A Review: Comparative Study of Enhanced Hierarchical Clustering Protocols in WSN

M. Sangeetha, A. Sabari, T. Shanthi Priya

Abstract—Recent advances in wireless networking technologies introduce several energy aware routing protocols in sensor networks. Such protocols aim to extend the lifetime of network by reducing the energy consumption of nodes. Many researchers are looking for certain challenges that are predominant in the grounds of energy consumption. One such protocol that addresses this energy consumption issue is ‘Cluster based hierarchical routing protocol’. In this paper, we intend to discuss some of the major hierarchical routing protocols adhering towards sensor networks. Furthermore, we examine and compare several aspects and characteristics of few widely explored hierarchical clustering protocols, and its operations in wireless sensor networks (WSN). This paper also presents a discussion on the future research topics and the challenges of hierarchical clustering in WSNs.

Keywords—Clustering, Energy Efficiency, Hierarchical routing, Wireless sensor networks.

I. INTRODUCTION

Modern advances in networking technology include the development of tiny, cost-effective, low-power, and intelligent sensor nodes in a wireless sensor network [1]. WSNs are used in different domains towards military applications, medical engineering, environment and industrial task automation [2]. Based on this serious expectation, in several significant WSN applications the sensor nodes are remotely deployed randomly in larger numbers and operated autonomously. In these unattended environments the sensors cannot be charged and replaced, thus energy constraints are the most serious problem so as to must be considered [3], [4].

Usually routing protocols on the basis of network structure are divided into three main groups: flat, hierarchical and location-based routing. Here, the hierarchical routing protocols also known as ‘cluster-based routing’, proposed in wireless sensor network. In order to support data aggregation through efficient network group, nodes can be partitioned into an amount of small groups called clusters [23]. All clusters have a manager, referred to as a cluster head and a number of member nodes. The cluster formation process eventually leads to a two-level hierarchy where the cluster head (CH) nodes from the highest level and the cluster-member nodes of the small level. Sensor nodes sometimes broadcast their data to the corresponding CH nodes. The CH nodes aggregate the data and transmit them to the base station (BS) either directly or through the intermediate communication with extra CH nodes. Though, because the CH nodes send all the time data to higher distances than the ordinary (member) nodes, they of course spend energy at higher rates. A general solution in order to balance the energy consumption between all the network nodes is to sometimes re-elect new CHs in each cluster [5]. Energy consumption in the network can be reduced by forming cluster structures in an efficient way; so many energy aware routing protocols are designed on the basis of the cluster configuration.

II. RELATED WORK

A. Leach

Low-Energy Adaptive Clustering Hierarchy (LEACH) [6], is an initial hierarchical clustering protocol that suggests both distributed and centralized schemes. The basic idea of LEACH has been an inspiration for many later clustering routing protocols. Leach utilizes the randomized rotation of cluster heads to uniformly distribute the energy consignment among the sensors in the network. The cluster heads have the responsibility of gathering data from their clusters, while also aggregate the collected data for decreasing the amount of messages to be sent to the BS, which outcome in less energy dissipation, to enhance the network lifetime. This protocol is divided into rounds; every round consists of two phases.

1. Set-up phase - organizing the cluster.
2. Steady-state phase - deals with the actual data transfers to the base station.

Initially, when clusters are being created, each node decides to become a cluster-head for the current round based on the given threshold value.

The threshold value calculated by

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

where n is the given node, P is the priori probability of a node being elected as a cluster head, or is the current round number and G is the set of nodes that have not been elected as cluster heads at the end 1/P rounds. Each node during the cluster head selection will generate a random number between 0 and 1. If this number is lower than the threshold (T(n)) the particular node will develop into a cluster head.
1) Merits
   a) Accounting for adaptive clusters and rotate the cluster head periodically.
   b) Completely distributed. It requires no global knowledge of the network.
   c) Increases the lifetime of the network.
2) Demerits
   a) Uses single-hop routing within the cluster.
   b) Highly dynamic clustering brings extra overhead.
   c) Leach randomly select cluster head, which may result faster death in few nodes

B. TL-Leach

Two-level Hierarchy LEACH (TL-LEACH) [7], is an extension to the algorithm of LEACH. It uses the following two techniques to get energy and delay efficiency: randomized, self-configuring cluster formation, adaptive and localized control for data transfers.

It is TL-LEACH introduces two - level hierarchy

1. Primary cluster heads - Top CHs called primary clusters heads.
2. Secondary cluster heads - Second level representation called on secondary cluster heads.

TL-LEACH algorithm the primary cluster head in all cluster communicatates with the secondary and similar to the secondary’s communicate with the node in their sub-cluster. Finally data aggregation performed at each level and collected data will be sent to the sink. Here additional communication within a cluster is still scheduled using TDMA time slots.

Communication of data from source node to sink is reached in two steps.

1. Secondary nodes gather data from nodes in their particular clusters. Data fusion can be done at this stage.
2. Primary nodes gather data from nodes in their particular secondary clusters. Data fusion can be implemented at the primary cluster head level.

The two-level structure of TL-LEACH reduces the amount of nodes that want to transmit to the base station. It is effectively reducing the total energy usage.

1) Merits
   a) TL-LEACH uses the localized coordination, which is conductive to scalability in the network.
   b) It efficiently reduces the total energy consumption.
   c) Transmission distance is decreased in comparison with LEACH.
2) Demerits
   a) TL-LEACH is a two-hop inter-cluster routing protocol, through its still not suitable for large-range networks.
   b) A CH election without energy consideration affects the network performance.
   c) It cannot ensure real load-balancing in case of nodes with different amount of initial energy.

C. Teen

Threshold sensitive Energy Efficient sensor Network protocol (TEEN) [8], is a hierarchical clustering protocol. It is first reactive network protocol. This protocol main goal is to manage with sudden changes in the sensed attributes such as temperature. It combines the hierarchical method in line with a Data-centric approach.

TEEN, is a 2-tier clustering topology and two thresholds they are hard threshold and soft threshold, are namely.

1. Hard Threshold – Threshold value for the sensed attribute. A cluster member only reports/sends data to CH by switching on its transmitter, only if data values are in the scope of interest.
2. Soft Threshold – Small change in the value of the sensed attribute. A cluster member only reports/sends data to CH by switching on its transmitter, but its value changes by at least the soft threshold.

In TEEN a CH sends its members a hard threshold and a soft threshold. This way, the hard threshold sanctions the nodes to transmit only when the sensed attribute is in the range of importance, this method reducing the number of transmissions considerably. Once a node senses a value at or beyond the hard threshold, it transmits data only after the value of that attribute changes by an amount equipollent to or more preponderant than the soft threshold, which points to a minute transmutation in the value of the sensed attribute and triggers a sensor to turn ON its transmitter and send it sensed data to the CH. So, soft threshold will further reduce the number of transmissions for sensing data if there is diminutive or no transmutation in the value of sensed attribute.

1) Merits
   a) Based on the two thresholds, data broadcast can be controlled estimable.
   b) TEEN has complemented for reacting to great changes in the sensed attributes.
2) Demerits
   a) If the thresholds do not reach, the nodes will never communicate; the user will not get any data from the network at all and will not come to know even if all the nodes die.
   b) It is not well suited for real time application where the user needs to get data on a regular basis.

D. Apteen

The Adaptive Threshold sensitive Energy Efficient sensor Network protocol (APTEEN) [9], is an expansion to TEEN. It combines the best features of both proactive and reactive networks as minimizing their limits to make a new type of network called a hybrid network. APTEEN is based on a query system which allows three types of queries: historical, on-time, and constant which can be used in a hybrid network. In APTEEN the cluster head first broadcasts the following parameters:

- Attributes -interested physical parameters.
- Thresholds -hard threshold value and soft threshold value.
- Schedule -time slot using TDMA.
- Count time -Maximum time period between two successive reports sent by a node.

1) Merits
   a) Outperform LEACH in terms of energy disappearing and total lifetime of the network.
2) Demerits
   a) Overhead and complexity
   b) Forming multiple level clusters.
   c) Implement threshold-based functions.
   d) Dealing with attribute-based naming of query.

E. EECS

Energy Efficient Clustering Scheme (EECS) [10] is a clustering algorithm in which improved the magazine data gathering applications in WSNs. In this, cluster head candidates used for the ability to raise to CH for a given round. EECS is a LEACH-like design, where the network is split into several clusters and single-hop communication linking the CH and the BS is performed. This method also produces a near uniform distribution of cluster heads. Additional in the cluster formation stage, a new approach is introduced to balance the load along with cluster heads. EECS is fully distributed and high energy efficient. EECS extends LEACH by the dynamic sizing of clusters based on cluster distance from the base station.

1) Merits
   a) It constructs balancing point between intra-cluster energy consumption and inter-cluster communication load.
   b) It is performed by changing sizing based on cluster distance from the BS.

2) Demerits
   a) Account of single-hop communications in EECS, so it is not suitable for large-range networks.
   b) It requires high global knowledge regarding the distances between the CHs and the BS.

E ECS produce much more control overhead complexity.

F. EEUC

Energy-Efficient Uneven Clustering (EEUC) algorithm, [11], is a periodical data gathering application in wireless sensor networks. It wisely organizes the network unequal clustering and multihop routing. It is a clustering and distributed competitive algorithm where cluster heads are elected by localized competition, which is different LEACH. The hot spot is the major problem in WSNs because of multi-hopping that occurs while CHs closer to the sink tend to die faster compare to another node in the WSNs, since they relay much more traffic than remote nodes. EEUC divider the all nodes in the cluster of unequal size, and cluster nearer to the sink have smaller sizes than those farther away from the sink. Thus, cluster heads (CHs) close to the sink can conserve some energy for the inter-cluster data forwarding.

1) Merits
   a) The unequal clustering method of EEUC improves the network lifetime compare to LEACH and HEED.

2) Demerits
   a) The more global data aggregation can result in a large amount of overhead for all nodes and deteriorate the network performance.

b) Performing of clustering in every round imposes significant overhead.

c) The routing method can outcome in new hot spots, in that only one of the two nodes whose communication costs are the least among the neighbor CHs can be relayed nodes, even though both of them have little residual energy.

G. Heed

Hybrid energy efficient Distributed Clustering (HEED) [12] is a multi-hop clustering algorithm for wireless sensor networks. It is considered to choose different cluster heads in a field, based on the amount of residual energy that is distributed in relation to a neighboring node. The random selection of cluster head is not suitable in HEED, because it decrease the lifetime of network.

The main goals of the network

1. Extend the network lifetime by distributing energy consumption.
2. Terminate the clustering process within a constant number of cycles/steps.
3. Minimizing control overhead.
4. Produce well-distributed cluster heads and compact clusters.

The most important of HEED is the method of cluster head selection. Cluster head is chosen based on two important parameters. One parameter depends on the node’s residual energy, and the other parameter is the intra-cluster communication cost as a function of cluster density.

1. The residual energy of each node is most heads. That parameter is normally used in many other clustering schemes.
2. Intra-cluster communication cost reflects the node degree or node’s nearness to the neighbor and is used by the nodes in deciding to join the cluster.

1) Merits
   a) HEED distribution of energy; improve the lifetime of the nodes within the network this method stabilizing the neighboring node.

b) Does not require particular node’s ability, such as location-awareness.

c) Does not create assumptions about node distribution.

2) Demerits
   a) The random selection of the cluster head may cause higher communication overhead.

b) The cyclic cluster head rotation or election needs extra energy to reconstruct clusters.

H. Pegasis

Power-Efficient Gathering in Sensor Information Systems (PEGASIS) [13], a near best chain-based protocol that is a development over LEACH. The PEGASIS is for each node to receive from and transmit to close neighbors and take turns being the leader in transmission to the BS. It will be distributing the energy load evenly among the sensor nodes in the network. It is a data-gathering and near-optimal chain-based algorithm that establishes the design that energy conservation can outcome from nodes not directly forming...
clusters. This PEGASIS, the nodes are organized to form a chain; they can either be concentrated assigned by the sink and broadcast to all nodes or terminated by the nodes themselves by a greedy algorithm. In the process of chain formation, it is idea that all nodes have overall knowledge of the network and the greedy algorithm is employed.

1. The performance of this protocol can be better by using a greedy algorithm for chain structure.
2. Not allowing nodes, which disappear more energy to become the leader.
3. Applying a threshold adaptive for the remaining energy levels in nodes.

HGMR starts with a hierarchical decomposition of a multicast group into subgroups of manageable size by means of the key concept of the mobile geographic hashing of HRPM. Contained by each subgroup, HGMR adopts the local multicast scheme of GMR to forward data packets along multiple branches of the multicast tree in one transmission. HGMR, the multicast group is separated by subgroups using the mobile geographic hashing idea: the deployment area is recursively partitioned into a number of $d^2$ equations-sized square sub-domains called cells, somewhere $d$ is the parameter for any node to become a bottleneck.

**HGMR**

Hierarchical Geographic Multicast Routing (HGMR) [14], proposed is a location-based multicast protocol. This protocol seamlessly incorporates the key aim concepts of the Geographic Multicast Routing (GMR) [17] and Hierarchical Rendezvous Point Multicast (HRPM) [18] protocols, and optimizes them by providing forwarding energy efficiency as well as scalability to large-scale WSNs.

HGMR starts with a hierarchical decomposition of a multicast group into subgroups of manageable size by means of the key concept of the mobile geographic hashing of HRPM. Contained by each subgroup, HGMR adopts the local multicast scheme of GMR to forward data packets along multiple branches of the multicast tree in one transmission. HGMR, the multicast group is separated by subgroups using the mobile geographic hashing idea: the deployment area is recursively partitioned into a number of $d^2$ equations-sized square sub-domains called cells, somewhere $d$ is the parameter for any node to become a bottleneck.

**Merits**

a) This protocol is able to better than LEACH for different network sizes and topologies.
b) The energy load is spread out completely in the network.
c) It increases the lifetime of network twice as much the lifetime of the network under the LEACH protocol

**Demerits**

a) It is the requirement of having a complete view of the network topology at each node for chain construction.
b) The communication manner suffers from too much delay caused by the single chain of distant nodes and a high probability for any node to become a bottleneck.
c) All nodes maintain a complete database about the location of all other nodes in the network. So it takes lot of time and energy.

I. HGMR

HGMR exist free of the scalability problem in that only the manageable destinations in a cell.

2) Demerits

a) All transmissions are concentrated to APs. APs can be changed to another cell by hash function, cell will be incomplete and might carry on unbalanced energy consumption.
b) HGMR makes routing paths inefficient, when number of multicast member density and scalability of network is high.

J. CBHRP

Cluster Based Hierarchical Routing Protocol (CBHRP) [15], is an optimum energy efficient cluster based hierarchical routing protocol for wireless sensor network. It is a two layer protocol anywhere a number of cluster cover the complete region. This protocol introduces a concept of head-set instead of a cluster head. By one time, only one member of head-set is active and the remaining are in sleep mode. A number of states of a node are to establish in this protocol such as- active state, associate state, passive associate state, candidate state, and non-candidate state. This protocol divides the network into a few real clusters include an active cluster head and several associate cluster heads. For a particular number of data collection nodes, the head-set members are thoroughly adjusted to reduce the energy consumption, which increase the network lifetime.

1) Merits

a) An effective routing protocol, considered different states for reducing the energy consumption.
b) Headset members are active during data collecting, transmitting time only.

2) Demerits

a) At the end of election phase, a few nodes only selected as members of the head-sets within a cluster.
b) All the head-set members divide the same time slot to transmit their frames.

K. Beenish

Balanced Energy Efficient Network Integrated Super Heterogeneous (BEENISH) Protocol [16]. It assumes WSN containing four energy levels of nodes. Here, Cluster Heads (CHs) are elected on the basis of the residual energy level of nodes. BEENISH implements the same concept as in DEEC, in terms of selecting CH which is based on the residual energy level of the nodes with respect to the average energy of the network. Though, DEEC is based on two types of nodes; normal and advance nodes. BEENISH uses the idea of four types of nodes; normal, advance, super and ultra-super nodes. In this reach longer stability, lifetime and extra effective messages than Stable Election Protocol (SEP) [19], Distributed Energy Efficient Clustering (DEEC) [20], Developed DEEC (DDEEC) [21] and Enhanced DEEC (EDEEC) [22]. In BEENISH ultra-super nodes are mostly elected as CH as different from super, advance and normal nodes. Hence, in this way energy consumed by all nodes is equally distributed.
1) Merits
   a) Election of CH based on average and the residual energy of the network.
   b) Stability, network lifetime and throughput better than DEEC, DEEDEC and EDEEC.

2) Demerits
   a) Difficult to recognize different levels of heterogeneous network having normal, advance, super and ultra-super nodes.
   b) Hard to achieve coordination among clusters.

<table>
<thead>
<tr>
<th>Hierarchical clustering routing protocols</th>
<th>LEACH</th>
<th>TL-LEACH</th>
<th>TEEN</th>
<th>APTEE N</th>
<th>EECN</th>
<th>EEUSS</th>
<th>HEED</th>
<th>HGMR</th>
<th>CBHRP</th>
<th>PEGASIS</th>
<th>BEENISH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cluster head selection</td>
<td>Cluster-head based</td>
<td>Cluster-head based</td>
<td>Cluster-head based</td>
<td>Cluster-head based</td>
<td>Cluster-head based</td>
<td>Cluster-head based</td>
<td>Cluster-head based</td>
<td>Cluster-head based</td>
<td>Non-cluster-head based</td>
<td>Heterogeneous Stationary</td>
<td></td>
</tr>
<tr>
<td>Difference of capabilities</td>
<td>Homo</td>
<td>Homo</td>
<td>Homo</td>
<td>Homo</td>
<td>Homo</td>
<td>Homo</td>
<td>Homo</td>
<td>Homo</td>
<td>Homo</td>
<td>Homo</td>
<td>Homo</td>
</tr>
<tr>
<td>Mobility</td>
<td>Generous</td>
<td>Generous</td>
<td>Stationary</td>
<td>Generous</td>
<td>Stationary</td>
<td>Generous</td>
<td>Stationary</td>
<td>Generous</td>
<td>Stationary</td>
<td>Generous</td>
<td></td>
</tr>
<tr>
<td>Role</td>
<td>Relay</td>
<td>Relay</td>
<td>Relay</td>
<td>Relay</td>
<td>Relay</td>
<td>Relay</td>
<td>Relay</td>
<td>Relay</td>
<td>Relay</td>
<td>Relay</td>
<td></td>
</tr>
<tr>
<td>Cluster size</td>
<td>n</td>
<td>n</td>
<td>n</td>
<td>n</td>
<td>Yes</td>
<td>n</td>
<td>n</td>
<td>n</td>
<td>n</td>
<td>n</td>
<td></td>
</tr>
<tr>
<td>Clustering process</td>
<td>Distribute</td>
<td>Distribute</td>
<td>Distribute</td>
<td>Distribute</td>
<td>Distribute</td>
<td>Distribute</td>
<td>Distribute</td>
<td>Distribute</td>
<td>Distribute</td>
<td>Distribute</td>
<td></td>
</tr>
<tr>
<td>Parameters for CH election</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Network type</td>
<td>Proactive</td>
<td>Proactive</td>
<td>Reactive</td>
<td>Reactive</td>
<td>Proactive</td>
<td>Proactive</td>
<td>Proactive</td>
<td>Proactive</td>
<td>Proactive</td>
<td>Proactive</td>
<td></td>
</tr>
<tr>
<td>Energy Efficiency</td>
<td>Very low</td>
<td>Low</td>
<td>Very high</td>
<td>Moderate</td>
<td>Moderate</td>
<td>High</td>
<td>Moderate</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
<td></td>
</tr>
<tr>
<td>Load Balancing</td>
<td>Moderate</td>
<td>Bad</td>
<td>Good</td>
<td>Moderate</td>
<td>Good</td>
<td>Moderate</td>
<td>Bad</td>
<td>Bad</td>
<td>Moderate</td>
<td>Moderate</td>
<td></td>
</tr>
</tbody>
</table>

III. CONCLUSION

In this paper, we have surveyed some of the existing ‘hierarchical clustering’ routing protocols, specifically with respect to their network lifetime and reliability requirements. In addition, we also provided a comparative study of 11 protocols reacting towards various networking parameters. As there is the infinite existence of routing protocols, we have considered only 11 samples of hierarchical routing protocols to explicate the concept with individual Merits and Demerits. The factors affecting CH communication and cluster formation are highlighted and those challenges are pinpointed for future research directions in this regard.

REFERENCES


Mrs. M . Sangeetha received B.E. degree in computer science and Engineering from K.S.R college of Technology, Tirchencode, Namakkal, Tamil Nadu, India in 2003 and the M.E degree in Computer Science and Engineering from Kongu Engineering College, Perundurai, Erode, Tamil Nadu, India in 2005. She is doing her Ph.D. programme under the area of wireless sensor network in Anna University, Chennai. She has a teaching experience of about 10 years. At present she is working as Assistant professor in the Information Technology department at K.S. Rangasamy College of technology. She has published two papers in the reputed International Journal and three papers in International conference. Her current research interest includes Adhoc networks, Wireless sensor Networks, Network security She is a life member of ISTE.

Dr.A.Sabari received B.E. degree in Computer Science and Engineering from Madras University, Tamil Nadu, India in 2000 and the M.Tech degree in Computer Science and Engineering from Vishweshwariya technological university, karnataka, India in 2002. He has completed Ph.D. Degree in Anna University, Chennai, India in 2011. He is currently working as a Professor in the department of Information Technology at K. S. Rangasamy college of Technology, Tamil Nadu, India from 2002 to till date. He has published 10 papers in the reputed international journals and 9 papers in the national and international conferences. He is supervising 11 research scholars under Anna University, Chennai. His current research interest includes Adhoc networks, Wireless sensor Networks, Network security and Data Mining. He is a life member of CSI and ISTE.

Ms.T.Shanthi Priya holds a B.Tech degree in Information Technology from Periyar Manimmai University of Technology, Thanjoru, Tamil Nadu, India in 2013. Now she is an M.Tech student of Information Technology department in the K.S. Rangasamy College of Technology. She has presented two papers in the National level technical symposium. Her Research interests include Wireless Sensor Networks Adhoc Network, Network Security and Embedded System.