Analytical Study on Energy Efficient Clustering Algorithms for Wireless Sensor Networks

Nilima Rani Das¹, Debashree Mishra Sar²
¹Dept of CA, ITER, SOA University, Bhubaneswar, Odisha
²Dept of CA, ITER, SOA University, Bhubaneswar, Odisha

Abstract: Wireless Sensor Networks (WSNs) were initially designed to facilitate military operations but its application has since been extended to health, traffic, and many other consumer and industrial areas. The size of the sensor nodes can also range from the size of a shoebox to as small as the size of a grain of dust. As such, their prices also vary from a few pennies to hundreds of dollars depending on the functionality parameters of a sensor like energy consumption, computational speed rate, bandwidth, and memory. A lot of research has been done to maximize the longevity, balance the load and improve the energy efficiency of the WSN with minimal additional overhead. This necessitates the efficient organization of the network topology. For achieving scalable and efficient communication and proper organization of the network topology WSN uses clustering. In this paper we have discussed the underlying design principles and objectives of some existing energy efficient clustering algorithms.

Keywords: Clustering, TDMA schedule, heterogeneous WSN.

1. INTRODUCTION

It is a general trend in computing that computers are becoming smaller and more interconnected and sensor networks represent a logical extreme of this trend. Wireless sensor networks (WSNs) are attracting an increasing degree of research interest, with a growing number of industrial applications starting to emerge. Some of these applications are area monitoring, air pollution monitoring, machine health monitoring, agriculture, disaster management and border protection [1]. Since sensors are of limited battery power the techniques that improve energy efficiency to enhance the lifetime of the network are very much required. Clustering mechanisms are very useful in improving the energy efficiency and lifetime of the network. The energy spent in transmitting a bit is higher than performing a computation [2]. So it is wise to organize the sensor nodes into clusters [3] where the data gathered by the sensors is transmitted to the sink through a hierarchy of cluster-heads and the sink node will process the data based on the received information. In this way a lot of energy can be saved as the sensors send data to their local cluster-head which is nearer to them in comparison to the sink node.

Many clustering algorithms have been proposed [4, 5, 6, 7, 8, 9, 10, 11] to achieve the goal of energy efficiency. We have made a comparative study of some of them in this paper. The rest of this paper is organized as follows.

Section 2 discusses the working principles of some energy efficient clustering algorithms. Section 3 gives some comparisons based on some specific attributes of the algorithms. Finally, section 4 contains the conclusion.

2. WHAT IS CLUSTERING AND WHY IT IS NEEDED

A cluster in a network consists of a set of loosely connected computers that work together so that they can be viewed as a single system. And dividing the nodes into virtual groups according to some rules is called clustering. Nodes belonging to a group can execute different functions compared to other nodes. Some of the objectives of clustering include allowing of data aggregation, reducing data transmission, facilitating reusability of resources. Cluster-heads (CHs) and gateway nodes can form a virtual backbone for inter-cluster routing. Cluster structure gives the impression of a smaller and more stable network. It helps in improving network lifetime, reducing network traffic and contention for the channel. It also provides network scalability, resource sharing and efficient use of constrained resources. It is responsible for efficient resource allocations and reduced communication overheads, thus decreasing the overall energy consumption and interference among sensor nodes.

Clustering enables bandwidth reuse which can improve the system capacity. Due to the fact that within a cluster, all the normal nodes send their data to the CHs, so energy saving is achieved by absence of flooding, multiple routes, or routing loops. Any changes in a node behavior within a cluster affect only that cluster but not the entire network; as a result the network is robust to these changes.

3. WORKING PRINCIPLES OF ENERGY EFFICIENT CLUSTERING ALGORITHMS

A lot of studies have been made on clustering and many algorithms have also been developed. But all of them cannot be applied on WSNs, as the operational characteristics of the networks are such that these algorithms are unsuitable for them. Since the batteries used in sensors have limited energy, the message overhead should be less in a cluster. So the parameters used by the algorithms should include dynamic parameters. The clustering algorithms that can save more energy without hampering the general working of a network are more suitable for the WSNs.
In this section we present a literature survey of some existing protocols for energy efficient clustering in WSNs. We will also discuss the general working principle and design approach of the protocols which includes the features like network model, cluster head election, cluster formation and communication scheme.

The algorithms that are presented in this section are: Energy efficient clustering algorithm based on residual energy and concentration degree in WSN (ECHC) [4], Energy efficient clustering and data aggregation protocol for heterogeneous WSN (EECED) [5], Energy Efficient Clustering Scheme in WSN (EECS) [6], Distributive Energy Efficient Adaptive Clustering Protocol for WSNs (DEEAC) [7], Energy Efficient Clustering and Data Aggregation Protocol for Heterogeneous WSNs (EEAC) [8], Energy and Distance Based Clustering (EDBC) [9], Coined MIMO (CMIMO) [10] and Energy Efficient Hierarchical Clustering Algorithm for WSNs (EEHC) [11].

**ECHC.** Basic assumptions: The sink node is located outside the monitoring area with infinite energy; sensor nodes are stationary within a certain period of time after deployment; all the nodes have same computing power; CHs perform data aggregation; sensor nodes can dynamically adjust the radio power according to the communication distance; communication between the nodes is reliable and symmetric and it is based on both single hop and multi hop transmission.

In CH election process, the sink node elects the CHs based on their residual energy. For this, the sink node sends the average residual energy of the whole network to each sensor node. If the residual energy of a node is more than the average value then it calculates its election weight based on its residual energy and concentration degree. The concentration degree of a node is the number of sensor nodes it can sense. After calculating the weight it sends the weight and its own identification number to the sink. The sink node chooses a fixed number of sensor nodes with maximum election weights as CHs.

To form a cluster each CH sends a message as ‘re-join the cluster’ to the non cluster sensor nodes. After receiving the message the sensor node checks whether it is in the sensing range of the CH or not, that means whether it has received the message over a single-hop path or multi-hop path. If a node is in the sensing range of any cluster-head then it chooses that CH which has more residual energy and smaller cluster size than other clusters.

In data transmission phase the CH aggregates the data received from its member nodes and sends the aggregated data to the sink or to another cluster-head at the next higher level.

**EEED.** Basic assumptions: All sensor nodes and the BS are stationary after deployment; the WSN consists of heterogeneous nodes in terms of node energy; communication is based on single hop transmission; the BS is not energy limited in comparison to the energy of other nodes in the network.

In CH election process there are three types of nodes called normal, advanced and super. Nodes with higher battery power are termed as advanced and super nodes. The rest of the nodes are called as normal nodes. The chances of becoming CH are more for the advanced and super nodes than the normal nodes. All the cluster-heads are elected by using some weighted election probability based on their residual energy.

Clusters are formed when all CHs are elected. Each CH estimates its residual energy and broadcasts this information with its CH role to the neighboring nodes. After receiving this information a non-CH node joins the CH having maximum residual energy.

In data transmission phase each member in a cluster sends its data to the associated CH which then sends the received data to the base station.

**EECS.** Basic assumptions: All sensors and the BS are stationary after deployment; each node is location unaware but it knows the location of the base station; each sensor can communicate with the BS directly; sensors can use power control to vary the amount of transmission power depending on the distance to the receiver; communication is based on single hop transmission; communication is symmetric and a sensor can compute the approximate distance based on the received signal strength if the transmission power is known; all sensors are of equal significance.

In CH election process the nodes with higher residual energy are elected as CHs. Each node broadcasts a ‘compete-head-message’ within a specific radio range to check whether there is a candidate node existing with more residual energy within that range. If it finds a node with higher energy then it refrains itself from the competition. Otherwise it is elected as the head finally.

To form a cluster a node chooses the CH taking into account the two distance factors. One is the distance between the CH and the node itself. Second is the distance between the CH and the base station. The node chooses that CH which is nearest to it as well as nearest to the base station.

In data transmission phase the member nodes send data to their CHs which send the collected data to the base station.

**DEEAC.** Basic assumptions: Base station is fixed and located far away from the sensor nodes; all nodes in the network are homogeneous and energy constrained; communication is based on single hop transmission.

In CH election process the nodes with high residual energy and greater hotness value are elected as CH. The hotness factor for a node is its relative data generation rate to that of the network. The node having higher residual energy and greater hotness factor has a better chance to become the CH. Every node calculates a threshold value depending upon its residual energy and hotness factor and using that threshold value it decides whether it can be the CH or not.

In cluster formation when the nodes elect themselves to be CHs they broadcast an advertisement message. Every
non-CH node chooses its CH that requires minimum communication energy, based on the received signal strength of the advertisement. After choosing the CH the node informs this to the CH by sending a join request message.

The CH assigns a TDMA schedule to the member nodes in its cluster according to which the nodes send their data to the CH. Then the data is sent to the base station by the CH.

**EECED.** Basic assumptions: The sink with enough memory and computing capability is located at the centre of sensor nodes; sink node is assumed to know all the node locations; all sensor nodes are fixed and have a limited energy; all nodes are equipped with power control capabilities to vary their transmission power; communication is based on single hop transmission; the protocol is assumed to be an event-driven protocol architecture.

The selection of CH is done by some elector nodes. The sink node selects some nodes as elector nodes by broadcasting elector advertisement message. The elector node broadcasts energy request message along with its own energy level to its surrounding nodes. After receiving this message the other nodes compare their own energy with the energy of the nearest elector node. A node, whose energy is higher than that of the elector node, sends energy reply message. After getting information about the energy of the surrounding nodes the elector node selects the node with maximum residual energy as the CH and the node with second maximum residual energy as the next elector node. The process of selecting a CH using some elector nodes is shown in Fig. 1.

![Diagram of CH selection in WSN](image)

**Figure 1** Selection of CH using elector nodes in WSN

After being selected by elector node, the CH broadcasts a CH advertisement message containing its own identification number to form a cluster. The non-CH nodes choose the CH based on the signal strength of the advertisement message received from the CHs. Then the nodes send join request message to join that CH.

The CH set up a TDMA schedule according to which the member nodes send their data to the CH. The CHs aggregate all the data and send that to the base station directly.

**EDBC.** Basic assumptions: The whole network’s terrain is divided into concentric circular segments around the base station; the number of CHs is different in each segment; it is assumed that the nodes are aware of their approximate distance from the base station; every node is equipped with some computing power; communication is based on single hop transmission.

The nodes which are nearer to the base station and have higher residual energy are appropriate candidates for CHs. Each node computes some threshold value based on its residual energy and the distance measure between the base station and that node. The nodes having higher threshold value can become CHs.

When the nodes elect themselves to be CHs they broadcast an advertisement message. Every non-CH node chooses the closest CH as its own CH which can be known based on the received signal strength of the advertisement. After choosing the CH the node informs this by sending a join request message to that CH. The member nodes send their data to their corresponding CHs and these CHs send the collected data to the base station.

**CMIMO.** Basic assumptions: All sensor nodes are fixed and have a limited energy; the nodes in the network are energy constrained but not homogeneous; communication is based on multi hop transmission.

A cluster may have two CHs. One is master cluster-head (MCH) and another is slave cluster-head (SCH). Every cluster must have an MCH but it is not mandatory to have an SCH. First of all each node discovers its neighbors. After discovering the neighbors the node compares its remaining energy to those of its one-hop neighbors. If the node has the highest remaining energy in its neighborhood, it declares itself as an MCH and announces that to its neighbors. Each MCH sends an SCH invitation message to the node whose neighbor list overlaps the most with that of the MCH neighbor list. An invited node associates itself with the closest MCH and responds with an SCH acceptance message. The MCH confirms this association through an SCH confirmation message.

Every non-CH node asks its closest MCH to join its cluster through membership request message. After receiving a membership request message the selected MCH waits for a fixed duration of time allowing other non-CH nodes to send their membership request messages. Then the MCH sends a membership list message announcing the identification numbers of the non-CH nodes that this MCH has accepted these nodes to be in its cluster. The selected MCH sends its membership list message periodically.
MCH assigns a TDMA schedule to the member nodes according to which the nodes send their data to the MCH and then it aggregates and sends it to SCH. MCH and SCH send the data to MCH and SCH nodes of other clusters. Exchange of data in between the clusters through MCH and SCH is shown in Fig. 2. This process continues until the data reach the base station.

**Figure 2** Inter-cluster communication through MCH and SCH

**EEHC.** Basic assumptions: All sensors transmit at the same power level so they have the same radio range r; if two sensors try to communicate data between them and they are not within each other’s range then the data is forwarded by other sensors; a distance of d between any sensor and its CH is equivalent to d / r hops; each sensor uses 1 unit of energy to transmit or receive 1 unit of data; a routing infrastructure is used so that when a sensor communicates data to another sensor then only the sensors on the routing path can forward the data; this algorithm supports hierarchical clustering; the communication environment is free from error and contention which avoids data retransmission.

Each node with a certain probability advertises itself as a CH to the sensors that are present within its radio range. This advertisement is forwarded to all the sensors that are within certain fixed number of hops from the CH. Any sensor that receives such advertisements and is not itself a CH joins the closest CH. Any sensor that is neither a CH nor has joined any cluster becomes the CH. The algorithm has calculated the optimal values for the probability of becoming a CH and maximum number of hops allowed between a member and its CH using Poisson distribution process so that the energy consumption is minimized.

The member nodes send their data to the CH which in turn send that data to the next higher level CH in the hierarchy of clusters. This process continues until the data reach the base station.

**4. COMPARISON OF CLUSTERING ALGORITHMS**

The following characteristics are taken into consideration for the purpose of comparison.

**Communication with the base station.** EEHC supports both single hop and multi hop transmission to the base station. EECDA, EECS, DEEAC and EECEC are single hop based algorithms. EDBC, CMIMO and EEHC support multi-hop transmission. In single hop algorithms more energy is spent by the CHs as they send data directly to the base station which may be located far away from the CHs. This does not happen in case of multi hop algorithms. Though EECS is a single-hop based algorithm it allows a node to become a CH if it is nearer to the base station.

**Factors responsible for election of CHs.** In EEHC it depends on nodes concentration value and residual energy of a node. In EECDA it uses some threshold value which is based on residual energy of the node. In EECS the cluster-head selection is based on the residual energy of the node and the distance between that node and the base station. In DEEAC it depends on some threshold value based on the hotness factor and residual energy of the node. EECED, CMIMO also use the residual energy of the node for this. EDBC calculates some threshold value based on the residual energy of the node and the distance between that node and the base station for cluster-head selection. In EEHC it depends on some probability that is based on minimum energy consumption.

**Approach followed in CH selection.** In EEHC the CH is elected by the base station centrally for which the energy information of the clusters has to be sent to the base station which may consume some extra energy. But the CHs use no separate transmission to send this information to the base station. The information is sent along with the collected data. In EECDA, EECS, CMIMO, EEHC and EDBC energy information is not sent to the base station as all of them follow a distributed approach in cluster-head selection process. But in this approach some extra overhead is created in the exchange of control messages between the sensors. DEEAC also follows distributed approach to select the CHs. But it is required to send the energy information of the cluster to the base station so that the base station can broadcast the residual energy of the whole network periodically. The cluster-heads send this information to the base station along with the aggregated data. In EECEC the energy information is not sent to the base station but to the elector node that belongs to the same cluster and has the responsibility to elect the CH.

All of the above algorithms select the CHs based on their residual energy. But the algorithms that use the distance and hotness factors are more efficient than others. If only energy parameter is considered then the CHs which are far away from the base station will spend more energy in data transmission and the CHs having lower hotness values cannot manage well in the regions where the data generation rate is very high.
Cluster count. In ECHC, EECED, EECS, DEEAC, EDBC and EEHC the cluster count is fixed. But in EECDA and CMIMO the cluster count is not constant. The algorithms which use a fixed number of clusters are better than others, because it evenly consumes the node energy [12].

5. CONCLUSION

The above clustering algorithms are very effective in improving the energy efficiency of the WSNs which enhances their lifespan. But none of them is applicable for every kind of WSN. Every algorithm has its strength and limitations that make it suitable for a particular category of networks. Based on the architecture and critical design goals of different WSNs an algorithm can be chosen that is most appropriate for a particular network.

References