WALIA journal 30(1): 87-95, 2014
Available online at www.Waliaj.com
ISSN 1026-3861


# Study methods for clustering in wireless sensor network

Ali azimi kashani*, Mohsen Ganj, Iman Mosavian

*Department of Computer Engineering, Shoushtar Branch, Islamic Azad University, shoushtar, Iran*

**Abstract:** A wireless sensor network (WSN) consisting of a large number of tiny sensors can be an effective tool for gathering data in diverse kinds of environments. However, a WSN is a power constrained system, since nodes run on limited power batteries which shorten its lifetime. Prolonging the network lifetime depends on efficient management of sensing node energy resource. In this paper, we explain clustering in WSN and survey different clustering algorithms for WSN and also compare currently clustering algorithms for wireless sensor networks. This paper should provide the reader with a basis for research in clustering schemes for wireless sensor networks.

**Key words:** *Wireless sensor networks; Clustering protocols; Energy efficiency*

## 1. Introduction

Recent development in Micro-Electro-Mechanical Systems (MEMS) has made the deployment of small size sensor nodes inexpensive. The deployment of sensor nodes for the monitoring/detection of different events in environment are known as Wireless Sensor Network (WSN). In typical WSN, the sensors are generally randomly deployed in region of interest. Placed near to these nodes, is a base station (BS)/sink which is further connected to internet. Base station give commands to these nodes and gather the sensed data from the nodes, which is then accessible through the internet. To avoid redundancy sensor nodes sense the physical phenomenon, do the local data aggregation and transmit the sensed data to the base station. A typical sensor node consists of processor, memory, sensing, power supply, transceiver and sometimes mobility module, location module and actuator. Over the years WSN's have been used for many applications like military surveillance, disaster management, forest fire detection, seismic detection , habitat monitoring, biomedical health monitoring, inventory tracking, animal tracking, hazardous environment sensing and smart spaces. Senor nodes have very limited battery power and random deployment in difficult terrain; make it almost impossible tore charge/replace the dead battery. So, battery power in WSN is considered as scarce resource and should be used efficiently. Sensor node consumes battery in sensing data, receiving data, sending data and processing data. Consequently, Out of all sensor node operation, ending/receiving data consumes more energy than any other operation. The energy consumption when transmitting 1 bit of data on the wireless channel is similar to energy required to execute thousands of cycles of CPU instructions (Yick et al., 2008). Therefore, the energy efficiency of the wireless communication protocol largely affects the energy consumption and network lifetime of wireless sensor networks. In large WSNs sensors are often grouped into clusters. Clustering is essential for sensor network applications where a large number of wireless sensors are deployed for sensing purposes. If each and every sensor starts to communicate and engage in data transmission in the network, great data congestion and collisions will be experienced. This will drain energy quickly from the sensor network. Clustering is a method used to overcome these issues. In clustered networks, some sensors are elected as cluster heads (CHs) for each cluster created. Sensor nodes in each cluster transmit their data to the respective CH and the CH aggregates data and forwards them to a central base station. Clustering 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 reelection of CHs within clusters based on their residual energy is a possible solution to balance the power consumption of each cluster. Clustering increases the efficiency of data transmission by reducing the number of sensors tempting to transmit data to the base station. Aggregating data at CHs via intra-cluster communication also helps in eradicating data duplication. Clustering is proposed because of its network scalability, energy saving and network topology stability. Clustering schemes reduce the communication overheads among the sensor nodes. Clustering algorithms however have some disadvantages such as additional overheads during Cluster head (CH) selection, assignment and cluster

* Corresponding Author.

formation rocess (Karenos et al., 2008) Looking at Fig. 1, we can see the architecture of a generic Wireless Sensor Network (Dechene et al., 2010), and examine how the clustering phenomenon is an essential part of the organizational structure.

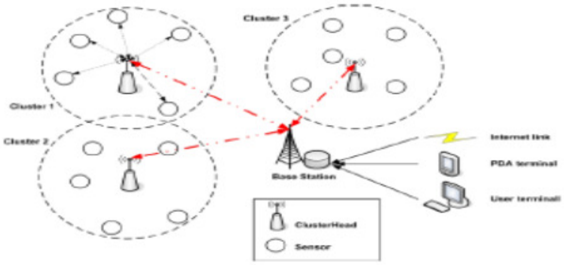


**Fig. 1:** General Sensor Network Architecture

The communication model that wireless sensor network uses is either single hop r multi hop. Since energy consumption in wireless systems is directly proportional to the square of the distance, single hop communication is expensive in terms of energy consumption. Most of the routing algorithms use multi hop communication model since it is more energy efficient in terms of energy consumption however, with multi hop communication the nodes which are closer to the cluster head are under heavy traffic and can create gaps near the cluster head when their energy terminates (Israr and Awan, 2006).

Some important clustering algorithms have appeared in the literature, the following are the components of a clustered WSN (Abbasi and Younis, 2007).

• Sensor Node: A sensor node is the core component of a WSN. Sensor nodes can take on multiple roles in a network, such as simple sensing; data storage; routing; and data processing.

• Clusters: Clusters are the organizational unit for WSNs. The dense nature of these networks requires the need for them to be broken down into clusters to simplify tasks such a communication.

• Cluster heads: Cluster heads are the organization leader of a cluster. They often are required to organize activities in the cluster. These tasks include but are not limited to data-aggregation and organization the communication schedule of a cluster.

• Base Station: The base station is at the upper level of the hierarchical WSN. It provides the communication link between the sensor network and the end-user.

• End User: The data in a sensor network can be used for a wide-range of applications. Therefore, a particular application may make use of the network data over the internet, using a PDA, or even a desktop computer. In a queried sensor network (where the required data is gathered from a query sent through the network). This query is generated by the end user. (Dechene et al., 2010) The rest of this paper is organized in the following manner: Section 2 will provide an overview of clustering schemes. Section 3 will introduce some important clustering protocols for WSNs and Comparison them. We will conclude this paper with Section 4.

## 2. Overview of clustering schemes

### 2.1. Basic concepts

In large scale WSNs, scalable architectural and management strategies are required to achieve goals such as extended lifetime, scalability, coverage, robustness and especially simplicity. To fulfill these requirements, it is necessary to design an efficient and scalable network layer protocol. There exist a lot of routing protocols in wireless sensor networks. Flat routing protocols are firstly proposed to solve the problem. However, flat routing protocols cannot scale well in large scale WSNs since they are mostly proactive or reactive. Consequently, researchers and engineers moved to clustering routing protocols in recently years. In clustering schemes, sensor nodes are often grouped into individual geographically disjointed and usually no overlapped clusters. All the adjacent nodes are in one ore more clusters according to different cluster formation mechanisms. In clustering schemes, there would be a leader, often called cluster-head (CH). CHs are responsible for coordination between inter-clusters and intra-clusters, such as cluster formation, data collection, data aggregation, and communication with base-stations. Non-cluster-head nodes in a cluster, e.g., cluster members serve as different roles associated with different status, functions and responsibilities according to different network usages and topologies (Jiang et al., 2009).

### 2.2. Advantages for Clustering Methods

#### 2.2.1. Less overheads

As all cluster members only send data to CHs and CHs only send data to base-stations or sinks after data aggregation and fusion within clusters, clustering schemes can dramatically alleviate flooding overheads while still fulfilling the network QoS requirements by decreasing the retransmission of broadcast or multicast packets. This feature significantly reduces transferring data, saves energy and bandwidth resources, and also scales well in routing path building phase and data transmitting phase.

#### 2.2.2. Easy maintenance

After cluster formation, clustering schemes make it easier for network topology control and responding to network changes caused by network dynamics, node autonomy, node mobility, local changes and unpredicted failures. Since all these changes only need to be detected, and managed within an individual cluster, not the entire network, the entire network is more robust and easy for maintenance and management. Basically, clustering schemes consist of two phases, cluster formation and cluster maintenance. Cluster formation is referred to how to construct a hierarchical cluster structure in the network initialization stage, while cluster maintenance is referred to how to update, control and manage the network topology changes caused by node mobility, failures, link breakage or some other reasons. Since cluster heads (CHs) serve as a central coordinator to perform the distributed sensing tasks in a local cluster, which node is chosen to be a CH is the key problem during the entire network initialization phase. When the first cluster heads are chosen according to some predefined rules, those CHs notify their neighbors by broadcasting some clustering information to permit their potential members to join. Moreover, a cluster head consumes more energy than the other nodes in a cluster because it has additional functions. Therefore, a cluster head cannot always act as a cluster head due to large energy consumptions. In case of failures or energy exhaustion, a new cluster head should be elected to provide self-organizing, self-healing, and robustness to network topology changes while maintaining the underlying network connectivity in a dynamic environment(Jiang et al., 2009).

### 2.3. Challenges for Clustering Methods

Clustering schemes can effectively improve the network performance. There are several basis limitations in WSNs that clustering schemes must consider.

1) 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 communication.

2) 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.

3) 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.

4) Secure Communication: The ability of a WSN to provide secure communication is ever more important when considering these networks for military applications (Akyildiz et al., 2002). The self-organization of a network has a huge dependence on the application it is required for. An establishment of secure and energy efficient intra-cluster and inter-cluster communication is one of the important challenges in designing clustering algorithms since these tiny nodes when deployed are unattended to in most cases.

5) 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.

6) Synchronization: When considering a clustering scheme, synchronization and scheduling will have a considerable effect on the overall network performance. Slotted transmission schemes such as TDMA allow nodes to regularly schedule sleep intervals to minimize energy used. Such schemes require synchronization mechanisms to setup and maintain the transmission schedule.

7) Data Aggregation: Data aggregation eradicates duplication of data. In a large network there are often multiple nodes sensing similar information. Data aggregation allows differentiation between sensed data and useful data. Many clustering schemes providing data aggregation capabilities (Heinzelman et al., 2000) must carefully select a suitable clustering approach.

8) 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. When designing clustering schemes, it is important to look for mechanisms that ensure link recovery and reliable data communication.

9) Quality of Service (QoS): From an overall network standpoint, we can look at QoS requirements in WSNs. Many of these requirements are application dependant such as acceptable delay and packet loss tolerance. Existing lustering algorithms for WSN mainly focus on providing energy efficient network utilization, but pay less attention to QoS support in WSN. QoS metrics must be taken into account in the design process.

### 2.4. Classification of Clustering Schemes

According to the cost of a clustering scheme in different aspects qualitatively or quantitatively, is grouped the clustering cost terms into five categories, i.e., the required explicit control message exchange, the ripple effect of re-clustering, the stationary assumption, constant computation round, and communication (message) complexity. Abbasi and Younis enumerated a rich set of attributes that can be used to categorize and differentiate clustering algorithms of WSNs. The major attributes and the relevant attributes are listed as follows:

1) Cluster properties, including cluster count, stability, intra-cluster topology and inter-CH connectivity;

2) Cluster-head capabilities, including mobility, node types and Role;

3) Clustering process, including methodology, objective of node grouping, cluster-head selection and algorithm complexity. Abbasi and Younis also summarized five clustering objectives, i.e., load balancing, fault-tolerance, increased connectivity and reduced delay, minimal cluster count, and maximal network longevity. They grouped existing clustering algorithms for WSNs into two subsections according to their convergence rate, i.e., variable and constant convergence time algorithms. In general, it is hard to set a common criterion for various underlying clustering schemes, including the similarities and differences between schemes in the same category. Based on the existing various classification criteria, we list the following attributes as classification criteria for WSNs according to cluster-head (CH) properties:

1) Existence. Depending on whether there exist cluster-heads within a cluster, clustering schemes can be grouped into CH-based clustering and non-CH-based clustering

2) Count variability. In some application environments, the set of cluster heads are predetermined and thus the number of clusters is preset. Therefore, clustering schemes can be categorized into fixed or variable cluster-heads clustering.

3) Selectivity. Ideally, all the cluster member nodes should be chosen to be a cluster head in a round-robin fashion to achieve load balancing, energy balancing, and topology reconfiguration. According to whether cluster-heads are pre-assigned or chosen from the deployed nodes set by certain cluster-head selection rules, clustering schemes can be grouped into reassigned or dynamic selected.

4) Role. In WSNs, cluster head can simple act as a local coordinator for its cluster members, perform intra-cluster transmission or serve as a backbone node for higher cluster hierarchy. Thus clustering schemes can be grouped into local or global ones.

5) Node mobility. Clustering schemes can be grouped into stationary or mobile ones according to the mobility attributes of cluster heads.

6) Hop distance. Depending on hop distance between node pairs in a cluster, clustering algorithms can be grouped into 1-hop clustering and multi-hop clustering.

7) Explicit control messages. During cluster formation or maintenance period, it may require explicit clustering-related information exchanged between node pairs, such as routing information or data packets. Therefore clustering schemes can be grouped into explicit control message-dependent or non-explicit control message ones. In the latter, cluster is proactive to receive data and cluster member switches from sleep mode to active mode when sensing objects reach a threshold.

8) Overlapping. In many applications, it is natural to group spatially close sensor nodes in the same cluster to eliminate redundancy and overlapping. However, in some cases when sensors scattered or placed not properly, there exist overlapping areas among clusters and nodes. Therefore, clustering schemes can be grouped into overlapping clustering and non-overlapping clustering. Based on the proposed classification criteria for clustering, we present a comprehensive survey on some existing clustering algorithms in the following section.

## 3. Clustering protocols for WSNS

In this section, classified and analyzed some popular and effective clustering algorithms for WSNs.

### 3.1. Existing clustering protocols

#### 3.1.1. LEACH protocol

Low-energy adaptive clustering hierarchy (LEACH) ( Heinzelman et al., 2002) is one of the most popular hierarchical routing algorithms for sensor networks. The idea is to form clusters of the sensor nodes based on the received signal strength and use local cluster heads as routers to the sink. This will save energy since the transmissions will only be done by such cluster heads rather than all sensor nodes. All the data processing such as data fusion and aggregation are local to the cluster. Cluster heads change randomly over time in order to balance the energy dissipation of nodes. This decision is made by the node choosing a random number between 0 and 1. The node becomes a cluster head for the current Round if the number is less than the following threshold: Where n is the given node, P is the a priori probability of a node

$$T(n) = \begin{cases} \dfrac{p}{1 - p * \left(r \bmod \dfrac{1}{p}\right)} & if \quad n \in G \\ 0 & Otherwise \end{cases}$$

Where n is the given node, P is the a priori probability of a node being elected as a clusterhead, r is the current round number and G is the set of nodes that have not been elected as cluster heads in the last 1/P rounds. Each node during clusterhead selection will generate a random number between 0 and 1. If the number is less than the threshold (T (n)), the node will become a cluster head.

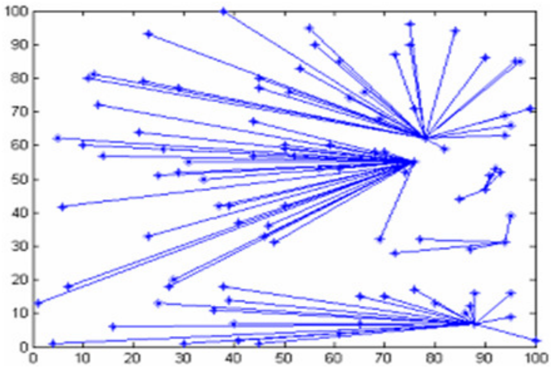


**Fig. 3:** Random selection of cluster head in LEACH

Following elevation to cluster head, the new cluster head will broadcast its status to neighboring nodes. These nodes will then determine the optimal cluster head (in terms of minimum energy required for transmission) and relay their desire to be in that particular cluster. The broadcast messages as well as cluster establishment messages are transmitted using CSMA (Carrier Sense Multiple Access) to minimize collisions. Following cluster establishment, cluster heads will create a transmission schedule and broadcast the schedule to all nodes in their respective cluster. The schedule consists of TDMA slots for each neighboring node. This scheduling scheme allows for energy minimization as nodes can turn off their radio during all but their scheduled time -slot.

In this way the nodes die randomly and dynamic clustering increases lifetime of the system. LEACH is completely distributed and requires no global knowledge of network. However, LEACH uses single-hop routing where each node can transmit directly to the cluster-head and the sink. Therefore, it is not applicable to networks deployed in large regions. Furthermore, the idea of dynamic clustering brings extra overhead, e.g. Head changes, advertisements etc., which may diminish the gain in energy consumption.

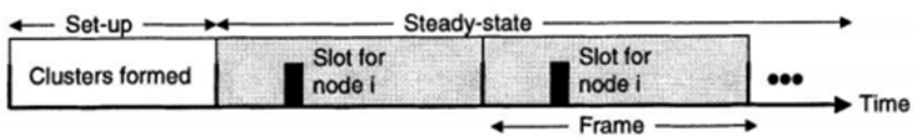


**Fig. 4:** Time line showing LEACH operation. Data transmissions are explicitly scheduled to avoid collisions and increase the amount of time each non-cluster head node can remain in the sleep state (Heinzelman et al., 2002).

### 3.1.2. LEACH-C protocol

LEACH offers no guarantee about the placement and/or number of cluster heads. In (Wendi et al., 2002), an enhancement over the LEACH protocol was proposed. The protocol, called LEACH-C, uses a centralized clustering algorithm and the same steady-state phase as LEACH. LEACH-C protocol can produce better performance by dispersing the cluster heads throughout the network. During the set-up phase of LEACH-C, each node sends information about its current location (possibly determined using GPS) and residual energy level to the sink. In addition to determining good clusters, the sink needs to ensure that the energy load is evenly distributed among all the nodes. To do this, sink computes the average node energy, and determines which nodes have energy below this average. Once the cluster heads and associated clusters are found, the sink broadcasts a message that obtains the cluster head ID for each node. If a cluster head ID matches its own ID, the node is a cluster head; otherwise the node determines its TDMA slot for data transmission and goes sleep until it's time to transmit data. The steady-state phase of LEACH-C is identical to that of the LEACH protocol.

### 3.1.3. HEED protocol

Hybrid Energy-Efficient Distributed Clustering (HEED) (Younis and Fahmy, 2004) is a distributed clustering scheme in which CH nodes are picked from the deployed sensors. HEED considers a hybrid of energy and communication cost when selecting CHs. Unlike LEACH, it does not select cell-head nodes randomly. Only sensors that have a high residual energy can become cell-head nodes. HEED has three main characteristics: The probability that two nodes within each other's transmission range becoming CHs is small. Unlike LEACH, this means that CHs are well distributed in the network.

• Energy consumption is not assumed to be uniform for all the nodes.

• For a given sensor's transmission range, the probability of CH selection can be adjusted to ensure inter-CH connectivity. In HEED, each node is mapped to exactly one cluster and can directly communicate w i^th its CH. The algorithm is divided into three phases:

1. Initialization phase: The algorithm first sets an initial percentage of CHs among all sensors. This percentage value, $C_{prob}$, is used to limit the initial CH announcements to the other sensors. Each sensor

sets its probability of becoming a cluster-head, $CH_{prob}$, as follows:

$$CH_{prob} = C_{prob} \times \frac{E_{residual}}{E_{max}}$$

Where $E_{residual}$ is the current energy in the sensor, and $E_{max}$ is the maximum nergy, which corresponds to a fully charged battery. $CH_{prob}$ is not allowed to fall below a certain threshold Pmin, which is selected to be inversely proportional to $E_{max}$

2. Repetition phase: During this phase, every sensor goes through several iterations until it finds the CH that it can transmit to with the least transmission power (cost).If it hears from no CH, the sensor elects itself to be a CH and sends an announcement message to its neighbors informing them about the change of status. Finally, each sensor doubles its $CH_{prob}$ value and goes to the next iteration of this phase. It stops executing this phase when its $CH_{prob}$ reaches 1. Therefore, there are 2 types of cell head status that a sensor could announce to its neighbors: Tentative status: The sensor becomes a tentative CH if its $CH_{prob}$ is less than 1. It can change its status to a regular node at a later iteration if it finds a lower cost CH. Final status: The sensor permanently becomes a CH if its $CH_{prob}$ has reached 1.3. Finalization phase: During this phase, each sensor makes a final decision on its status. It either picks the least cost CH or pronounces itself as CH.HEED uses a combination strategy of energy and communication cost to generate CHs. Since the energy is non-uniform distributed among all nodes, it is approximately avoided that two nodes within each other's transmission range have the same probability to become CHs in HEED. Moreover, the probability of CH election is flexible to provide inter-CH connectivity for a certain sensor's transmission range.

### 3.1.4. DWEHC protocol

Ding et al. (P. Ding, J. Holliday, A. Celik, 2005) have proposed DWEHC (Distributed Weight-Based Energy-Efficient Hierarchical Clustering) to achieve more aggressive goals than those of HEED. Basically, generating balanced cluster sizes and optimizing the intra-cluster topology. DWEHC proceeds in a distributed manner and has O (1) time complexity. Each sensor calculates its weight after locating the neighboring nodes in its area. The weight is a function of the sensor's energy reserve and the proximity to the neighbors. In a neighborhood, the node with largest weight would be elected as a CH and the remaining nodes become members. At this stage the nodes are considered as first-level members since they have a direct link to the CH. A node progressively adjusts such membership in order to reach a CH using the least amount of energy. Basically, a node checks with its non-CH neighbors to find out their minimal cost for reaching a CH. Given the node's knowledge of the distance to its neighbors, it can assess whether it is better to stay a first-level member or become a second-level one; reaching the CH over a two-hop path. It is worth noting that by doing so the node may switch to a CH other than its original one. The process continues until nodes settles on the most energy efficient intra-cluster topology. To limit the number of levels, every cluster is assigned a range within which member nodes should lay. Fig. 5, redrawn from (Ding et al., 2005), illustrates the structure of the intra-cluster topology. Both DWEHC and HEED are similar in many ways; every node in the network participates in the clustering process, they do not make any assumption about the network size and consider energy reserve in CH selection. Despite such resemblances, there are many performance differences between DWEHC and HEED. For example, clusters generated by DWEHC are more well-balanced than HEED. DWEHC also achieves significantly lower energy consumption in intra-cluster and inter-cluster communication than HEED.

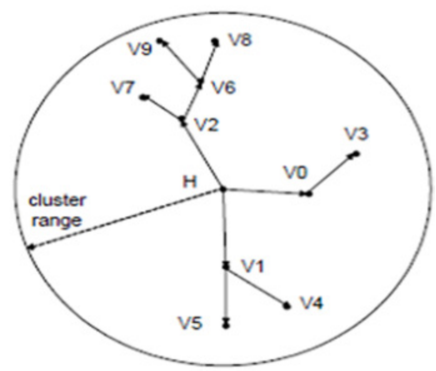


**Fig. 5:** DWEHC generates a multi-hop intra-cluster topology with the CH at the root and member nodes are ordered in breadth-first order.

### 3.1.5. TL-LEACH protocol

Two-Level Hierarchy LEACH (TL-LEACH) is a proposed extension to the LEACH algorithm. It utilizes two levels of cluster heads (primary and secondary) in addition to the other simple sensing nodes as shown in the Fig. 5. In this algorithm, the primary cluster head in each cluster communicates with the secondary's, and the corresponding secondary's communicate with the nodes in their sub-cluster. Data-fusion can also be performed as in LEACH. In addition, communication within a cluster is still scheduled using TDMA time-slots. The organization of a round will consist of first selecting the primary and secondary cluster heads using the same mechanism as LEACH, with the a priori probability of being elevated to a primary cluster head less than that of a secondary node.

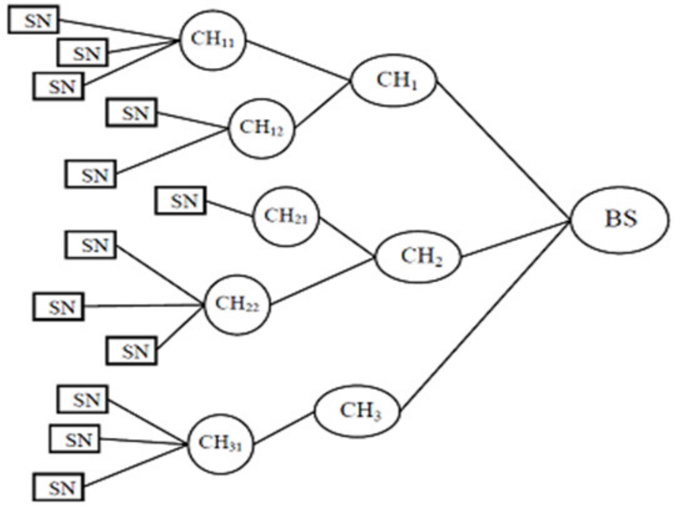


**Fig. 5:** The topology network after the setup cluster phase is complete

Communication of data from source node to sink is achieved in two steps:

1) Secondary nodes collect data from nodes in their respective clusters. Data-fusion can be performed at this level.

2) Primary nodes collect data from their respective secondary clusters. Data-fusion can also be implemented at the primary cluster head level .

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

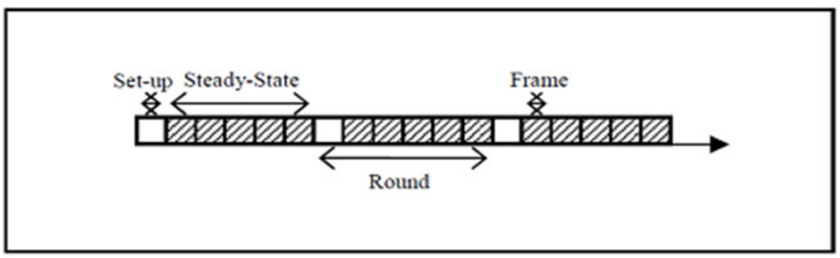


**Fig. 7:** Time-line showing TL-LEACH operation. Adaptive cluster are formed during the set-up phase

And data transfers occur during the steady state phase. We obtained that, this protocol is able to deliver more data packets than the original LEACH protocol. Particularly, the TL-LEACH has allowed us to reach a 20% increase as it regards the delivered packets and an increase of even the 30% of the lifetime of the network.3.1.6. A-LEACH protocol in this section we describe A-LEACH (Abdellah et al., 2010) which is an extension of the LEACH, which improves the stable region of the clustering hierarchy and decrease probability of failure nodes using the characteristic parameters of heterogeneity. Routing in A-LEACH works in rounds and each round is divided into two phases, the Setup phase and the Steady state; each sensor knows when each round starts using a synchronized clock.

*A-LEACH protocol has the following advantages:*

- Self-configuration of clusters is independent of the base station (distributed algorithm).
- The data are merged to reduce the amount of information transmitted to the base station.
- The Gateways reduces energy consumption and extends the lifetime of the cluster head in network.
- The Gateway decrease probability of failure nodes and prolong the time interval before the death of the first node and increasing the lifetime in heterogeneous WSNs
- The use of techniques TDMA / CDMA allows a hierarchy and makes clustering on several levels. They can save more energy.
- When all normal nodes death the CAG nodes continues to send data to the sink.

### 3.1.6. PEACH protocol

PEACH (Yi et al., 2007), is a power-efficient and adaptive clustering hierarchy protocol for wireless sensor networks. Most existing clustering protocols consume large amounts of energy, incurred by cluster formation overhead and fixed level clustering, particularly when sensor nodes are densely deployed in wireless sensor networks. To solve this problem, PEACH (Power-Efficient and Adaptive Clustering Hierarchy) protocol is proposed for WSNs to minimize the energy consumption of each node, and maximize the network lifetime. In PEACH, cluster formation is performed by using overhearing characteristics of wireless communication to support adaptive multi-level clustering and avoid additional overheads. In WSNs, overhearing a node can recognize the source and the destination of packets transmitted by the neighbor nodes. PEACH is applicable in both location-unaware and location-aware wireless sensor networks. This protocol can significantly save energy consumption of each node, prolong the network lifetime, and is less affected by the distribution of sensor nodes compared with existing clustering protocols.

### 3.1.7. EEHC protocol

Energy Efficient Hierarchical Clustering (EEHC) (Bandyopadhyay and Coyle, 2003) is a distributed, randomized clustering algorithm for WSNs. CHs collect data from the non-cluster head node in different clusters and send an aggregated report to the base-station. This technique 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 neighboring 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. Any node that receives such announcements and is not itself a CH becomes the member of the closest cluster. Forced CHs are nodes that are neither CH nor belong to a cluster. If the announcement does not reach to a node within a preset time interval t that is calculated based on the duration for a packet to reach a node that is k hops away, the node will become a forced CH assuming that it is not within k hops of all volunteer CHs. 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. The CHs closest to the base station are at a disadvantage because they are relays for other CHs.

### 3.1.8. DEBC protocol

Distributed Energy Balanced Clustering (DEBC) (Bajaber and Awan, 2010), is a distributed energy balanced clustering for wireless sensor networks. DEBC forms clusters, selects cluster heads, selects cluster senders and determines appropriate routings in order to reduce overall energy consumption and enhance the network lifetime. This protocol is used to collect data from distributed sensor nodes and transmit the sensed data to the base station using single hop or multi hop communication. The operation of the DEBC protocol is organized into rounds. Some rounds begin with a setup phase when the clusters are formed followed by steady state phase. The steady state phase is divided into several frames and sensor nodes transmit and receive the data at each frame. In next round, the cluster senders sense their environment, collect sensed data and transmit the data to the next neighbors counterclockwise. Each sensor node receives data from previous neighbor, aggregates with its own data, and transmits to the next neighbor on the ring. Upon receiving the aggregated data from previous neighbors, cluster senders transmit it to the base station using multi hop or single hop communication as shown in Fig8. In order to distribute energy consumption over sensor nodes, the cluster sender's role should be rotated among the sensor nodes to prevent their exhaustion. DEBC uses the remaining energy for cluster sender's rotation

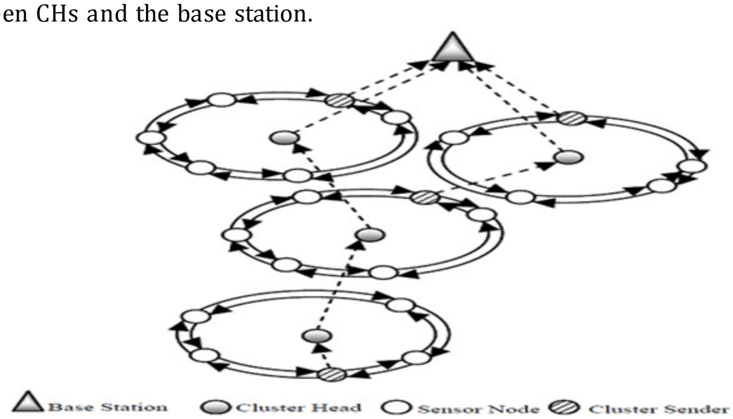


**Fig. 8:** Examples of multi hop clustering in which cluster senders transmit data to the base station.

### 4. Conclusion

Clustering is most suitable for large scale wireless sensor networks and a useful topology management approach to reduce the communication overhead and exploit data aggregation in sensor networks. There exists a large number of clustering algorithms and some are reviewed in this paper. We have focused mostly on distributed and hierarchical clustering approaches, because they are more suitable for large-scale sensor networks. However energy consumption during cluster formation and maintenance is still high; also the compelling challenges for clustering algorithms which not consider here are how to schedule concurrent intra-cluster and inter cluster transmissions, how to compute the optimal cluster size, and how to determine the optimal frequency for cluster head rotation in order to maximize the network lifetime..

**References**


Zhao, W., Chellappa, R., Phillips, P. J. and Rosenfeld, A. (2003). Face Recognition: A Literature Survey. ACM Computing Survey, December Issue 399-458.

J. Yick , B. Mukherjee , D. Ghosal,(2008), "Wireless sensor network survey", Computer Networks.

K. Karenos, V. Kalogeraki, S. Krishnamurthy,(2008) "Cluster-based congestion control for sensornetworks," ACM Transactions on Sensor Networks , vol. 4, pp. 1-39, January.

D. J. Dechene, A. El. Jardali, M. Luccini, A. Sauer ,(2010)," A Survey of Clustering Algorithms for Wireless Sensor Networks".

N. Israr , I. Awan ,(2006),"Multihop Clustering Algorithm for load balancing in wireless sensor networks ",I. J. of Simulation, Vol. 8 No. 1,

A. A. Abbasi , M. Younis ,(2007)," A survey on clustering algorithms for wireless sensor networks" ,Elsevier 2007

D. J. Dechene , A. El. Jardali , M. Luccini, A. Sauer, (2010) ," A Survey of Clustering Algorithms for Wireless Sensor Networks".

C. Jiang, D.Y.Grid, Y. Zhao ,(2009),"Towards Clustering Algorithms in Wireless Sensor Networks-A Survey", IEEE Communications Society subject matter experts for publication in the WCNC.

I. F. Akyildiz et al,(2002) "Wireless sensor networks: asurvey", Computer Networks.

W. Heinzelman, A. Chandrakasan, H. Balakrishnan,(2000)," Energy- Efficient Communication Protocol for Wireless Micro sensor Networks", Proceedings of the 33rd Hawaii International Conference on System Sciences (HICSS '00), pp. vol.2:10.

W. R. Heinzelman, A. P. Chandrakasan, H. Balakrishnan,(2002)"An application-specific protocol architecture for wireless microsensor networks", IEEE Transactions on Wireless Communications , vol.1,no. 4, pp. 660–670.

Wendi B. Heinzelman, Anantha P. Chandrakasan, and Hari Balakrishnan,(2002)," An Application-Specific Protocol Architecture for Wireless Microsensor Networks ", IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, VOL. 1, NO. 4, OCTOBER.

O.Younis , S.Fahmy (2004) ,"HEED: A Hybrid, Energy-Efficient, Distributed Clustering Approach for Ad-hoc Sensor Networks", Mobile Computing, IEEE Transactions on, vol.3,no.4, pp. 366- 379, 2004Clustering Approach Distributed/centralized Distance CH to non-CH Overhead in CH selection Residual EnergyDelay BalancedclusteringLEACH Distributed Yes Low No Low OKLEACH-C centralized Yes High Yes High GoodHEED Distributed No igh Yes High GoodDWEHC Distributed No High Yes Med Very good TL-LEACH Distributed Yes Med No Med OKA-LEACH Distributed Yes Med No Low GoodPEACH Distributed Yes Low No Med OKEEHC Distributed Yes Low No Low OKDEBC Distributed Yes Med Yes Med Very good

P. Ding, J. Holliday, A. Celik,(2005), "Distributed energy efficient hierarchical clustering for wireless sensornetworks", in: Proceedings of the clustering for wireless sensor networks, IEEE International Conference on Distributed Computing in Sensor Systems(DCOSS'05), June.

V. Loscrì , G. Morabito , S. Marano,(2005)" A Two-Levels Hierarchy for Low-Energy Adaptive Clustering Hierarchy (TL-LEACH) ", IEEE.

E. Abdellah, S. Benalla, A. B. Hssane,(2010) "Advanced Low Energy Adaptive Clustering Hierarchy",(IJCSE) International Journal on Computer Science and Engineering, Vol. 02, No. 07.

S. Yi, J. Heo ,Y.Cho, J. Hong ,(2007)," PEACH: Power-efficient and adaptive clustering hierarchy protocol for wireless sensor networks ", Computer Communications ,Elsevier.

S. Bandyopadhyay, E. Coyle,(2003) "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.

F. Bajaber and I. Awan,(2010)," Distributed Energy Balanced Clustering for Wireless Sensor Networks",10th IEEE International Conference on Computer and Information Technology.