selection by few information provided by the BS and few by the nodes themselves [6].

I. The main parameters in selecting CHs are:



- Clustering method is to be considered.
- Distance between CHs and the BS to ensure that CHs are not too far from the BS, making inter- CH and BS expensive
- Uniform CH distribution so that CHs are not untidy. Cluttered CHs can cause long distance between non-CH nodes and their corresponding CH. In other words causing high energy consumption for intra-cluster communication
- CHs perform extra tasks for WSNs such as: data aggregation and forwarding therefore energy at the CHs might be depleted quickly. CH re-selection or rotation is another concern in clustering. In reselecting a CH, one must consider how much overhead in terms of energy consumption and time it takes.
- Residual Energy in a sensor node to be elected as a CH.

II. The main parameters in joining a CH for a sensor node are:

- Distance between a node and a CH. This is often presented as signal strength between CHs and the node itself.
- Number of hops from a node to its CH

Table 1: Comparison of clustering algorithms

	CH Selection					**Cluster Formation for Member Nodes**			**Energy Efficient**
Clustering Algorithm	Clustering Method	Distance (CH-BS)	CH Distribution	Overhead in CH Selection	Residual Energy	Distance(Member node - CH)	Cluster Size	No. of Hops	
LEACH	Distributed	-	Non-Uniform	Low	No	Yes	No	1	No
LEACH-C	Centralized	Yes	Uniform	High	Yes	Yes	Yes	1	Yes
TL-LEACH	Distributed	-	Non-Uniform	Medium	No	Yes	No	2	Yes
HEED	Distributed	-	Uniform	High	Yes	-	Yes	2	Yes
EECS	Distributed	Yes	Uniform	Low	Yes	Yes	Yes	1	Yes
EEHC	Distributed	-	Non-Uniform	Low	No	Yes	No	K	Yes
EEUC	Distributed	Yes	Uniform	Low	Yes	Yes	Yes	K	Yes
S-WEB	Hybrid	Yes	Uniform	Low	Yes	Yes	Yes	1	Yes

MRPUC	Distributed	Yes	Uniform	Low	Yes	Yes	No	1	Yes
PEACH	Distributed	-	Non-Uniform	Low	No	Yes	No	K	No
DDAR	Distributed	Yes	Non-Uniform	Medium	Yes	Yes	No	1	Yes
THC	Distributed	Yes	Non-Uniform	Medium	Yes	Yes	No	2	Yes

- Cluster size: the decision whether or not a node joins a cluster also depends on the size of a cluster. The number of nodes in a cluster represents the accumulated energy in a cluster and also extra energy consumption for the CH to serve its cluster. Table 1 shows comparative study of clustering algorithms based on CH Selection parameters and few parameters of cluster formation for member nodes with energy efficiency.

3. Proposed System

*Assumptions:*The proposed system would create some assumptions for energy efficient clustering and prolonging the network lifetime. Following are the assumptions made:

- The network area is fixed.
- All sensors nodes are homogeneous in the network.
- All nodes and BS are stationary.
- BS knows location of all nodes.
- The intra-cluster communication isat most two-hop and inter-cluster communication is multi-hop.

Proposed Statement: To design a clustering approach which will increase the energy saving and improve the network lifetime. The proposed method would use concept of two-hop for cluster formation, which will reduce number of CH than the 1-hop clustering CH which will helpful to increase energy saving in WSN. Also it reduces amount of energy required by member node to communicate with CH nodes.

Proposed System Design:
The proposed system design actions are divided into 4 steps as shown in figure3

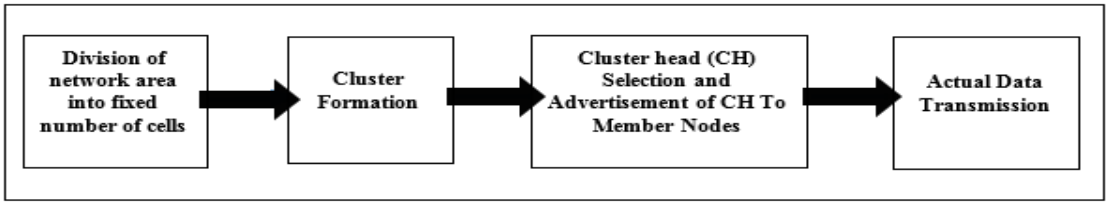


Fig.3. Proposed System Design

- Division of network area into fixed number of cells: Divide network area into fixed number of rectangular cells. Statically we divide sensor area intorectangular cells
- Cluster Formation:One cluster is formed within each cell, by allowing at most two-hop which cover more elements compared to single hop clustering.

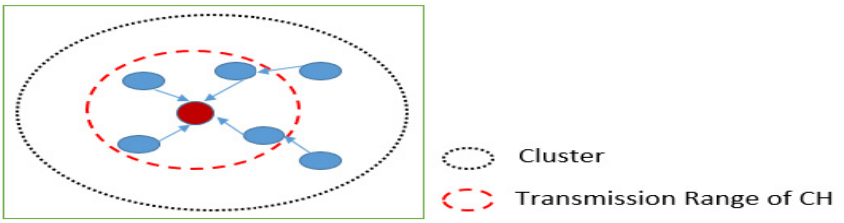


Fig.4. Formed Cluster in fixed Cell

- **C**H Selection and advertisement to member nodes: CH selection is based on energy of node and node degree. In this step CH selection is taken place and advertisement of that CH to nodes. After certain time

new CH selection taken place for next round. Here single hop member is formed within the transmission range of the cluster node. We use shortest distance to join single hop nodes.

- Actual Data Transmission: Data is aggregated at CH from all member nodes which is transmitted to BS is taken place in this phase.

4. Conclusion

Clustering is most suitable toreduce communication overhead forlarge scale wireless sensor networks. This paper includes survey of various clustering algorithms. We have mainly focused on distributed clustering approaches, because they are more suitable for large-scale sensor networks.So, the comparative study is useful to find out parameters needed in cluster head selection, overhead in CH selection and shows whether the algorithms is energy efficient or not.Energy consumption during cluster formation and maintenance is high; so there is thechallenges for clustering algorithms include scheduling concurrent intra-cluster and inter-cluster transmissions, how to compute the optimal cluster size, and how to decide the optimal frequency for cluster head rotation in order to increase the network lifetime. The clusterformation overhead includes packet transmission cost of theadvertisement, announcement, joining, and scheduling messages from sensor nodes. Therefore, nodemobility and distribution, overhead and lifetime are majorconcerns for clustering policies in WSNs.Moreover, to support dynamic configuration and nodedistributions, adaptive multi-level clustering is expected. Hence, in our proposed system we will use two hops clustering, in which our intra-cluster communication is at most, two-hop and inter-cluster communication is multi-hop. Hence there is less number of CH compared to 1-hop so as to improve energy saving and network lifetime in WSN.

References

1. Naveen Sharma and Anand Nayyar, “A Comprehensive Review of Cluster Based Energy Efficient Routing Protocols for Wireless Sensor Networks”, IJAIEM,Volume 3, Issue 1, January 2014.
2. Wendi B. Heinzelman, Anantha P.Chandrakasan, Hari Balakrishnan, “An Application-Specific Protocol Architecture for Wireless Microsensor Networks”, IEEE Transactions on Volume 1, Issue 4. 2002, pp.660 –670.
3. M. Aslam, N. Javaid, A. Rahim, U. Nazir, A. Bibi, Z. A. Khan, “Survey of Extended LEACH-Based Clustering Routing Protocols for Wireless Sensor Networks”, IEEE 2012.
4. Vijay Kumar, Sanjeev Jain, Sudharshan Tiwari, “Energy Efficient Clustering Algorithms in Wireless Sensor Networks: A Survey”, IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 5, No 2, September 2011.
5. Liliana M. Arboleda C. and Nidal Nasser, “Comparision of Clustering Algorithms And Protocols for Wireless Sensor Networks”, IEEE CCECE/CCGEI, Ottawa, May 2006.
6. Congfeng Jiang, Daomin Yuan, Yinghui Zhao, “Towards ClusteringAlgorithms in Wireless Sensor Networks-A Survey”, IEEE 2009.
7. P. Kumarawadu, D. J. Dechene, M. Luccini, A. Sauer, “Algorithms for Node Clustering in Wireless Sensor Networks: A Survey”, IEEE 2008.
8. Ameer Ahmed Abbasi, Mohamed Younis, “A survey on clustering algorithms for wireless sensor networks”, published by Elsevier B.V. 2007.
9. G. Vennira Selvi, R. Manoharan, “A Survey of Energy Efficient Unequal Clustering Algorithms for Wireless Sensor Networks”, International Journal of Computer Applications, Volume 79- No. 1, Oct 2013.
10. V. Loscri, G. Morabito and S. Marano, “A Two-Level Hierarchy for Low-Energy Adaptive Clustering Hierarchy”, Proceedings of Vehicular Technology Conference 2005, vol3, 1809-1813.
11. S. Bandyopadhyay, E. Coyle, “An energy efficient hierarchical clustering algorithm for wireless sensor networks”, Proceedings of the 22nd Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM 2003), San Francisco, California, April 2003.
12. Navin Gautam, Won-Il Lee and Jae-Young Pyun, “Dynamic Clustering and Distance Aware Protocol for Wireless Sensor Networks” , ACM Spain, pages 9-14, 2009-10.
13. Babu N V, Puttamadappa. C, Bore Gowda S B, ”Energy Efficient Two Hop Clustering for Wireless Sensor Networks”, IJCSNS International journal of Computer Science and Network Security, Vol. 13 No.9, September 2013.
14. Le Hanh, Doan Hoang and Ravi Poliah, “S-Web: an efficient and self-organizing Wireless Sensor Network Model”, 2nd International Conference on Network-Based Information Systems, Turin, Italy 2008.