A Survey on various Clustering Algorithms in Wireless Sensor Network

Karampal Singh
Department of Computer Science
Guru Nanak Dev University Campus
Jalandhar
karampalsingh16@gmail.com

Varinder Attri
Department of Computer Science
Guru Nanak Dev University Campus
Jalandhar
varinder2002@yahoo.com

Abstract *— Wireless Sensor Network has gained the attention of scientists and researchers because of its potential for applications in various fields of science and technology and is one of the most rapidly advancing technologies. Wireless sensor network faces some severe resource and computational constraints due to which their implementation becomes a difficult task. Among the resource constraints, the major challenge for WSN is energy limitation. Energy efficient algorithms must be implemented so as to prolong the lifetime of network. Network lifetime can be maximized through clustering algorithms, where cluster is responsible for sending the data to the base station and not all the nodes are involved in transmission of data. In this paper, we summarized various clustering algorithms for WSNs and draw comparison between them based on clustering properties, cluster head capability and clustering process and briefly discuss types of clustering as well as discuss various performance metrics, challenges and issues faced during the formation of cluster head.*

Keywords: Cluster, load balancing, energy efficiency, network lifetime, LEACH, HEED.

1. INTRODUCTION

Clustering is the process of grouping up of data, where the grouping is determined by finding similarities between data based on their characteristics and therefore such groups are termed as clusters. In a computer system, a cluster is described as a group of servers that act like a single system and enable high availability. [1] Clustering is used for various purposes like parallel processing, load balancing and fault tolerance. A cluster is defined as a collection of objects which are "similar" between them and are "dissimilar" to the objects belonging to other clusters [13]. Thus clustering involves grouping of nodes into clusters and electing a clusterhead such that: (i) members of a cluster can communicate with their clusterhead directly, (ii) clusterhead can forward the gathered data to the central base station through other clusterheads.

Cluster is defined as grouping of similar objects or sensors in context to (i) distance or proximity, means how far or near the sensors nodes be placed to form a cluster, (ii) logical organizing, choose a logical way to organize the sensor nodes within a cluster. There are basically two topology control approaches: (i) load balancing, some strategies are proposed that balanced the energy consumption of sensor nodes and ensure maximum network lifetime by balancing the traffic load equally as possible, (ii) network scalability, the ability to handle growing amount of work and capability of system to increase the total throughput under the increased load when resources are added.

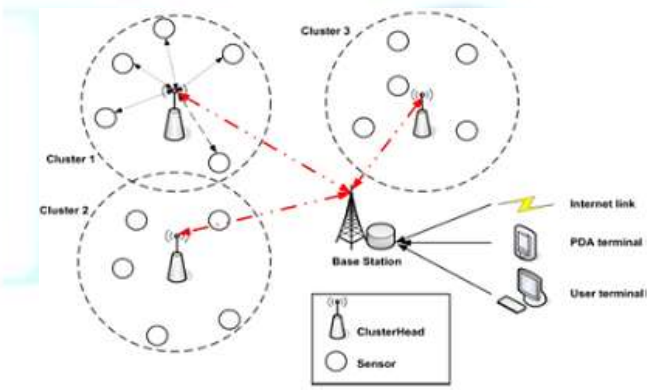

Figure 1. General Cluster Architecture

2. CLUSTERING TYPES

2.1 Dynamic Clustering

In dynamic clustering the clusters are formed and the cluster heads are selected. In dynamic clustering, the operation is divided into rounds and each round is separated into two phases known as the setup phase and the steady-state phase as shown in figure 2. Each round starts with the setup phase in which clusters are formed and cluster heads are selected while in the steady-state phase, the cluster heads collect data from sensor nodes and send it to the base station. [3] The dynamic clustering consists of two phases:

(i) **Setup Phase:** Each node sends its energy status and location to the base station; the nodes get their current location by using a global positioning system (GPS) receiver that is

activated at the beginning of each setup phase [3]. The base station uses this information to find number of cluster heads and also it makes sure that only nodes with enough energy are participating in the cluster heads selection. The base station broadcasts a message containing the cluster head ID for each node.

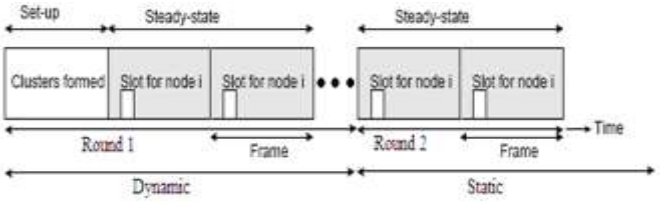

Figure 2. Setup phase and steady-state phase in our scheme

(ii) **Steady-State Phase:** The steady-state operation is divided into frames. The nodes start sensing the events and transmits the data to the cluster heads according to the TDMA scheme. Once the cluster heads receive all the data, cluster heads aggregate the received data and transmit the information to the base station. At the end of steady-state phase, the cluster heads receive data and energy status from sensor nodes. [3] The cluster heads aggregate the received data and transmit it to the base station also cluster heads use energy status to select new cluster heads for next round and inform the base station. The dynamic clustering completes and the static clustering operation starts.

2.2 Static clustering

In static case, the clusters are fixed for 10 rounds and cluster head position rotates among the nodes within the cluster as shown in figure 3. The operation of static clustering is divided into rounds and each round contains the steady-state phase only. During the steady-state phase, the cluster heads receive data and energy status from sensor nodes [3]. The cluster heads aggregate the received data and transmit it to the base station. Also cluster heads use energy status to select new cluster heads for next round.

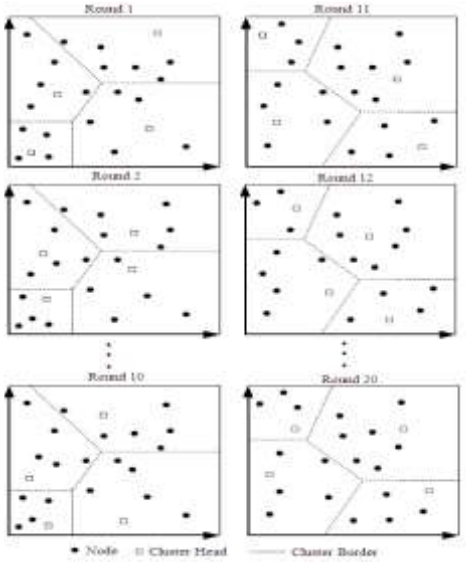

Figure 3: Clusters

2.3 Single-hop clustering

In computer networking, a hop is one portion of the path between destination and source. The data packets are passed through routers and gateways on the way. Each time the packets are passed to the next device, a hop occurs. In this type of clustering, the sensor node use single hopping or path to reach the cluster head (CH).

2.4 Multi-Hop clustering

In this type of clustering, the sensor nodes use multiple hopping or path to reach the cluster head. Figure 4 shows single-hop and multi-hop clustering.

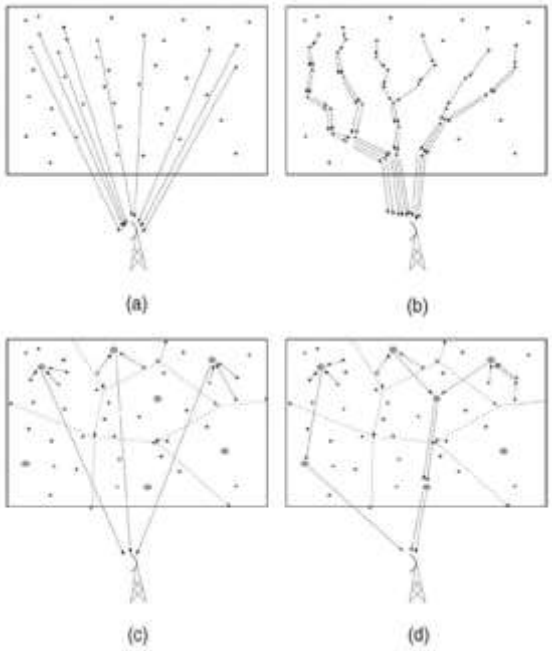

Figure: 4 Sensor information forwarding with and without clustering and aggregation. (a) Single hop without clustering. (b) Multihop without clustering. (c) Single hop with clustering. (d) Multihop with clustering.

3. CLUSTERING ALGORITHM TAXONOMY

The clustering algorithm can be classified as follows:

3.1 Heuristic algorithms

3.1.1 Linked Cluster Algorithm (LCA): The first clustering algorithms developed was LCA, initially developed for wired sensors, but later on implemented in wireless sensor networks [2] [11]. In LCA, a unique ID number is assigned to each node and has two methods to form a clusterhead [2]. The first method is the node with the highest ID number in the set including the node itself and all neighbor nodes, becomes a cluster head. The second method is to assume that none of its neighbors are clusterheads, then it would become a clusterhead.

3.1.2 Linked Cluster Algorithm 2 (LCA2): It was a proposed algorithm in order to eliminate the unnecessary election of clusterheads. In LCA2, we introduce the concept of covered node and non-covered node [11]. A node is covered if one of its neighbor is a clusterhead whereas in non-covered, the election of clusterhead begins by starting with the lowest ID node [2].

3.1.3 Highest-Connectivity Cluster Algorithm (HCCA): Highest-Connectivity Cluster algorithm is much similar to LCA. In this, the node neighbor number is broadcast to the surrounding nodes and the result is that, the connectivity of the node is considered instead of looking at the ID number [11]. The node with the highest connectivity is elected as clusterhead [2].

3.2 Weighted algorithm

3.2.1 Weighted Clustering Algorithm (WCA): This algorithm tries to find an architecture during the election of first cluster-head [11]. When a sensor loses its connection with any clusterhead, we find a new cluster head by invoking the election procedure. This algorithm is based on a combination of metrics that takes into account several parameters such as transmission power, mobility, ideal node degree and the remaining energy of the nodes. [7] Another important feature is that it is fully distributed which means that all the nodes in the mobile network shares the same responsibility acting as clusterheads.

3.3 Hierarchical algorithm

3.3.1 Low Energy Adaptive Clustering Hierarchy (LEACH): In this, nodes transmits the data to the cluster heads. The cluster head in LEACH [4], [2] is selected on the basis of a stochastic algorithm in every round i.e. a node can only become a cluster head once and cannot become a cluster head again. [14]

3.3.2 Two-Level Hierarchy LEACH (TL-LEACH): It is an extension to the LEACH algorithm. It make use of two levels of cluster-heads namely primary and secondary [14]. In primary cluster-head, each cluster communicates with the secondary cluster-head and the corresponding secondary cluster-head communicates with the nodes in their sub-cluster.

3.3.3 Energy Efficient Clustering Scheme (or EECS): This algorithm involves broadcasting of the residual energy from clusterhead candidate to neighboring candidates [8]. If a given node does not find a more residual energy node, it becomes a clusterhead. Cluster formation in EECS is different from that of LEACH. In EECS, clusters are formed on dynamic sizing based on cluster distance from the base station while in LEACH, clusters are formed based on the minimum distance of nodes to their corresponding clusterhead.

3.3.4 Hybrid Energy-Efficient Distributed Clustering (HEED): It is a multi-hop clustering algorithm for WSNs that focus on efficient clustering by proper selection of clusterheads based on the distance between nodes [2]. Therefore, the main objectives of HEED are [10]: (i) to distribute energy consumption so as to prolong network life-time; (ii) to minimize energy during the clusterhead selection phase; (iii) to minimize the control overhead of the network.

3.4 GRID algorithms

3.4.1 Power-Efficient Gathering in Sensor Information Systems (PEGASIS): It is an energy efficient protocol [12], which provides improvements over LEACH. In PEGASIS, nodes exchange data only by communicating with a nearby neighbor. The nodes transmit information to the base station turn by turn due to which the amount of energy spent per round reduces. PEGASIS has better performance as compared to LEACH, but as the nodes are grouped into chains so it causes redundant data transmissions.

3.4.2 GROUP: It is a grid-based clustering algorithm in which one of the sinks dynamically and randomly builds the cluster grid [2]. The clusterheads are arranged in a grid-like manner and the data queries are forwarded from the sink to source node and propagates to the clusterheads via Grid Seek.

4. PERFORMANCE METRICES

Some performance measures that are used to evaluate the performance of clustering protocols are listed below:

Clustering Approach	Clustering Properties			Cluster Head Capability			Clustering process		
	Cluster Count	Intra-cluster Topology	Inter-cluster Topology	Mobility	Node Type	Role	Methodology	CH selection	Algorithm Stages
LCA	Variable	Single-hop	Single-hop/multi-hop	Mobile	Sensor	Aggregation	Distributed	Adaptive	Cluster construction
LEACH	Variable	Single-hop	Single-hop	Stationary	Sensor	Relaying	Distributed	Adaptive	Cluster construction
TL-LEACH	Variable	Single-hop	Multiple-hop	Stationary	Sensor	Relaying and aggregation	Distributed	Adaptive	Cluster construction
EECS	Variable	Single-hop	Single-hop	Stationary	Sensor	Aggregation and relaying	Distributed	Adaptive	Cluster construction
HEED	Variable	Single-hop	Single-hop/multiple-hop	Stationary	Sensor	Relaying and aggregation	Distributed	Adaptive	Cluster construction
PEGASIS	Variable	Multiple-hop	Single-hop	Stationary	Sensor	Relaying and aggregation	Distributed	Adaptive	Data transmission
GROUP	Variable	Multiple-hop	Single-hop	Stationary	Sensor	Relaying and aggregation	Distributed	Adaptive	Data transmission

4.1 **Network lifetime**: It is the time until the first sensor node or group of sensor nodes in the **network** runs out of energy.

4.2 **Number of cluster heads per round**: Instantaneous measure reflects the number of nodes which would send directly to the base station, information aggregated from their cluster members.

4.3 **Number of alive nodes per round:** This instantaneous measure reflects the total number of nodes and that of each type that has not yet expended all of their energy.

4.4 **Throughput:** It is the total data rate sent over the network, the data rate sent to the base station from the cluster head as well as the data rate sent to their cluster heads from the nodes.

5. RESEARCH GAP

Some of the major issues faced while clustering the wireless sensor network are discussed in this section. These are:

5.1 Network deployment

The network deployment in wireless sensor networks is either fixed or random [9]. In random deployment, the resulting distribution can either be uniform or non-uniform whereas in fixed deployment the network is deployed on predetermined locations. So, it is necessary that the energy consumption is uniform across the network and entire area is being covered.

5.2 Single-hop or Multi-hop communication

The communication model that WSNs uses is either single hop or multihop. The single hop communication is expensive in terms of energy consumption because the energy consumption is directly proportional to the square of the distance whereas in multihop communication [9], the nodes that are closer to the clusterhead are under heavy traffic and when their energy terminates they create gaps near the cluster head. So, multihop communication model is more energy efficient than single hop in terms of energy consumption.

Table 1. Classification of Different Clustering Algorithms in WSNs

5.3 Network scalability

When a wireless sensor network is deployed, sometime there is a need to add new nodes to the network in order to cover more area and to prolong the network lifetime. In both cases the clustering scheme should be able to adapt such changes.

5.4 Uniform energy consumption

The transmission in WSNs is more energy consuming as compared to sensing, therefore the clusterheads that transmits the data to the base station consumes more energy as compared to the rest of the nodes [9]. The clustering schemes should ensure that energy diffusion across the network should be balanced and the cluster head should be rotated to balance the network energy consumption.

5.5 Attribute based Addressing

It is not possible to assign IDs to nodes in wireless sensor networks due to sheer number of nodes. The data is accessed from the nodes via attributes not by IDs [9]. So, this makes implementing a security mechanism difficult n intrusion into the system easier.

5.6 Cluster Dynamics

The term cluster dynamics means how to determine different parameters of the cluster [9]. In some cases, the number is dynamic and in some cases the number is pre-assigned. The cluster head performs the function of data transmission and compression. The major issue is the distance between the clusterhead. It can be set in accordance with some minimum value or can be dynamic. In dynamic, there is a possibility that unbalanced clusters are formed. The number of clusters might be dynamic or fixed. Fixed number of clusters cause less overhead that is the network will not have to repeatedly go through the setup phase in which formation of clusters is done.

6. CONCLUSION

In this paper we have examined various clustering algorithms, with respect to their reliability and power requirements. In WSNs, the energy limitations of nodes play a vital role in designing of any clustering algorithm for implementation. There are also some issues related while designing recovery mechanisms for clustering algorithms such as delay, data loss tolerance and network lifetime. Many energy improvements have been focused with reducing of energy associated in the selection process of cluster-head. In terms of energy efficiency, optimal clustering should minimize all overhead associated not only with the selection process of cluster-head, but also with node association to their respective cluster-heads. Some performance metrics like network lifetime, number of alive nodes, number of cluster-head and throughput are also discussed in this paper.

7. REFERENCES

[1] Abbasi, A.A. and Younis, M., "A survey on clustering algorithms for wireless sensor networks. Computer Communications. Vol 30: pp 2826–2841.

[2] A. El Jardali, D. J. Dechene M. Luccini, and A. Sauer, "A Survey of Clustering Algorithms for Wireless Sensor Networks", The University Of Western Ontario, London, Ontario, Canada.

[3] Fuad Bajaber and Irfan Awan, "Dynamic/Static Clustering Protocol for Wireless Sensor Network", Second UKSIM European Symposium on Computer Modeling and Simulation.

[4] G. Morabito, S. Marano and V. Loscri, "A Two-Level Hierarchy for Low-Energy Adaptive Clustering Hierarchy", 2005.

[5] Li Qing, Qingxin Zhu and Mingwen Wang, "Design of a Distributed Energy-Efficient Clustering Algorithm for Heterogeneous Wireless Sensor Networks", 2006.

[6] Liyang Yu, Neng Wang, Wei Zhang, and Chunlei Zheng, "GROUP: A Grid-Clustering Routing Protocol for Wireless Sensor Networks", 2006.

[7] Mainak Chatterjee, Sajal K. Das and Damla Turgut, "WCA: A Weighted Clustering Algorithm for Mobile Ad Hoc Networks", Clustering Computing, Vol. 5, pp. 193–204, 2002.

[8] Mao Ye, Guihai Chen, Chengfa Li and Jie Wu, "EECS: An Energy Efficient Clustering Scheme in Wireless Sensor Networks", In *Proceedings of the 24th IEEE International Performance, Computing, and Communications Conference (IPCCC)*, Phoenix, AZ, USA, 7–9 April 2005; pp. 535–540.

[9] Nauman Israr and Irfan Awan, "Multihop Clustering Algorithm For Load Balancing In Wireless Sensor Networks", 2007.

[10] Ossama Younis and Sonia Fahmy, "HEED: A Hybrid Energy-Efficient Distributed Clustering Approach for Ad Hoc Sensor Networks", IEEE Transactions on Mobile Computing, Vol. 3, No. 4, Oct-Dec 2004, 366–379.

[11] Rudranath Mitra and Diya Nandy, "A Survey on Clustering Techniques for Wireless Sensor Network", International Journal of Research in Computer Science, Volume 2 Issue 4, 2012.

[12] S. Lindsey and C. S. Raghavendra, "PEGASIS: Power-Efficient Gathering in Sensor Information Networks", 2004.

[13] Vivek Katiyar, Narottam Chand and Surender Soni, "Clustering Algorithms for Heterogeneous Wireless Sensor Network", International Journal of Applied Engineering Research, Volume 1, No 2, 2010.

[14] W. R. Heinzelman, H. Balakrishnan and A. Chandrakasan, "Energy-Efficient Communication Protocol for Wireless Sensor Networks", Proceedings of the 33th Hawaii International Conference on System Sciences, 2000.