International Journal of Advanced Science and Research
ISSN: 2455-4227, Impact Factor: RJIF 5.12
www.newresearchjournal.com/science
Volume 1; Issue 5; May 2016; Page No. 41-45

Analysis and exploration of probabilistic and non-probabilistic clustering approaches with topology control mechanisms

Er. Kapil Mehta

Assistant Professor & Head, Department of Computer Science & Engineering Gian Jyoti Group of Institutions, Shambhukalan, Rajpura, Punjab, India.

Abstract

The topic of Wireless Sensor Networks (WSNs) has gained considerable attention in the research community due to the variety of applications and interesting challenges in developing and deploying such networks. The typical WSN is significantly energy constrained and often deployed in harsh or even hostile environments, resulting in sensor nodes that are prone to failure [1]. Failing nodes alter the topology of the network resulting in segmented routing paths and lost messages, ultimately reducing network efficiency. These issues spur the desire to develop energy efficient, Fault-Tolerant (FT) algorithms that enable the network to persist in spite of the failed nodes. Clustering protocols have extensively been used so that they can greatly contribute to overall system scalability, lifetime, and energy efficiency. In this paper, the state of the art in corresponding clustering approaches for large-scales environments is presented. The basic advantages, objectives, and design challenges are also briefly explored.

Keywords: WSN, CH, BS, Probabilistic, Non-Probabilistic

1. Introduction

The use of wireless sensor networks (WSNs) has grown enormously in the last decade, pointing out the crucial need for scalable and energy-efficient routing and data gathering and aggregation protocols in corresponding large-scale environments. The WSN is built of "nodes" – from a few to several hundreds or even thousands, where each node is connected to one or more sensors. Each such sensor network node has typically several parts: a radio transceiver with an internal antenna or connection to an external antenna, a microcontroller, an electronic circuit for interfacing with the sensors and an energy source, usually a battery or an embedded form of energy harvesting. A sensor node might vary in size from that of a shoebox down to the size of a grain of dust. In many significant WSN applications the sensor nodes are often deployed randomly in the area of interest by relatively uncontrolled means and they form a network in an ad hoc manner [2, 3]. Moreover, considering the entire area that has to be covered, the short duration of the battery energy of the sensors and the possibility of having damaged nodes during deployment, large populations of sensors are expected; it's a natural possibility that hundreds or even thousands of sensor nodes will be involved. In addition, sensors in such environments are energy constrained and their batteries usually cannot be recharged. Therefore, it's obvious that specialized energy-aware routing and data gathering protocols offering high scalability should be applied in order that network lifetime is preserved acceptably high in such environments. Naturally, grouping sensor nodes into clusters has been widely adopted by the research community to satisfy the above scalability objective and generally achieve high energy efficiency and prolong network lifetime in large-scale WSN environments. CH nodes aggregate the data and transmit them to the base station (BS) either directly or through the intermediate communication with other CH nodes.

Factors affecting Network Management and Connectivity

The wireless sensor network is smaller in size and having limited memory, computation and battery power [4]. Related to design of wireless sensor network different factors are discussed as:

a) Energy Efficiency: To make a Wireless Sensor Networks efficient, there is a need of perfect coordination between hardware and software. Most of sensor nodes use battery power as their energy source. The sensor network can be deployed in hazards conditions so it becomes difficult to change their batteries or provider the energy so there is requirement of developing the networks which efficiently use the battery as energy [3].The energy consumption depends upon major operations of the sensor nodes like [2]. Sensing, Communication and Data processing.

b) Security: Risks related to security are unavoidable as nodes are of mobile nature in a network. Normally, data transmitted over a network can be suffered from denial of service, node capture jamming attacks, dropping packets etc. So, Security mechanisms must ensure that when sink collects from sensor node no malicious node is allowed to interfere.

c) Application Dependency: When designing clustering and routing protocols for WSNs, application robustness must be of high priority and the designed protocols should be able to adapt to a variety of application requirements.

d) Synchronization: Slotted transmission schemes such as TDMA allow nodes to regularly schedule sleep intervals to minimize energy used. Such schemes require corresponding synchronization mechanisms and the effectiveness of this mechanisms must be considered.

e) Data Compression & Aggregation: Because this process makes energy optimization possible it remains a fundamental design challenge in many sensor network schemes nowadays.

However its effective implementation in many applications is not a straightforward procedure and has to be further optimized according to specific application requirements. Data Compression and aggregation techniques aid in reducing the amount of data transferred. The use of a robust strategy to manage distributed data flow, query and analysis is important to sensor networks.

2. Probabilistic & Non Probabilistic Approaches

There have been several different ways (based directly on the above-mentioned parameters or not) to initially distinguish and further classify the algorithms used for WSNs clustering, [5].

2.1 Probabilistic

In the category of probabilistic selection clustering algorithms, a priori probability assigned to each sensor node is used to determine the initial CHs. The probabilities initially assigned to each node often serve as the primary (random) criterion in order for the nodes to decide individually on their election as CHs; however other secondary criteria may also be considered either during CH election process or during the cluster formation process in order achieve better energy consumption and network lifetime. Beyond the high energy efficiency, the clustering algorithms of this category usually achieve faster execution/convergence times and reduced volume of exchanged messages.

2.2 Non Probabilistic

In the category of non-probabilistic clustering algorithms [25, 43] more specific (deterministic) criteria for CH election and cluster formation are primarily considered, which are mainly based [25, 36]. on the nodes‟ proximity (connectivity, degree, etc.) and on the information received from other closely located nodes. The cluster formation procedure here is mainly based on the communication of nodes with their neighbours (one or multi-hop neighbours) and generally requires more intensive exchange of messages and probably graph traversing in some extent, thus leading sometimes to worse time complexity than probabilistic/random clustering algorithms. On the contrary these algorithms are usually more reliable toward the direction of extracting robust and well-balanced clusters. In addition to node proximity, some algorithms [37, 40]. also use a combination of metrics such as the remaining energy, transmission power, mobility, etc. (forming corresponding combined weights) to achieve more generalized goals than single-criterion protocols. In the same category we also address a relatively new and quite challenging class of clustering algorithms for WSNs, namely, the biologically inspired protocols [41, 43]. (Based on swarm intelligence) which are probably the most promising alternative approaches for clustering in WSNs nowadays.

3. Popular Probabilistic Clustering Protocols

3.1 Low Energy Adaptive Clustering Hierarchy (LEACH)

One of the first and most popular clustering protocols proposed for WSNs was LEACH (Low Energy Adaptive Clustering Hierarchy) [6, 7]. It is probably the first dynamic clustering protocol which addressed specifically the WSNs needs, using homogeneous stationary sensor nodes randomly deployed, and it still serves as the basis for other improved clustering protocols for WSNs. Its main objectives (a) to improve the lifetime of WSNs by trying to evenly distribute the energy consumption among all the nodes of the network and (b) to reduce the energy consumption in the network nodes . It forms clusters based on the received signal strength and also uses the CH nodes as routers to the BS. All the data processing such as data fusion and aggregation are local to the cluster. LEACH forms clusters by using a distributed algorithm, where nodes make autonomous decisions without any centralized control. All nodes have a chance to become CHs to balance the energy spent per round by each sensor node. Initially a node decides to be a CH with a probability "p" and broadcasts its decision [8]. In Figure 12.2 the cluster formation scheme is given in a more clear view.

1.	Suppose here is a node 'i'.
2.	If 'i' is a Cluster-Head then advertise it to the other nodes present in the neighbourhood *then* Wait for Combine-Request messages and create a TDMA Schedule to send the request message to the neighbour nodes.
3.	*Else* Wait for Cluster-Head announcements and send the request message to the selected Cluster-Head.
4.	Operation Executed

A common solution in order balance the energy consumption among all the network nodes is to periodically re-elect new CHs in each cluster. . The BS is the data processing point for the data received from the sensor nodes, and where the data is accessed by the end user. It is generally considered fixed and at Generally, LEACH can provide a quite uniform load distribution in one-hop sensor networks. Moreover, it provides a good balancing of energy consumption by random rotation of CHs. Furthermore, the localized coordination scheme used in LEACH provides better scalability for cluster formation, whereas the better load balancing enhances the network lifetime. However, despite the generally good performance, LEACH has also some clear drawbacks. Because the decision on CH election and rotation is probabilistic, there is still a good chance that a node with very low energy gets selected as a CH. Due to the same reason, it is possible that the elected CHs will be concentrated in one part of the network and some nodes will not have any CH in their range. Also, the CHs are assumed to have a long communication range so that the data can reach the BS directly. It cannot be used effectively on networks deployed in large regions.

3.2 Energy-Efficient Hierarchical Clustering (EEHC)

The main objective of this algorithm was to address the shortcomings of one-hop random selection algorithms such as LEACH by extending the cluster architecture to multiple hops. It is a distributed, k-hop hierarchical clustering algorithm aiming at the maximization of the network lifetime [9]. Initially, each sensor node is elected as a CH with probability "p" and announces its election to the neighbouring nodes within its communication range. The above CHs are now called the "volunteer" CHs. Next, all the nodes that are within "k"-hops distance from a "volunteer" CH, are supposed to receive the election message either directly or through intermediate forwarding. Consequently, any node that receives such CH election message and is not itself a CH, becomes a member of the closest cluster. Additionally a number of „forced‟ CHs are elected from nodes that are neither CHs nor belong to a cluster. Specifically, if the election messages do not reach a node within a present time interval t, the node becomes a “forced” CH assuming that it is not within k hops of all volunteer CHs. However, the most challenging feature of the EEHC algorithm is the direct extension to a corresponding multi-level clustering structure. Considering the overall performance of EEHC, the energy consumption for network operations (data gathering, aggregation, transmission to the BS, etc.) clearly depends on the parameters p and k of the algorithm.

3.3 Hybrid Energy-Efficient Distributed Clustering (HEED)

Another improved and very popular energy-efficient protocol is HEED (Hybrid Energy- Efficient Distributed Clustering [14]). HEED is a hierarchical, distributed, clustering scheme in which a single-hop communication pattern is retained within each cluster, whereas multi-hop communication is allowed among CHs and the BS. The CH nodes are chosen based on two basic parameters, residual energy and intra cluster communication cost. Residual energy of each node is used to probabilistically choose the initial set of CHs. On the other hand, intra cluster communication cost reflects the node degree or node’s proximity to the neighbor and is used by the nodes in deciding to join a cluster or not. Thus, unlike LEACH, in HEED the CH nodes are not selected randomly. Only sensors that have a high residual energy are expected to become CH nodes. Also, the probability of two nodes within the transmission range of each other becoming CHs is small. Unlike LEACH, this means that CH nodes are well distributed in the network. Moreover, when choosing a cluster, a node will communicate with the CH that yields the lowest intra cluster communication cost. In HEED, each node is mapped to exactly one cluster and can directly communicate with its CH. Also, energy consumption is not assumed to be uniform for all the nodes. It also clearly outperforms LEACH with regard to the network lifetime and the desired distribution of energy consumption. However, synchronization is required and the energy consumed during data transmission for far away cluster heads is significant, especially in large-scale networks. Also, acknowledge of the entire network is normally needed to determine reliably the intracluster communication cost and configuration of those parameters might be difficult in practical world.

In conclusion, the probabilistic protocols can be regarded as the leading class of clustering algorithms for WSNs due to their simplicity and their high energy efficiency. The basic disadvantage (however not critical in large scale environments) of these protocols is that due to their probabilistic nature, the CHs are not always distributed well and the CH role is not always rotated uniformly, which sometimes influences the distribution of energy consumption.

4. Non Probabilistic Clustering Approaches

4.1 Popular Non Probabilistic Clustering Protocols

In Non-Probabilistic Clustering Algorithms, more specific criteria for CH election and cluster formation which are primarily considered are mainly based on the nodes‟ proximity i.e. connectivity, degree, distance etc. and on the information received from other closely located nodes. In addition to node proximity, some algorithms conjointly use a combination of metrics like the residual energy, transmission power, mobility, etc. to achieve more generalized goals than single-criterion protocols [9]. The cluster formation procedure here is mainly based on the communication of nodes with their neighbour’s i.e. in one or multi-hop neighbours and generally requires more intensive exchange of messages and probably graphs traversing in some extent. On the contrary these algorithms are usually more reliable toward the direction of extracting robust and well-balanced clusters.

4.2 Highest-Connectivity Cluster Algorithm (HCC)

HCC is a Non-Probabilistic Clustering Algorithm. It is a distributed multi-hop hierarchical clustering algorithm which also efficiently extends to form a multi-level cluster hierarchy [15, 16].

HCC proceeds in two phases, one is Tree Discovery” and other one is “Cluster Formation [9]. Any node in the WSN can initiate the cluster formation process. Each node broadcasts the number of neighbor’s it have, connectivity of node is considered, the node with highest connectivity is elected as CH, but in the case of a tie, the node with the lowest connectivity persuaded. Node which has already selected a CH withdraws its intention to be a CH. The connectivity based heuristic used in this scheme elect’s the sensor with maximum number of 1-hop neighbours as the CH. The creation of one-hop cluster and clock synchronization requirement limit the practical usage of the algorithm [3].

4.3 Biologically Inspired Clustering Algorithm

In the last few years some new algorithms have also been proposed based on swarm intelligence techniques which model the collective behavior of social insects such as ants [17]. Swarm intelligence clustering algorithm based on the Antclust method which is a model of an ant colonial closure to solve problems in cluster formation. In colonial closure model, once two objects meet along they acknowledge whether or not they belong to identical cluster or not. Within the case of a WSN, in the start the device nodes with a lot of residual energy become CHs severally. Then, every which way chosen nodes meet one another, exchange info, and clusters are created, merged, and discarded through these native conferences and comparison of their info. Every node with less residual energy chooses a cluster supported specific criteria, just like the residual energy of the CH, its distance to the CH, and estimation of the cluster size. Eventually, energy economical clusters are fashioned that end in Associate in Nursing extension of the lifespan of the WSN. Generally, biologically inspired clustering algorithms show that they can dynamically control the CH selection while achieving quite uniform distribution of CHs and energy consumption.

4.5 Weight-Based Clustering Algorithm

WCS is a Non-Probabilistic Clustering Algorithm, which uses distributed scheme for cluster formation in this single-hop communication pattern. In this CH is elected non-periodically. For purpose of power saving, it invoked a new election every time a sensor loses the connection with any CH. It is invoked on demand, every time a reconfiguration of the network’s topology is inescapable [18, 19]. WCA is based on a combination of metrics that take into account several system parameters such as: the ideal node degree; transmission power; mobility; and the remaining energy of the nodes. Depending on the specific application, any or all of these parameters can be used as a metric to elect CHs. The election procedure is based upon a global parameter that is called combined. This algorithm attempted to provide better load balancing through reduced number of sensors in a cluster but the requirement of clock synchronization limits its applications. [9, 14]. Priya Vyas *et al.* / (IJCSIT) International Journal of Computer Science and Information Technologies, 2014; 5(5):6614-6619 www.ijcsi.

5. Topology Construction and Maintenance

Energy Saving and improve the lifetime of the sensor node is main focus. The most important technique used in WSN to reduce energy consumption is topology control. The main component of WSN is battery powered sensors having low cost and power. So in order to utilize limited energy resources, it has to be managed carefully in order to extend the life time of the network [16]. The sensor nodes communicating with each other using the various types of topology like mesh, tree, chain etc. Therefore it is required to form the efficient topology. Lately, topology control algorithms have been divided into two sub problems: topology construction, in charge of the initial reduction, and topology maintenance, in charge of the maintenance of the reduced topology so that characteristics like connectivity and coverage are preserved.

This is the first stage of a topology control protocol. Once the initial topology is deployed, especially when the location of the nodes is random, the administrator has no control over the design of the network; for example, some areas may be very dense, showing a high number of redundant nodes, which will increase the number of message collisions and will provide several copies of the same information from similarly located nodes. However, the administrator has control over some parameters of the network: transmission power of the nodes, state of the nodes (active or sleeping), role of the nodes (Cluster head, gateway, regular), etc. By modifying these parameters, the topology of the network can change. Upon the same time a topology is reduced and the network starts serving its purpose, the selected nodes start spending energy: Reduced topology starts losing its "optimality as soon as full network activity evolves. After some time being active, some nodes will start to run out of energy. Especially in wireless sensor networks with multihopping, intensive packet forwarding causes nodes that are closer to the sink to spend higher amounts of energy than nodes that are farther away. Topology control has to be executed periodically in order to preserve the desired properties such as connectivity, coverage, density.

In topology construction phase, an efficient methodology is used that constructs a topology dynamically without violating network scenarios and thereby providing both connectivity and coverage. We start constructing the topology T (v) where each node „n" selects its neighbor v based on range, energy and distance. Initially the neighbour list of a node is set to NULL. In order to select the neighbour we determine energy of each node E (v). Now the node list E (v) is sorted in decreasing order. After computing the energy we find the distance of each node v from „n" in order to determine the neighbour list of every node. The nodes which are not present in the neighbour list N (v) are in inactive state.

In topology maintenance phase, as node start communicating with each other the energy depletes and some nodes fail rendering the network useless. Hence to make network functioning optimal we need to provide topology maintenance. This process involves periodic restoration, rotation and recreation of network topology. Here the node periodically broadcasts the hello messages to its neighbour and waits for the acknowledgement. The hello message carries the following information <node id, neighbour id, energy) If the node does not receive back acknowledgement then it involved topology construction function.

6. Conclusion

The lifetime of a Wireless Sensor Network operation on battery power is critical to its usefulness. Generally, clustering in such networks have been of high interest in the last decade and there is already a large number of related published works. Throughout this paper, we tried to present the main characteristics of the most significant protocols that were proposed till now in the literature. As it was pointed out, grouping nodes into clusters, thus leading to various representations and protocols, have been regarded as the most efficient approach to support scalability in WSNs. It also facilitates the efficient data gathering and aggregation independent to the growth of the WSN, and generally reduces the total amount of communications as well as the energy spent. To increase the network life time, topology control plays significant role which improves the energy saving with the help of intervention of topology construction and maintenance approaches.

7. References

1. Kulothungan K J. Angel Arul Jothi, Kannan A. An Adaptive Fault Tolerant Routing Protocol with Error Reporting Scheme for Wireless Sensor Networks, European Journal of Scientific Research, 2011, 16:19-32:1
2. Sohrabi K. et al. Protocols for self-organization of a wireless sensor network, IEEE Personal Communications, 2000; 7(5):16-27.
3. Min R. et al. Low power wireless sensor networks, in Proceedings of International Conference on VLSI Design, Bangalore, India, 2001.
4. Abbasi A A, Younis M. A survey on clustering algorithms for wireless sensor networks, Computer Communications, 2007. Conference on VLSI Design, Bangalore, India, 2001, 30:2826-2841.
5. Liliana M, Arboleda C, Nidal N. Comparison of clustering algorithms and protocols for wireless sensor networks, in Proceedings of IEEE CCECE/CCGEI Conference, Ottawa, Ontario, Canada, 2006, 1787-1792.
6. Heinzelman WR, Chandrakasan AP, Balakrishnan H, Energy efficient communication protocol for wireless microsensor networks, in Proceedings of the 33 rd Hawaaian Interantional Conference on System Sciences, 2000.
7. Heinzelman WB, Chandrakasan AP, Balakrishnan H. An application specific protocol architecture for wireless microsensor networks, IEEE Transactions on WirelessCommunications, 2002; 1(4):660-670.

8. Flowchart of the cluster formation process of LEACH. (Redrawn from Heinzelman WB. et al. IEEE Trans. Wireless Commun, 2002. 1:660.
9. Younis O. Fahmy S. HEED: A hybrid, energy-efficient, distributed clustering approach for Ad Hoc sensor networks, IEEE Transactions on Mobile Computing, 2004; 3(4):366-379.
10. Parekh AK. Selecting routers in ad-hoc wireless networks, in Proceedings of SBT/IEEE International Telecommunications Symposium, ITS, 1994.
11. Amis R. Prakash T. Vuong, Huynh D. Max-Min DC luster formation in wireless Ad Hoc networks, 2000. IEEE INFOCOM.
12. Selvakennedy S, Sinnappan S, Shang Y. A biologically inspired clustering protocol for wireless sensor networks, Computer Communications, 2007. 30:2786-2801.
13. Basagni S. Distributed clustering for Ad Hoc networks, in Proceedings of Intl. Symp. On Parallel Architectures, Algorithms & Networks (ISPAN) 1999, 310-315.
14. Chatterjee M, Das SK, Turgut D. WCA: A weighted clustering algorithm for mobile ad hoc networks, Clustering Computing, 2002. 5:193-204.
15. Cook DJ, Das SK. Smart environments; technologies, protocols and applications, New York: John Wiley 2004, 13-15.