A Review on Data Collection Techniques in Wireless Sensor Networks

A.Prasanth
Assistant Professor, Department of Electronics and Communication Engineering
Akshaya College of Engineering and Technology
Coimbatore, India
aprasanthdgl@gmail.com

***Abstract* — Data collection is a fundamental task in wireless sensor networks. Due to the constraints in communication bandwidth and energy budget data collection is a wise choice in many applications. In object tracking applications, an object is often detected by sensor nodes in multiple clusters, leading to redundant data transmissions through multiple paths from the cluster heads to the data sink. Due to minimum no of sensor nodes, it is valuable to reduce the amount of data transmission so that the average sensor lifetime is reduced. The main goal of data gathering algorithms is to gather and aggregate data in an energy efficient manner so that network lifetime is enhanced. This paper describes a review on data collection techniques in wireless sensor networks. We also studied advantages and disadvantages of each method and the comparison of each method with different constraints such as amount of gathered packets, energy efficiency, network lifetime etc.**

***Keywords* — wireless sensor network, object tracking, data collection, energy efficiency, network lifetime.**

I. INTRODUCTION

Wireless sensor networks are usually composed of several inexpensive, low-powered sensing devices with limited memory, computational, and communication resources. Energy efficiency and time efficiency are two major considerations for sensor data gathering in wireless sensor networks. Clustering is an important technique to organize a densely deployed network. The clustering algorithms keep only a portion of nodes (CHs) active and save the energy for the remaining nodes. Nevertheless, in a target tracking application, due to the random movement of the target and large overlapped sensing area between clusters, the target will often be detected by the nodes in multiple clusters at the same time. When all the nodes report their data to the respective CHs, each involved CH has to send the data to the sink. This way may produce much redundant data that causes unnecessary energy consumption. The cluster head should closer to the sink [5]. While collecting the data, load balance should be considered to balance energy cost [1]. The nodes should also avoid mutual transmission and loop transmission [2]. Sometimes mobility is not considered in the data collection process [3]. The main goal of data collection is to increase the network lifetime by reducing the resource consumption of sensor nodes (such as battery energy and bandwidth). After data collection process the data should be aggregated to reduce redundant data transmission. Data aggregation means the process of aggregating the data from numerous sensors to eliminate redundant transmission and provide aggregated information to the base station. Data aggregation usually involves aggregating the data from several sensors at intermediate nodes and transmitting the aggregated data to the base station (sink). The node which aggregates the data is called aggregator node. It transmits the data to the cluster head or the next level node based on residual energy, communication cost etc.

The paper is organized as follows: Section 2, 3 presents the data collection and its types in wireless sensor networks. Section 4 presents the application of data collection. In Section 5 the data collection techniques in wireless sensor networks and the advantages and disadvantages of each scheme are discussed. In Section 6,7,8 issues, objectives and future work are discussed. The conclusion of the paper is given in section 9.

II. DATA COLLECTION

Data collection is the fundamental function of wireless sensor networks, but also a challenging task due to limited battery power of those tiny sensor nodes. Among all activities of sensor nodes, it is well-known that data communication causes the maximum energy drain. Therefore, data gathering methods should avoids abundant communication overhead yet keeps the quality of data, becomes the effective method to achieve a longer network lifetime of wireless sensor networks for data-driven applications, which require sensor nodes to perform data sampling and transmit data to sink at regular time interval, such as environmental monitoring. Figure 1 shows the data collection task. The sensor nodes gather data from the object and it transmits to the sink node. Then the sink node transmits the data to the base station via internet. To gather the data efficiently it should consider the optimal path for data transmission. Data is gathered and aggregated before transmission [10]. The node which aggregates the data is called aggregator node. It reduces the redundant data transmission and increases the energy efficiency of the sensor nodes.

Extensive research work has been done to conserve the finite resources, such as network bandwidth, data packets energy and CPU usage, and various energy-saving protocols and algorithms have been proposed for these data-driven applications.

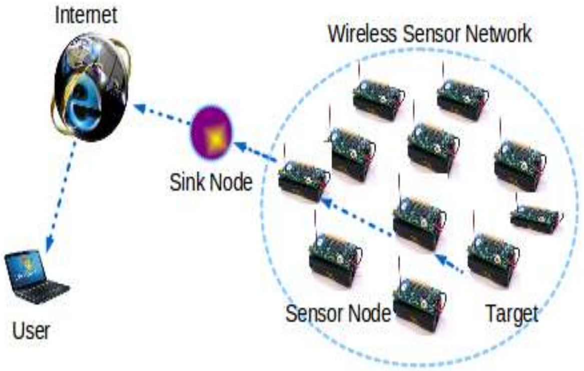


Fig. 1. Data collection

III. TYPES OF DATA COLLECTION

The data collection technique is used to collect the aggregated data from the sensor node to the sink node. The main objective of the data collection process is to reduce the delay and improves the network's lifetime. There are some types of data collection to gather data. They are tree based technique, cluster based technique, multipath technique and hybrid technique.

A. Tree Based Technique

The tree based technique defines data collected from an aggregation tree. The hierarchy model is shown in Fig. 2. Here, sink node (base station) considers as a root node and source node consider as a leaves nodes. The data is flowing initiated from leaves node up to sink (base station). Disadvantage of this technique is wireless sensor network are not free from collapse in case of data packet failure at any level of tree, the data will be misplaced not only for single level but also for entire related sub tree as well.

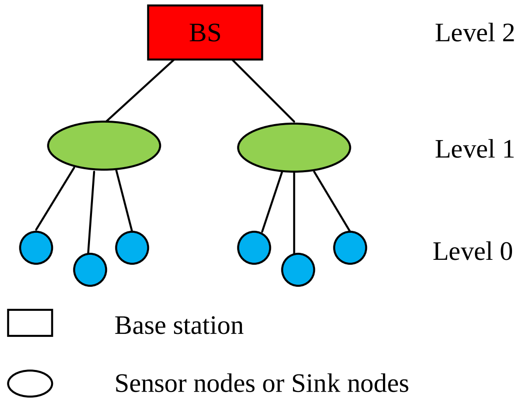


Fig. 2. Tree based technique

B. Cluster based technique

It is not suited for sensors to broadcast the data directly to the base station, in energy constrained sensor networks of large size. In such process, Cluster based technique is hierarchical technique. Fig. 3. show cluster based technique, in this technique the whole network is separated in to numerous clusters. Cluster member selects Cluster head for each cluster. Cluster heads do the role of data collectors which gather the data received from cluster members locally and then convey the result to base station (sink). Recently, numerous cluster based network organization and data gathering protocols have been proposed for the wireless sensor networks. The cluster heads are capable to communicate through the Base station directly via lengthy range transmissions or multi hopping through other cluster heads.

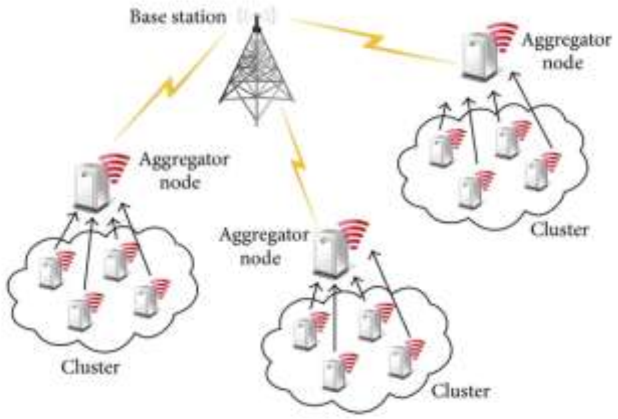


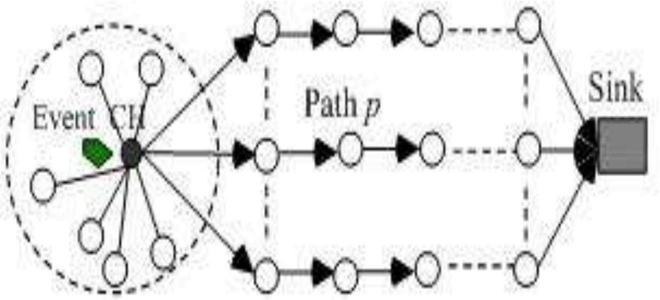
Fig. 3. Clustering

Data collection methods for Cluster-based WSN

Typical data collection methods for cluster-based WSNs can be categorized into: 1) CH directly sends data to the sink (Dynamic Location-cluster), and 2) CH sends data to the sink through other CHs in a multi-hop way (Multi Hop-cluster). In Dynamic Location-cluster each CH aggregates the gathered data from its CMs and directly transmits the data to the sink, while in a Multi Hop-cluster a hierarchical model is built for data aggregation among clusters. According to the cluster hierarchy, the aggregated data on CHs will be further aggregated on a CH at a higher level. Therefore, the gathered data can be aggregated hop by hop through multiple intermediate CHs to the sink. Here the hierarchy of the network is fixed at design time and the CMs in a cluster always send the sensed data to respective CHs. Then, each CH involved in data aggregation performs aggregating the data and transmits the aggregated data either directly (Dynamic Location-cluster) or hop by hop (Multi Hop-cluster) to the sink. These approaches do not work efficiently for target tracking due to the changing position of the randomly moving target.

C. Multipath technique

The disadvantage of tree based technique is the imperfect robustness of the system. To overcome this disadvantage, a new technique was proposed by many researchers .in which sending partially gathered data to single parent node in aggregation tree, a node could send data over numerous paths. Fig. 4. illustrates the multipath technique, in which each and every node can send data packets to it's possibly numerous neighbors. Hence data packet flow from source node to the root node along numerous paths, lot of intermediate node between leaves node to root node so gathering done in every intermediate node. The instance of this technique like ring topology, where network is separated in to concentric circle with defining level levels according to hop distance from root. These strategies have both issues: energy efficiency and robustness. In which solitary path to connect every node to the sink node it is energy saving but high risk of link collapse. But on the other head multipath technique would require more nodes to participate with consequent waste of energy.



○ Node or Intermediate node

● Node or Intermediate node

■ Node or Intermediate node

→ Node or Intermediate node

Fig. 4 Multipath technique

D. Hybrid technique

Hybrid technique followed among tree, cluster based and multipath method. In which the data gathering structure can adjusted according to particular network situation and to some performance statistics.

IV. APPLICATION

Wireless Sensor Networks are mainly used in Military, Traffic monitoring, Environmental monitoring, monitoring of weather conditions, Health applications etc. Data gathering is one of the most important applications in wireless sensor networks is. In data gathering the sensed data are collected at sensor nodes and forwarded to a central base station. In most of the applications of wireless sensor networks, the data gathering is an intelligent choice due to the constraints in communication bandwidth and the energy budget. The key idea of data gathering approach is to divide a sensor network into clusters, and determine local data correlations on each cluster head. By using wireless infrastructure and battery powers, sensor nodes can be very small and easily attached at particular locations and without disturbing the surrounding environments. It makes WSN a competitive approach for data collection comparing with its wired counterpart.

V. DATA COLLECTION TECHNIQUES

A. Load balance data gathering algorithm

This paper [1] proposes a load balance data gathering algorithm that classifies sensor nodes into different layers according to their distance to sink node and furthermore, divides the sense zone into several clusters. Routing trees are established between sensor node and sink depending on the energy metric and communication cost. For saving energy consumption, the target of data aggregation scheme is adopted as well.

Through the approach of load balance to the sensor nodes, particularly for the sensor nodes closer to the sink node, we can prevent the data packet collision, network congestion and packet lose happening and balance the energy consumption among sensor nodes, furthermore, prolong the networks lifetime.

The algorithm describes with the following three steps: (1) Nodes are divided into several layers according to the hop counts between sensor nodes and sink node. (2) Routing tree generation and data gathering mechanism. (3) Data aggregation mechanism. Advantages of this approach are achieved load balancing among inter-cluster nodes and it deals with the energy consumption hot spot problem. Disadvantage of using this method is, it is suitable only for low dynamic wireless sensor network.

B. Energy efficient Routing Algorithm to Prolong Lifetime (ERAPL)

This paper [2] proposes a routing algorithm termed Energy-efficient Routing Algorithm to Prolong Lifetime (ERAPL), which is able to dramatically prolong network lifetime while efficiently expends energy. In the ERAPL, a data gathering sequence (DGS), used to avoid mutual transmission and loop transmission among nodes, is constructed, and each node proportionally transmits traffic to the links confined in the DGS. In addition, a mathematical programming model, in which minimal remaining energy of nodes and total energy consumption are included, is presented to optimize network lifetime. Moreover, genetic algorithms are used to find the optimal solution of the proposed programming problem.

The ERAPL is a centralized algorithm and runs at the sink. Recall that the sink and the nodes in the WSN are stationary and the sink knows the topology of the WSN. Therefore, in practice, the DGS can be constructed upon the WSN is deployed. Then, the sink uses the procedure, referred as Announce_msg(), to inform all the nodes in the WSN of a packet that contains the constructed DGS, which guides all the nodes to transmit traffic to their respective neighbors so that mutual transmission among nodes and route loop is avoided and accordingly energy is saved. The procedure is given in the following, in which pkt stands for a packet delivered by the sink, List_nodes is a list to keep nodes, jList nodesj represents the number of nodes in List_nodes, Done_nodes is used to keep the set of nodes having received the pkt, and N(a) is the set of neighbors of node a.

When the sink starts collecting data, it also uses the above procedure to deliver a REQ packet, i.e., Announce_msg(REQ), to inform all the nodes in theWSN. As soon as a node gets the REQ, it sends a REP packet to the sink through the route contained in the DGS received previously. The REP includes the information about the amount of data to be delivered to the sink. The ERAPL is invoked when the REP packets from all the nodes arrive at the sink. Upon the ERAPL ending, the optimal OTP matrix is worked out. Then, each node is sent a packet containing corresponding OTPs of the node. Each node only transmits data to the links confined in the DGS, and the traffic in each link is determined by the received OTPs. Advantage of this method is it expends least energy in delivering all the packets kept in nodes to the sink. Disadvantages of using this method are, it is particularly suitable for such scenario as environment monitoring in which data are generated in a constant rate and network lifetime fluctuates when remaining energy and energy consumption changes.

C. Minimum spanning tree approach

This paper [3] proposes a Minimum Spanning Tree approach for Cluster and Super Cluster formation. In this approach the energy wastage in distance CH node transmission to the sink node is reduced by having multi hop communication between cluster head node to the sink node and having Super cluster head nodes, which aggregates the information from different cluster heads and transmits it to the sink node.

The Proposed algorithm has three phases. There are cluster formation, cluster head selection and data transmission using shortest path. First phase is based on Minimum Spanning Tree (MST) concept. A tree is a connected graph without cycles. The spanning tree is „minimal‟ when the total length of the edges is the minimum necessary to connect all the vertices in the graph. MST may be constructed using kruskal‟s (or) Prim‟s algorithm. In second phase the newly formed clusters, the node with the highest energy level is selected as the cluster head and the next higher energy level node is selected as the next CH node. To maintain the stability within the clusters, next CH nodes were selected. Once the cluster head are selected, it generates the TDMA schedule for its cluster members and broadcasts to its cluster members. In third phase, in order to reduce further energy wastage due to data transmission between the long distanced Cluster head and sink node, multi-hop data transmission takes place. The data from the nearby cluster heads to the sink node will be directly transmitted to the sink node whereas the data from the distanced cluster head will be transmitted through the shortest multi-hop path. Advantages of using this method are node leader is selected to avoid fault tolerance and delay is avoided by sending packets through, shortest path. Disadvantages of this method are mobility is not considered and network is static.

D. A chain-cluster based routing algorithm

This paper [4] proposes an algorithm called CCM (Chain-Cluster based Mixed routing). It divides a WSN into a few chains and runs in two stages. In the first stage, sensor nodes in each chain transmit data to their own chain head node in parallel, using an improved chain routing protocol. In the second stage, all chain head nodes group as a cluster in a self organized manner, where they transmit fused data to a voted cluster head using the cluster based routing.

Disadvantages of LEACH are 1) The voting process spends non-negligible additional energy and causes serious overhead in network traffic. 2) Cluster heads need to notify member nodes in a broadcast way. Each member node detects the signal strength of different cluster heads and responds to the cluster head with the highest residual energy. This process increases the communication overhead. To reduce the energy overhead for clustering, PEGASIS uses a greedy algorithm that organizes sensor nodes as a chain. Compared with LEACH, PEGASIS avoids the overhead for setting up clusters and uses less energy because the distance of packet transmission is reduced significantly. At the same time, sensor nodes act as the chain head in turn such that the energy load is evenly distributed among the sensor nodes in the network.

CCM exploits the full advantages of LEACH and PEGASIS and to some extent alleviates their weakness, i.e., extra energy overhead for voting cluster heads periodically in LEACH, and long transmission delay in PEGASIS. It has the advantages over both chain based routing and cluster based routing in terms of energy consumption, transmission delay and especially the energy and delay metrics. Disadvantages of using this method are it needs extra energy overhead for voting cluster heads periodically in LEACH and long transmission delay in PEGASIS.

E. Benders decomposition algorithm

This paper [5] consider a mixed-integer linear programming (MILP) model to optimally determine the sink and CH locations as well as the data flow in the network. This model effectively utilizes both the position and the energy-level aspects of the sensors while selecting the CHs and avoids the highest-energy sensors or the sensors that are well-positioned sensors with respect to sinks being selected as CHs repeatedly in successive periods. For the solution of the MILP model, an effective Benders decomposition (BD) approach is developed that incorporates an upper bound heuristic algorithm, strengthened cuts, and an -optimal framework for accelerated convergence.

This approach separates the original formulation into two smaller easier-to-solve problems called a *master problem* and a *sub problem*. The master problem accounts for all the integer variables and the associated portion of the objective function and the constraints of the original problem. It also embodies the information regarding the sub problem portion of the problem via use of an additional (continuous) auxiliary variable and a set of constraints called *Benders cuts*. On the other hand, the sub problem includes all continuous variables and the associated constraints in the original problem. Solving the dual of the sub problem provides information about the sub problem portion of the original objective function, and this information is communicated to the master problem via Benders cuts. Advantages of using this method are it deals with routing problem in cluster based WSNs and it provides a uniform energy consumption profile in the network.. Disadvantages of this method are it is not suitable for time critical applications and the location of the CH is not closer to the sink that takes some time delay.

F. Velocity Energy-efficient and Link-aware Cluster-Tree (VELCT)

This paper [6] proposes a Velocity Energy-efficient and Link-aware Cluster-Tree (VELCT) scheme for data collection in WSNs which would effectively mitigate the problems of coverage distance, mobility, delay, traffic, tree intensity and end-to-end connection.

The VELCT algorithm constructs the Data Collection Tree (DCT) based on the cluster head location. The Data Collection Node (DCN) in the DCT does not participate in sensing on this particular round, however, it simply collects the data packet from the cluster head and delivers it to the sink. The designed VELCT scheme minimizes the energy exploitation, reduces the end-to-end delay and traffic in cluster head in WSNs by effective usage of the DCT. The strength of the VELCT algorithm is to construct a simple tree structure, thereby reducing the energy consumption of the cluster head and avoids frequent cluster formation. It also maintains the cluster for a considerable amount of time.

The VELCT scheme consists of set-up phase and a steady-state phase. In the set-up phase, cluster formation and data collection tree construction is initiated to identify the optimal path between cluster member and sink. It is denoted in intra cluster and DCT communication. Then, the steady-state phase is initiated to transfer the data from the cluster members to the sink. Advantage of using this method is, it exploits the network lifetime, connection time, residual energy, RSSI, throughput, PDR and stable link for mobile sensor nodes and it is adaptable to high mobility environment and provides a better quality communication. Disadvantages are for each Cluster Head is elected for aggregation which takes long transmission to base station and for every time new Dynamic cluster tree is generated for choosing cluster head.

G. Distributed Self-Organization based Clustering Algorithm (DSBCA)

This paper [7] proposes a DSBCA algorithm to generate clusters with more balanced energy and avoid creating excessive clusters with many nodes. The clusters near the base station also forward the data from further clusters (all clusters need to communicate with the base station, but long-distance wireless communication consumes more energy), and as we all know, too many members in a cluster may bring about excessive energy consumption in management and communication. Hence, based on the above concerns, DSBCA algorithm considers the connectivity density and the location of the node, trying to build a more balanced clustering structure.

The basic idea of DSBCA is based on the connectivity density and the distance from the base station to calculate k (clustering radius). The clustering radius is determined by density and distance: if two clusters have the same connectivity density, the cluster much farther from the base station has larger cluster radius; if two clusters have the same distance from the base station, the cluster with the higher density has smaller cluster radius. DSBCA can be divided into three stages: cluster head selecting phase, clusters building phase and cycle phase.

In Cluster head selecting phase the node with the highest weight will become the cluster head. In Cluster building phase the DSBCA sets the threshold of cluster size. The number of cluster nodes cannot exceed the threshold to avoid forming large clusters, which will cause extra overhead and thus reduce network lifetime. In Cycle phase the cluster head gathers the weight of all member nodes, and then selects the node with highest weight as the next head node. In this way, the communication costs are decreased. The reelecting of cluster head occurs in the ‘old’ cluster. Advantage of using this method is it is scalable and works for different network sizes. Disadvantage of this method is it causes communication overhead.

H. A novel differential evolution (DE) based clustering algorithm

In clustering technique, improper formation of clusters can make some CHs overloaded with high number of sensor nodes. This overload may lead to quick death of the CHs and thus partitions the network and thereby degrade the overall performance of the WSN. This paper [8] proposes a novel differential evolution (DE) based clustering algorithm is proposed for WSNs to prolong lifetime of the network by preventing faster death of the highly loaded CHs.

A DE based clustering algorithm for WSNs to prolong network lifetime is used. There are three phase used in this algorithm. In First phase, an efficient vector encoding scheme for complete clustering solution. In second phase, Mathematical derivation of the fitness function for DE based solution. In third phase, Incorporation of a local improvement phase into the traditional DE for faster convergence and better performance of the algorithm.

In Clustering algorithm Network setup is performed in two phases: bootstrapping and clustering. During the bootstrapping process, all the sensor nodes and gateways are assigned unique IDs. Then the sensor nodes broadcast their IDs using CSMA/CA MAC layer protocol. There-fore, the gateways within the communication range of these sensor nodes can collect the sensor IDs and finally send the local net-work information to the base station. Thus for each sensor node, the number of gateways within its communication range can be calculated by the base station. In clustering phase base station executes the clustering algorithm. When the clustering is over, all the gateways provide their IDs to their member sensor nodes by single hop communication. Then the gateways provide a TDMA schedule to their member sensor nodes for intra cluster communication.

© 2015 JCT. All Rights Reserved

Advantage of this method is takes care of energy consumption of both the gate-ways and the sensor nodes. Disadvantages of this method are it is not realistic for the large area networks and multi hop between cluster and base station is not done.

VI. Issues in data collection

There are some other issues in Wireless Sensor Networks. Some of the design challenges are scalability, mobility, bandwidth, power consumption and production cost, reliability and responsiveness. The production cost and power consumption of the multiple sensor nodes will be high and also there is a limited computational power and memory size of each sensor node. As the number of node increases, the overhead of the network gets increases.

By achieving the high responsibility, scalability and reliability can be achieved automatically. By implementing the mobility concepts, the lifetime of the network can be greatly increased. Limited bandwidth results in congestion that affects the normal data exchange.

VII. Objectives

To improve these methods:

It should collect sensing readings without compromising too much data accuracy loss.

It must avoid abundant communication overhead to improve the data quality.

It is to achieve a longer network lifetime of WSNs for data driver applications.

It must conserve the finite resources, such as network bandwidth, energy, and CPU usage.

It should effectively accomplish the ubiquitous temporal-spatial correlation in most natural phenomena for energy efficient data gathering for WSNs.

VIII. Future work

The future work will focus on developing a new node selection technique for data gathering and aggregation. Here instead of fusing data into cluster head a weight based node selection method is introduced to reduce the redundant data transmission and balance the energy cost. Our approach should confront with the difficulties of energy consumption, data collision, and accuracy of data gathering, data redundancy by including several optimization techniques that further decrease the energy consumption and increase the network life time. Later, we will simulating our developed technique and compare it with some protocols to prove its efficiency.

IX. Conclusion

The time delay and energy consumption in data collection is the major issue in wireless sensor network. This paper presents the overview of data gathering methods in WSN and the performance is based on the statistical parameters such as energy consumption, packet collision, retransmission, and time delay are discussed clearly. Most of the fundamental issues regarding data gathering in WSN are explained. Further, this paper will help the researcher to invent novel methods in order to collect the data with minimum time delay and energy consumption in wireless senor network.

References

[1] Xin Guan, L. Guan, X.G.Wang and Tomoaki Ohtsuki, “A new load balancing and data collection algorithm for energy saving in wireless sensor networks,”Springer Science+Business Media, pp. 313-322, 2010.

[2] Yi-hua Zhu, Wan-deng Wu, Jian Pan and Yi-ping Tang, “An energy-efficient data gathering algorithm to prolong lifetime of wireless sensor networks” Elsevier-Computer Communications, pp.639-647, 2010.

[3] B. Baranidharan and B. Shanthi , “An Energy Efficient Clustering Protocol Using Minimum Spanning Tree for Wireless Sensor Networks,” springer- Verlag Berlin Heidelberg, pp.1-11, 2011.

[4] Feilong Tang, Ilsun You, Song Guo and Minyi Guo Yonggong Ma, “A chain-cluster based routing algorithm for wireless sensor networks,” Springer, pp. 1305-1313, 2012.

[5] Hui Lin and Halit Üster, “Exact and Heuristic Algorithms for Data-Gathering Cluster-Based Wireless Sensor Network,” IEEE/ACM TRANSACTIONS ON NETWORKING,2013.

[6] Velmani Ramasamy, Kaarthick Balakrishnan, “An Efficient Cluster-Tree based Data Collection Scheme for Large Mobile Wireless Sensor Networks,” IEEE SENSORS JOURNAL, 2013.

[7] Ying Liao, Huan Qi, and Weiqun Li, “Load-Balanced Clustering Algorithm With Distributed Self-Organization for Wireless Sensor Networks,” IEEE SENSORS JOURNAL, VOL. 13, NO. 5, 2013.

[8] Pratyay Kuila and Prasanta K. Jana, “A novel differential evolution based clustering algorithm for wireless sensor networks,” Elsevier-Applied Soft Computing, 2014.

[9] Efe Karasabun, Ibrahim Korpeoglu, Cevdet Aykanat, “Active node determination for correlated data gathering in wireless sensor networks,” Elsevier - Computer Networks, pp. 1124–1138, 2012.

[10] Aubin Jarry A Pierre Leone A, Sotiris Nikoletseas B, Jose Rolim A (2010), ‘Optimal data gathering paths and energy-balance mechanisms in wireless networks’, Elsevier – Adhoc Networks., pp.1036-1048.