A Survey on Hierarchical Clustering Algorithm for Wireless Sensor Networks

Nirnaya Pradhan
Department of CSE, SMIT
Sikkim Manipal University

Kalpana Sharma
Department of CSE, SMIT
Sikkim Manipal University

Vikash Kumar Singh
Department of CSE, SMIT
Sikkim Manipal University

ABSTRACT

The main constraint of Wireless Sensor Networks (WSNs) is irreplaceable power source of the sensor nodes. In recent years, a lot of research has been carried out to investigate different aspects like, cluster formation, load balancing among sensor nodes, network establishment, routing protocol and coverage problem in wireless sensor network to prolong network lifetime. In this paper, we examine hierarchical clustering algorithms mentioned in the literature of Wireless Sensor Networks. We will briefly discuss the operations of these algorithms and also examine the performance in terms of power consumption.

Keywords

Wireless Sensor Networks, Cluster Head, Hierarchical Clustering, single-hop, multi-hop, Base Station.

1. INTRODUCTION

Wireless Sensor Networks (WSNs) is a collection of sensor node deployed randomly at some area of interest to collect information from the environment to execute specified task. Sensor node is coupled with irreplaceable power with the limited memory, communication range and processing speed. A node in a sensor network perform two demanding tasks sensing and communicating.

Some of the features of sensor networks are: i. nodes are densely deployed. ii. Sensor nodes are prone to failures. iii. The topology of the sensors network changes very frequently. iv. Sensor nodes are limited in power, computational capabilities and memory. v. Sensor nodes may not have global identification (ID) because of large amount of overheads and large number of sensors.

WSN protocol stack consist of physical layer, data link layer, network layer, transport layer, application layer, power management plane, mobility plane, and task management plane[1]. Energy efficiency can be improved at various layer of the communication protocol stack. Therefore each in layer there are numerous open area for research.

Sensor nodes are operated with limited power sources which is the main limitation of WSNs. In numerous application such as battlefields, dense forest etc. sensor nodes are not accessible easily therefore they cannot be recharged. Thus, the most challenging issues in WSNs is energy conservation of the sensor nodes to maximize the network lifetime. The efficient way to overcome these challenges is clustering sensor nodes.

In this paper various energy efficient hierarchical clustering based routing algorithm are discussed and compared. The rest of the paper is organized in following manner: Section 2 will briefly explain energy efficient hierarchical clustering structures in WSNs. In section 3, the energy efficient hierarchical cluster based routing algorithms are discussed and compared. We will conclude the paper with section 4.

2. ENERGY EFFICIENT HIERARCHICAL CLUSTERING STRUCTURES IN WSNs

One of the most efficient techniques to save the energy consumption in wireless sensor network is clustering sensor nodes. The main objective is to minimize the total transmission power aggregated over the node and to balance the nodes for increasing the network lifetime.

In this process sensor nodes are grouped into cluster, each cluster has a leader sensor node referred to as cluster head which is responsible for coordinating the data transmission activity of all sensors. Remaining nodes within the cluster act as a cluster members which sends their sensed data to corresponding cluster head. Cluster head aggregate the received data from its member node and using single hop or multi hop communication it transfer data to the base station.

The hierarchical cluster structures facilitate the efficient data gathering and aggregation independent to the growth of the wireless sensor networks, and generally reduce the total amount of communications as well as the energy spent by the sensor nodes. The architecture of a generic wireless sensor network is shown in Figure 1.
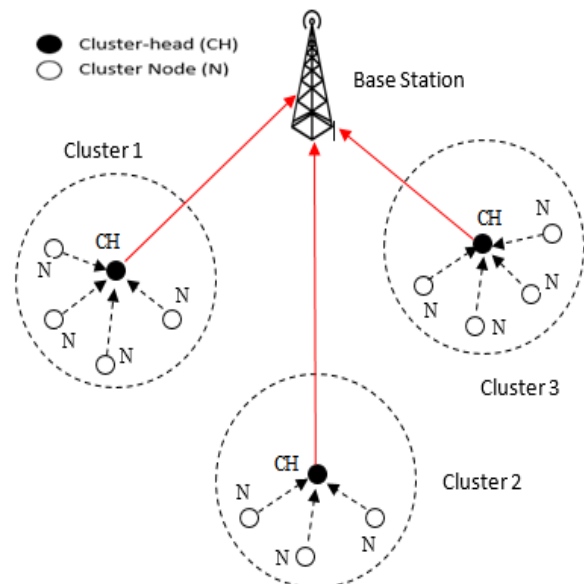

Figure 1. General Sensor Network Architecture

3. ENERGY EFFICIENT HIERARCHICAL CLUSTER ROUTING ALGORITHM

3.1 Low-Energy Adaptive Clustering Hierarchy (LEACH)

LEACH [3] is the first energy efficient hierarchical clustering algorithm for WSNs proposed by Heinzelman et. al. and has become the most popular in hierarchical clustering by reducing power consumption. LEACH distributes the energy load evenly among all sensor nodes in a network by dividing the network into a several clusters and selecting cluster head from each cluster randomly. If cluster head are chosen as a priori and fixed throughout the network lifetime then nodes which are selected as cluster heads die quickly, this effects the member of cluster as well. In LEACH cluster head responsibility is randomly rotated among all the sensors such that all the nodes drain its battery equally. Cluster head aggregate (or Fusion) the original data into the smaller size of data that carry only meaningful information from the cluster to the base station which also helps to prolong network lifetime.

The operation of LEACH is divided into rounds, where each round starts with a set-up phase where clusters are organized, cluster head advertisement, transmission schedule creation followed by steady-state phase for data aggregation, compression and data transfer to the base station from cluster head.

Cluster head advertise their status using CSMA MAC protocol using same transmit energy. In order to hear advertisement sent by the CH all non-cluster head sensor node must keep their receiver ON in start-up phase. The decision for a sensor node whether or not to become a cluster head for the current round is independent without any negotiation with the other sensor node i.e. based on the suggested percentage *P* of the cluster heads for the network and the number of times the node has been a cluster head so far.

A sensor node becomes a cluster head by choosing a random number between 0 and 1. If the random number is less than T (n), then the node will be selected as a cluster head for the current round. Where T (n) is the threshold given by

$$T(n) = \begin{cases} \frac{P}{1-P(r \bmod \frac{1}{P})} & \text{if } n \in G \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

P is the desired percentage of cluster heads, r is the current round, and G is the set of sensor nodes that has become cluster head in the last 1/P rounds. A sensor node cannot be a cluster head for the next 1/p rounds if the nodes have already become cluster head in round 0. During round 0 each sensor node has an equal probability P of becoming a cluster head.

At the end of the startup phase, each non cluster head node decides which cluster it will belong for this round based on the received signal strength of the cluster head advertisement. After taking the decision it informs the cluster head that it will be a member of the cluster using CSMA MAC protocol. Cluster head computes a TDMA schedule transmission time slot for its member nodes such that each member node can handover its data to cluster head in allocated transmission time slot.

The non-cluster-head node turns OFF the radio until the node's allocated transmission time to minimize the energy dissipation. A sensor node turns its radio ON, only during its scheduled time slot and transmit data to the cluster head. At the same time cluster head must keep their receiver ON to receive the data from its member node. Cluster head aggregate the original data into smaller size and then sends them to the sink. Cluster head spends more energy to transmit data to the base station if base station is far away, but non cluster head node saves power because the sensed data to be transmitted to the cluster head is closer to it.

LEACH minimizes the global energy by randomly distributing the load to all the node at different point in time and reducing the communication distance between sensor node and their cluster heads.

Main drawback of LEACH is unequal and high energy depletion in the cluster head due to the different transmission distance from each cluster head to base station.

3.2 Energy Efficient Extended LEACH (EEE LEACH)

EEE LEACH [4] uses multilevel clustering technique. Unlike LEACH, EEE LEACH maintains two layers of clustering information. Level 1. Cluster Heads are formed and receives sensed data from the cluster member sensor node. Level 2. Master cluster heads are formed and receives the aggregate data from the nearest cluster head, aggregates all received data and forwards compressed data to the base station. In EEE LEACH cluster head and master cluster head for the networks is determined as a priori.

In LEACH communication distance is calculated by using the following formula

$$D = (n1*y) + (n2*x) \quad (2)$$

And EEE LEACH communication distance is calculated by using the following formula

$$D' = (n1*y) + (n2*z) + (n3*x) \quad (3)$$

Where x is the distance from the sensor field to base station, y is the distance between nodes, n1 is the number of normal nodes, n2 is the number of cluster heads, and n3 is the number of master cluster heads. The value of D should always be higher than D'. Thus multilevel clustering conserve energy consumption compared to single hop LEACH.

3.3 Energy Efficient Clustering Scheme (EECS)

EECS [5] uses novel approach in cluster formation phase to balance the load among cluster heads. EECS starts with cluster head election phase where sensor node locally compete for the cluster head based on their residual energy and constant number of cluster heads are selected.

EECS works differently than LEACH, at network deployment phase, by broadcasting bacon from the base station to all the nodes by taking a certain power level. Using the received signal strength each node compute the distance to the base station and select the power level to further communicate with the base station.

In cluster head election phase nodes are nominated for candidate node with some probability and broadcasts complete message within the radio range. Each candidate node will check for the nodes having more residual energy within the radio range. If no powerful nodes exists in its radio range then node will be elected as cluster head node. Otherwise it will give up the competition.

In Cluster formation phase of EECS cluster head broadcast the cluster head advertisement message over the network. To select the cluster head EECS introduce a weighted function cost (j,i) for the non-cluster-head node j, which is *cost (j,i)=w X f(d(P_j , CH_i)) + (1-w) X g(d(CH_i , BS))* and P_j chooses CH_i with min{cost} to join.

Non cluster head node wisely takes the decision to select its cluster head. It selects cluster head by considering workload balancing of the cluster head and also saving its own energy.

3.4 Distance Aware Intelligent Clustering Protocol (DAIC)

DIAC [6] conserve the energy and prolong network lifetime by reducing the transmission distance between the cluster heads and the base station. It divides the network into number of layers. From each layer one cluster heads and gateway cluster head is selected. Cluster head receives all the sensed data from the non-cluster-head and forwards it to the gateway cluster head of immediate upper layer.

Similar to LEACH, DIAC operation is divided into rounds and each round has two phase, setup phase and steady state phase.

Setup phase consist of network setup, routing path construction and schedule creation phase and steady state phase consist of data transmission phase.

Network setup phase: This phase is responsible for layer formation in the network region, cluster head selection, gateway cluster head selection, creating cluster, identifying cluster head node for every non cluster head node, and creating time slot for every individual non cluster node in each cluster.

In the starting of this phase each node sends its details about its location and residual energy to the base station. Base station divides the network region into layers using the location information. For each node, base station calculates the vertical distance and compares with the average vertical distance from the base station. If the vertical distance of a node is lesser then the average vertical distance then node is assigned as a member of primary layer otherwise the node is assigned as a member of secondary layer. To calculate the distance co-ordinate geometry is used and all the sophisticated calculation is executed in the base station so that the sensor nodes does not get affected by this calculation. Now, the base station calculates average energy level of all the nodes from the primary layer. Then the base station chooses a set of nodes from the primary layer whose energy level is greater than the average energy level and declares it as primary cluster heads. These primary cluster heads are later used as a cluster head for each round. From the primary cluster head nodes, the nodes with greater vertical distance from the base station is selected as a gateway cluster head for secondary layer. Similarly, base station selects secondary layer cluster head from the secondary layer members whose residual energy is more than the average residual energy. Secondary layer members having residual energy less than the average residual energy will become the non-cluster head node.

Finally, base station broadcasts a message containing primary cluster head IDs, secondary cluster head IDs, and gateway IDs for each node in primary and secondary layer. The node compares its own ID with cluster head IDs if it matches with its own ID then the node is acts as a cluster head for specific layer otherwise it is just a member of one of the cluster of its layer. Similarly if gateway ID matches with the node ID in

primary layer then the node is set as a gateway cluster head. All the non-cluster-head associates themselves with the respective cluster head and forms a cluster of specific layer.

Routing Path Construction Phase: In this phase all the Secondary layer cluster heads are connected to gateway cluster head of primary layer. Gateway cluster head are connected to primary layer cluster head.

Schedule Creation Phase: To minimize the collision TDMA slot for data transmission are determined by the cluster head to each node in a cluster.

Data Transmission Phase: Data sensing, Data Processing, Compression, and data fusion are done in this phase.

3.5 Partition Energy Balanced and Efficient Clustering Scheme (PEBECS)

PEBECS [7] is a Hierarchical clustering protocol, which divides the entire sensing region of wireless sensor network in n partition. Each partition is labeled with the partition ID(PID) from 1 to n. PID 1 is located closer to the Base Station(BS) and PID n is the farthest from the BS. Since, PEBECS allows cluster head of each partition to use multi-hop communication to transfer data between the cluster head or to the BS, except the cluster head of nth partition will not be able to relay the data.

PEBECS uses distributed algorithm to form clusters and divides the process into two phases.

   a. Cluster head election- in this phase cluster head are randomly selected based on their residual energy, node degree difference and relative location in networks.

   b. Cluster formation phase- PEBECS carefully assign the cluster member nodes to the respective cluster head. Clusters closer to the base station are expected to have smaller cluster sizes, while cluster in the partition farther away have the larger sizes.

Constraints of PEBECS include network delay, admissible degree restriction and the impact of partition number.

3.6 Energy-Aware, Cluster-Based Routing Algorithm (ECRA)

ECRA [8] is a cluster based routing protocol, where each sensor are location aware. ECRA construct the Voronoi diagram by selecting the nodes as cluster head. It includes three phases: clustering, data transmission, and intra cluster head.

In clustering phase first ECRA define voronoi diagram and nodes located nearest to their corresponding centroidal point become the cluster head. ECRA knows every sensor node's location because sensor nodes are location aware and construct CVT for sensing field.

Base station directs each cluster head which nodes are its member and cluster head informs all its member by broadcast and advertisement. After receiving the advertisement from the cluster head member node acknowledges to its cluster head and confirms it will be a member of the cluster.

Data transmission phase: this phase is similar to LEACH, all member nodes are assigned with time slots and member node communicate with the cluster head using single hop. Cluster head aggregates the received data to a single message and transmit it to the base station using single hop.

Inter cluster head rotation: base station periodically rotates the role of a cluster head among member node of same cluster to achieve the load balancing.

3.7 Energy-Aware Distributed Unequal Clustering Protocol (EADUC)

EADUC [9] allows cluster head to use multi-hop communication to minimize the distance between cluster head and base station. To reduce the imbalanced energy consumption among the cluster head. It selects the cluster head base on the ratio between the average residual of neighbor nodes and the residual energy of node itself and also construct cluster of uneven sizes by using uneven competition range. EADUC construct larger cluster size to cluster head located far away from the base station and smaller size to cluster head closer to the base station to balance the energy consumption among cluster head.

Just like LEACH, EADUC divide the whole operation into rounds, where each round starts with setup phase and data transmission phase.

Setup phase divided into three sub phases neighbor node information collection phase, cluster head competition phase, and cluster formation phase and in transmission phase cluster head receive all the sensed data from the member nodes and aggregate the received data and send aggregate data based on the routing information constructed by EADUC to the next hop nodes.

3.7.1 Cluster setup phase

3.7.1.1 Network Deployment Phase:
It starts with broadcasting a signal at a certain power level by the base station. Based on the received signal strength each node computes approximate distance to the base station. There are three sub-phases in the cluster setup phases: neighbor node identification collection phases, duration given T1; cluster head competition phase, duration given T2; cluster formation phase, duration given T3;

Neighbor node identification collection phases: In this phase each node broadcasts node_message within its radio range. Node_message consist of two values node_id and node residual energy. At the same time it also receives node_message from its neighbor node. Each node calculates the average residual energy of its neighbor node. Each node calculate its waiting time for broadcasting node_message. After T1 expires, cluster head competition phase starts, in this phase if a node receive no node_message when time t1 expires, then node will become a cluster head within its radio range. EADUC also generates unequal cluster size to prolong the network life. Cluster head with high residual energy and farther away from base station will assign larger cluster area. Cluster head having lower residual energy control smaller area.

Cluster formation phase: In this phase each non cluster head selects the nearest cluster head and cluster head assign transmission time slot for each member.

3.7.2 Data Transmission phase:
Cluster member communicate with its cluster head directly to transfer sensed data.

Intra Cluster communication: If the cluster head distance to the base station is less then distance threshold, cluster head directly communicates to the base station. Otherwise it selects forwarding nodes from its neighboring nodes with high residual energy. In this way cluster head constructs a tree with the base station.

3.8 Energy-Aware Clustering Algorithm (EADC)

EADC [10] is a cluster based routing protocol it consist of an energy aware clustering algorithm and a cluster based routing algorithm. To elect a cluster head EADC uses ratio between the residual energy of node itself and the average residual energy of neighbor nodes. It construct cluster of even size which leads to energy consumption of cluster head imbalance. To balance energy consumption among cluster head EADC uses intra cluster energy efficient multi hop routing protocol in which cluster head selects the next hop cluster head with higher residual energy and a smaller number of cluster member to balance the energy consumption among cluster head.

Cluster set-up phase is similar to the EADUC [9] the entire process is divided into three phases: information collection phases, whose duration is t1; cluster head competition phases, whose duration is t2; cluster formation phases, whose duration is t3.

Cluster based routing algorithm: This phase takes t4 duration to measure cluster head distance to the base station. If it is less then distance threshold, cluster head directly communicates to the base station otherwise, routing tree are constructed by selecting the neighbor cluster head using following parameters like residual energy, the number of cluster member and distance to base station of the neighbor node.

3.9 Energy Aware Routing (ERA)

ERA [11] is a cluster based routing for a wireless sensor network. To be elected as a cluster head each node initiates a time delay and starts the campaign to become a cluster head, time delay is dependent on its residual energy. Timer of the sensor node is derived from the following formula:

$$t(i) = \frac{E_m(i) - E_r(i)}{E_m(i)} \times T_{CH} \qquad (4)$$

Where $E_m(i)$ is the initial maximum energy of a sensor node i, $E_r(i)$ is the residual energy of a sensor node i, T_{CH} is the maximum allotted time for selection of cluster head. A sensor node whose residual energy is more will be selected as a cluster head, since it has shorter time. Once the timer gets expired sensor node declares itself as a cluster head. Cluster head then generates the announcement message and broadcasts it. The announcement message includes its identification number, residual energy, and location information. When non-cluster head receives the announcement message from the neighboring cluster head it cancels its timer and withdraws its selection of cluster head candidature. Each node will keep track of all the source node from which it receives the cluster head announcement message. Finally, to form a cluster each non cluster head node computes the average residual energy of the cluster head. Node becomes the member of cluster head whose residual energy is greater than or equal to average residual energy.

Table 1: An overview of hierarchical clustering algorithms.

Clustering Algorithm	Description
LEACH:	Forms cluster to minimize energy dissipation.
EEE LEACH:	Reduces the communication distance from cluster head to base station.
EECS:	Balances the node among cluster using the weighted function.
DAIC:	Divides the network into layers to prolong network life.

PEBECS:	Adjusts the cluster size by partitioning the sensing region.
EADUC:	Reduces the imbalance in the energy consumption among the cluster head.
EADC:	Balances the energy consumption among the cluster head using intra cluster energy efficient multi hop routing protocol.
ERA	Balances energy consumption of the cluster heads during data routing process.

Table 2. Classification of Survey algorithms based on clustering attributes

Clustering Algorithm	Intra-cluster communication	Inter-cluster Communication	Clustering process	Cluster head selection	Sensor node
LEACH	Single hop	Single hop	Distributed	Random	Location unaware
EEE LEACH	Single hop	Two hops	Distributed	Random	Location unaware
EECS	Single hop	Single hop	Distributed	Random	Location unaware
DAIC	Single hop	Two hops	Distributed/Base Station Guided	Random	Location aware
PEBECS	Single hop	Multi-hop	Distributed/Base Station Guided	Random	Location unaware
ECRA	Single hop	Multi-hop	Distributed/Base Station Guided	Random	Location aware
EADUC	Single hop	Single hop / Multi-hop	Distributed	Random	Location unaware
EADC	Single hop	Multi hop	Distributed	Random	Location unaware
ERA	Single hop	Multi hop	Distributed	Random	Location aware

4. CONCLUSION

In this paper we have surveyed and briefed about research work on energy efficient hierarchical cluster based routing protocol for wireless sensor networks with respect to their power requirement. The main objective of most of the existing hierarchical protocols lies on how to extend the network lifetime and how to make a more efficient use of the critical resources, such as battery power. A lot of studies has already been done by researchers and also proved that clustering efficiently increases the stability and lifetime of wireless sensor networks. But it also imbalances the energy consumption among the nodes. To design any protocol for WSN, energy limitation of a node play a very crucial role. Each protocols discussed in this paper have individual advantages and pitfalls. To conserve energy in a balanced way among the sensor node many multi hop based data transmission has been considered. Based on the evaluation and discussion, it is concluded that multi hop communication are more suitable for addressing most of the issues such as energy efficiency, coverage and balance energy consumption among sensor nodes in wireless sensor networks.

5. REFERENCES

[1] I. F. Akyildiz,W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A Survey on Sensor Networks," IEEE Communications Magazine, vol. 40, no. 8,pp. 102–114, 2002.

[2] K. Akkaya, M. Younis, "A survey on routing protocols for wireless sensor networks," Elsevier Journal of Ad Hoc Networks 3, 325–349, 2005.

[3] W. Heinzelman, A. Chandrakasan and H. Balakrishnan, "Energy-Efficient Communication Protocol for Wireless Microsensor Networks," *Proc. 33rd Hawaii Int'l. Conf.Sys. Sci.*, Jan. 2000.

[4] M. Sharma and K. Sharma "An energy efficient extended leach (EEE leach), "Communication Systems and Network Technologies (CSNT) 2012 International Conference on. IEEE, 377-382, 2012.

[5] Ye M, Li CF, Chen GH, Wu J. "EECS: an energy efficient clustering scheme in wireless sensor networks," IEEE international performance computing and communication conference; p. 535–40, 2005.

[6] Gautam N, Pyun J-Y "Distance aware intelligent clustering protocol for wireless sensor networks," KICS J Commun Netw, 12(2):122–12, 2010.

[7] Y. Wang and T. L. X. Y. and D. Zhang, "An energy efficient and balance hierarchical unequal clustering algorithm for large scale sensor networks," *Information Technology*, vol. 8, no. 1, pp. 28–38, 2009.

[8] J. Chang, "An energy-aware cluster-based routing algorithm for wireless sensor networks," J*ournal of Information Science and Engineering 26*, pp. 2159–2171, 2010.

[9] Yu Jiguo, Qi Yingying, Wang Guanghui, Guo Qiang, Gu Xin. "EADUC: an energy-aware distributed unequal clustering protocol for wireless sensor networks," Int J Distrib Sensor Netw (Article ID 202145), 2011.

[10] Jiguo Yu et aI., "A cluster-based routing protocol for wireless sensor networks with non-uniform node distribution," into J. Electron. Communication, Vol. 66, pp. 54-61, 2012.

[11] Amgoth, T., & Jana, P. K. "Energy-aware routing algorithm for wireless sensor networks," Computers and Electrical Engineering, 41, 357–367, 2014.