The Survey of the Clustering Algorithms for Wireless Sensor Networks

Elnaz Siamie Tekmedash (Corresponding author) 1, Masoud Mohammadi [2]

[1.] Islamic Azad University, Zanjan science Research Branch
[2.] Islamic Azad University, Zanjan science Research Branch

Abstract: In recent years, interest in the use of wireless sensor networks has been increased in some applications such as disaster management, identifying the areas of the wars, border protection and security monitoring. It is expected that sensors can be increased long-distance in large numbers that they can act in non-regulatory environments autonomously. The nodes are often grouped in the discrete and non-overlapping clustering for support Scalability. We survey the different clustering algorithms for WSN and specify the goals, features, complexity and the others.

[Elnaz Siamie Tekmedash, Masoud Mohammadi. **The Survey of the Clustering Algorithms for Wireless Sensor Networks.** *Rep Opinion* 2014;6(2):65-70]. (ISSN: 1553-9873). http://www.sciencepub.net/report. 9

Key words: Survey, Clustering Algorithms, Wireless Sensor Networks.

1. Introduction

Recent advances in miniaturization and low-power designer led to development of small sensors that they get their power from battery and are able to recognition circumstances surrounding like temperature and sound. The sensors usually have the communication abilities and data processing. Sensor circuits get the parameters from the environment where is the around the sensor and transfer them by electrical signal. The process of such a this signal causes that the characteristics of the object are cleared. Every sensor is a portable radio that can use it to send the data collected to requirement part. The development of this technology is interesting and although the sensors have limited abilities in huge and unsupervised networks individually, it is possible to collect the data. We can use the network of the sensors for collecting the variables which are related to meteorological like temperature and pressure, then we can use these measurements for predicting the harsh natural phenomena. In disaster management, we can use sensor network to map the affected Areas by guidance emergency response units to survivors in such a Status like earthquake.

One of the Advantages of WSN is acting in unsupervised environments at adverse conditions. If these plans were available, there would be likelihood of regulatory risk and even it is sometimes impossible and deficient. So it is expected that the sensors are developed in related areas by non-controlled tools randomly. Since some sensors which work with battery in short period of life and the possibilities of the node damaging during the developing time, it is expected that many sensors are necessary in some application of ESN. Assume we have hundreds or thousands sensors. The design and the operation of these huge networks need the strategy management and Scalable Architecture. In addition, sensors have energy limitation in these environments and we cannot charge their batteries. So the design of the aware algorithms is an important factor for developing the period of life of the sensors. The most important goal of design in other applications is the recognition of the correctness of identifying the target and grouping. Every cluster can have a leader named Cluster head (CH). A Cluster head may be determined with the sensors of the clusters or network designers. A cluster head is one of the nodes or sensors that have more resources. The numbers of the cluster can be fixed or variable. Cluster heads may form a two layers network and just transfer the data to the related parts. This related part can be a major station or a command center. In addition to network scalability, clustering has many advantages. By means of clustering, we can set the paths in the cluster locally and so the size of the saved routing table is decreased in single nodes. Since the range of the interacting within cluster with cluster head are limited and prevent from the additional messaging in sensor nodes, Clustering can also protect the contact bandwidth. The clustering can fix the network topology at the level of sensors and so we do not need the overload of protecting the topology. The sensors should just communicate only with the cluster heads themselves which don't effect on the changes at level of the layers which are within cluster. The cluster heads can implement their optimal management strategies of the life period of the single batteries of the sensors and for more promoting the network actions and the increasing the life period of the network. The cluster heads can time the activities. So the nodes can switch to sleep mode and decrease the rate of the energy consumption. The sensors can act in the ambulatory intermittent way and their time of sending and receiving are determined. So reconnecting of the nodes are prevented and it is possible to limit the power redundancy at the covered

level and prevent from conflict.

1- The categories of clustering features

1-2- the categories clustering techniques

1-1-2 the network model

The different goals and architecture/ the limitation of the design are survived for different application

- The dynamism of network: WSNs usually consist of three major parts. The sensor nodes, the major station and events observed. Except of the motion sensor which has small setting, it is assumed in much architecture that the sensor nodes are fixed. Sometimes it is assumed that it is necessary to support the motion of the cluster head or the major station. The movement of the nodes can be very challenging. Because the group membership changes dynamically and the clusters have to update during the time. In the other hand, the events supervised by the sensor can be either intermittent or continuous or be according to the application. It is possible that the load of the cluster would be inappropriate for the nodes within the cluster and if the cluster head are chosen randomly among the sensors, it is possible that the role of the cluster head gets ambulatory. If the number of the events fluctuates significantly, the intermittent event may occur much in the consistent clustering strategies.
- The data processing within a network: far as Mach as the sensor nodes may produce significant redundant data, we can collect the similar packs of the multiple nodes. So the number of the transiting are decreased. Collecting the data combines the different resource data by using the functions like omit, the minimum, the maximum and the average. Some of these functions can be implemented in every sensor node and it decreases the network data. It is determined that we can consume less energy and save more energy by means of collecting data. These techniques are used for obtaining energy efficiency and the optimum traffic in the number of the routing protocol.
- The development of the nodes and the abilities

The development of the nodes is our analysis. This application is dependent and effected on network clustering. This development is done either definite or self-organized. In definite situations the sensors are placed manually and the data comes from the routes where are determined before. So clustering is unnecessary in predefined routes and in self organized systems, the sensor nodes intersperse randomly and design the substructure in the ad hoc way. In this substructure, the place of the major station or the cluster head is important in the subject of the efficiency when the distribution of the nodes is equally.

2-1-2- the clustering goals

Many of the clustering goals facilitate the applicable requirements. For example, if an application sensory to the delay is given, the external and internal connection and the length of the path in routing are used as choosing criterion and grouping the nodes. The following goals are defined for network clustering.

- The load blanching (Traffic): the distributing of the sensors among the clusters are usually determined as a goal and cluster head processes the data and does the within cluster management tasks. Beside the duties of the cluster head, it is obvious that the load balancing should be done that we can obtain our efficiency. The load balancing is a complex issue in WSN that cluster heads are chosen from accessible sensors. In these situations the clusters in the same size are determinant for development of the network life period and avoiding from wasting untimely energy in the high Rate and also the distribution of the sensors may cause delay in data.
- The error tolerance: in many applications, WSNs are in an abnormal operating and the nodes are usually in the prone to injuries or improper performance. Usually in such these applications, the tolerance of the errors of the cluster heads is important to avoid wasting the important sensor data. The best way to data recovery from a damaged cluster head is clustering the network again. Although not only the re clustering consumes the resource budget of the nodes, but also it is usually unbalanced for developing operations. So the contemporary techniques for fault tolerance can be more appropriate. The allocation of the supporting cluster heads is the most important plan for data recovery from a damaged cluster head.
- Increased connection and decreased delay: the satellite connection and the connection within cluster are an important need in most applications. The connection goal can be just limited to the availability guarantee of a path from to the major station. When it is assumed that some of the sensors have the roe of cluster head, the connection goal cause cluster the network. On the other hand, when the data delay is seen. The connection within clustering is a goal or the limitation of design
- The minimum number of clusters: when cluster heads are the nodes with many resources, this goal occurs. The network designer usually wants use from the minimum nodes. The size of these nodes can be more than the sensors significantly that cause the nodes would be recognizable easily. Observing the node in the more applications of WSN is undesirable and this is because of security issues and border protection.
- The most network life period: since the sensor nodes have energy limitation, the network life

period is very important, especially for WSN application in the harsh situations. When the cluster heads have more resources in comparison to the sensors, it is necessary that the connection within clustering gets its minimum level. If it is possible, we can put the cluster heads next to the most sensors in their clusters. On the other hand, when the cluster heads are usual sensors, their life period can be developed by limiting their load. The adaptive clustering is a critical choice for achieving the longer network life period.

3-1-2- The categories of clustering features

Clustering features:

- The number of the cluster: in some existence procedure the set of the clusters are pre-defined and the number of the clusters are preset, the random choosing the cluster heads from the developed sensors causes the number of the clusters would be variable.
- Stability: when the number of the clusters is variable and the memberships change during the time, it is Saied that the plan of the clustering is stable. Otherwise we assume that it is fixed. Although the sensors don't move, the numbers of the clusters are fixed during the network life period.
- Topology within a cluster: some plans of the clustering are based on the direct connection between a sensor and the designed cluster head. Sometimes it is necessary that a sensor joins to a cluster head, especially when the communication range of the sensor would be limited or the number of the clusters would be bounded.
- The connection within a cluster: when the cluster head doesn't have the connection ability, the connection between the cluster head and the major stations should be provided. In these situations, the clustering plan guarantees the conceivability of the path within a cluster. In some papers, it is assumed that the cluster heads can arrive at the major station directly.

1- The cluster abilities

- Motion: when a cluster head is mobile, the sensor membership changes dynamically and it is requirement that the clusters are protected continually. On the other hand, the fixed cluster head causes provide the fixed clusters and the network management will facilitate into the cluster and out of the cluster. Sometimes the cluster heads can revolve to reform the network efficiency.
- The node type: in some settings a subset of the developed sensors are designed as a cluster head. Otherwise in another plans, the cluster heads have the significant connection and computational resources.
- The role: a cluster head can act as a role for the traffic produced by the sensors into the cluster or can accomplish the Combination/collecting sensor data. Sometimes a cluster head act as a link or a major station that does the operation based on the goal or phenomenon detected.
- Clustering process: Full synchronization process clustering and the characteristics of the algorithm have significant differences among the clustering plans. The mentioned characteristics are:
- The method: when the cluster heads are usual nodes, the clustering is done in distributed manner and without any especial coordination.
- The clustering goals: there are several goals for providing the clusters. The A examples are the fault tolerant, the load balancing, the network connection and others.
- Choosing the cluster head: cluster heads can be pre-defined or choose them from the developed node set randomly.
- The algorithm complexity: many clustering algorithms are proposed according to the goals and the methods. The complexity and the convergence rate of these algorithms can be constant or depended on the number of cluster heads or sensors.

1- The WSN clustering algorithms

Usually, WSNs consist of many sensors which themselves include hundreds or thousands sensors. The clustering is an efficient tool for managing these numbers of nodes. Scalability is an important advantage for network clustering. The survived algorithms are grouped as the algorithms which have fixed or variable schedule according to the convergence rate.

1-3- The algorithms with the variable convergence time

Time is an important factor in the convergence of clustering algorithms. Some proposed clustering algorithms like LCA, RCC, and CLUBS have convergence time O (n) that n is the number of the present sensor nodes.

The hybrid clustering algorithms (LCA): [the presented paper by 1694-1701-Baker,Ephremides]LCA is presented to maximize network connections.in this algorithm, it is assumed that the nodes are simultaneously and have an access based on time. A node is allocated in a track frames that is adopted with ID. First every node broadcasts its ID and pays attention to the other node transfers. Second a node broadcasts a set of neighbors which finds their transfers. So every node can identify its neighbors which are in the distance, a jump or double jump. a X node become a cluster head when has highest ID and doesn't have the highest ID among its single jump neighbors. LCA produces many clusters [this procedure is in Ephremides, Wiese ltheir, Barker-1987] the new idea is that choosing randomly the X node as a first cluster head and allocation its

neighbors to the first cluster head.

The random competition based on RCC: Although RCC is proposed for ad hoc mobile network, but it can be used for WSNs. If a node claims that it is cluster head, RCC algorithm uses presented first winners low in which every node can proctor on the other radio covered nodes. After this claim which is broadcasted from the first node, neighboring nodes join its cluster as a member and accept that it is cluster head. Every cluster head broadcasts a claim cluster head pack periodically. Since there is a delay time between the broadcasting a claims pack and receiving it, it is possible that the simultaneous broadcasting would be incompatible with each other. If we don't know the Process of these claims, the number of the neighbor nodes may broadcast several claim cluster head packs simultaneously. RCC uses random timing for avoiding these problems and proctor the nodes ID.

CLUBS: [Nag pal, core-October 1998]this algorithm is used for making the cluster by the local broadcasting. Making the clusters is based on three characteristics:

- Every node should join to the cluster in the network.
- The diameters of the clusters should be equal in the network.
- The cluster should support the connections within a cluster. This means the nodes should be able connect with each other.

This algorithm builds the clusters with two jumps. Every node cooperates by choosing a random number in the range of the fixed integers during the construct process in the network. Then it begins to account the choosing number inversely slowly. If during the accounting, a break doesn't occur and this number is zero, the node introduces itself as a cluster head and broadcasts an employment message. When the neighbor nodes which are placed in the territory Qatar receive the massage, stop the accounting and accept the invitation and join to the cluster. The nodes which joined to the cluster are called follower and aren't allowed to compete for getting cluster head for a long time. Since CLUBS allow overlap. The follower nodes also pay attention to the other employment message and can follow more than one clustering. If a node identifies the conflict during the cluster head competition or receives the incomplete massage, a node seems a follower node and it is assumed that several cluster heads are trying to employment at a time. This algorithm doesn't finish until the whole of nodes join to a cluster head or cluster follower. Figure 3 shows the final clustering plan.

The most important problem in CLUB algorithm is that the cluster heads are in the range of single jump in the clusters. If the cluster wastes in this situation, the process of the choosing the cluster head re begins.

The energy efficiency hierarchical clustering (EEHC): it is suggested that EEHC is random and distributed clustering algorithm for WSNs to maximum the network life period. The cluster heads receive the data from every cluster and send it to the major station like a collected report. This technique is based on two levels. They are the primary level and the expanded one. In the primary level named first level clustering, every sensor node introduces itself as a cluster head to the present neighbor nodes in its connecting area with P probability. These cluster heads are called volunteer cluster heads. The all of the nodes placed in the K jumping rang receive this massage either by direct connection or sending from the other nodes. The node which receives the information and also isn't a cluster head, joins to the nearest cluster. The mandatory nodes are the ones which aren't a cluster head or don't belong to the any clusters. If a pack is defined by K jumping during the predefined time and a node doesn't receive any information, it is a mandatory node.

In the second level, this process expands to have a multilevel clustering, so n levels related to the clustering hierarchy, are provided. This algorithm guarantees the h jumping connection between the cluster head and the major station. It is assumed that level h is the highest level and it sends the collected sensor nodes to the lowest level clustering. The lowest level clustering sends the collected data to the second level cluster head and so on. At the highest level hierarchy clustering, the cluster heads send the collected data report to the major station.

EEHC has a time complexity such as O $(K_1+K_2+\ldots K_n)$ which is improvable. The energy consumption for network operation is depended on the parameters like K and P. The authors calculate Mathematical expression for K and P. This Extraction is based on the sensor data production and periodically transfer and using the stochastic Geometry for estimating the connecting energy. The simulated result confirms that we can decrease the energy consumption significantly by the optimum parameters.

2-3- The algorithms with the fixed convergence time

The clustering algorithms which cover the fixed number of repeat completely, except the size and the number of the nodes, are called the algorithms with the fixed convergence time.

Low energy compatible clustering hierarchical (LEACH): LEACH is one of the most famous clustering algorithm for WSNs. This algorithm makes the clusters according to the receiving signal power and uses the cluster head nodes as a router to the major station. All of the data process such as the data combination and data collection are done locally.

LEACH builds the clusters by means of the distributed algorithms.

The fast local clustering service: [Demirbas, Arora, Mittal-June 2004] it is a distributed technique which produces the same size clusters. The assumed radio model categories the nodes according to the cluster head distance in I band or O band. The band I nodes have less communication with the cluster head while in the band O nodes, the massages may be missed. A node stays in expectation mood till it receives an invitation from the cluster head. If this node receives no invitation, it gets the cluster head candidate and broadcast a candidate message. Once the massage is received by K node, it promulgates its membership to the candidate cluster head. Then the candidate cluster head finds the conflict and join to the CKas O band node. If the candidate cluster head receives no conflict massages, it itself get to the cluster head and begins to invite its members. An idle node joins to the cluster as an O band node when it receives no invitations from a nearer cluster head and if that node receives an invitation from the nearer cluster head later, these decisions are changed and so the node joins to the better cluster.

These algorithms have self-treatment ability, because the O band nodes can Become to the I band node in another cluster.

Hybrid energy efficiency distributed algorithms (HEED): [younis, fahmy-2004]: HEED is a distributed clustering plan in which the cluster heads nodes are chosen from the sensor nodes. HEED surveys the combination of connection cost and energy when the cluster heads are being chosen. HEED opposite of LEACH; don't choose the head cell nodes randomly. Just the nodes which have much energy can form to the head cell. HEEDs have three characteristics:

- There is low possibility that two nodes change to the cluster head and this means that the nodes are distributed well in the network.
- All of the nodes don't have the same energy consumption level.
- The probability of the cluster head choosing is set to guarantee the connection within the cluster.

In HEED, every cluster is mapped into the cluster and can communicate with its cluster directly. This algorithm consists of three phases:

1- First the algorithm sets the cluster head primary percent among the sensor nodes. The amount of this percent C_{prob} is used to limit the cluster heads to other nodes. Every sensor sets the cluster head changing probability CH_{prob} like below:

$CH_{prob}=C_{prob}*E_{residual}/E_{max}$

$E_{residual}$is present sensor energy and E_{max} is the maximum energy and it is like a full charged battery. CH_{prob} shouldn't be less than the particular threshold P_{min}.

2- Repeat phase: during this phase, every sensor does several repeats till it can connect to the cluster head with the minimum cost. If it doesn't find any cluster heads, this sensor introduces itself as a cluster head and sends a message to its neighbors and informs them about its changing situation. It stops this phase when CH_{prob} is one. So there are two statues for the cluster head:

- Temporarily statues: when the sensor gets to the temporarily cluster head, the amount of $CH_{prob}$is less than 1.
- Final statues: when the sensor forms to the cluster head constantly, the amount of $CH_{prob}$is 1.

3- Final phase: during this phase, every sensor decide for its final statues. This node or cluster either choose a cluster head with the minimum cost or chooses itself as a cluster head.

Wu, Huang developed the primary algorithm HEED, but they don't surrender in front of the nodes which don't receive any alarms from the cluster head. In the top final phase, these nodes themselves form to the cluster head. This small change has significant decreasing the number of the cluster head that causes the routing tree size decreasing.

The weight energy efficiency hierarchy clustering (DWEHC): the basic is producing the cluster with balanced size and optimum the topology within the cluster. This algorithm acts in a distributed manner and has the time complexity O (1). Every sensor calculates its weight after defining the place of the neighbor nodes. In the neighborhood, the nodes with the highest weight are chosen as a cluster head and the rest of the nodes are the memberships. It is necessary to pay attention that every node can switch to the different cluster head. This process continues until the nodes find the highest energy efficiency.

DWEHC and HEED are like each other in many cases. Every node can attend in the clustering process in the network. And there are no theories about the network size. Although there are some similarities between these two algorithms, they are different in efficiency.

Clustering based on features: Wang and his friends promoted the WSN clustering idea based on questions and features. The major motivation is access to data efficiency broadcasting in the network. This concept is like WSN data plan model. We can do clustering by data features hierarchy mapping. This procedure is based on the correct recognition of leadership choosing. Major station begins this process for building cluttering with asking the nodes. The nodes which hear this request, decide according to the energy that do they get candidate for clustering or not?

After getting the major station request, the sensor nodes tend to be a cluster head. The nodes which have

higher energy wait more. If a node gets candidate, it broadcasts an alarm that are sent from one node to another. Then the node joins to the cluster head that can get it with the minimum jumping.

During the prospect, if a node claim that it has a pack from a neighbor node, the node rejects the cluster head request.

This procedure causes promote the role of the cluster head among the others. The damage cluster head is identified easily. A cluster head sends a message to members periodically. If the neighbors don't update periodically, their clutter head isn't operational and one of the members gets the cluster head.

Conclusion

WSNs have been very impotent in several recent years. We can list the WSN military and non-military application. The examples are disaster management and border protection. It is expected that there are many sensors in these applications and we need the network management and architecture. Grouping the nodes in the clustering is the most famous procedure of scalability in WSN. There are significant attentions to the cluttering strategies and the algorithms which have many nodes. We have done the different developed categories for related characteristics. We categories different plans according to the goals, clustering process and good features of cluster and make clear the network model effects on the procedure and summarize the plans with their weakness and strength points.

Correspondence to:
Elnaz Siamie Tekmedash
Islamic Azad University, Zanjan science Research Branch

References

1. [D. J. Barker, A. Ephremides. The architectural organization of a radio network via distributed algorithm, IEEE transaction on communications, com-29(11)(1981) 1694-1701]
2. [D. J. Barker, A. Ephremides, J. A. Flynn, the design and simulation of a radio network with distributed control, IEEE journal on selected areas in communications (1984)-237]
3. [A. Ephremides, J. E. Wieseithier, D. J. Barker, A design concept for reliable radio networks with frequency hopping signaling. Processing of IEEE 75(1987) 56-73]
4. [K. Xu, M. Gerla, A heterogeneous routing protocol based on a new stable clustering scheme, in: Procceding of IEEE military communications conference (MIL COM 2002), Ano him.CA, October 2002]
5. [R. Nagpal, D. core. An algorithm for group formation in an amorphous computer, in: Proceeding of the 10th international conference on parallel and distributed systems (PODC 2002), Monterey CA, Julay 2002]
6. [S. Bandyopadhyay, E, coyle, An energy efficient hierarchical clustering algorithm for wireless sensor networks, in: proceeding of the 22nd annual joint conference of the IEEE computer and communication societies (IN FOCOM 2003), San Francisco, cail-fornia, April 2003]
7. [M. Demirbas, A. Arora, V. Mittal, Floc: a fast local clustering service for wireless sensor networks, in: proceeding of workshop on Dependability Issues in Wireless Ad Hoc Networks and sensor Networks (DIWANS'04), Palazzadei congress, Italy, June 2004]
8. [O. Younis, S. Fahmy, HEED: AHybrid, Energy-Efficient, distributed clustering approach for Ad Hoc sensor networks, IEE transaction on the computing 3(4)(2004) 366-379]
9. [P. Ding. J. Holiday, A. celik, Distributed energy efficient hierarchi Cal clustering for Wireless sensor networks, in: proceedings of the IEEE International conference on Distributed Computing in sensor systems (DCOSS'05), Marina Del Rey, CA, June 2005].

2/19/2014