An n-Level Optimal Spanning Tree Based Routing Approach for Wireless Sensor Networks

Harendra S. Jangwan , Ashish Negi

Abstract— **Energy conservation is most important issue prevailing in various real world systems. Wireless sensor networks (WSNs) also suffer with this difficulty as the sensor nodes deployed in the regions being monitored cannot be re-energized. Thus, it is required to attain higher performance of systems while lowering the cost in terms of energy. We have proposed new n-Level Tree-based Routing Approach (NLTRA) for data routing for homogenous as well as heterogeneous WSNs in order to meet the system requirements. Mathematical modeling has been done for the distribution of residual energy among the sensor nodes of the network using normal distribution. A graph theory based approach is employed for the representation of network as a set of nodes and edges. A tree based path is determined for data routing in the network. The hierarchy is maintained on the basis of energy levels of sensor nodes and the optimal spanning tree (OST) is obtained.**

Index Terms—**Optimal spanning tree, data routing, NLTRA.**

I. Introduction

Recent advances in wireless sensor network technologies have enabled us to extract reliable information from real world and physically inaccessible environments. Sensors are tiny low-power electronic devices that sense or gather some specific information of interest from the surrounding environment and send it to base station (BS) through wireless medium in the form of electrical signals. The BS is a control point that monitors and controls the functioning of sensor network. The sensor nodes are deployed randomly in a particular region that has to be monitored [1]. The region can be a two dimensional area or a three dimensional space what so ever is required by the application. The BS may be situated inside as well as outside the network field. The nodes collectively form a network that are connected through a wireless medium, hence the name wireless sensor network. The process of transmitting the sensed data by sensor nodes to BS is referred to as routing. A sensor node has several tasks to perform i.e. sensing data, data aggregation, transmission and receiving data. To perform all these tasks, the nodes consume energy form the batteries mounted on them. But the sensor nods are generally deployed at random in the regions which are nearly physically inaccessible to humans, it makes this impossible to re-energize the sensor nodes after their energy gets depleted and the nodes are said to die out. Hence, there arises a need of developing such routing protocols that are energy efficient i.e. the protocols that minimize the energy consumption of sensor nodes in the whole network and maximize the network lifetime. There are also certain issues faced by WSNs, for instance, if the sensed data is transmitted directly to BS by each and every sensor node (direct transmission), it will lead to more energy consumption as the total distance of transmission is increased [2]. If the data is transmitted via other nodes, it will also lead to inefficient utilization of energy as the number of transmissions and receiving operations are increased. So, certain routing protocols must be developed which optimize the cost of routing process in terms of energy utilization. One way to achieve this is through clustering. LEACH was developed as a hierarchical clustering protocol that divides the whole network into clusters. Each cluster having a cluster head (CH) that performs the task of aggregating data collected by member nodes of its cluster and sending it to the BS. Various other protocols such as HEED and PEGASIS were also developed as modifications of LEACH to improve energy efficiency and network lifetime of WSNs. Data aggregation helps in removal of similar type of data gathered by the sensor nodes. The data is sent in the form of packets and size of packet is defined by number of bits. The energy consumption during data routing depends on the size of data and distance between communicating nodes. A sensor node consists of an amplifier that increases the signal strength to reach up to longer distances. A major issue faced by WSN is load balancing. In a node-to-node transmission, the nodes receiving more data are more prone to die out faster as compared to other nodes of the network. Therefore, a routing protocol should be designed in such a way that it balances the energy load among the various nodes of the network. If some nodes die out early, it will lead to coverage and connectivity problems within the network. An efficient working of a WSN depends on the fact that the maximum area of the sensing region should be covered for sensing to achieve better interpretations of the phenomenon occurring into that region. Also, if some intermediate nods die, there may be a chance that the connectivity between the other nodes is lost and this will end up in loss of important data as well. Therefore, the protocol is expected to be fault tolerant i.e. if due to some unexpected reasons, some nodes stop functioning at some point of time, the overall working of the network should not be affected.

Heinzelman et al. [3] developed protocol for wireless micro sensor networks known as low-energy adaptive clustering hierarchy (LEACH). This emerged as a combination of an approach of energy-efficient for energy

Manuscript received September, 2017*.*
Harendra S. Jangwan, Dept. of Computer Science & Application, G.B. Pant Engineering College, Pauri Garhwal, India.
Ashish Negi, Dept. of Computer Science & Application, G.B. Pant Engineering College, Pauri Garhwal, India.

efficiency and media access. Data aggregation task was assumed to be specified according to the application. LEACH outperformed in terms of latency and overall system lifetime. LEACH was based on the technique of cluster formation in a distributed manner in which the sensor nodes of the network organize themselves to form clusters. The cluster head role is rotated among the nodes of clusters in order to achieve even distribution of energy load among all the nodes of that cluster which in turn results in conservation of energy resources while communication. Their results showed that the approach improved improve overall as compared to simple multi-hop approaches.

In [4] Khan et al. Tried to minimize the data transmission energy along with energy dissipated in data aggregation and fusion. This was achieved by minimizing transmission energy by creating an MST using all the sensor nodes of the network. The node on the top tier which has the highest energy then transfers the data in aggregated form to the BS. If the node dies out then the second highest energy node from top tier is chosen. The BS is located at the centre of the network field. The network topology doesn't change throughout the rounds of communication. Results show that this technique outperforms LEACH in view of overall network lifetime.

In order to achieve energy efficiency and reduce network delay, Seelam et al. [5] proposed a new cluster-based routing protocol by combining the ideas of MST, LEACH and Clustering with One-Time Setup (COTS). Distance Energy-based MST (DEMST) technique was adopted for communication among Cluster heads (CH) of different clusters and for effective channel utilization; a new approach was introduced known as pipelining technique. The proposed work was simulated and compared with existing protocols such as LEACH, COTS and some clustering techniques based on MST.

Saravanan and Madheswaran [6] proposed Bee Algorithm-Simulated Annealing Weighted Minimal Spanning Tree (BASA-WMST) which is a hybrid evolutionary algorithm for data routing in WSNs. In this approach there is random deployment of sensor nodes in the field. The nodes are divided into clusters and the best possible number of clusters is estimated along with the optimal route of data. CHs transmit data to BS and Weighted MST technique is employed is for determining the shortest path between the member nodes of the cluster. The weights of the edges of the network are not constant and change according to the energy levels of sensor nodes and the tree is optimized using bee algorithm simulated annealing algorithm proposed by the authors.

Djenouri [7] introduced an approach based on the evaluation of spanning tree of the network viewed as edge weighted graph. The data sent to each node is proportional to its so called effective energy. This paper considered the availability of green energy resources form environment. The intermediate nodes that fail to meet the definition of survivability are accompanied by relay nodes (RNs). Integer programming optimization technique is used to evaluate a lower-bound of the model. The technique is simulated using a simple model for the resulting solution and comparisons are made with the single solution from some existing work. Simulation results show that employing lower number of relay nodes ensure the survivability of intermediate nodes of the spanning tree of the network.

Vijay and Gupta [8] presented MST based clustering along with divide and conquer approach. The MST based clustering was first proposed in 1970's for large databases. While the approach of divide and conquer is employed in WSNs keeping the constraints as concerned with the network model. The paper emphasises on conserving the computational energy by not computing complete MST before cluster formation. The edges belonging to a part of MST are determined and clusters are formed. The edges that can be removed from the network are determined on the basis of certain parameters.

Based on improved Kruskal's algorithms, energy efficient and security-high routing (EEASHR) protocol was introduced by Zeng and Jiang [9]. This scheme was proposed as an effort for solving the problems relating to information overlap, low utilization rate and network transmission security in routing protocols of WSNs. In this work, the weight of an edge is determined on the basis of reliability between the nodes and amount of energy required by the transmitting node for data transmission. Improved Kruskal's algorithm is applied to determine the optimal path between the nodes for data routing to the sink in the form of MST. The NS2 simulation results represent that this schemes provides better results in terms for energy efficiency, reduction in packet data loss rate and increasing network lifetime.

A survey of existing literature in the field of cluster based routing protocols for WSNs was done by Baranidharan and Shanthi [10] together with an overview and analysis of factors that affect clustering and their significance. As an extension, they proposed another clustering technique by determining shortest path for data routing using MST. They also discussed pros and cons for the proposed method.

Yamuna and Elakkiya [11] considered the vertex and edge connectivity attributes of networks for determining an appropriate weighted MST when there is a fault occurrence in the system.

For multi-criteria degree-constrained minimum spanning tree problem (mcd-MST), Guo et al [12] introduced a particle swarm optimization approach which is an improved PSO. The aim of the paper was to improve the underlying factors affecting WSNs such as provisioning of QoS, energy consumption by sensor nodes, reliability of the system etc. The principles of mutation and crossover operator in the genetic algorithm (GA) are incorporated into the proposed PSO algorithm to achieve a better diversity and break away from local optima. Comparison of results are done with enumeration method and the results after simulation results represent the efficiency of the proposed scheme and better solutions the problem of mcd-MST are obtained.

Gagarin [13] discussed a method for determining minimum weight and balanced MST for data routing in the network. This work employed the distributed searching methods by cluster based hierarchy and modified kruskal's algorithm for determining the MST. The MST was formed with greater diameter and the maximum degree of nodes was lessened so that the consumption of energy in communication can be decreased. This approach can be employed to both parallel and distributed algorithms. Simulations based on transmission energy matrix show that the network lifetime was increased three to four times for the same transmission energy.

Bagaa et al. [14] considered energy harvesting wireless sensor networks (EH-WSN) and introduced an approach to prolong network lifetime. Data aggregation in the network is also considered and an aggregation tree is constructed to achieve the goal. This helps in reducing the work load on energy non-harvesting nodes to some extent. The methods of integer programming and MST are employed that reduce the runtime complexity of the system. Simulation results show that the design goals are achieved and are feasible.

The work presented in this paper is based on the concepts of normal distribution and minimum spanning tree of the graph. A tree based structure is determined for data routing in the network where the BS is considered to be a root node. As the sensed data is aggregated at each step (level), this approach is best suited when there is a large amount of redundancy in data. To balance the energy load among the various nodes, their residual energy is considered as a parameter to determine how much work load should be assigned to that particular node in the current round. The sensor nodes are assumed to have different energy levels and the energy graph is expected to follow normal curve. The following section deals with the proposed methodology in detail.

II. Procedure for Paper Submission

This section discusses in a brief some of the preliminary concepts in relevance to our proposed approach:

A. Minimum Spanning Tree

A network can be represented by a graph i.e. a set of nodes and edges G=(N, E), where N is the set of nodes and E is the set of edges. Each edge in E connects any two vertices in E [15]. The length of an edge specifies its weight as shown in the Fig. 1. A minimum spanning tree T of a graph G is a connected acyclic sub-graph of G [15]. The main objective is to cover all the nodes in G such that the total weight of all the edges in T is least. The number of edges in a tree is one less than the number of nodes. The MST of a graph or network can be obtained by first deleting all the edges in G and then adding them one by one starting from the minimum weight edge such that no cycle is formed until all the nodes are connected. This method for determining MST is known as Kruskal's algorithm.

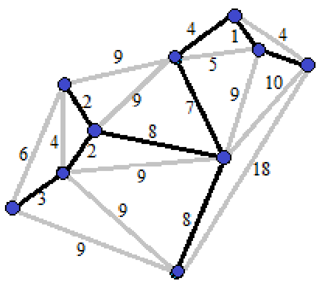

Fig. 1 Minimum spanning tree of a graph

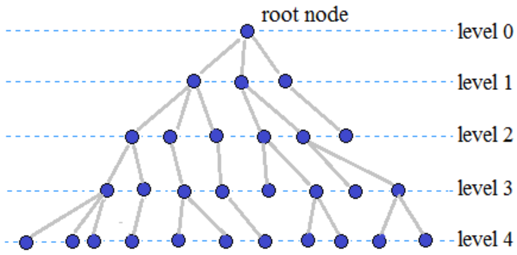

Fig. 2 Level wise representation of a tree

The level wise representation of a tree is shown in Fig. 2. Node at the highest level is known as root node and the ones at the lowest level are termed as pendent nodes. For two connected nodes, the higher level node is called as parent node, while the other as child node. The information can travel via intermediate nodes of a tree either from top to bottom or bottom to top as required by the application.

B. Normal Distribution

In statistical theory and data analysis, normal distribution plays a very important role due to of several reasons. Most of the sampling distributions tend to normality for large samples [16]. Even if the data is not normally distributed, it can be brought to normal form by simple transformation of variables. Normal curves are best suited for large sampling distributions existing in nature. In order to fit normal distribution, first the mean (μ) and standard deviation (σ) of the given data are calculated using $\mu = \frac{\sum x_i}{N}$ and $\sigma = \sqrt{\frac{\sum(x_i-\mu)^2}{N}}$ where N is the number of entities and x_i is the attribute of i^{th} entity i.e. i =1,2,3,4…..N. The normal curve fitted to the given data is given by following function

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\left(\frac{x-\mu}{\sigma}\right)^2} \quad (1)$$

where, $z = \frac{X-\mu}{\sigma}$ is termed as standard normal variate [16]. On the basis of the value of the z, the area under curve can be determined accordingly. Such as percentage area under curve for some particular z-values: z = ±1, ±2, ±3 are 68.23, 95.44 and 99.73 respectively, though for the other values of z, the area under curve can be obtained from normal distribution table.

The mean, median and mode of normal distribution are coincident. The graph of this distribution is represented with help of Fig. 3.

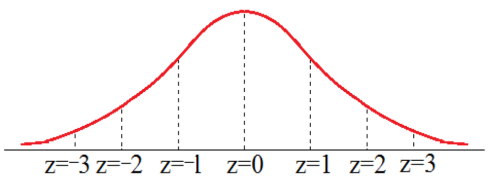

Fig. 3 Graphical representation of normal distribution

We can see from the graph that the maximum number of data values lie around the mean value μ.

C. Radio Model

The first order energy model is considered as the radio energy dissipation model. In a tree like network structure, the distance between the parent node and child nodes is smaller as compared to the distance between the nodes and the base station as in direct transmission as well as distances between any arbitrary nodes. Hence, there is no need to consider any threshold distance as considered in [2]. The free-space model for energy dissipation considered here is represented with the help of Fig. 4.

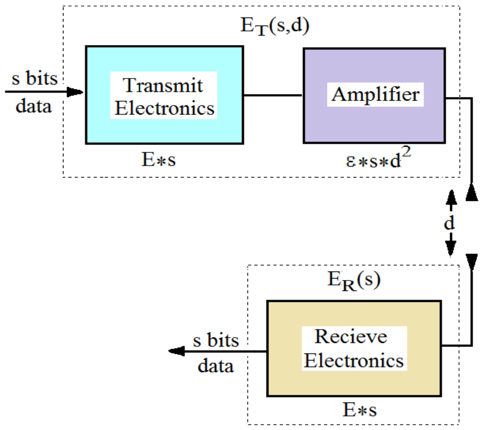

Fig. 4 First order radio energy model

The radio energy model of a sensor node represents the energy dissipated in transmitting the data by one node and receiving it by the other node. The energy of transmitter and receiver electronics is same but a transmitter has to amplify a signal also so that it may reach the other node or BS. The amplification task depends on the distance between the nodes in communication. Size of a data packet is determined by the number of bits in the message. The symbols and notations used in the model are summarized here:

NOMENCLATURE

s	size of data/length of message in bits	
d	distance between transmitting and receiving node	m
E	energy of transmitter/receiver electronics	nJ
ε	energy of amplifier	pJ
E_T(s,d)	energy dissipated in transmitting ‘s’ sized data through a distance ‘d’	nJ
E_R(s)	energy dissipated in receiving ‘s’ sized data	nJ

The mathematical equations representing the above process of communication can be written as:

$$E_T(s,d) = E \times s + \varepsilon \times s \times d^2 \quad (2)$$

$$E_R(s) = E \times s \quad (3)$$

D. Network Model

The sensor nodes of a WSN are limited in energy but the energy of BS is limitless. However, mostly some nodes are not directly reachable by the BS due to limited transmission range. Hence a routing protocol for sending the data via other intermediate nodes is required. The network model considered here comprises of sensor nodes that are randomly deployed into the region having random initial energies also. This makes the network heterogeneous in terms of energies. Fig. 5 shows the structure of a network area and the base station.

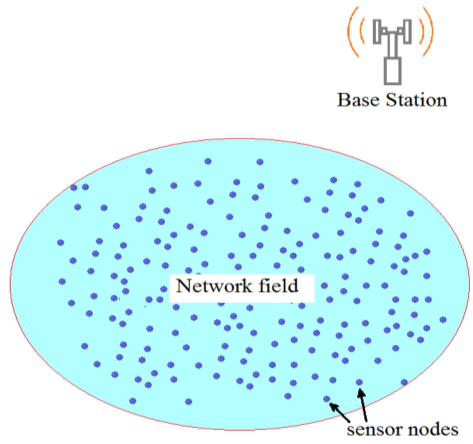

Fig. 5 Wireless sensor netwok

III. PROPOSED SCHEME

To improve the energy efficiency of a WSN an MST based routing protocol is proposed here. Each routing protocol for WSNs has some advantages as well as disadvantages of their own and is opted on the basis of the network topology and application. Tree based routing has main advantage of reduction in redundancy of data through data aggregation.

This scheme also emphasises on efficient network coverage of network field, as it utilizes the concept of residual energy of sensor nodes. After each round of communication, the sensor nodes are left with different amounts of energy. So they are not assigned the same role in different rounds. The work load is distributed in such a way that the nodes with less amount of residual energy have to perform lesser amount of work. In other words, an effort has been done to balance the average work done by each sensor node per round. In this manner, each node is expected to remain alive for a longer time and when the nodes death starts, they die out at nearly same time.

In some of the network applications, the sensed data is not that much useful if it is taken from only some parts of the region and not the whole region under study. The main aim is to keep the network region under coverage as much as possible for longer periods of time.

Following assumptions are made for the proposed scheme:

- the network is heterogeneous in view of energy
- the sensor nodes are stationary.
- region of deployment can be two or three dimensional
- sensor nodes always remain within the region
- nodes are homogenous in view of architecture
- transmission and receiving power of each node is equal
- nodes can get faulty due to technical or environmental issues.

Routing of data in the proposed scheme is done by following a tree like path. Various MST based routing schemes have been proposed earlier by authors in which the network is considered to be a weighted graph where the weights of the edges represent the distance between the sensor nodes of the network. In the proposed work, the spanning tree of the network is formed on the basis of residual energy of the sensor nodes

The sensor nodes are considered to be heterogeneous in nature, i.e., the initial energy levels of the nodes are not same and are considered to be random. Also, the nodes are deployed randomly into the network field. Data sensing is done at the same rate by each sensor node in the network. An MST based approach for data routing is best suited when there exist a lot of redundancy in data. This technique of MST based clustering evolved in 1970's for accessing large databases. In order to route the data back to the BS, the approach of following a tree like path is employed. As energy distribution among the nodes is not same and is assumed to be random, we have considered that the normal distribution is best suited for such a type of situation.

The aim is to construct an n-level routed tree which spans the network, on the basis of residual energy of sensor nodes. The root node of the optimal tree represents the root node of the tree and the nodes with less amount of energy are taken to be at the lower levels of the tree. The number of levels in the tree is user dependent. Data is routed from lower level nodes to higher level nodes and finally to the root node, i.e., the BS.

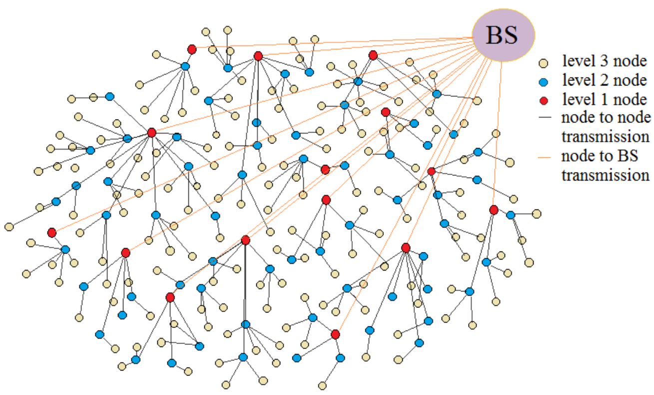

Fig. 6 MST based routing path

The proposed approach for routing is represented in Fig. 6, and termed as Optimal Spanning Tree(OST), which is a three level tree (n=3). The data is directed towards the BS starting from the lowest level nodes and passing through the intermediate level nodes. The nodes are divided into three groups according to their energy levels.

We have considered here that the energy levels of sensor nodes are random and can be modeled with the help of normal distribution due to the fact that every natural data in large amount tends to normality. The number of groups in which the nodes are divided is not restricted to some particular value. It is user dependent and varies with the network application accordingly. An illustration is for n=3 is done in Fig. 7.

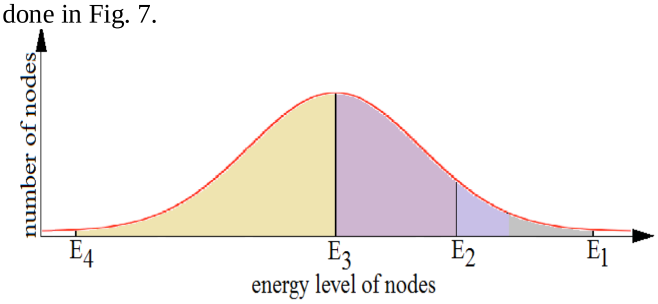

Fig. 7 Division of nodes into groups

E_1 denotes the energy of the nodes with highest energy and E_4 with minimum energy. The division is done at E_2 and E_3. Nodes with energies between E_1 and E_2 are placed at level one, E_2 and E_3 at level 2 and E_3 and E_4 at level 3. In this way, the nodes are expected to have a balanced work load. Also, our aim is to achieve proper coverage of the network. In certain applications, only the data gathered from the whole region is useful. So, it is necessary that if the nodes die out, they die out at nearly same times. In other words, the difference between the first node and last node death is least while maintaining the aim of prolonged lifetime of the network. In some existing protocols, the results show that the death of nodes start very early and the system keeps on running with very less amount of nodes which is not fruitful. So with the present approach both of these goals can be achieved.

In addition to coverage issues, the present approach keeps account of nodes failure also. If due technical reasons or some environment factors a node at any intermediate level dies, then the data of its child nodes is directed to its respective parent node as shown in Fig. 8. In this manner, we can have a fault tolerant approach and consequently the connectivity of the network is not lost.

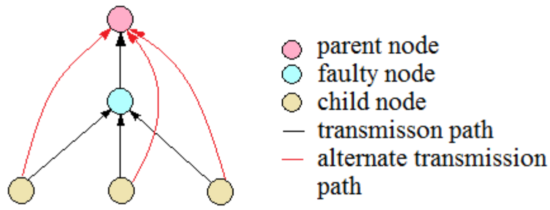

Fig. 8 Transmission by child nodes in case of failure of intermediate node

The flowchart of the proposed approach is represented in Fig. 9.

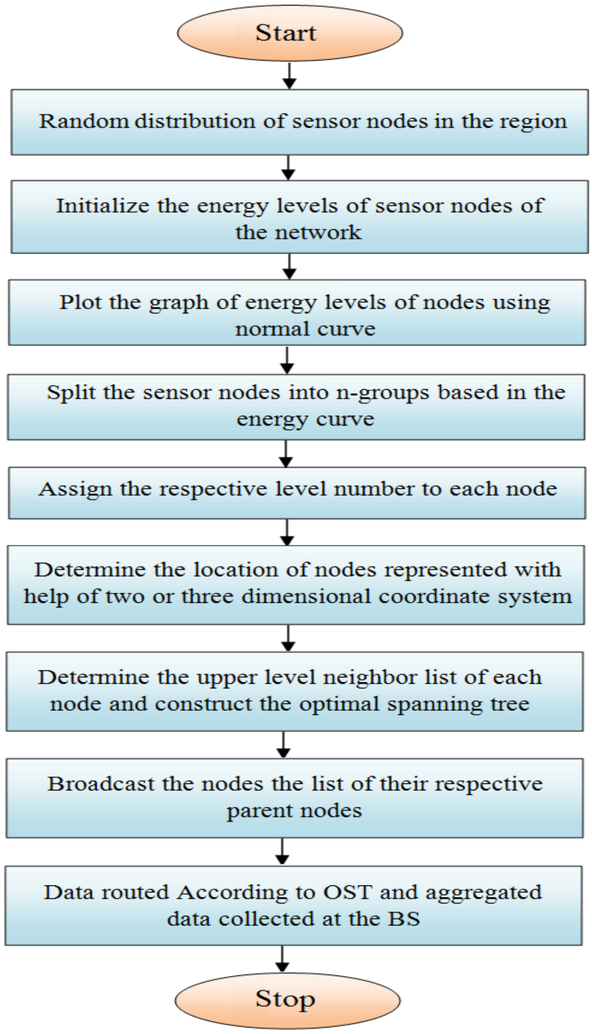

Fig. 9 Flowchart of NLTRA

IV. CONCLUSION

The paper presents a heuristic approach for determining optimal path for data routing in a network. Due to energy constrained nature of sensor, an effort has been made to balance the work load among sensor nodes. The method of MST can be useful when there is lot of redundancy in data and the networks size is huge as well as when the distance between the nodes plays an important role in incrementing data transmission energy. This method is not area specific and can be implemented in both two and three dimensional network regions. This approach also proves to be advantageous in view of scalability, coverage, connectivity and fault tolerance which are major issues in WSNs. The main focus of this approach is to obtain a routing path which optimal in terms of energy efficiency. Considering the criteria of residual energy for constructing the MST apart from the traditional distance based weights is a novel approach in the field of cluster based routing.

REFERENCES

[1] H. Z. Abidin, N. M. Din, N. A. M. Radzi, and Z. I. Rizman, “A Review on Sensor Node Placement Techniques in Wireless Sensor Networks,” *International Journal on Advanced Science, Engineering and Information Technology*, vol.7, no.1, pp. 190-197, 2017.

[2] Harendra S. Jangwan and Ashish Negi, “A Swarm Optimization Based Power Aware Clustering Strategy for WSNs,” *International Journal on Advance Science, Engineering and Information Technology*, Vol. 7, 2017.

[3] W. Heinzelman, A. Chandrakasan, and H. Balakrishnan, “An Application-Specific Protocol Architecture for Wireless Microsensor Networks,” *IEEE Trans. Wireless Comm.*, vol. 1, no. 4, pp. 660-670, 2002.

[4] Gulista Khan, Gaurav Bathla, and Wajid Ali, “Minimum Spanning Tree based Routing Strategy for Homogeneous WSN," *International Journal on Cloud Computing: Services and Architecture(IJCCSA)*, vol. 1, no. 2, pp. 22-29, 2002.

[5] Koteswararao Seelam and M. Sailaja, “An Efficient Distance-Energy-based Minimum Spanning Tree (DE-MST) for Wireless Sensor Networks,” *International Journal of Computer Applications*, vol. 126, no. 10, pp. 1017, 2015.

[6] Matheswaran Saravanan, and Muthusamy Madheswaran, “A hybrid optimized weighted minimum spanning tree for the shortest intrapath selection in wireless sensor network,” *Mathematical Problems in Engineering*, pp.1-8, 2014.

[7] Djamel Djenouri , Miloud Bagaa, Ali Chelli, and Ilangko Balasingham , “Energy Harvesting Aware Minimum Spanning Tree for Survivable WSN with Minimum Relay Node Addition,” *Globecom Workshops (GC Wkshps)*, 2016 IEEE, 2017.

[8] Uttam Vijay and Nitin Gupta, “Clustering in WSN based on minimum spanning tree using divide and conquer approach,” *International Journal of Computer Science and Engineering*, vol.7, no.7, pp. 926-930, 2013.

[9] Meimei Zeng and Hua Jiang. “A New Security Routing Algorithm Based on MST for Wireless Sensor Network,” *International Conference on Intelligent Information Processing*, Springer Berlin Heidelberg, pp. 331-336, 2012.

[10] B.Baranidharan and B.Shanthi, “A New Graph Theory based Routing Protocol for Wireless Sensor Networks,” *International journal on applications of graph theory in wireless ad hoc networks and sensor networks (GRAPH-HOC)*, vol.3, no.4, pp. 15-26, 2011.

[11] M. Yamuna, and A. Elakkiya, “Development of Alternate Fault Tolerant Minimum Weighted Spanning Tree using Vertex, Edge Connectivity,” *International Journal of Pharmacy & Technology*, vol.8, no.2, pp. 11810-11819, 2016.

[12] Wen Zhong Guo, Hong Lie Gao, Guo-Long Chen and Lun Yu, “Particle swarm optimization for the degree-constrained MST problem in WSN topology control,” *IEEE International Conference on Machine Learning and Cybernetics*, vol. 3, pp.1793-1798, 2009.

[13] Andrei Gagarin, Sajid Hussain, and Laurence T. Yang, “Distributed search for balanced energy consumption spanning trees in wireless sensor networks,” *IEEE International Conference on Advanced Information Networking and Applications Workshops*, pp. 1037-1042, 2009.

[14] Miloud Bagaa, Mohamed Younis, and Ilangko Balasingham, “Data aggregation tree construction strategies for increasing network lifetime in EH-WSN,” *IEEE Global Communications Conference (GLOBECOM)*, pp. 1-6, 2015.

[15] Narsingh Deo, “Graph Theory with Applications to Engineering and Computer Science,” *Prentice Hall, India,* 2004.

[16] S.C. Gupta, V.K. Kapoor, “Fundamental of mathematical Statistics,” *SULTAN CHAND & SONS,* India, 2007.

Harendra S. Jangwan is a research scholar in the department of computer science and engineering at G.B. Pant engineering college, Ghurdauri, Pauri Garhwal. He has received Master degree in Computer Science &Application from H.N.B Garhwal University. His research interest includes wireless sensor networks, network security and fault tolerant systems.

Dr. Ashish Negi is a Professor in the department of computer science and application at G.B. Pant engineering college, Ghurdauri, Pauri Garhwal. He has received M.Sc. Physics, MCA, Ph.D in computer science from H.N.B. Garhwal University, MMM engineering college Gorakhpur and Gurukul Kangri Univrsity Haridwar respectively. His research Areas includes fractal networks, computer graphics and image compression.

