Energy Efficient Data Transmission with Tree Clustering Algorithm

Sonali khose
Department of Electronics & Telecommunication
G.H. Raisoni College of Engineering & Management
Savitribai Phule Pune University
Pune, Maharashtra, India

Vijay Joshi
Department of Electronics & Telecommunication
G.H. Raisoni College of Engineering & Management
Savitribai Phule Pune University
Pune, Maharashtra, India

ABSTRACT
Wireless sensor networks (WSNs) collects a amount of sensory data. WSN networks can not avoid the problems like unbalanced energy consumption. As in various data gathering schemes a hotspot problem occurs which is responsible for this unequal energy consumption. In this energy efficient data gathering scheme we constructs the tree with n number of sensor nodes and depending upon its weight we construct a tree cluster for data gathering. It balances the network load and improves network lifetime.

Keywords
WSN, WSNs, DSNs

1. INTRODUCTION
WSN networks are widely applicable in different outdoor and military areas. A WSN is consists of large number of small sensor nodes which are of limited storage and low power. Most of the energy of these sensor nodes are utilised during data collection phase while collecting huge amount of data. As sensor nodes nearer to the base station are receiving as well as transmitting the data to the whole network hence consumes more energy which causes the non-uniform energy consumption within a network. That means transceiver consume more energy than any other component of a sensor node. It is being great challenge to design an energy efficient routing scheme that prolongs the network lifetime.

Initially in a network all tiny sensor nodes are deployed with same amount of energy weight at random location. Clustering is carried out according to the location of each node. These clusters are then connected for the data gathering within a network.

2. RELATED WORK
Literature survey on data gathering algorithm gives a brief idea of different communication techniques. Those techniques are preferred according to the area of network range. BS can directly collects the information from sensor nodes is nothing but a single hop communication, while BS collects information from sensory node via various subset of a network is called multi hop communication. In multi hop communication BS only visit the subset of nodes which have already gathered data from nodes routed towards them.

While designing any energy efficient network various issues needs to be operated are data transmission latency, energy spend while transmission, end to end delay etc. Collection of data is also categorized according to the requirement i.e. in direct data collection only interested data is requested to collect. In full data collection technique the whole data of a network is collected which further decreases the network lifetime.

To make our categories more clearly and well understood, we describe the related literatures in the following three classifications: direct communicating collection, local data collection, and full data collection.[1]

A. *Direct data collection:*
BS collects data from each node when it comes in it's communication range by only single hop communication.it reduces the network lifetime. Latency of data gathering can also be reduced by accepting shortest path communication scheme. Hence a proper scheduling algorithms are required to reduce the data gathering latency and for that BS is required to collect data from each node when it is in it's communication range. This type of single hop communication is impractical in some aspects hence; most of the time multi hop communication is favourable.

B. *Local data collection:*
This communication technique follows an agent based approach. One or more number of agents are actually tracking the location of BS and construct a path for data collection towards root node.

C. *Full Data Collection:*
In full data collection scheme a BS requests for collection of data of a whole network, which further decreases the network lifetime. Also it causes congestion while data gathering.it is further follows following approaches while data collection.

Root point selection: Root point is most weighted sensor node in the network which is responsible data transmission to the BS. It collects the data from each cluster via communicating with each cluster head. This approach is more robust and will never cause the whole network to fail due to failure of any cluster node.

Cluster formation: It is based on multi hop clustering technique. It make use of cluster head only for routing a data, hence chances of depletion of energy of CHs is more than other sensor nodes within a network due to heavy load of transmission.

Area based data collection: BS follows different routes or paths while collecting data in a network. It monitors data in circular manner which is repeated for n number of times. In each repetition BS calculates the energy consumption of each and every node. Unbalanced energy consumption can be reduced by defining a circular area while tracking a path. A radius of that circle is to be known. A circle is always moving along the network is called as buffer area. The buffer area is dealing with the Maximum data transmission within a network. Also the radius o f a moving circle should be known, so that sensor nodes selected for a buffer area for specific round will deliver their data to the BS.

3. OVERVIEW

In our proposed work, BS starts collecting data from each subset node which is itself gathering information data from its child nodes which are located at one hop distance from the root node. Our whole network design will follow few steps like selecting a root node, designing a tree cluster based network and finally collection of information data from various distributed sensor nodes within a network for balancing a load of a network.

The selection of root nodes totally depends upon its weight of residual energy. Some levels we are going to add in our network which introduces some subsets of nodes in between root node and leaf nodes at one hope distance. Depending upon the load of transmission, number of levels are required to add to the network will be decided.

A. *Selection of root point*

Root points are required to be selected to avoid hotspot problem which introduces non uniform energy consumption. Root point is supposed to be chosen such that it reduces the number of stop points in the network and distance from BS will also be reduced. In order to pursue a trade-off between the energy saving and data gathering latency, which achieves a balance between the relay hop count for local data aggregation and the moving tour length of the mobile sink, the relay hop count 'd' should be constrained to a small level (2 or 3) to limit the energy consumption at sensors.[2]. Hence we are only dealing with the calculation upto two hop distance far nodes.

Initially each node is supposed to calculate the distance from it's two hop neighbour as well as it's weight. Each node broadcasts a message with it's identity value, its residual energy as well as its time to live value. The node which receives this broadcast message further forward it to its neighbour nodes until the time to live value reaches to zero. The number of neighbour nodes of each sensor node is calculated depending upon how many times a node receives this broadcast message and also its time to live value.

Residual energy of one hop neighbour of any node can be calculated by the following equation,

$$E_i^{1-hop} = \left(\sum_{1}^{N_i^{1-hop}} E_r\right) / N_i^{1-hop}$$

Weight of each node can be calculated by,

$$W_i = N_i^{2-hop} \times E_i \times E_i^{1-hop} / D_i$$

Residual energy is denoted by Ei .

Distance from the BS is denoted by Di .

By exchanging weight information derived by the equation above with its one-hop neighbours, every node maintains a list of weights of its one-hop neighbors. By comparing its own weight with its one-hop neighbours, every node selects its appropriate parent node respectively. If neighbour nodes is having weight bigger than node *i*, then node *i* makes the node having the maximum weight as its parent node, and node *i* itself is the child node. Here, if situation occurs where decision to be made for more than one number of nodes to be a parent node at that time selection is carried out depending upon the identity number of nodes.

If node i is not having any parent node, and the maximum weight value of node i is highest than them, then node i is an isolated node. To avoid such situation, node i still makes the neighbour node with the highest weight as its parent node as shown in figure 2.

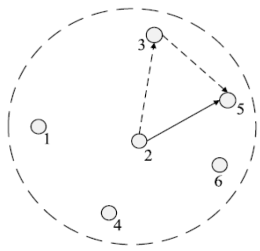

Figure 1 Selecting parent node with a bigger ID number

The selection of parent node 1 which is having maximum weight in the network will be based on maximum weight of its one hop neighbour even if it is less than node1.

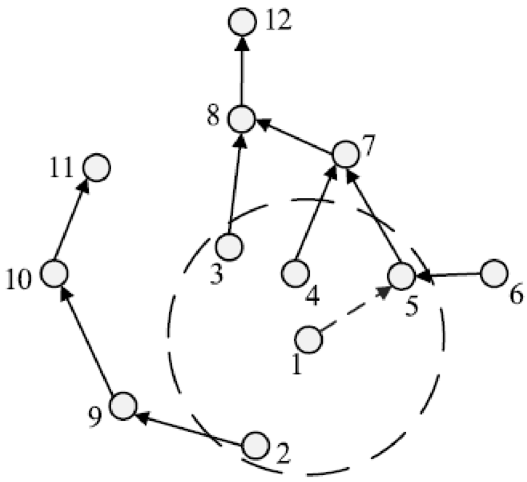

Figure 2 Selecting proper parent nodes

B. *Identification of leaf point:*

The sensor nodes can be divided into three categories: root nodes, leaf nodes, and subset nodes. When node is having only child node and no parent node, it's a root node. A node with no child node and only parent node is a leaf node. Remaining are subset nodes. Eventually, each tree only has one root node, and every node except root node on the tree has only one parent node respectively [1].

4. DATA UPLOADING, STORING AND DATA COLLECTION

- Data uploading and storing

After all root points and subset nodes are selected, the mobile sink begins to perform its data gathering. When mobile sink moves to a root points or subset nodes, it polls each sensor within its transmission range. The sensors that receive the polling messages upload packets directly to the mobile sink via single-hop transmissions instead of sending large amount of data they received to the corresponding Root points or subset nodes, which reduces the burden of Root points and subset nodes significantly. Thus, child nodes in the one-hop range of root points or subset nodes need to be identified before data gathering process. After that, other nodes only need to upload the packets to its corresponding parent node. So all packets will be forwarded in the direction of the corresponding root points or subset nodes until being received and stored by the one-hop child nodes of Root points or subset nodes.

Hence, root points and subset nodes only need to send a notification to their one-hop child nodes to inform them to be data storage nodes (DSNs) at first. When packets from child nodes have been forwarded to these DSNs, the data uploading process stops. Other nodes except these DSNs, may be two or more hops away from their corresponding root points or subset nodes, transmit their data to their corresponding parent nodes respectively. Obviously, all packets are stored by these DSNs and then can be uploaded to the mobile sink when it arrives.

To store all data of a tree in these DSNs, we can not only take advantage of the high residual energy of them, but also reduce the heavy load of the root points or subset nodes caused by receiving too large amount of data from the whole tree.

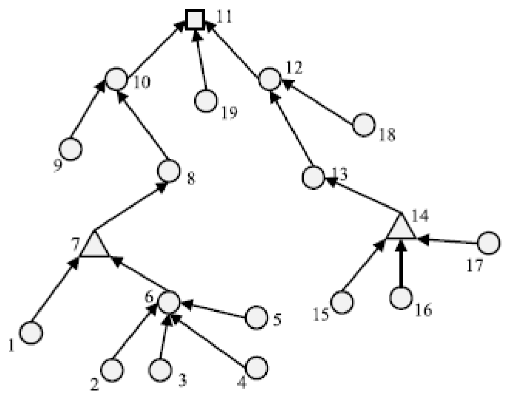

Figure 3 Tree Structure

- **Data collection**

A data collection stage consists of many data collection round. After each data collection, the mobile sink decides whether to reconstruct the whole network trees based on some conditions

In each data collection round, the mobile sink needs to move in the network deployment area periodically to perform data gathering. The selected root points and subset nodes report their location information to the BS firstly. After getting these location information from BS, mobile sink invokes an approximation algorithm for the TSP (Traveling Salesman Problem) to plan an approximately optimal route for these Root points and subset nodes. It then moves to these nodes one by one to gather data from sensors in the prox- imity via single-hop transmissions. When arriving at these root points or subset nodes, the mobile sink firstly sends a polling message to these DSNs and root points or subset nodes in its one-hop range, stops at these Root points and subset nodes for data gathering, and then moves to another position until all data of the tree have been uploaded. In this manner, the mobile sink visits all the trees and finishes its one data collection round.

The energy consumption of DSNs significantly influences the network lifetime, since they have almost the heaviest load of their corresponding trees. Hence in each data collection round, after all data of the corresponding tree of the root points or subset nodes is uploaded to the mobile sink, the root points or subset nodes retrieve the residual energy of their one-hop child nodes and then decide whether to send a request of performing tree reconstruction to the mobile sink. The mobile sink calculates whether to perform tree reconstruction according to the ratio of the accumulate value of requests to the number of Root points and subset nodes. The details are as follows:

1) In each data collection round, after all data of the corresponding tree is received by the mobile sink, the corresponding root points or subset nodes retrieves the residual energy of its one-hop child nodes. If at least one of these nodes has residual energy below its initial energy recorded in the first data collection round after the tree constructing or reconstructing, then the root points and subset nodes needs to send a request of tree reconstruction to the mobile sink before it moves to next position.

2) After these Root points and subset nodes are visited by the mobile sink the mobile sink calculates the ratio of the accumulate value of requests to the number of root points and subset nodes before the next tree reconstruction stage. If it is higher than *n*% i.e. threshold value, then the tree reconstruction is performed and new root points and subset nodes are selected as the stop points of the mobile sink for data gathering. Otherwise, in the next round the mobile sink still traverses the original Root points and subset nodes of last round for data collection.

3) Every time after tree reconstruction, the new Root points and subset nodes need to retrieve the residual energy of its one-hop child nodes and record it in the first data collection round. Then the value is a benchmark for these root points and subset nodes to make a decision about whether to send a tree reconstruction request to the mobile sink after each data collection round. And the mobile sink continues to calculate whether to perform a tree reconstruction based on the value of *m* and *n* i.e. residual energy and threshold value respectively.

From the procedures above, we believe that there is a trade-off between the value of *m* and *n* and the cost of tree reconstructing. If *m* is too large, which means root points or subset nodes need to send a request when the residual energy of the DSNs is still very high, in order to reduce the cost of frequently tree reconstruction, *n* needs to be set with a large value, which implies that the mobile sink performs tree reconstruction when it receive a large proportion of requests. On the contrary, when *m* is set too small, which means the residual energy of the DSNs is very low, in order to avoid the fast depletion of them, *n* should be constrained to a small value to make the mobile sink perform tree reconstruction as soon as possible, which means to sacrifice the cost of tree reconstruction to prolong the whole network lifetime. We validate this by simulations in the next section.

5. SIMULATION AND RESULTS

First the node formed to transmit and also receive the packets. Here, set more number of nodes for clustering and sending and receiving the packets from sender to receiver. Then 50 nodes are formed for processing. The node formation is the first step of our process in which nodes are added in to the network. The nodes are in mobile nature and are free to move.

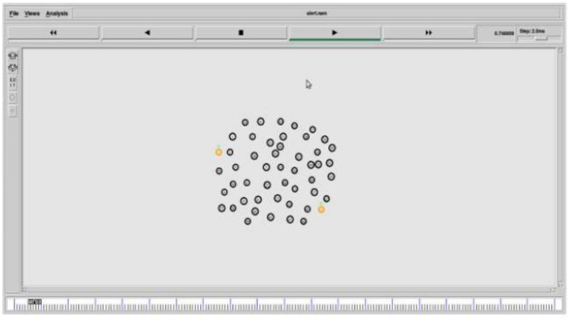

Figure 4 Node Formation

After selection of source and destination topology formation is constructed. Topology formation means that the each node will send their information to the neighbour nodes.. In this stage we predict the location of each node.

Then create the mobile node for the purpose of packet transmission. Nodes are used to send the packet and make communication. After node formation then we select source and destination to make packet transmission. Then after we construct topology to knows the neighbour information. Each node sends the message to its neighbour to know node details. This topology formation mainly used to find the node location for packet transmission.

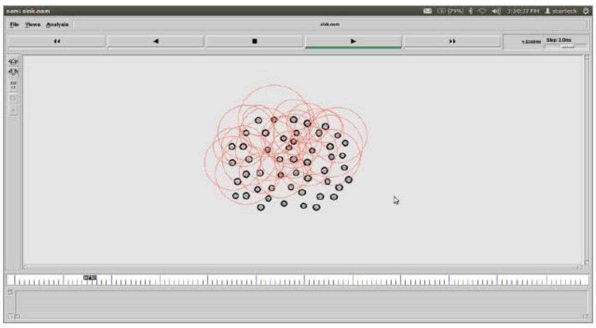

Figure 5 Topology Creation

A comparison of the residual energy of proposed algorithm and LEACH is shown in Figure 6. We can find that energy consumption decreases with round (or time). The rate of decrease of LEACH is much faster than that of our proposed algorithm with clustering technique. The energy consumption rate is similar up to some rounds. After that, the energy consumption of our proposed algorithm is much less than that of LEACH. Consequently, the energy in LEACH network gets drained away much earlier, and our proposed routing algorithm outperforms LEACH.

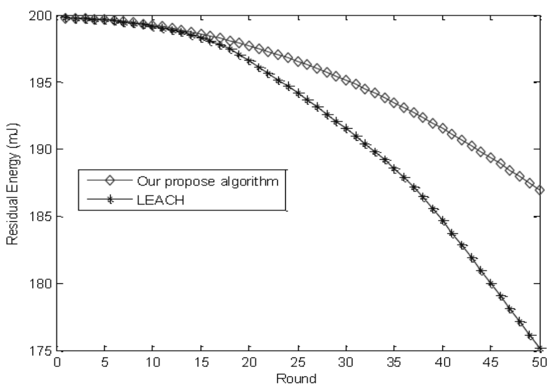

Figure 6 Residual Energy Comparison

6. CONCLUSION

In order to alleviate the hot spot problem and further balance the energy consumption, we use mobile sink nodes for data collection. We first compare the residual energy of sensor network respectively using our proposed routing algorithm and LEACH. Based on the comparison and analysis, we propose a tree-cluster based data gathering algorithm with a mobile sink for WSNs. We also introduce an adjusting method for the root points and subset nodes, in order to avoid the heavy burden around them.

We are comparing our results with LEACH with some other network parameters like throughput, energy spent and end to end delay.

We aim to improve the energy efficiency in future by increasing the number of mobile sink nodes which will further improves the overall performance of a WSN.

7. REFERENCES

[1] Kyung Tae Kim, Man Youn Kim, Ji Hyeon Choi, Hee Yong Youn,"An Energy Efficient and Optimal Randomized Clustering for Wireless Sensor Networks" IEEE Trans. College of information and Communication Engineering Sungkyunkwan University.June 2015.

[2] M. Zhao and Y. Yang, ``Bounded relay hop mobile data gathering in wireless sensor networks,'' IEEE Trans. Comput., vol. 61, no. 2,pp. 265 277, Feb. 2012.

[3] CHUAN ZHU1, SHUAI WU1, GUANGJIE HAN1, (Member, IEEE), LEI SHU2, (Member, IEEE),AND HONGYI WU3, (Member, IEEE) "A Tree-Cluster-Based Data-Gathering Algorithm for Industrial WSNs With a Mobile Sink" IEEE. Trans. 2169-3536, volume 3 ,May 2015.

[4] DAISUKE TAKAISHI1, (Student Member, IEEE), HIROKI NISHIYAMA1, (Senior Member, IEEE), NEI KATO1, (Fellow, IEEE), AND RYU MIURA2" Toward Energy Efficient Big Data Gathering in Densely Distributed Sensor Networks"Wireless Network Research Institute, National Institute of Information and Communications Technology, 2168-6750 IEEE. Trans., volume 2, no. 3,september 2014.

[5] JUN LONG1,MIANXIONG DONG2, KAORU OTA2, ANFENG LIU1,AND SONGYUAN HAI "Reliability Guaranteed Efficient Data gathering in Wireless Sensor Networks", IEEE. Trans. volume 3, 2169-3536 ,7 May 2015.

[6] www.wikipedia.org/wiki/Wireless sensor network