A Research on Decentralized Clustering Algorithms for Dense Wireless Sensor Networks

S.R.Boselin Prabhu
Assistant Professor
SVS College of Engineering
Tamilnadu, India.

S.Sophia, PhD.
Professor
Sri Krishna College of Engineering and Technology
Tamilnadu, India.

P.D.Manivannan, S.Nithya
R.Mahalakshmi
Final Year Students
SVS College of Engineering,
Tamilnadu, India.

ABSTRACT

A wireless sensor network with large number of sensor nodes can be used as an effective tool for collecting data in various situations. Recent advancement in wireless communications and electronics has enabled the development of low-cost sensor network. Wireless sensor networks 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. The conventional routing protocols cannot be used here due to its battery powered nodes. To support scalability, nodes are often grouped into non-overlapping clusters. This paper gives a brief introduction on clustering process in wireless sensor networks. A comparative analysis of different decentralized clustering algorithms used in wireless sensor networks is elaborated based on some metrics such as node mobility, cluster count, cluster head selection, etc.

General Terms

Wireless communication, energy efficiency, network lifetime.

Keywords

Wireless sensor network (WSN), clustering, decentralized clustering algorithms.

1. INTRODUCTION

Wireless Sensor Network (WSN) is an emerging and very interesting technology applied to different applications [1]. WSNs are networks in which thousands of small and battery powered nodes communicate with each other. A WSN consists of number of spatially distributed nodes which are interconnected without the use of wires. Each node is connected with one or more sensors. Each sensor node consists of a radio receiver, a microcontroller and a battery. The WSN is mainly used in area monitoring, environmental monitoring, industrial monitoring, waste water monitoring, structural monitoring, etc (Figure 1). The major advantages of these networks: they avoid wiring problems and it can be accessed through centralized control.

In order to reduce the data transmission time and energy consumption [6], the sensor nodes are grouped into a number of small groups called clusters. The grouping of sensor nodes is known as clustering (Figure 2). Every cluster has a leader which is known as cluster head (CH). A CH is also one of the sensor nodes which have higher capabilities than other sensor nodes. The cluster head is selected by the sensor nodes in the respective cluster. CHs may also be pre-assigned by the user. The advantages of clustering are that it transmits the aggregated data [5] to the sink or base station [3]. It provides scalability for large number of nodes and reduces energy consumption.

Clustering can be classified into three types: centralized clustering, distributed clustering and hybrid clustering. The centralized clustering is the one in which the cluster head is fixed. The remaining nodes in the cluster act as a member nodes. Distributed clustering is the one in which the cluster head is not fixed. The cluster head keeps on changing form node to node based n on some parameters. Hybrid clustering is the one which is formed by the combination of the centralized clustering and the distributed clustering.

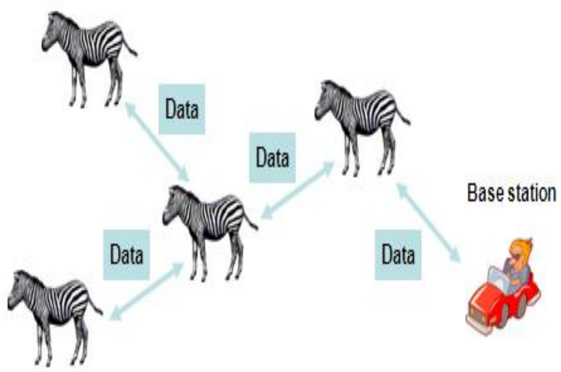

Figure 1: Application fields of Wireless Sensor Network for habitat monitoring (ZebraNet)

In centralized clustering, the cluster head is fixed. Because of this, when the cluster head becomes dead, the whole cluster gets collapsed. Hence, there is no reliability in centralized clustering. Distributed clustering provides reliability of data even when the cluster head fails. The network fails only when all the nodes fail. The distributed clustering is commonly used in WSNs because it provides better collection of data [4, 7] and reliability. In addition to clustering algorithms, there are many routing algorithms available in wireless sensor networks. The routing protocols are divided into flat-based routing, hierarchical-based routing and location-based routing depending upon the network structure (Few examples are: MCFA, IDSQ, CADR, COUGAR, ACQUIRE, MECN, SOP, VGA, HPAR, TTDD, GAF, GEAR, MFR, DIR, GEDIR, GOAFR, SPAN, etc.).

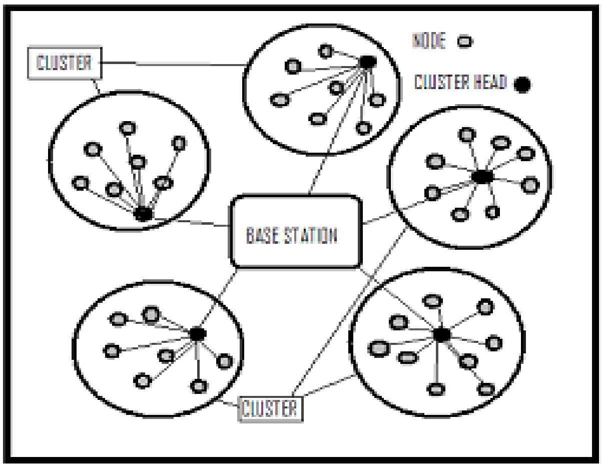

Figure 2: Mechanism of Cluster Formation

The clustering mechanisms and the decentralized clustering algorithms are discussed briefly in this paper. Their features are discussed in the first section. In the second section, the different available decentralized clustering algorithms for WSNs are compared. And finally the third section concludes the paper.

2. DECENTRALIZED CLUSTERING ALGORITHMS IN DENSE WIRELESS SENSOR NETWORKS

Decentralized clustering is the one in which the cluster head changes from one node to another depending on the resources of the nodes. In this section, a comparative study of various available decentralized clustering algorithms for WSNs is presented.

Linked Cluster Algorithm (LCA): The Linked Cluster Algorithm (LCA) is a decentralized clustering algorithm which avoids communication collisions among nodes. It uses TDMA frames for inter-node communication. Each frame has slot for each node in the network for communication. In LCA, every nodes requires 2n time slots, where 'n' is the number of nodes in the network [16, 17, 25]. 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 use in the networks having less than 100 nodes. In such small networks, the delay between the node transmissions is minor and may be accepted.

Algorithm for Cluster Establishment (ACE): ACE [14] is a highly uniform cluster formation, self-organizing, efficient coverage, lesser overlapping and emergent cluster forming algorithm for WSNs. This 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 concept 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 are "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 sends an invitation 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 find which node is the best candidate to become a CH for the cluster. The best candidate is the node which has 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, clustered 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.

An enhancement to the migration process in ACE algorithm was proposed in [15]. The idea is to further iterate in order to increase the regularity of cluster layout. 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 compared to emergent algorithms in large-scale distributed systems.

Hausdroff Clustering (HC): In a decentralized clustering algorithm nodes make autonomous decisions. HC assumes that nodes use traditional RTS/CTS based collision avoidance mechanisms. In this algorithm, once cluster formations take place it remains same throughout the network lifetime [19]. Moreover, to evenly use the energy among all the nodes, CH is rotated among cluster members. At the beginning, each node sends a topology discovery message with the lowest power level to find all its neighbors. Then, the base station (BS) appoints an initiator for starting the clustering operation. The initiator sends a clustering message and waits for join-requests from neighboring nodes. It then admits cluster members according to the clustering conditions. If the applying node receives an admission message, it sends back a confirmation message. On receiving the confirmation message, the initiator updates its membership list and broadcasts a membership update message with a higher power level for informing neighboring clusters. If the candidate node is rejected by all its neighboring clusters, it organizes a new cluster with itself being the initiator. For each cluster, a node with maximum residual energy is selected as the CH, which is also the one with minimum root mean-square distance to its neighbors. In each beginning of a round, the old CH carries out the greedy algorithm and selects the new one. After the new CH is selected, the old one will announce with the lowest power level within the cluster. The new CH broadcasts a message with a higher power level to its neighboring CHs and gets itself connected for routing. This algorithm maximizes the lifetime of each cluster in order to increase the lifetime of the system.

Ring-structured Energy-efficient Clustering Algorithm (RECA): Clustering methods have reduced the conservation of energy in WSNs. RECA mainly focuses in prolonging the network lifetime [22]. This algorithm uses deterministic CH management algorithm to evenly distribute the workload among the nodes within a cluster. Nodes within a cluster make local decisions on the length of their duty cycle according to their remaining energy supply. This shows that all nodes deplete their energy supply at approximately the same time regardless of the initial amount of energy in their batteries. This algorithm is efficient in managing energy in a wide range of networks settings.

Low Energy Adaptive Clustering Hierarchy (LEACH): *LEACH [18]* is a clustering mechanism that distributes energy consumption all along its network, the network being divided into clusters, CHs which are purely distributed in manner and the randomly elected CHs, collect the information from the nodes which are coming under its cluster. It forms clusters based on the received signal strength (RSS) and 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 [8], where nodes make autonomous decisions without any centralized control. Initially a node decides to be a CH with a probability ‘p’ and broadcasts its decision. Each non-CH node determines its cluster by choosing the CH that can be reached using the least communication energy. The role of being a CH is rotated periodically among the nodes of the cluster in order to balance the load. The rotation is performed by getting each node to choose a random number ‘T’ between 0 and 1. A node becomes a CH for the current rotation round if the number is less than the threshold described in equation 1.

$$T(i) = \begin{cases} \frac{p}{1-p} * \left(\frac{rmod1}{p}\right) & if \rightarrow i\epsilon G \\ 0 & otherwise \end{cases} \quad \textbf{(1)}$$

where, ‘p’ is the desired percentage of CH nodes in the sensor population, ‘r’ is the current round number and ‘G’ is the set of nodes that have not been CHs in the last (1/p) rounds. Since the decision to change the CH is probabilistic, there is a good chance that a node with very low energy gets selected as a CH. When this node dies, the whole cell becomes dysfunctional. Also, the CH is assumed to have a long communication range so that the data can reach the BS from the CH directly. This is not always a realistic assumption since the CHs are regular sensors and the BS is often not directly reachable to all nodes due to signal propagation problems. This algorithm also forms one-hop intra-cluster and inter-cluster topology where each node can transmit directly to the CH and thereafter to the BS. Consequently, it is not applicable to large scale networks.

Two-Level LEACH (TL-LEACH): A new version of LEACH called TL-LEACH, the CH collects data from other cluster members as in original LEACH, but rather than transferring data to the BS directly, it uses one of the CH that lies between the CH and the BS as a relay station. It has two levels of cluster heads (primary CH and secondary CH). The primary CH in each cluster communicates with the secondary CH, and the corresponding secondary CH communicates with the nodes in their sub-cluster. Data fusion can be performed as in LEACH. Additionally, communication within a cluster is scheduled using TDMA time-slots. The organization of a round will consist of first selecting the primary and secondary CHs using the same mechanism as in LEACH, with the probability of being elevated to a primary CH less than that of a secondary node. Communication of data from source node to sink is done in two steps: secondary nodes collect data from nodes in their respective clusters. Data fusion can be performed at this level. Primary nodes collect data from their respective secondary clusters. Data-fusion can also be implemented at the primary CH level. The two-level structure of this algorithm reduces the number of nodes that need to transmit to the BS, thereby effectively reducing the total energy usage.

CLUBS: CLUBS is an algorithm that forms clusters through local broadcast and converge in a time proportional to the local density of nodes. 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 [12], which means that every node within the cluster should be able to communicate with each other using only nodes within that same cluster. 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 [11] a “recruit message“. The nodes that get recruited are generally called “followers”. Since CLUBS allows overlapping, follower nodes keep listening to additional recruit messages and can be follower of more than on CH. If a node that is competing to become a CH detects a collision or received a garbled message, it becomes a follower node and assumes that multiple CHs attempted to recruit it at the same time. It can find out its CH later. CLUBS can be implemented in the asynchronous environment without losing efficiency and simplicity. Furthermore, CLUBS satisfies many constraints that are common in other distributed environment such as limited or no topology knowledge or access to global IDs. The major problem of CLUBS algorithm is the clusters having their CHs within one-hop range of each other. If this is the case, both clusters will collapse and CH election process will restart.

Hybrid Energy-Efficient Distributed Clustering (HEED): HEED [13] is a decentralized algorithm which selects the CH based on both residual energy and communication cost. Basically HEED was proposed to avoid the random selection of CHs. This algorithm 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 2.

$$CH_{prob} = \left(C_{prob}\right) * \left(E_{residual}/E_{max}\right) \quad \textbf{(2)}$$

where, Eresidual is the residual energy of the concerned node, Emax is the 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 and finalization phase, in which the selection of CH will be properly finalized.

Multi-hop Overlapping Clustering Algorithm (MOCA): MOCA is a randomized, distributed multi-hop overlapping clustering algorithm introduced for organizing the sensors into overlapping clusters [24]. The goal of this clustering process is to ensure that each node is either a CH or within k-hops from at least one CH, where ‘k’ is a preset cluster radius. The algorithm initially assumes that each sensor in the network becomes a CH with probability ‘p’. Each CH then advertises itself to the sensors within its radio range. This advertisement is forwarded to all sensors that are no more than k-hops away from the CH. A node sends a request to all CHs to join their clusters. In the join request, the node includes the ID of all CHs it heard from, which implicitly implies that it is a boundary node [9]. The CH nomination probability (p) is used to control the number of clusters in the network and the degree of overlap among them. By choosing an appropriate value of 'p' this algorithm achieves an excellent cluster count and overlapping degree.

Fast Local Clustering Service (FLOC): This decentralized clustering algorithm [20] is suitable for clustering large-scale WSNs. It is fast, scalable, produces non-overlapping and approximately equal-sized clusters. FLOC achieves locality: effects of cluster formation and faults or changes at any part of the network are contained within a predetermined radius. FLOC exploits the double-band nature of wireless radio-model. 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. I-band nodes will be affected by very little interference communicating with the CH, while message from o-band nodes may have maximum probability of getting lost. FLOC favors i-band membership in order to increase the robustness of the intra-cluster traffic. FLOC 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 i-band node in another cluster. Also it achieves clustering in constant time regardless of the network size and in a local manner. It also achieve locality, in that each node is only affected by the nodes within two units.

Distributed Weight-Based Energy-Efficient Hierarchical Clustering (DWEHC): Distributed Weight-Based Energy-Efficient Hierarchical Clustering [21] is a decentralized 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 node which is having more weight 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. A node progressively adjusts to such membership in order 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. 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: 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, 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. The leading advantage of DWEHC over HEED is that, it shows a great improvement in both intra-cluster energy consumption and inter-cluster energy consumption [21].Table 1, shows the comparison of various presented decentralized clustering algorithms for WSNs.

Decentralized Clustering Algorithm	Time Complexity	Node Mobility	Cluster Overlap	In-Cluster Topology	Cluster Count	Cluster Head Selection
LCA	Variable	Possible	No	1-hop	Variable	Random
ACE	Constant	Possible	Yes	K-hop	Variable	Random
HC	Variable	Possible	No	1-hop	Variable	Connectivity
RECA	Constant	No	No	1-hop	Variable	Random
TL-LEACH	Constant	Possible	No	1-hop	Variable	Random
CLUBS	Variable	Possible	Yes	2-hop	Variable	Random

HEED	Constant	Stationary	No	1-hop	Variable	Random
LEACH	Constant	Fixed BS	No	1-hop	Variable	Random
MOCA	Constant	Stationary	Yes	K-hop	Variable	Random
FLOC	Constant	Possible	No	2-hop	Variable	Random

Table 1: Comparison of presented decentralized clustering algorithms [2]

3. CONCLUSION

A growing list of civil and military applications can employ WSNs in hostile and remote areas. It can be used by the military for a number of purposes such as monitoring militant activity in remote areas and force protection. WSNs are also used for the collection of data for monitoring of environmental information. Most recent researches have begun to consider a wider range of aspects such as wireless link reliability, real-time capabilities and quality-of-service (QoS). Clustering provides scalability, energy saving, reliability, etc. In this paper, a comparative analysis of different available decentralized clustering algorithm has been done. A comparative analysis of various decentralized algorithms with their parameters has been detailed elaborately. Future works may concentrate on developing a better clustering algorithm.

4. ACKNOWLEDGMENTS

Our sincere thanks to the management of SVS Educational Institutions and all the experts who have guided us to come out with this research work.

5. REFERENCES

[1] Chia Hung Tsai, Yu Chee Tseng, (2012), "A Path-Connected-Cluster Wireless Sensor Network and Its Formation, Addressing, and Routing Protocols", IEEE Sensors Journal, Volume 12, Number 6.

[2] S.R.Boselin Prabhu, S.Sophia, (2011), "A survey of adaptive distributed clustering algorithms for wireless sensor networks", International Journal of Computer Science and Engineering Survey, Volume 2, Number 4, Pages 165-176.

[3] C.T. Cheng, C.K. Tse, and F. C. M. Lau, (2011), "A clustering algorithm for wireless sensor networks based on social insect colonies," IEEE Sensors Journal, Volume 11, Number 3, Pages 711–721.

[4] Pedro A. Forero, Alfonso Cano, Georgios B.Giannakis, (2011), "Distributed clustering using wireless sensor networks", IEEE Journal of Selected Topics in Signal Processing, Volume 5, Number 4, Pages 707-724.

[5] Yajie Ma, Yike Guo, Xiangchuan Tian, and Moustafa Ghanem, (2011)," Distributed Clustering-Based Aggregation Algorithm for Spatial Correlated Sensor Networks", IEEE Sensors Journal, Volume 11, Number 3.

[6] Yang Yang, Rick S. Blum, and Brian M. Sadler, (2010)," A Distributed and Energy-Efficient Framework for Neyman-Pearson Detection of Fluctuating Signals in Large-Scale Sensor Networks", IEEE Journal, Volume 28, Number 7.

[7] Y. Yang, R. S. Blum, and B. M. Sadler, (2009), "Energy-efficient routing for signal detection in wireless sensor networks", IEEE Transactions on Signal Processing, Volume 57, Number 6, Pages 2050-2063.

[8] H. Jeong, C.S. Nam, Y.S. Jeong, and D.R. Shin, (2008), "A mobile agent based LEACH in wireless sensor network", Proceedings of the International Conference on Advanced Communication Technology (ICACT), Pages 75–78.

[9] X. Fan and Y. Song, (2007), "Improvement on LEACH protocol of wireless sensor network," IEEE Sensor Communication, Pages 260–264.

[10] Tony Q.S. Quek, Davide Dardari, and Moe Z. Win, (2007)," Energy Efficiency of Dense Wireless Sensor Networks: To Cooperate or Not to Cooperate", Volume 25, Number 2.

[11] R. Nagpal, D. Coore, "An algorithm for group formation in an amorphous computer", Proceedings of the 10th International Conference on Parallel and Distributed Systems (PDCS'98).

[12] A Werbuch, Berger, Cowen and Peleg, (1996), "Fast distributed network decompositions and covers", Journal of Parallel and Distributed Computer", Pages 105-114.

[13] O. Younis, S. Fahmy, (2004), "HEED: A hybrid energy-efficient distributed clustering approach for Ad Hoc sensor networks", IEEE Transactions on mobile computing, Pages 366-379.

[14] H. Chan, A. Perrig, (2004), "ACE: An emergent algorithm for highly uniform cluster formation",

Proceedings of the 1st European Workshop on Sensor Networks (EWSN).

[15] H. Chan, M. Luk, A. Perrig, (2005), “Clustering information for sensor network localization”, Proceedings of the International Conference on Distributed Computing in Sensor Systems (DCOSS’05).

[16] D.J. Barker, A. Ephremides, (1981), “The architectural organization of a mobile radio network via a distributed algorithm”, IEEE Transactions on Communications, Pages 1694-1701.

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

[18] W.B. Heinzelman, A.P. Chandrakasan, H. Balakrishnan, (2002), “Application specific protocol architecture for wireless microsensor networks”, IEEE Transactions on Wireless Networking.

[19] X. Zhu, L. Shen and T. Yum, (2011), “Hausdorff clustering and minimum energy routing for wireless sensor networks”, IEEE Transaction on Vehicular Technology, Volume 58, Number 2, Pages 990–997.

[20] M. Demirbas, A. Arora and V. Mittal, (2004), “FLOC: A fast local clustering service for wireless sensor networks”, Proceedings of Workshop on Dependability Issues in Wireless Ad Hoc Networks and Sensor Networks (DIWANS’04).

[21] P. Ding, J. Holliday, A. Celik, (2005), “Distributed energy efficient hierarchical clustering for wireless sensor networks”, Proceedings of the IEEE International Conference on Distributed Computing in Sensor Systems (DCOSS’05).

[22] Guang Feng Li, Taieb Znati, (2007), “A ring-structured energy-efficient clustering architecture for robust communication in wireless sensor networks”, International Journal of Sensor Networks, Volume 2.

[23] V. Loscri, G. Morabito, and S. Marano, (2011), “A Two-Level Hierarchy for Low Energy Adaptive Clustering Hierarchy”, DEIS Department, University of Calabria

[24] A. Youssef, M. Younis, M. Youssef, A. Agrawala, (2006), “Distributed formation of overlapping multi-hop clusters in wireless sensor networks”, Proceedings of the 49th Annual IEEE Global Communication Conference (GLOBECOM’06).

[25] H. Karkvandi, E. Pecht, and O. Yadid Pecht, (2011), “Effective lifetime-aware routing in wireless sensor networks”, IEEE Sensors Journal, Volume 11, Number 12