An Energy Efficient Clustering Approach based on K-means ++ Algorithm with Leach Protocol for WSN

Prashuma Singh
M.Tech Scholar
Department of Information Technology
NRI Institute of Information Science and Technology, Bhopal, India

Angad Singh
Associate Professor
Department of Information Technology
NRI Institute of Information Science and Technology Bhopal, India

ABSTRACT
Wireless Sensor Network (WSN) is a wireless network of thousands of inexpensive miniature devices capable of computation, communication and sensing. Nodes in the WSN have restrictions of memory, storage, processing and energy. Sensors nodes in WSN are used to measure the environmental parameters like temperature, pressure, humidity, sound, vibration etc. WSNs are assumed to be energy restrained because sensor nodes operate with small capacity or may be placed such that replacement of its energy source is not possible. Due to these limitations several routing protocols have been proposed to utilize sensor's energy to prolong the life time of deployed WSN.

An effective routing protocol is desirable which is able to manage communication among energy restrained sensor nodes and able to provide load in uniform way such that difference between life times of nodes is not very large.

In this paper, K-means++ with Adaptive leach based routing algorithm has been implemented. The proposed methodology is used to find life time of sensor nodes in terms of rounds in network. Sensor nodes drops to zero energy ignored for next round of CH (Cluster Head) election in network. Adaptive LEACH is used to improve the hop-count of transmitted data in the transmission phase of cluster head nodes to the BS (Base Station) so as to keep the balance of energy consumption and prolong the survival time of network. Proposed routing protocol outperforms the LEACH-CKMEANS, CH-LEACH algorithms with improved Average Throughput, improved number of transmitted data packets and improved Network lifetime

General Terms
Improvement in energy efficiency of the wireless sensor network

Keywords
Clustering in WSNs, Energy Efficiency, minimizing energy consumption, LEACH, K-means++.

1. INTRODUCTION
Wireless Sensor Network (WSN) is a wireless network of thousands of inexpensive miniature devices capable of computation, communication and sensing. The emerging field of wireless sensor networks combines sensing, computation, and communication into a single tiny device. Through advanced mesh networking protocols, these devices form a sea of connectivity that extends the reach of cyberspace out into the physical world. The mesh networking connectivity will seek out and exploit any possible communication path by hopping data from node to node in search of its destination. While the capabilities of any single device are minimal, the composition of hundreds of devices offers radical new technological possibilities. Energy efficiency is a considerable factor when designing a sensor network. Clustering is an efficient data gathering technique that effectively reduces the energy consumption [1]. In the coming years, as advances in micro-fabrication technology allow the cost of manufacturing sensor nodes to continue to drop, increasing deployments of wireless sensor networks are expected, with the networks eventually growing to large numbers of nodes (e.g., thousands). Potential applications for such large-scale WSN exist in a variety of fields, including medical monitoring [1], environmental monitoring [2], surveillance, home security, military operations, and industrial machine monitoring.

Wireless sensor network may contain different types of sensors including seismic, thermal, magnetic and visual, that are able to track changes of environmental conditions such as humidity, pressure, sound, light and movement [2]. The nodes of WSN have limited computing power, energy supply, and short communication range. To effectively cope with this issue, the sensor network should be designed so that scalability and energy efficiency can be achieved. Clustering the sensor nodes is one of the most effective solutions. There exist various WSNs employing the cluster structure, which efficiently allocate the resource and energy and thereby maximize the network life time. Here each cluster of sensor nodes is monitored and controlled by a node, called Cluster Head (CH). Each CH aggregates the data sent from the sensor nodes belonging to its cluster, and then transmits them to the BS. Forming the clusters, especially CH selection, is one of the most critical tasks in the management of WSNs since CHs consume much larger energy than other bodes in the network [3]. It is usually difficult to recharge or replace the sensor nodes which have limited battery capacity. Energy efficiency is thus a primary issue in maintaining the network.

2. RELATED WORK
This section contains related work done on wireless sensor network, .Existing research work also discussed that is performed by various researchers. Routing in WSN is very challenging due to the inherent characteristics that separate WSN from other wireless networks like mobile ad hoc networks or cellular networks[4]. Many new algorithms have been proposed for the routing problem in WSNs. These routing mechanisms have taken into consideration the inherent features of WSNs along with the application and architecture requirements. The task of finding and maintaining routes in WSNs is non-trivial since energy restrictions and sudden changes in node status like failure cause unpredictable topological changes. To minimize energy consumption, routing techniques proposed in the literature for WSN described below:

[5, 6] Introduced LEACH protocol stands for Low Energy Adaptive Clustering Hierarchy. The LEACH Network is made up of nodes, some of which are called cluster-heads. The job of the cluster-head is to collect data from their surrounding nodes and pass it on to the base station. LEACH is dynamic because the job of cluster-head rotates. In this, CH are not fixed and position CH rotates. Clusters are divided randomly, which results in uneven distribution of clusters.

Djamila Mechta[8] proposes LEACH-C protocol using sink mobility, in which the formation strategy of clusters and the selection of clusters-heads are accomplished using the non-supervised K-means technique. Sink moves along two trajectories, the first one is the density center for each cluster while the second is the center of each cluster to gather the data collected by the clusters-head during all rounds but the use of one hop transmission intra and inter cluster leads nodes and CHs to consume more energy.

Hassan echoukairi1[11] proposes A Novel Centralized Clustering Approach based on K-Means Algorithm for Wireless Sensor Network, to enhance the clustering process in LEACH-C protocol by applying k-means algorithm in order to produce a new cluster scheme and therefore lighten the lifetime of the sensor network**.** The issue is, account the mobility of the nodes and integrate some variants of k-means. In this, nodes cannot communicate their coordinates and residual energies to the BS which may result in a clustering that is not efficient. This can degrade the performance of the network due to data loss.

G. Park et al.presented K-means clustering based routing protocol in WSN which forms the clusters of objects based on the Euclidean distances between them. K-means algorithm is executed for cluster formation with the target WSN [3]. Assume that the WSN of *n* nodes is divided into *k* clusters. First, *k* out of *n* nodes is randomly selected as the CHs. Each of the remaining nodes decides its CH nearest to it according to the Euclidean distance. After each of the nodes in the network is assigned to one of *k* clusters, the centroid of each cluster is calculated.

Yang [7] introduced k-means clustering based fault detection algorithm(k-CFD), where fault detection based on two voting, define a two round detection and decision process in which each node has 2-hop neighbors clustering views.

Seifemichael B. [4] introduced Genetic algorithm inspired clustering hierarchy(GAICH), this protocol makes use of genetic algorithm to create optimum clusters in terms of energy consumption. In this Issues such as delay of delivering data packets to BS and overhead of running the algorithms are not taken considerations.

Walid Abushiba[12] proposes An Energy Efficient and Adaptive Clustering for Wireless Sensor Network (CH-leach) using Leach Protocol in which, CH-Leach selects a certain number of nodes as these nodes are nearest to centroid in their cluster region, cluster head will be responsible for transmit the data to Base Station. The cluster-heads are randomly selected and in every round each of the nodes that will become cluster-head is assigned to nearest centroid. As the nodes that already chosen before have energy under the threshold, then the cluster-head will updated by chosen another node, these steps are depending on location of the nodes which are randomly deployed.

3. PROBLEM STATEMENT

In WSN it is difficult to recharge or replace the sensor nodes which have limited battery capacity. Sensor nodes are usually deployed in a random fashion, and collect the context information and perform the given mission through the cooperation with other nodes. To mitigate the problem, one effective way is to separate the planes of gathering information from sensor node and sending this information to BS. There are many routing techniques in WSN based on clustering like LEACH, HEED, K-MEANS. But the problem with the method is to decide on what basis two planes of communication will be separated. Hence to improve the performance of the WSN, an improved routing algorithm is required.

4. METHODOLOGY

The proposed methodology is used to find life time of sensor nodes in terms of rounds in network. Sensor nodes drops to zero energy ignored for next round of CH election in network. In each round remaining energy calculated on deducting consumed energy from initial energy of node.

The steps of proposed approach are as follows:

i. Classify a network into specified number of clusters by using adaptive LEACH with K-Means++ algorithm.

ii. Calculate centroid in each clusters and choose a node as CH which is nearest to the

centroid and contains high energy.

iii. Assign size of the packet to perform communication between sensor nodes and CH.

4.1 Initialization of WSN

This is the very first step of proposed system; in this step WSN is initialized. In network, predefined energy is assigned to nodes and positions of these nodes are present in the network in random fashion. In proposed approach the WSN is assumed to have the following features:

i. Unique ID number is assigned to each node.

ii. All nodes are pseudo static.

iii. All nodes are able to control their energy consumption.

iv. All nodes are able to send the data to the BS.

v. Initial energy of all nodes are same.

vi. All CH are aware of their remaining energy.

vii. The sensor nodes are randomly distributed in the target area.

4.2 Cluster formation

For clusters formation, we opted for the unsupervised classification method K-means++.It's principle is the first initial cluster center selected uniformly and randomly from all sensor nodes, after that each successive cluster center is selected based on a probability function. According to it, a probability function the probability should be proportional to its squared distance from the existing cluster center closets to the point. K-means++ prevents all the hurdle of k-means and sure to provide a better result. The k-means++ initialization method includes two secondary phases. The first one is choosing one point as a center based on the probability and the second one is updating the sum of distances that points to their nearest centers.

For the k-means++ algorithm, at any time of instant, Let d(x) is a set of data point x that has a little distance from the closest center, we have already chosen the k-means++ initialization algorithm is illustrated in following steps:

1) Select the first center that is called initial center c1, orderly at random from x.
2) Select the next center ci, by picking c2 from x' ϵ X with help of the probability of $dx^2/\Sigma x \in xd(x)^2$, here d(x) is the data point with closer distance with cluster centroid.
3) Perform again and again Step 2 until k number of centers not achieved.
4) Continue with the basic k-means algorithm.

4.3 Selection of CH

After formation of cluster, a unique ID is assigned to each node of a cluster according to the distance from the centroid. The ID number plays an important role because that indicates the order to be selected as the CH. Each cluster is managed by a CH. Average energy of all nodes calculated along with average distance to BS.

In each round threshold T(n)= (p(i)/(1-p(i)*mod(r,round(1/p(i)))))

determined using remaining energy with optimal count of CHs. Random number generated for each node which is compared with T(n).If result is less than T(n) then node will be elected as CH. If elected CH distance is more than average distance then next nearby CH of Base Station is selected. Finally data transmitted with multi-hop using TDMA (Time Division Multiple Access) to the BS.

Algorithm

%design Network Architecture

Input: Node of Point

Step 1: Initialize node

Step 2: For i=1:1:n

Step 3: Design K=cluster set

Step 4: Place k centroids in random places{c1,…ck}c1,..ck}:ck=∑i:C(i)=kxiNk,k=1,..K.

Step 5: end for

Step 6: Do find closest center Ck in k

Step 7: For j=1:nj=1:n(CH nodes)

Step 8: Data=datagenerate(i)

Step 9: Set>threshold

Step 10: Set>cluster and communicate(K)

Step 11: Set>energy

Step 12: checks each nodes energy against 0 to establish node dying

Step 13: Removes dead nodes from all lists

Step 14: For i=1:ni=1:n(alive nodes)

Step 15: End for

Step 16: End for

%all nodes sent data directly to BS

5. SIMULATION AND RESULT ANALYSIS

We use MATLAB R2013a to simulate the algorithm. The nodes are randomly distributed in the 100×100m environment. The parameters used in the experiment are

Shown in table 1.

Table 1: Simulation Parameters

PARAMETER	VALUE
Number of nodes(N)	100
Number of clusters	20
Network size	500
Initial energy(E_0)	0.5J
Probability of CH(P)	0.1
Max. number of Rounds(r max)	3000
EDA(Data Aggregation Energy)	5*0.000000001
EFS(Amplifier energy for free space)	10*0.000000000001
EMP(Amplifier energy for multipath fading)	0.0013*0.000000000001
Network time	0.0150

5.1 Result Analysis

The Table shown below is the table of analysis and comparison of Average Throughput on Leach-ckmeans and Proposed work. The proposed Methodology implemented provides better Average Throughput.

Note: The table given below compares Average Throughput up to 140 nodes while in my proposed work number of nodes taken up to 200 and in nodes 160, 180 and 200 with respective throughput values are 3.96, 4.32 and 4.38.

Table 5.1: Average Throughput of Leach-ckmeans and Proposed Work

Number of Nodes	Average Throughput	
	Leach-Ckmeans	Proposed
60	2.85	2.82
80	3.25	3.60
100	3.32	3.83
120	3.56	3.91
140	3.54	4.10

The Table shown below is the table of analysis and comparison of packet delivery ratio on Leach-ckmeans and proposed work. The proposed Methodology implemented provides better Packet Delivery Ratio.

Note: The table given below compares Packet Delivery Ratio up to 140 nodes while in my Proposed work number of data packets taken up to 200 and in nodes 160, 180 and 200 with respective values are 2.55, 2.71 and 2.88.

Table 5.2: Packet Delivery Ratio of Leach-ckmeans and Proposed Work

Number of Data Packet	Packet Delivery Ratio	
	Leach-Ckmeans	Proposed
60	1.23	1.38
80	1.20	1.70
100	1.22	1.97
120	1.24	2.18
140	1.30	2.36

The Table shown below is the table of analysis and comparison of number of Dead nodes on CH-Leach and proposed work. The proposed Methodology implemented provides better Number of Dead Nodes.

Table 5.3: Number of Dead Nodes of CH-Leach and Proposed Work

Number of Rounds	Number of Dead Nodes	
	CH-Leach	Proposed
1000	18	84
1500	57	90
2000	95	91
2500	NA	92
3000	NA	94

The Graph shown below is the comparison of Average Throughput for Existing and proposed work. Average Throughput is the ratio of the total number of data received at base station to the network lifetime.

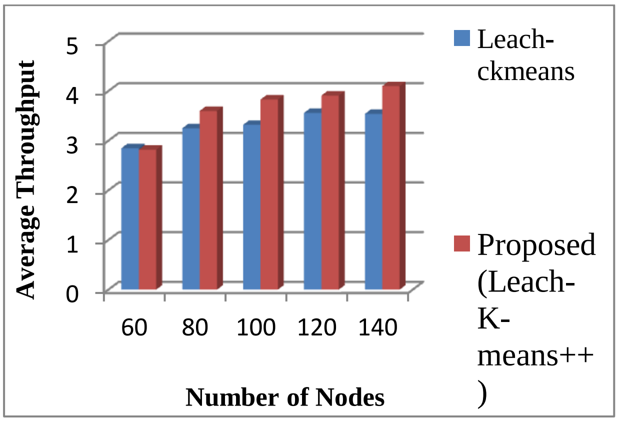

Figure 5.1: Comparison of Average Throughput for Leach-ckmeans and proposed

The Graph shown below is the comparison of Packet Delivery Ratio for Existing and proposed work. The Packet delivery ratio (PDR) is the ratio of number packets that received by the base station and number of packets sent from the source. When it is higher, the performance is better

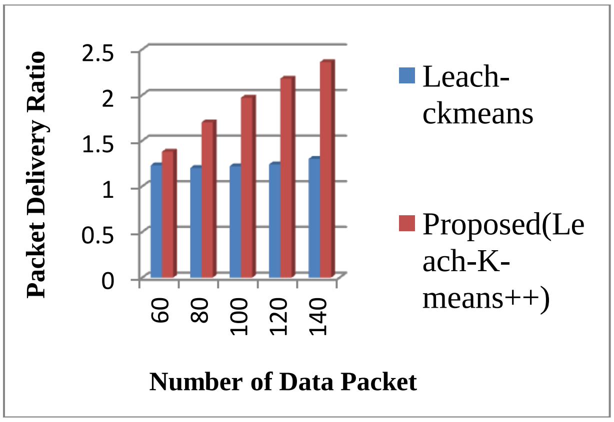

Figure 5.2: Comparison of Packet Delivery Ratio for Leach-ckmeans and proposed

The Graph shown below is the comparison Number of dead nodes for Existing and proposed work.

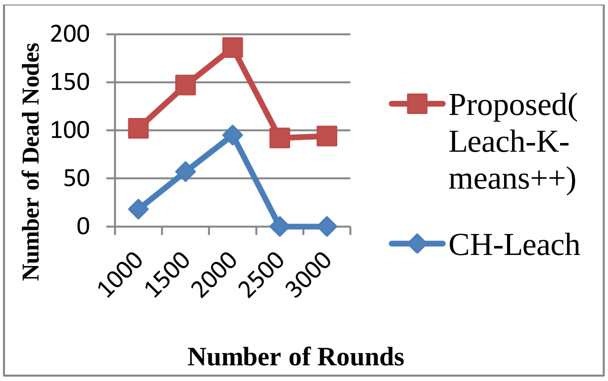

Figure 5.3: Comparison of Number of Dead Nodes for Ch-Leach and Proposed

6. CONCLUSION

In this paper, a new cluster based routing protocol used to improve the life time of WSN is implemented successfully. In this new approach K-means++ algorithm is used to form clusters in optimized manners so that communication between sensor nodes will cause less power consumption and some features of Adaptive LEACH is used to improve the hop-count of transmitted data in the transmission phase of cluster head nodes to the BS (Base Station) so as to keep the balance of energy consumption and prolong the survival time of network. Simulation results show that the proposed k-means++A-leach routing protocol outperforms the LEACH-CKMEANS, CH-LEACH algorithms with improved Average Throughput, improved number of transmitted data packets and improved Network lifetime.

7. FUTURE WORK

Although implemented approach used here is feasible and efficient but there are some future directions that can be implemented further.

(i) Implemented approach will not work properly if sensor nodes are not static (fixed) because clusters have to form repeatedly for mobile nodes. It will further increases power consumption.

(ii) Impact of implemented approach is less on WSN with small number of nodes.

8. ACKNOWLEDGEMENT

With great pleasure and sense of obligation I express my heartfelt gratitude to my project guide Mr.Angad Singh Associate Professor, Department of information Technology NIIST Bhopal. His contributions are beyond the purview of the acknowledgement.

9. REFERENCES

[1] Huili jia, Xunli fan, yang qian, kun liu, “An energy efficiency routing algorithm for wireless sensor network”, IEEE International conference on computational science and engineering and IEEE International conference on embedded and ubiquitous computing, pp.735-739, 2017.

[2] Kadir tohma,yakup kutlu,Ipek abasikeles-turgut, “A new cluster head based dynamic base station positioning for wireless sensor networks”, IEEE, 2017.

[3] Geon yong Park, Heeseong Kim, Hwi Woon Jeong, and Hee yong youn, “A Novel cluster head selection method based on k-means algorithm for energy efficient wireless sensor network”, IEEE 27th international conference on advanced information networking and applications workshops,pp.910-915,2013.

[4] Seifemichael B.Amsalu*, Wondimu K.Zegeye, Dereje Hailemariam, Yacob Astatke, “Design and performance evaluation of an energy efficient routing protocol for wireless sensor networks”. IEEE Annual conference on information science and systems, 2016.

[5] Krishnakumar A, Dr. Anuratha V, “An energy-efficient cluster head selection of LEACH protocol for wireless sensor networks”, IEEE, International conference on Nextgen Electronic Technologies, pp.57-61, 2017.

[6] Hairong Zhao, Wuneng Zhou, Yan Gao , “Energy Efficient and cluster based routing protocol for WSN”, IEEE Eight International conference on computational intelligence and security, pp.107-111, 2012.

[7] Yang yang, Qian liu, Zhipeng gao, Xuesong Qui, and LanlanRui, “Data clustering-based fault detection in WSNs”, IEEE 7th International conference on advanced computational intelligence, pp.334-339, 2015.

[8] Djamila Mechta, Saad Harous,”Improving LEACH-C using sink mobility”IEEE, 2016.

[9] A.T.e. Andradel,2, e. Montezl, R. Moraesl, A.R. Pintol, Francisco Vasques3, G. L. da Silval, “Outlier Detection Using k-means Clustering and Lightweight Methods for Wireless Sensor Networks”,IEEE, pp.4683-4688,2016.

[10] Vaibhav vijay, manoj singh, “Energy efficient cluster head selection technique for homogeneous wireless sensor networks”, IEEE, 2017 2nd International conference on communication Systems, computing and IT applications,pp.54-59, 2017

[11] Hassan echoukairi1, Amine kada1, Khalid bouragba1, Mohammed ouzzif1, “ A Novel Centralized Clustering Approach based on K-Means Algorithm for Wireless Sensor Network”, IEEE, pp.1259-1262, 2017.

[12] Walid Abushiba, Princy Johnson, Saad Alharthi, Colin Wright, “An Energy Efficient and Adaptive Clustering for Wireless Sensor Network (CH-leach) using Leach Protocol”, IEEE, pp.50-54, 2017.