Some Issues in Clustering Algorithms for Wireless Sensor Networks

Prakashgoud Patil
B. V. Bhoomaraddi College of Engineering. & Technology, Vidyanagar, Hubli - 580031
Karnataka, India

Umakant Kulkarni
S. D. M. College of Engineering. & Technology, Dharwad - 580002
Karnataka, India

N. H. Ayachit
B. V. Bhoomaraddi College of Engineering. & Technology, Vidyanagar, Hubli - 580031
Karnataka, India

ABSTRACT
Wireless Sensor Networks (WSNs) present new generation of real time embedded systems with limited computation, energy and memory resources that are being used in wide variety of applications where traditional networking infrastructure is practically infeasible. In recent years many approaches and techniques have been proposed for optimization of energy usage in Wireless Sensor Networks. In order to gather information more efficiently, wireless sensor networks are partitioned into clusters. However, these methods are not without problems. The most of the proposed clustering algorithms do not consider the location of the base station. This situation causes hot spots problem in multi-hop wireless sensor networks. Unequal clustering mechanisms, which are designed by considering the base station location, to some extent solve this problem. In this paper, we present issues related to these approaches.

Keywords
Wireless Sensor Networks, Aggregation, Fuzzy Logic, Fuzzy Clustering, Probabilistic Clustering, Unequal Clustering.

1. INTRODUCTION
There have been recent advances in micro-electro-mechanical systems (MEMS) technology, wireless communications, and digital electronics. These advances have enabled the development of low-cost, low-power, multifunctional sensor nodes that are small in size and communicate with each other using radio frequencies [1]. A single sensor node has limited capability in sensing and is not sufficient for gathering useful information from a specific domain. This data gathering process can be accomplished by the collective work of a number of sensor nodes. In many applications the number of sensor nodes could be hundreds or thousands. These collaboratively working sensor nodes form a network which is called a wireless sensor network (WSN).

Wireless sensor networks have plenty of advantages. The deployment of WSNs is easier and faster than the wired sensor networks or any other wireless networks [10], because they do not need any fixed infrastructure [22]. Since sensor nodes are densely deployed in most of the cases, they are able to tolerate the network failures. Wireless sensor networks do not require a central organization and they are self-configuring [10].

There are several types of wireless sensors such as seismic, low sampling rate magnetic, thermal, visual, infrared, and acoustic and radar sensors [1]. These sensor nodes can monitor various environmental conditions. Some of these conditions are temperature, pressure, humidity, soil makeup, vehicular movement, noise levels, lighting conditions, the presence or absence of certain kinds of objects and mechanical stress levels on attached objects [3].

Wireless Sensor Networks provide unforeseen applications in this new field of design [1]. From military applications such as battlefield mapping and target surveillance, to creating context aware homes [2] where sensors can monitor safety and provide automated services tailored to the individual user; the number of applications are endless.

In wireless sensor networks, each sensor node receives signal from a limited region. This signal is processed in that sensor node and sensed information is generally transmitted to the observers (e.g. base stations) [21]. Sensor nodes consume energy while receiving information, processing information and transmitting information. In most of the cases, these sensor nodes are equipped with batteries which are not rechargeable. Therefore, energy efficiency is a major design goal in wireless sensor networks [21].

Nodes can be partitioned into a number of small groups, called clusters, for aggregating data through efficient network organization [21]. In general, each cluster has a cluster-head which coordinates the data gathering and aggregation process in a particular cluster. Each cluster member forwards its data packets to the cluster-head. Clustering in wireless sensor networks guarantees basic performance achievement with a large number of sensor nodes [17] [2]. In other words, clustering improves the scalability of wireless sensor networks [14]. This is because clustering minimizes the need for central organization and promotes local decisions.

Benefits of Clustering: The major benefits of clustering in wireless sensor networks are listed [22] below:

1) Clustering provides the spatial reuse of resources to increase system capacity. For example, if the clusters are not neighbors, they can use the same frequency for wireless communication.

2) Routing information of a cluster is shared with only other cluster-heads or cluster gateways. This restriction reduces the number of transmissions performed for distributing routing information. By using this advantage of clustering, more energy efficient routing protocols have been implemented.

When cluster structure is used in a WSN, the local changes need not be reflected to entire network. This reduces the information processed by sensor nodes and data stored in sensor nodes.

2. CLASSIFICATION OF CLUSTERING ALGORITHMS

There have been substantial amount of research on clustering protocols for WSNs. These clustering protocols are classified according to different criteria. The classification of clustering protocols according to their objectives is given [22] below:

Dominating-set-based clustering: This type of clustering protocols try to find a weakly connected dominating set which is responsible for searching route and maintaining routing table. Thus, table-driven routing and on-demand routing can be applied easily.

Low-maintenance clustering: These types of clustering protocols aim to provide a stable cluster structure to upper layer protocols. To achieve this goal, they try to limit reclustering situations or reducing the control messages for clustering.

Mobility-aware clustering: Mobility-aware clustering protocols take the mobility of sensor nodes into consideration. They try to group the mobile nodes that move with similar speed. The clusters that consist of mobile nodes moving with similar speed build a more stable cluster structure for wireless sensor networks.

Energy-efficient clustering: Energy-efficient clustering protocols try to use the battery energy of the sensor nodes more wisely, because sensor nodes have limited battery energy, and they are generally not rechargeable. Energy consumption of sensor nodes can be reduced by eliminating redundant energy consumption and balancing the energy usage of sensor nodes over the network. The main goal of this type of clustering protocols is prolonging the network lifetime.

Load-balancing clustering: This type of clustering protocols tries to limit the number of sensor nodes in each cluster. This approach produces clusters with similar sizes. If the clusters are similar in size, loads can be more evenly distributed within each cluster.

Combined-metrics-based clustering: As the name implies, this type of clustering protocols consider different metrics together. These metrics can be node degree, battery energy, cluster size, mobility speed, etc. These types of metrics are generally used in cluster-head election phase of clustering protocols.

3. OVERVIEW OF CLUSTERING ALGORITHMS

There are several proposed clustering algorithms for WSNs in recent years. In this section, we review Probabilistic Clustering Algorithms, Fuzzy Clustering Algorithms and Unequal Clustering Algorithms.

Probabilistic Clustering Algorithms:

In probabilistic clustering approaches, each node in the wireless sensor network decides its role by itself. This type of clustering algorithms aims to minimize the communication between sensor nodes. Probabilistic clustering algorithms guarantee rapid convergence and provide balanced cluster sizes [21]. Basically, each node assigns itself a probability which is a number between 0 and 1. If this probability is less than a predefined threshold, then that node becomes a cluster-head. Based on this principle, various probabilistic clustering algorithms are proposed. Here we overview LEACH (Low Energy Adoptive Clustering Hierarchy) [9], HEED (Hybrid Energy-Efficient Distributed Clustering) [20] and the algorithm proposed by Kuhn et al. [12].

The objective of LEACH protocol is to minimize energy dissipation in sensor networks. LEACH has distributed coordination and control mechanisms for cluster set-up and operation processes [9]. Static clustering algorithms select cluster-heads for WSNs only once, and these cluster-heads operate as cluster-head until they die. Since cluster-heads consume much more energy than ordinary sensor nodes, energy consumption over the network cannot be distributed evenly by using static clustering. Therefore, WSN can quickly move to a useless state, because the number of cluster-heads decreases drastically. In LEACH protocol, cluster heads are rotated in randomized manner, and cluster-head election is done periodically. The interval between two consecutive cluster formation processes is called as round. A single round consists of two phases which are set-up and steady-state phases [7]. The cluster-head election and cluster formation are done during set-up phase. In steady-state phase, the data, which is gathered from cluster member nodes, is aggregated at local cluster-head and transmitted to the base station.

In HEED protocol, residual energy of each sensor node is the primary parameter for probabilistic election of cluster-heads [21]. As stated in [20], there are four primary goals of HEED. These are listed below:

- Prolonging the lifetime of the wireless sensor network by evenly distributing energy consumption
- Selecting cluster-heads in a constant number of iterations
- Minimization of control overhead
- Formation of well-distributed cluster-heads and compact clusters

In case of a tie in cluster-head election, node degree or average distance to neighbors parameters are used to determine the cluster-head. HEED protocol is implemented in TinyOS, which is an operating system developed for Berkeley motes. Experimentations that are employed for evaluating HEED protocol show that clustering and data aggregation at least double the lifetime of the wireless sensor network [21].

Kuhn et al: studied initializing newly deployed adhoc and sensor networks, and proposed a probabilistic cluster-head election algorithm. In this approach, the probability of each node depends on the node degree [21]. Kuhn et al: showed that their proposed clustering approach computes an asymptotically optimal clustering in poly-alogarithmic time [12]. This algorithm tries to find a dominating set of nodes which will be assigned as cluster-heads. Sensor nodes compete to become dominators by exponentially incrementing their sending probability on a specified channel. Three different channels are used in this algorithm. Remaining two channels are used to keep the number of dominators small in a vicinity of an emerging dominator [12].

Pure probabilistic clustering algorithms (e.g. LEACH) have some disadvantages, which are listed [11] below:

Since pure probabilistic clustering algorithms only depend on probability, they can produce cluster-heads closer to each other.

They do not consider the residual energy of the sensor nodes. Therefore, the nodes that have lower energy levels may become cluster-heads.

These algorithms may randomly elect cluster-heads in vicinities that have low node density.

Pure probabilistic clustering approaches are useful for cluster-head election, but they are not sufficient. In order to make a more accurate cluster-head election, some additional parameters such as node degree, residual energy and local distance should be taken into consideration.

Fuzzy Clustering Algorithms: Fuzzy logic is useful for making real time decisions without needing complete information about the environment. On the other hand, conventional control mechanisms generally need accurate and complete information about the environment [5]. Fuzzy logic can also be utilized for making a decision based on different environmental parameters by blending them according to predefined rules.

Some of the clustering algorithms employ fuzzy logic to handle uncertainties in the wireless sensor networks. Basically, Fuzzy Clustering algorithms use fuzzy logic for blending different clustering parameters to elect cluster-heads. They assign chances to tentative cluster-heads according to the defuzzified output of fuzzy if-then rules. The tentative cluster-head becomes a cluster-head if it has the greatest chance in its vicinity. There are distributed and centralized fuzzy logic clustering approaches. Here we are going to overview the centralized approach of Gupta et al: [5] and the distributed approach of Kim et al: [11] which are abbreviated as CHEF.

In the fuzzy clustering approach proposed by Gupta et al. the cluster-heads are elected at the base station. In every round, each sensor node forwards its clustering information to the base station. There are three fuzzy descriptors which are considered by the base station during cluster-head election. These fuzzy descriptors are node concentration, residual energy in each node and node centrality [5]. The definitions of these fuzzy descriptors are given [5] below:

Node Concentration: Number of the nodes in the vicinity.

Residual Energy: Remaining battery energy of each sensor node.

Node Centrality: A parameter that indicates how central the node is to the cluster

There are 27 fuzzy if-then rules which are defined at the base station. The base station elects the cluster-heads according to these fuzzy rules. After the base station elects the cluster-head, it forwards the election results to entire network. This algorithm is a centralized clustering algorithm, because all clustering decisions are made at the base station. Gupta et al: claims that a centralized clustering approach will produce more accurate cluster-heads, because the base station has all clustering information about the network and base stations are more powerful than ordinary nodes [5]. However, this centralized approach has some disadvantages [11]:

The base station must collect all clustering information from the network. Repeating this process in every round brings a high overhead to sensor nodes. Thus, the battery levels of the sensor nodes may run low quickly.

In this approach the simulation is done for electing only one cluster-head per round. Therefore, this simulation is not a realistic one.

CHEF is a similar approach to that of Gupta et al: [5], but it performs cluster-head election in a distributed manner. Cluster-head election is done locally. Thus, the base station does not need to collect clustering information from all sensor nodes [11]. In every round, each node generates a random number between 0 and 1. If the random number is smaller than the predefined threshold, then that node becomes a tentative cluster-head. There are two fuzzy descriptors that are used in cluster-head election. These are residual energy of each node and local distance. Local distance is the total distance between the tentative cluster-head and the nodes within predefined constant radius r. There are 9 fuzzy if-then rules that are defined in all sensor nodes. Tentative cluster-heads calculate their chances to be a cluster-head using these fuzzy rules. If the chance of a tentative cluster-head is greater than the other tentative cluster heads' chances in radius r, then that tentative cluster-head becomes an actual cluster-head. Afterwards, it sends a cluster-head advertisement message to the nodes in the vicinity. The nodes that are not elected as cluster-head join to the closest cluster by sending a message to that cluster-head. CHEF guarantees that any two cluster-heads cannot exist within r distance [11].

Unequal Clustering Algorithms: The sensor nodes closer to the base station consume more energy, because the network traffic increases as we get close to the base station [21]. Therefore, the nodes closer to the base station quickly run out of battery. In order to balance energy consumption over the network, unequal clustering approach is introduced. This approach is based on the idea of decreasing the cluster sizes as we get close to the base station. If a cluster-head closer to the base station has less intra-cluster work, then it can contribute to inter-cluster data forwarding more. Unequal clustering is meaningful even in the cases where each cluster-head forwards its aggregated data to the base station directly. Here, we overview two unequal clustering approaches. These are the approaches that are proposed by Shu et al: [18] and Li et al: [11] which are abbreviated as EEUC.

If a cluster-head is closer to the base station, it has to relay more data forwarding traffic than the sensor nodes which are far from the base station [18]. Each sensor node in the network tries to send its data to the base station. Therefore, as we get close to the base station, the data forwarding traffic increases. Shu et al: proposed an approach that aims to achieve optimal power allocation over the sensor network. This approach assigns larger cluster sizes to cluster-heads that take fewer roles in data forwarding process. This approach is illustrated in Figure 1.

The proposed network model in this approach assumes a circular sensing region. However, generally sensor nodes are deployed randomly by throwing them to the target region. Therefore, this approach is not a practical one for real environments in most of the cases. This model should be improved to handle non-circular regions.

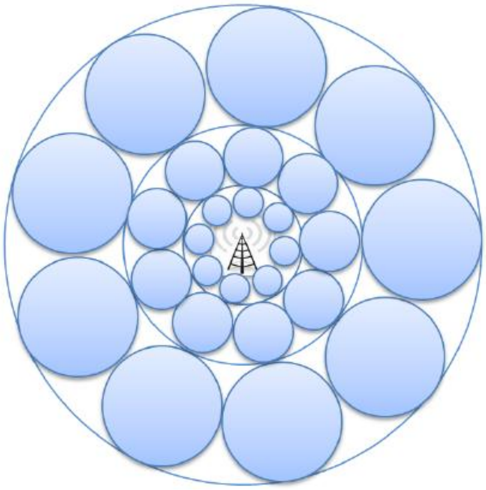

Figure 1: Cluster size distribution of Shu et al: approach

EEUC (Energy-Efficient Unequal Clustering) is a distributed competitive unequal clustering algorithm where cluster-heads are elected by local competition [11]. Every node has a pre assigned competitive range. This range gets smaller as we get close to the base station. This makes EEUC an unequal clustering algorithm. EEUC algorithm is also a probabilistic clustering algorithm, because in each cluster formation round, each node generates a random number between 0 and 1 to decide whether it is going to participate to the cluster-head election competition or not. If a sensor node has decided to participate to the competition, then it becomes a tentative cluster-head. Tentative cluster-heads in local regions compete in order to become an actual cluster-head. This competition is based on the residual energy of each tentative cluster-head. After cluster-head election is completed, the remaining sensor nodes join to the closest cluster.

4. OBSERVATIONS

Most of the clustering algorithms utilize two techniques which are selecting cluster-heads with more residual energy and rotating cluster-heads periodically to balance energy consumption of the sensor nodes over the network [13]. These clustering algorithms do not take the location of the base station into consideration. This lack of consideration causes the hot spots problem in multi-hop wireless sensor networks. The cluster-heads near the base station die earlier, because they will be in heavier relay traffic than the cluster heads which are relatively far from the base station.

In order to avoid this problem, some unequal clustering algorithms are proposed in literature [18] [13]. In unequal clustering, the network is partitioned into clusters with different sizes. The clusters close to the base station are smaller than the clusters that are far from the base station. EEUC mechanism for periodical data gathering partitions the sensor nodes into clusters of unequal size, and clusters closer to the base station have smaller size.

The different clustering algorithms, types and classifications and methods are studied. The Table 1 below summarizes the different types of clustering algorithms along with advantages and drawbacks.

5. CONCLUSION AND FUTURE WORK

In this paper we have examined the current state of proposed clustering protocols. Several clustering protocols have been proposed for wireless sensor networks. Reviewing the protocols presented in this paper is possible to observe that all of them are concerned on how to prolong the WSN life time and how to make a more efficient use of the critical resources located at the sensor nodes. The most of the proposed clustering algorithms do not consider the location of the base station. This situation causes hot spots problem in multi-hop wireless sensor networks. Unequal clustering mechanisms, which are designed by considering the base station location, solve this problem. The Unequal Clustering Algorithm proposed by Shu et al: approach assumes a circular sensing region.

However, generally sensor nodes are deployed randomly by throwing them to the target region. Therefore, this approach is not a practical one for real environments in most of the cases. Some of the clustering algorithms employ uncertainties in the Wireless Sensor Network.

Basically, a fuzzy clustering algorithms use a fuzzy logic for blending a different clustering parameters to elect cluster heads according to the defuzzified output of fuzzy if-then rules. Clustering algorithm presented in this paper offer a promising improvement over a conventional clustering; however, there is a still much work to be done. Optimal clustering in terms of energy efficiency should eliminate all overhead associated not only with the cluster head selection process, but also with node association to their respective cluster heads.

6. REFERENCES

[1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci. Wireless sensor networks: a survey. Comput. Netw., 38(4):393–422, 2002.

[2] E.M. Belding-Royer. Hierarchical routing in ad hoc mobile networks. Wireless Communications and Mobile Computing, 2(5):515–532, 2002.

[3] D. Estrin, R. Govindan, J. Heidemann, and S. Kumar. Next century challenges: Scalable coordination in sensor networks. In Proceedings of the 5th annual ACM/IEEE international conference on Mobile computing and networking, page 270. ACM, 1999.

[4] I. Gupta. Cluster head election using fuzzy logic for wireless sensor networks. Master's thesis, Dalhousie University, Halifax, Nova Scotia, March 2005.

[5] I. Gupta, D. Riordan, and S. Sampalli. Cluster-head election using fuzzy logic for wireless sensor networks. In Communication Networks and Services Research Conference, 2005. Proceedings of the 3rd annual, pages 255–260, 2005.

[6] Y. Han, S. Park, J. Eom, and T. Chung. Energy-E_cient Distance Based Clustering Routing Scheme for Wireless Sensor Networks. LECTURE NOTES IN COMPUTER SCIENCE, 4706:195, 2007.

[7] MJ Handy, M. Haase, and D. Timmermann. Low energy adaptive clustering hierarchy with deterministic cluster-head selection. In IEEE MWCN. Citeseer, 2002.

[8] W. Heinzelman, A. Chandrakasan, and H. Balakrishnan. An application-specific protocol architecture for wireless microsensor networks. IEEE Transactions on Wireless Communications, 1(4):660–670, 2002.

[9] W.R. Heinzelman, A. Chandrakasan, and H. Balakrishnan. Energy-e_cient communication protocol for wireless microsensor networks. In Proceedings of the 33rd Hawaii International Conference on System Sciences, volume 8, page 8020. Citeseer, 2000.

[10] J. Ibriq and I. Mahgoub. Cluster-based routing in wireless sensor networks: issues and challenges. In Proceedings of the 2004 Symposium on Performance Evaluation of Computer Telecommunication Systems (SPECTS), 2004.

[11] J.M. Kim, S.H. Park, Y.J. Han, and T.M. Chung. CHEF: cluster head election mechanism using fuzzy logic in wireless sensor networks. In Proceedings of the ICACT, pages 654–659, 2008.

[12] F. Kuhn, T. Moscibroda, and R. Wattenhofer. Initializing newly deployed ad hoc and sensor networks. In Proceedings of the 10th annual international conference on Mobile computing and networking, pages 260–274. ACM New York, NY, USA, 2004.

[13] C. Li, M. Ye, G. Chen, and J. Wu. An energy-e_cient unequal clustering mechanism for wireless sensor networks. In Proceedings of the 2nd IEEE International Conference on Mobile Ad-Hoc and Sensor Systems, page 8. Citeseer.

[14] M. Lotfinezhad and B. Liang. E_ect of partially correlated data on clustering in wireless sensor networks. In Proc. of the IEEE International Conference on Sensor and Ad hoc Communications and Networks (SECON). Citeseer, 2004.

[15] V. Mhatre and C. Rosenberg. Design guidelines for wireless sensor networks: communication, clustering and aggregation. Ad Hoc Networks, 2(1):45–63, 2004.

[16] M. Negnevitsky. Artificial intelligence: A guide to intelligent systems. Addison-Wesley, Reading, MA, 2001.

[17] C.E. Perkins and E.M. Royer. The ad hoc on-demand distance-vector protocol. 2001.

[18] T. Shu, M. Krunz, and S. Vrudhula. Power balanced coverage-time optimization for clustered wireless sensor networks. In Proceedings of the 6th ACM international symposium on Mobile ad hoc networking and computing, page 120. ACM, 2005.

[19] T. Voigt, A. Dunkels, J. Alonso, H. Ritter, and J. Schiller. Solar-aware clustering in wireless sensor networks. In Proceedings of the Ninth IEEE Symposium on Computers and Communications, 2004.

[20] O. Younis and S. Fahmy. HEED: a hybrid, energy-e_cient, distributed clustering approach for ad hoc sensor networks. IEEE Transactions on Mobile Computing, pages 366–379, 2004.

[21] O. Younis, M. Krunz, and S. Ramasubramanian. Node clustering in wireless sensor networks: Recent developments and deployment challenges. IEEE Network, 20(3):20–25, 2006.

[22] JY Yu and PHJ Chong. A survey of clustering schemes for mobile ad hoc networks. IEEE Communications Surveys & Tutorials, 7(1):32–48, 2005.

[23] H.J. Zimmermann. Fuzzy set theory–and its applications. Kluwer Academic Pub, 2001.

Table 1 Different types of clustering algorithms along with advantages and drawbacks.

Category	Protocol	Methods / Advantages	Drawbacks
Probabilistic Clustering Algorithm	**LEACH**	• Cluster Head Election is done periodically • Cluster head rotated randomized manner	• Does not guarantee good cluster head distribution. • Assumes uniform energy consumption for CH. • Cost of the overhead to form the clusters is expensive. • Since pure probabilistic clustering algorithms only depend on probability, they can produce cluster-heads closer to each other. • These algorithms may randomly elect cluster-heads in vicinities that have low node density.
	HEED	• Residual energy of sensor node is the primary parameter for probabilistic election of cluster-heads • Selecting cluster-heads in a constant number of iterations. • Minimization of control overhead • Formation of well-distributed cluster-heads and compact clusters • Prolonging the lifetime of the wireless sensor network by evenly distributing energy consumption	• Localized communication overhead and energy consumption during the data transmission for far away cluster heads is significant, especially in large scale networks.

Table 1 Different types of clustering algorithms along with advantages and drawbacks.

Category	Protocol	Methods / Advantages	Drawbacks
Fuzzy Clustering Algorithms	**CHEF**	• Centralized clustering algorithm -all clustering decisions are made at the base station. • These fuzzy descriptors are node concentration, residual energy in each node and node centrality. • There are 27 fuzzy if-then rules which are defined at the base station. The base station elects the cluster-heads according to these fuzzy rules.	• The base station must collect all clustering information from the network. • Repeating this process in every round brings a high overhead to sensor nodes. Thus, the battery levels of the sensor nodes may run low quickly.
Unequal Clustering Algorithms	**EEUC**	• Probabilistic clustering algorithm, • Based on the idea of decreasing the cluster sizes as we get close to the base station • cluster-heads are elected by local competition	• Assumes circular sensing region for clustering formation. However, generally sensor nodes are deployed randomly by throwing them to the target region. Therefore the approach is not a practical one for real environments in most of the cases