

AN ELABORATED COMPARATIVE ANALYSIS OF DISTRIBUTED CLUSTERING ALGORITHMS FOR DENSE WIRELESS SENSOR NETWORKS

A. Yashwanth Reddy¹, N. Abhishek² ^{1,2}Assistant Professor, Department of CSE, Sree Dattha Group of Institutions, Hyderabad, Telangana, India

ABSTRACT

WSNs are web of sensor nodes with a set of processors and limited memory unit embedded in it. Reliable routing of packets from sensor nodes to its base station is the most important task for these networks. Routing in WSN is bit more complex than other wired or wireless networks. The conventional routing protocols cannot be used here due to its battery powered nodes. To support scalability, energy efficiency and efficient routing, nodes are often grouped in to non-overlapping clusters. This paper gives a crisp introduction on clustering process in WSNs. The survey of different distributed clustering algorithms used in WSNs, based on some metrics such as cluster count, cluster stability, cluster head mobility, cluster head role, clustering objective and cluster head selection is done. The study concludes with comparison of few distributed clustering algorithms in WSNs based on these metrics.

Keywords: Wireless sensor networks, Clustering methodology, distributed clustering algorithms.

1. INTRODUCTION

A WSN consists of a group of spatially distributed sensor nodes which are interconnected without wires. Each of the distributed sensor nodes typically consist of one or more sensing elements, a data processing unit, communication components and a power source which is usually a battery (Fig. 2). The sensed data is collected, processed and then routed to the desired end user through a designated sink point, referred as base station. WSNs are originally motivated for the use in military applications, such as border monitoring (Fig. 1). Now it is mainly focused on civilian applications such as environment monitoring, object tracking, biomedical applications, gathering meteorological variables like temperature and pressure, disaster management, etc. The major advantage of WSNs is their ability to operate in unattended environments, where human life is infeasible. Given the vast area to be covered, the short lifespan of the battery-operated sensors and the possibility of nodes to be destroyed while deployments, huge population of sensor nodes are expected in most of the wireless sensor network applications. Researchers reveal hundreds or even thousands of sensor nodes to be involved. These sensor nodes are energy constrained, thereby designing energy-aware algorithms becomes an important factor for extending the lifetime of sensors.

In order to support data aggregation through efficient network organization, nodes can be partitioned into a number of small groups called clusters. This phenomenon of grouping sensor nodes into clusters (Fig. 3) is called clustering. Every cluster would have a leader, commonly referred to as cluster-head (CH). A CH may be elected by the sensor nodes in the cluster or pre-assigned by the network designer [1]. A CH may also be just one of the sensors or a node that is generally richer in resources. The cluster membership may be fixed or variable. There are several advantages in clustering phenomenon. The basic advantage is that, it supports network scalability. It can localize the route setup within the cluster [2]. Clustering can also conserve communication bandwidth [3].

Moreover, clustering can stabilize the network topology at the level of sensors and thus cuts on topology maintenance overhead [4]. The CH can also implement optimized management strategies to prolong the battery life of the individual sensors and to maximize the network lifetime [3].





A CH can schedule activities in the cluster so that the nodes can switch to the low-power sleep mode most of the time and reduce the rate of energy consumption. Furthermore, a CH can aggregate the data collected by the sensors in its cluster [5].



Fig. 2. An articulation of the various components of a sensor node.



Fig. 3. Schematic diagram of clustering mechanism.

Very basically clustering can be classified in to centralized, distributed and hybrid clustering methodologies. Centralized clustering is the one in which, a centralized architecture is used in the clustering process i.e. a fixed CH and the remaining nodes in the cluster act as member nodes [6,7]. Distributed clustering is one in which, there is no fixed central CH and this keeps on changing from node to node based on some parameters, for instance residual energy [8]. Hybrid clustering is one which is formed as the resulting combination of both the above mentioned mechanisms [9]. If a centralized architecture is used in a WSN and the central node fails, the entire network will collapse and hence there is no guarantee for reliability in centralized clustering mechanism. Hence, the reliability of a WSN can be much improved by using distributed architecture. Distributed architecture is used in WSNs for some specific reasons like sensor nodes prone to failure, better collection of data and provide backup in case of failure of the central node. Also, nodes sensing and forwarding the redundant information can be minimized. Since there in no centralized body to allocate the resources, they have to be selforganized. Focusing on these projected advantages of distributed algorithms over centralized algorithms, some of the distributed clustering algorithms are discussed in this paper with their parameters.

Literature survey of available routing algorithms for WSNs

Apart from clustering algorithms, many routing algorithms were developed for wireless networks in general. All major routing protocols proposed for WSNs may be divided into several Location-based Protocols like categories: MECN, SMECN, GAF, GEAR, TBF, BVGF and GeRaF. Data-centric Protocols like SPIN, Directed Diffusion, Rumor Routing, COUGAR, Information-Directed ACQUIRE, EAD. Routing, Gradient Based Routing, Energy-aware Ouorum-Based Information Routing. Home Dissemination and Agent Based Mobility-based Information Dissemination. Protocols like SEAD, TTDD, Joint Mobility and Routing, Data MULES and Dynamic Proxy Tree-Base Data Dissemination. Multipath-based Protocols like Sensor-Disjoint Multipath, Braided Multipath and N-to-1 Multipath Discovery. Heterogeneity-based Protocols like IDSQ, CADR and CHR. QoS-based protocols like SAR, SPEED and Energy-aware routing.

The clustering mechanisms and the advantages of distributed clustering mechanisms over other mechanisms in WSNs is elaborately discussed in this paper. Their features, necessities and applications in the state of research manner is briefed in the first section. In the second section, the different classifications distributed clustering techniques of are summarized and a set of parameters for categorizing published algorithms has been enumerated and the third section concludes the paper.

2. Distributed clustering algorithms for wireless sensor networks

Distributed clustering is the mechanism in which, there is no fixed central CH and the CH keeps on changing from node to node based on some pre-assigned parameters. In this section, literature survey of various published distributed clustering algorithms for WSNs is presented, based on some advantages like efficient utilization of communication bandwidth within the clusters, avoiding redundant message transfer between the sensor nodes, localizing energy efficient route setup within the clusters, reduction in energy consumption [1], etc.

Low Energy Adaptive Clustering Hierarchy (LEACH): LEACH [10] is a clustering mechanism that distributes energy consumption all along its network, the network being divided into clusters and CHs which are purely distributed in manner and the randomly elected CHs, collect the information from the nodes which are coming under its cluster. LEACH protocol involves four main steps for each round: Advertisement phase, cluster set-up phase, schedule creation and data transmission. In the first step, the advertisement phase the eligible CH nodes will be issuing a notification to the nodes coming under them to become a cluster member in its cluster. The nodes will be accepting the offer based on received signal strength (RSS). In the cluster set-up phase the nodes will be responding to their selected CHs. In schedule creation step, as the CH receives response from the nodes it have to make a TDMA scheme and send back to its cluster members to intimate them when they have to pass the information to it. In the data transmission step, the data collected by the individual sensors will be given to the CH during their time intervals. The main constraint here is that, the radio of the cluster members will be turned off to reduce energy consumption after the data transmission during particular slot is finished.

Here in LEACH protocol, multi-cluster interference problem was solved by using unique CDMA codes for each cluster. The energy drain is prevented for the same sensor nodes which have been elected as the cluster leader using randomization, for each time CH would be changed. The CH is responsible for collecting data from the cluster members and fusing it. Finally each CH will be forwarding the fused data to the base station. When compared with the previous protocols (discussed in section 1), LEACH has shown a considerable improvement mainly in terms of energy-efficiency.

Hybrid Energy-Efficient Distributed Clustering (HEED): HEED [11] is a distributed algorithm which selects the CH based on both residual energy and communication cost. Basically HEED was proposed to avoid the random selection of CHs. Though LEACH protocol is much more energy efficient when compared with its predecessors (discussed in section 1), the main drawback in this approach is the random selection of CH. In the worst case the CH nodes may not be evenly distributed among the nodes and it will have its effect on the data gathering. The HEED protocol gets executed in three subsequent phases: Initialization phase, repetition phase and finalization phase.

Initialization phase, in which the initial CH nodes percentage will be given to the nodes. It is represented by the variable *Cprob*. Each sensor node computes its probability to become CH using equation 1.

CHprob = Cprob * Eresidual/Emax(1) Where,

Eresidual - residual energy of the concerned node.

Emax - maximum battery energy.

Since HEED supports heterogeneous sensor nodes, *Emax* may vary for different nodes according to its functionality and capacity. Repetition phase, in which until the CH node was found with least transmission cost, this phase was iterated. If the node cannot find the appropriate CH, then the concerned node itself was selected as CH. Finalization phase, in which the selection of CH was finalized. In general, the tentative CH now becomes the final CH node.

Energy Efficient Hierarchical Clustering (EEHC): EEHC [12] is a distributed, randomized clustering algorithm for WSNs, in which the CHs collect the information about the individual clusters and send the aggregated report to the base-station. Their technique is based on two stages: Initial and extended. The initial stage which is also called single-level clustering, in which each sensor node announces itself as a CH with a probability p to the neighboring nodes within its communication range. These CHs are named as 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 that announcements and is not itself a CH becomes the member of the closest cluster. Forced CHs are nodes that are neither CHs nor belong to a cluster. If the announcement does not reach to a node within a preset time interval tthat is calculated based on 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.

In the second stage, the process is extended to allow multi-level clustering and generally builds h levels of cluster hierarchy. Thereby the clustering process is recursively repeated at the level of CHs to form an additional tier [13]. The algorithm ensures h-hop connectivity between CHs and the base-station. Assumed that level h is highest, sensor nodes transmit the collected data to level-1, the lowest level CHs. The CH at the level-1 transmits the aggregated data to the level-2 CHs and so on. At the top level of the clustering hierarchy, CHs transmit the aggregated data report to the basestation.

The time complexity of EEHC is which shows $O(k_1 + k_2 + + k_h),$ a drastic improvement over many O(n) clustering algorithms such as LCA, thereby making it adaptive for network with large number of Energy consumption for network nodes. operations such as sensor data collection, aggregated information transmission to the basestation generally depends on parameters p and kof the algorithm. The values of *p* and *k* should be made in such a way that, minimum energy consumption is attained. Simulation results [12] showed that by using optimal parameter values, clustering could be done effectively such that the minimum energy consumption factor is met significantly.

Linked Cluster Algorithm (LCA): The Linked Cluster Algorithm (LCA) is a distributed clustering algorithm that avoids communication collisions among nodes and uses TDMA frames for inter-node communication, with each frame having a slot for each node in the network for communication. Proposing cluster formation and CH election algorithms, many papers focuses on single-hop clustering and guarantees that no node will be more than one hop away from leader [14,15].

In LCA, every nodes requires 2n time slots, where *n* is the number of nodes in the network, to have knowledge of all nodes in its neighborhood. If a node *x* has the highest identity among all nodes within one wireless hop of it or does not have the highest identity in its one hop neighborhood, but there exists at least one neighboring node *y* such that *x* is the highest identity node in *y*'s one hop neighborhood, it becomes a cluster-head. Basically, the LCA approach was designed to be used in the networks with less than 100 nodes. In such small networks, the delay between the node transmissions is minor and may be accepted.

CLUBS: This algorithm uses the advantage of local communication to efficiently aggregate the nodes into clusters, in which the time taken for convergence is proportional to the local density of nodes [16]. In order that the clusters to be useful for resource allocation and self-organization, the clustering phenomenon in

CLUBS is characterized by the following: First, every node in the network must belong to some cluster. Second, every cluster should be of same diameter. Third, a cluster should have local routing, which means that every node within the cluster should be able to communicate with each other using only nodes within that same cluster.

The **CLUBS** algorithm forms overlapping clusters, with a maximum cluster diameter of two hops. Every nodes starts competing to form a cluster by choosing random numbers from a fixed integer range [0, R]. Each node counts down from that number silently. If it reaches zero without being interrupted, the node becomes a CH and recruits its local neighborhood in to its cluster by broadcasting a "recruit message". The nodes that get recruited are generally called "followers". Once a node has been recruited as a follower, it stops counting down and listens for additional recruit messages. In CLUBS the main characteristic feature that should be remembered, it supports cluster overlap. If a node detects a collision while counting down, it assumes that more than one of its neighbors tried to recruit it at the same time and becomes a follower. At the end of Rsteps, all nodes in the network are either leaders or followers.

The CLUBS formed generally satisfies the cluster requirements. In addition, the algorithm satisfies several other constraints that occur in large distributed environments. For example, the algorithm does not need global IDs, limited or no topology knowledge of the network. The simplicity of local CH selection mechanism allows them to perform in asynchronous environments without loss of efficiency. The main advantage of CLUBS is that CH conflict is probabilistically lower.

Fast Local Clustering Service (FLOC): FLOC [17] is a distributed clustering technique that produces non-overlapping clusters and approximately equal-sized clusters. FLOC achieves locality: effects of cluster formation and faults/changes at any part of the network within almost two units distance. FLOC exhibits a double-band nature of wireless radio-model for communication [18]. A node can communicate reliably with the nodes that are in the inner-band (i-band) range and unreliably with the nodes in its outer-band (o-band) range. Hence the i-band nodes suffer very little interference communicating with the CH, thereby it is a reliable communication. Messages from o-band nodes are unreliable during communication and hence it has the maximum probability of getting lost during communication.

FLOC is fast and scalable, hence it achieves clustering in O(1) time regardless of the size of the network. It also exhibits self-healing capabilities since o-band nodes can switch to iband node in another cluster. It also achieves reclustering within constant time and in a local manner. It also achieve locality, in that each node is only affected by the nodes within two units. These features stimulate FLOC algorithm to be suitable for large scale WSNs.

Algorithm for Cluster Establishment (ACE): ACE [19] is a highly uniform cluster formation, self-organizing, efficient coverage, lesser overlapping and emergent cluster forming algorithm for WSNs, which is scale-independent and completes in time proportional to the deployment density of the nodes regardless of the overall number of nodes in the network. ACE requires no knowledge of geographic location and requires only small amount of communication overhead.

The main idea of ACE is to assess the potential of a cluster node as a CH before becoming a CH and steps down if it is not the best CH at the moment. The two logical steps in ACE algorithm is "*spawning*" of new clusters and "*migration*" of existing clusters. Spawning is the process by which a node becomes a CH. During spawning, when a node decides to become a CH it broadcasts an invitation message to its neighbors. The neighboring nodes accept such invitation and become a follower of new CH.

The main characteristic feature of ACE is that, a node can be a follower of more than one CH. During migration, the best candidate for being CH is selected. Each CH will periodically check all its neighbors to determine which node is the best candidate to become a CH for the cluster. The best candidate is the node which, if it were to become a CH, would have greatest number of follower nodes with lesser amount of overlap with the existing clusters. Once the best CH is determined by the current CH, it will promote the best candidate as the new CH and steps down from its CH position. Thus, the position of the cluster tends to migrate towards the new CH and some of the former follower nodes of the old CH are no longer part of the clusters while some new nodes near the new CH becomes new followers of the cluster. Each time that an action can be initiated for a node is called *node's iteration*.

In ACE, a node can have three possible states: it can be unclustered (not a follower of any cluster), clustered (a follower of one or more clusters) or it may be a CH. In the beginning of the protocol, all nodes are unclustered. In further iterations the node decides and becomes either a clustered node or a CH. The overall effect would appear as clusters are applying a repulsive force to spread out and reduce their overlap. An enhancement to the migration process in ACE algorithm was proposed in [21]. The idea is to further iterate inorder to increase the regularity of cluster layout. In addition to the repulsive effect, an attraction between clusters that are far apart is provisioned by factoring in the degree of overlap between neighboring clusters.

ACE exhibits perfect scalability. The protocol takes a fixed amount of time O(d) to complete regardless of the total number of nodes in the network, where d-is the estimated average degree (number of neighbors) of a node in the network. ACE is fast, robust against packet loss and node failure thereby efficient in terms of communication. It uses only local communication between the nodes and shows a good demonstration of flexibility of emergent algorithms in large-scale distributed systems.

ACE exhibits perfect scalability. The protocol takes a fixed amount of time O(d) to complete regardless of the total number of nodes in the network, where d-is the estimated average degree (number of neighbors) of a node in the network. ACE is fast, robust against packet loss and node failure thereby efficient in terms of communication.

DWEHC: Distributed Weight-Based Energy-Efficient Hierarchical Clustering [20] is a welldistributed clustering algorithm, which generates well balanced clusters and shows drastic improvements in performance over HEED. The clustering process terminates in O(1) iterations and does not depend on network topology on size. Each node first locates its neighbors, and then calculates its weight which is based on its residual energy and distance to its neighbors. The largest weight node in a neighborhood may become a CH. Neighboring nodes will join the clustered hierarchy as member nodes. At this stage the nodes are considered as first-level members since they have a direct link to the CH. node progressively adjusts Α to such membership inorder to reach a CH with minimum amount of energy.

Basically, a node checks with its non-CH neighbors to find out their minimal cost for reaching a CH. Given the knowledge of the node about the distance to its neighbors, it can assess whether it is better to stay as a first-level member or become a second-level one by reaching the CH over a two-hop path [20]. It is worth nothing that by doing so, the node may switch to a CH other than its original one. The process is iterated until the nodes settle in most energy-efficient topology.

DWEHC shows some features on forming a clustered network [20]: A node is either a CH or a member in the cluster but the level of the node depends on the cluster range and the minimum energy path to the CH, CHs are well distributed over the sensor field, each cluster has a minimum energy topology, and a CH has a limited number of member nodes. The algorithm constructs multilevel clusters and the nodes in each cluster reach the CH by relaying through other intermediate nodes. Fig. 4, which is redrawn from [20], depicts a node *s* reaches the node *d* by relaying through the intermediate node *r*.



Fig. 4. Node *s* relaying through node *r* to reach node *d*.

The leading advantage of DWEHC over HEED is that, it shows a great improvement in both intra-cluster energy consumption and intercluster energy consumption. Table 1 which is referred from [18], compares the presented distributed clustering algorithms for WSNs, enumerating their parameters listed in the abstract.

| Distributed Clustering Protocols | Cluster count | Cluster stability | Cluster- head mobility | Cluster- head role | Clustering objective | Cluster- head selection |
|--|------------------|----------------------|------------------------------|-----------------------------|-------------------------------------|-------------------------------|
| LEACH | Variable | Provisioned | Stationary | Relaying | Energy saving | Random |
| HEED | Variable | Assumed | Stationary | Aggregation and relaying | Energy saving | Random |
| ЕЕНС | Variable | Assumed | Stationary | Aggregation and relaying | Energy saving | Random |
| LCA | Variable | Provisioned | Mobile | Aggregation | Connectivity | Random |
| CLUBS | Variable | Assumed | Re- locatable | Aggregation and relaying | Scalability & management | Random |
| FLOC | Variable | Provisioned | Re- locatable | Aggregation and relaying | Scalability & fault tolerance | Random |
| ACE | Variable | Provisioned | Re- locatable | Aggregation and relaying | Scalability & load balancing | Random |
| DWEHC | Variable | Provisioned | Stationary | Aggregation and relaying | Energy saving | Random |

Table 1. Comparison of the presented distributed clustering algorithms for WSNs describing their parameters.

3. CONCLUSION

A growing list of civil and military applications can employ WSNs for increased effectiveness, especially in hostile and remote areas. Examples include disaster management, border protection, combat field surveillance, etc. In these applications, huge populations of wireless sensor nodes are needed with proper network management. Grouping of various nodes in to clusters has been vital to support some energetic functions like scalability, energy saving, etc. In this paper, the survey of some frequently used distributed clustering algorithms, stating their advantages and disadvantages has been done. A comparative analysis of various presented algorithms in the state of research, with various parameters has been sketched in a profound manner. Future works may concentrate in achieving better clustering mechanisms for mobile WSNs.

References

[1] B.Baranidharan, B.Shanthi, A survey on energy efficient protocols for wireless sensor networks, International Journal of Computer Applications, 11(10) (2010) 0975-8887.

[2] K.Akkaya, M.Younis, A survey on routing protocols for wireless sensor networks, Elsevier

Journal of Ad Hoc Networks 3(3) (2005) 325-349.

[3] M.Younis, M.Youssef, K.Arisha, Energyaware management in cluster-based sensor networks, Computer Networks 43(5) (2003) 649-668.

[4] Y.T.Hou, Y.Shi, H.D.Sherali, On energy provisioning and relay node placement for wireless sensor networks, IEEE Transactions on Wireless Communications 4(5) (2005) 2579-2590.

[5] K.Dasgupta, K.Kalpakis, P.Namjoshi, An efficient clustering-based heuristic for data gathering and aggregation in sensor networks, in: Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC 2003), New Orleans, LA, March 2003.

[6] S.D.Murugananthan, D.C.F.Ma, R.I.Bhasin, A.O.Fapojuwo, A centralized energy-efficient routing protocol for wireless sensor networks, IEEE Transactions on Communication Magazine 43(3) (2005) S8-13.

[7] F.Bajaber, I.Awan, Centralized dynamic clustering for wireless sensor networks, in: Proceedings of the International Conference on Advanced Information Networking and Applications, 2009.

[8] Noritaka Shigei, Hiromi Miyajima, Hiroki Morishita, Michiharu Maeda, Centralized and distributed clustering methods for energy efficient wireless sensor networks, in: Proceedings of the International Multiconference of Engineers and Computer Scientists (IMECS 2009), Volume I, March 2009.

[9] Boselin Prabhu S. R. and Balakumar N., "Evaluation of Quality in Network and Interoperable Connectivity between IP Networks", International Journal of Current Engineering and Scientific Research, Volume 3, Issue 9, pp. 81-85.

[10] Boselin Prabhu S.R. and Sophia S., "Bio-Medical Application of Wireless Power Transmission System", International Journal of Research and Engineering, Volume 3, Number 7, July 2016.

[11] Boselin Prabhu S.R. and Sophia S., 'Evaluation of clustering parameters in WSN fields using distributed zone-based approach', ASTM Journal of Testing and Evaluation, 43(6), 2015.

[12] Boselin Prabhu S. R. and Gajendran E., "An Investigation on Enlightening Performance in an Overburdened Highway System by Integrating Roadside Technologies", International Journal of Advances in Engineering Research (IJAER), Vol. No. 12, Issue No. V, November 2016, pp. 06-15.

[13] Boselin Prabhu S. R. and Balakumar N., "Methodology for Improving Security Issues and Reducing Vulnerability in Microprocessors", International Journal of Advances in Agricultural Science and Technology, Vol.3, Issue.6, November- 2016, pp. 60-65.

[14] D.J.Barker, A.Ephremides, The architectural organization of a mobile radio network via a distributed algorithm, IEEE Transactions on Communications, COM-29 (11) (1981) 1694-1701.

[15] D.J.Barker, A.Ephremides, J.A.Flynn, The design and simulation of a mobile radio network with distributed control, IEEE Journal on Selected Areas in Communications (1984) 226-237.

[16] R.Nagpal, D.Coore, An algorithm for group formation in an amorphous computer, in: Proceedings of IEEE Military Communications Conference (MILCOM 2002), Anaheim, CA, October 2002.

[17] M.Demirbas, A.Arora, V.Mittal, FLOC: A fast local clustering service for wireless sensor networks, in: Proceedings of Workshop on Dependability Issues in Wireless Ad Hoc Networks and Sensor Networks (DIWANS'04), Palazzo dei Congressi, Florence, Italy, June 2004.

[18] A.A.Abbasi, M.Younis, A survey on clustering algorithms for wireless sensor networks, Elsevier Journal of Computer Communications 30(2007) 2826-2841.

[19] H.Chan, A.Perrig, ACE: An emergent algorithm for highly uniform cluster formation, in: Proceedings of the 1st European Workshop on Sensor Networks (EWSN), Berlin, Germany, January 2004.

[20] P.Ding, J.Holliday, A.Celik, Distributed energy efficient hierarchical 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.

[21] H.Chan, M.Luk, A.Perrig, Using clustering information for sensor network localization, in: Proceedings of the International Conference on Distributed Computing in Sensor Systems (DCOSS'05), Marina Del Rey, CA, USA, June 2005.S