Improvement in K-Medoids using Shortest Path in Wireless Sensor Network

Garima Sharma, Praveen Kumar and Laxmi Shrivastava

Abstract: ***A Wireless Sensor Network (WSN) is a collection of a large number of sensor nodes (SNs) & at the smallestsingle base station (BS). The SN is an independent tiny device that comprises of predominantly four units that are sensing, processing, communication & power supply. Clustering is a competent method for limiting node energy consumption and amplifying lifetime of the network. K-medoids algorithm is a kind of K-means algorithm, wherever the hubs are selected from the set of data facts. The node is closest to the nearest distance of the K-medoids in the node groups, finding the center of the same clusters as groups of the same cluster. Formerly created K-Medoids clusters are allergic to clusters and cluster heads, and then send data through cluster heads to reach up to BS. The BS is used to collect altogether nodes and detecting data finished cluster head bulges. They harvested the node coordinates and the rest of the energy, and then cluster number is detected. Central Circle-based points also reduce the time by calculating the remaining energy. But the nodes are randomly selected and the distance in the middle of the nodes as well as the base position is collective, which is collective time. Cluster heads connectstraight with the base position. So, the specific algorithm can override it. In this paper, Dijkstra Algorithm used for the transmission of data by generating the shortest path among the cluster heads. The tree has been formed to transmit the data from the source node to the BS through he intermediate cluster heads. The simulation takes place on the MATLAB device to show specific actions. In effect, indicates the effectiveness of the mission's capabilities in operation and living capacity.***

Keywords: Wireless Sensor Network, Clustering, K-Means, K-Medoids, Dijkstra Algorithm.

I. INTRODUCTION

At present, WSN is most motivating topics in the research area. It won't be a distortion to consider WSNs as a standout amongst the most investigated areas in the most recent decade. A WSN is described as a web of smaller devices, called SNs, and distributes and distributes information from the monitor field through the wireless links to the group. Data collected using various types of hubs are sent to a locality or information on the Internet is associated with different systems. One of the largest SNs combined was a BS. SN is an autonomous little gadget, most of which contains four components, which are handled, handled, burned and control supplied. These sensors utilized to gather the data from the earth and pass it on to BS. A BS gives an association with the wired reality where the gathered information is handled, broke down and exhibited to helpful applications. Subsequently, by inserting handling and correspondence inside the physical world, WSN can be utilized as an instrument to connect genuine and virtual condition [1].

The respite of the paper is prearranged as follows, LEACH, as well as Section 2, was chosen as Clustering Algorithm. Then a detailed literature survey of the selected clustering algorithm is presented in Section 3. The algorithm set in section 4 is explained. In Section 5, the performance is compared to the analogy suggested by the simulations. Finally, in Section 6 we can see the conclusion.

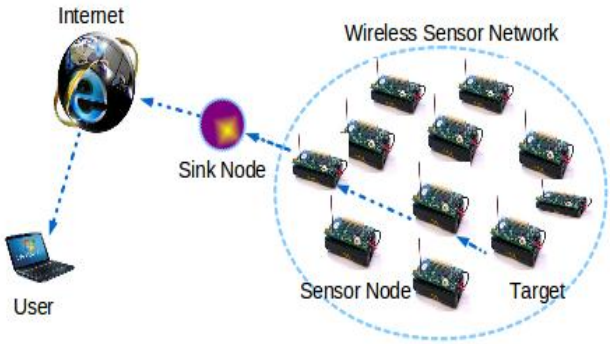

Fig.1. WSN [12]

II. CLUSTERING IN WSNS

Ns are different or unrestricted quantities are added. This strategy was identified as clustering. Clusters of clusters are known as clusters. Cluster head (CH) for each group is selected. This CH from all the hubs of a group. CH is used to select differentiation options. The CEO needs a qualification to retain the groups. A.The rest of the members are members Note (MN). The main task of CH is to summarize infoafter the nodes of the members and send them to the BS. While various centers are permanently disturbed and sent to CH. The life expectancy of CNN is very low compared to MNCs. Subsequently, the hub was selected and the high energy level for high CH speculation. Low-energy hubs are unsuccessful. In clustering, there are 2 types of communication It's like an inter-cluster. Some people in the group can speak with CH and CH, who speak with different CHs. The Group estimate also includes a clustering strategy. The hub should be slightly different from the close BS distance. The BS should be set up to detect remote bunch or any other data will be lost. Clustering techniques appeared in image 2 on RPs in the following sequence.

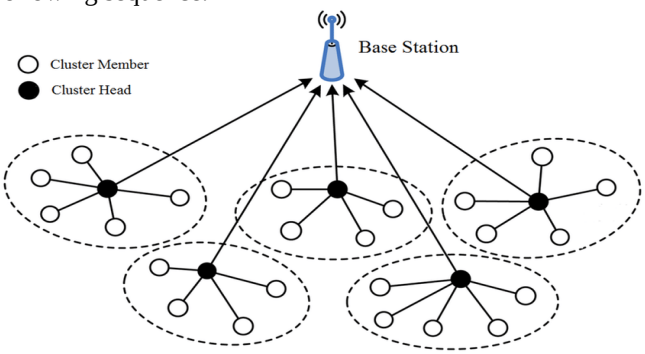

Fig.2. Clustering approach [13]

A. LEACH protocol

The LEACH Protocol is the greatestprevalent cluster-based routing protocol on many categorized route protocols with energy efficiency. In this sub-section, we briefly present this protocol. LEACH is a hierarchical pathway protocol, aimed at reducing energy consumption and thus increasing network life. LEACH works around with each stage set for the radius and constant nationalstage. In the setup phase, the sensor nodes will be selectedin the middle of 0 and 1. Through a random number. The uncertainty this is less than the random number T (N), formerly the associated sensor node will act as a cluster-head at this stage, which is named a round. Depending on the compilation value computed by Lash Below (1), the nodes of the cluster-head also dispersed in the nodes of the associates of the cluster.

$$T(n) = \begin{cases} \frac{p}{1-p*(r \bmod \frac{1}{p})} & \text{if } i \in G \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

Wherever P is the percentage of cluster-heads, R is the present round, set of animate nodes which are not designated as cluster head inside the preceding (1 / p) round. Works in two steps: Systemstage and arena step. Each CH in the setup stage transmits an ad message to make its cluster. A cluster is added to the RSS based on the message received by each message other than each CHCH. Once the clusters have formed, TDMA (Time Division Multiple Access) agendas will be assumed as association nodes in each cluster. In each cluster, each user communicates his shared data in his set time slot, as well by way ofgathers all the data of memberships from one packet through a package and sends it straight to the remote base position. To escape
Each CH is selected by interacting with cluster levels and CDMAC (coding division multiple access codes). It helps to communicate with basic stages from other clusters.

B. K-Means technique

K-Mean sets the most straightforward simple desktop. This is the case that the K-MAX in the simplest method of determining specific data through a set of numbers (eg K clusters). K. The main idea is that each of the centroids is defined for each cluster. This centroid should be kept away from each other. The next step is to get a given set of data sets If there is nothing to link it to the nearest infrastructure, the first phase is finished, the clusters are shaped. In this stage, we essential to compute k new centroids as a calculation of clusters from the preceding stage. Subsequently these different chemists, new bindings need to be done at the same data set point and the new center close. A loop is producedAs a consequence of this circle, k centroid does not make any more changes, we can see that their strokes are in their phases in phases. Finally, this method aims to reduce the object's function, in this equation a quadratic function in the equation (2).

$$J = \sum_{j=1}^{k} \sum_{i=1}^{n} \|x_i^j - c_j\|^2 \quad (2)$$

Where$\|x_i^j - c_j\|$the distance selected is the expanseamongst a data point of xij and the cluster middle between which the distance of data marks from their cluster centers are indicated. K-Yards is made up of the following steps:

1) K objects in the space represented by the clustering materials. These points represent the initial group strokes.
2) Allocateevery entity to the latest instrumental collection.
3) When all objects are prescribed, it is again calculating the positions of K.
4) Repeat steps 2 and 3 until the cells are no lengthiertouching. This can be separated into matrix groups that are divided into fragments [4].

III. LITERATURE SURVEY

In this paper, [5] The effective energy conservation program aims at reducing energy consumption and expanding the network throughout life. Technology is the technology to reduce energy ingesting for WSNs. The core idea of this method is to decrease the communiqué distance of sensor nodes using clustering techniques. To become the best result of the group, we use KMDOS algorithm to detect optimal media as central nodes. Formerly let's select the suitable cluster capacities. The network load balance amid the clusters of the method can recover energy efficiency, the more effective the lifetime of the network. Simulation results give us an extraordinary performance in our suggested algorithm rather than the energy consumption of the network and other popular algorithms.
Armin Mazinani et al. [6] The fundamental point The result is a method to reduce the CH elections and improve message transmitted at each stage. The rest of the energy is not. Each node is considered as criteria for selecting nodes and distances. Different calculations of the gamers are dialectic systems such as FMCR-CT, for example, network life, FD, HND & LND in every round. The results from the reconstruction reveal whether FHRC-CT can force different technologies.
Saber AMRI et al. [7] propose and execute another system for geographical routing. In this way, the specific component is dependent on a weighted centered control mechanism, where multiple nodes are used to determine the critical logic strategy. For this, we suggest a Fuzzy Localization Algorithm utilizing stream destination by remote channels to promote the separation between living and sensor nodes. Hence,Mamdani and Sugeno adversely affect the framework of the conclusions of the state of the precursor to the accuracy of our work and evaluation. Strict calculation If the division of nodes is changed to an invisible location, the chosen option is selected by choosing a successor CH, reducing the vantage scattering of sensor nodes and expanding the life span of the system. Acquisition of entertainment acquired by the specific system is based on the number of packages transferred to the base station, current solutions such as energy consumption, local time (localization time) and localization errors.
P.S. Mehra, M.N. Doja and B. Alam [8] CHF Shell Algorithm (FBESC) suggests a Fuzzy-based expenditure, remaining energy and a solid non-thick in its area contributing to the company. The eligibility criteria forselecting the CH function for each job is assigned. This convention assumes the load adjudication chosen for the convenience of the Group facilitator by considering the possibilities raised in all sensor nodes. Experimental results, based on FBESC, BCSA, leach, etc., weapons of mass trust, load balancing and load balancing,

Zhang Siqinga et al. [9]A clustered protocol is proposed based on the Fuzzy Layout of Multi-Hope WSNs (FLCMN) to growth the distance of life of Wireless Sensor Networks (WSNs) and reduce energy consumption. The algorithm's capacity for nodes is about energy. The energy with the regular capacity of neighboring nodes. According to the predetermined fuzzy round base, the Fuzzy frame is used to evaluate the fulfillment of the selected CA. In addition, the usualinfluence of adjoining nodes increases the control of the hotspot and equilibrium between the groups. In the meantime, so improvement of hot chassis among monopolies, and intended to delay the lifetime of multi-hop transmission plot systems that depend on on the Fibonacci group. By reproduction, the FLCAMN algorithm uses the LEACH, EAMMH and DFLC calculations to determine the time and power of system retention.

In this paper [10] produce a better life on the architecture of WSN, LEACH partition topology (LEACH-PT) is proposed in this paper. Base Station (BS) is selected by Cluster Head (CH) on LEACH-PT. Ensure the number of clusters fixed at each round. In addition, this activity reduces energy consumption per node, as there will be a better network life than leaflets. In addition, the specific provisions for the distribution node from the current random topology of the line are similar. The LEACH-PT Algorithm is an important focus. The simulation results using Network Simulator 2 indicate better performance in the node period than the leak-pad lead.

Fan et al. [11] The energy-based clustering model (EBCM) scheme proposed for reducing energy ingesting of all bulges and decrease of network energy consumption. EBCM is based on LEACH, considering some factors, such as the energy level and degree in the clustering process. Reproductionconsequences show that the IBM algorithm will reduce the average power ingesting.

IV. PROPOSED METHODOLOGY

A. Overview

In this research, we have used the clustering algorithm for data transmission. Communication amongst the improperposition and the cluster head performed which is created on the Dijkstra algorithm that is implemented in this article. So, firstly we define the K-Medoids and then Dijkstra algorithm and its steps.

K-Medoids

K-Medoids algorithm is a version of the K-means algorithm, K-medoids associated with the C-node find the most optimal center, so the K-medoids groups are the lowest distance between the clusters. There are two ways to start the algorithm: the first one chooses to choose nodes from their nodes and group them in network groups, and the latter algorithm chooses the initials of CH.

B. Proposed Methodology

Formerly [5] created K-Medoids clusters are allergic to clusters and cluster heads, and then send data through cluster heads to reach up to cluster.The base station is used to collect all nodes and sensing data through cluster head nodes. They cut node coordinates and the rest of the energy, and then cluster number was detected. K- Medoids al-algorithm has been optimized to reduce repetition time, but the nodes are randomly selected and the distance between the nodes and the base station increases.The cluster heads directly communicated with the base station which is very time-consuming. So it can be overcome by the proposed algorithm. There is Dijkstra Algorithm used in the proposed work which is described below:

Dijkstra Algorithm

Dijkstra's algorithm is a set of rules for discovery the bottommost possible route amongst nodes in a single graph. This procedure is changed. The unique variant containsdiscovery the lowest pathway amongst a source node as well as a node. An alternate variant ensures that the source is "source" and sees shortcuts on all other nodes from the source. We assume that the source is the graph G (X, A, C), which is the source and the CH and the active groups, where X is the set of searches, A is the set of curves and the variable values of Care variables. X is separatedaddicted to two groups, a set of group visits, a short path to side node, and sessions that have not been unselected, and a node set that does not have a fixed number of shortcuts. Dijkstra algorithm calculates the lowest paths between two bulges on one chart:

Step 1: Set limitless range for source node and every node.

Step 2: Consider all the neighbors and calculate their expenses. Attach a small one while comparing the designation value for a new unique. For example, if a node is noticeable with the price of N1 5 and adds it to N2 with length 3, N2 (via N1) 5 + 3 = 8. It may be greater than N2 or greater 8. Keep current or current value.

Step 3: Eliminate this node after the unwanted sets then add it to the visited set.

Step 4: Step 2 and Step 3 before marking as a node.

C. Proposed Work

In the proposed work, Initially, the sensor nodes are configured over the network. The K-medoids clustering algorithm used to form clusters of nodes and cluster heads is transmitted using clusters. Dijkstra algorithm is now used to create a tree by joining cluster levels.

In this nearest cluster heads are combined to generate the tree towards the base station. The data transmission is done by the tree to move the data at the base station. This method generally improves network performance and reduces the consumption of energy. The proposed algorithm described below to show the overall working of the proposed work:

Proposed Algorithm:

// initialization of network

Step 1: Initialize the network

Step 2: Place base station at coordinate (50, 180)

Step 3: Put all the sensor nodes randomly

// formation of cluster

Step 4: Cluster formation performed by using K-Medoids Algorithm

a. Start of the centers (CH)
b. Associate each node with its nearest CH
c. Calculate the conformation cost (sum of the detachments of each CM to its CH)
d. whereas the shape's cost reductions do
   i. for each cluster do
   ii. Exchange the CH purpose with a CM node

iii. Recomputed the novelprice
Uncertainty the cost increases then
Open the exchange (keep the precedingconfirmation)
End if
iv. end for
e. end while

// data transmission start
Step 5: If the node has data
Step 6: The minimum path must be found using Dijkstra Algorithm
Step 7: Find nearest Cluster head
Step 8: Repeat until tree formed towards the source node to a base station
Step 9: Data transmission performed by moving through the tree
Step 10: Exit

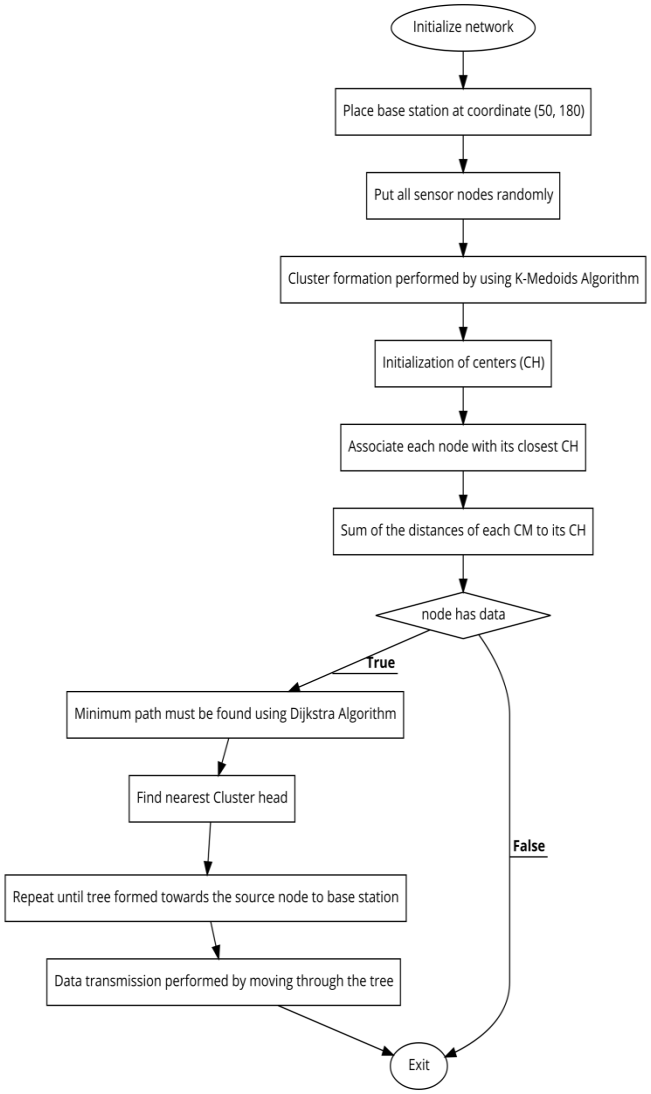

Fig.3. Proposed Work Flowchart

Table 1: Comparison between K-Medoids and Proposed Algorithm

Specification	K-Medoids	Proposed Algorithm
Basic Idea	The data transmission should be performed in the form of clusters	The data transmission should be performed by clusters and then a tree.
Main Difference	The data is traveled directly from cluster head to base station	The data is travelled from cluster head to nearest cluster head and then towards the base station
Similarity	The main motive is to transmit from source to destination by forming clusters in the network	

V. SIMULATION AND RESULTS

MATLAB is used in a variety of functions, as well as incoming use by imaging transfer, transportation, organizing design, sizing experiment, computational computing, financial modeling, and analysis. MATLAB, the trending early for rapid prototyping, is an array of languages that is now in progress for progress structure code secondhand for arithmetical and logical requests. Experimental studies were performed to analyze the given scenario. The simulator was programmed using MATLAB R2018a. In this, we are comparing different parameters in different scenarios. Then again, factors, for example, the rest of the energy of the system, the no. 1st node passes on in all rounds i.e. First DeadNodeand Last Dead Node are observed in each circumstance.

The reference network comprises of 50 nodes arbitrarily disseminated ended a region of 100×100 m^2. The base position is situated at the midpoint of Area-of-Interest (50, 180). In the first phase of simulation, each node has 2J energy, the data packet for this model is 2000 bytes. The beginning of a network with a base station and device node is exposed in Fig. 4.4. In this paper, we compare the work we have suggested to Wang [5] jobs. First, we demonstrate our work and then our specific work.

Table 2: Simulation parameters

Parameters	Value
Area	100×100 m2
Number of Nodes	50
Coordinate of BS	(50,180)
The initial energy of the node	2J
Packet size	2000 bits
Energy for data aggregation	5nJ/bit/signal
Energy consumption on the circuit	50nJ/bit

A. Wang [5]

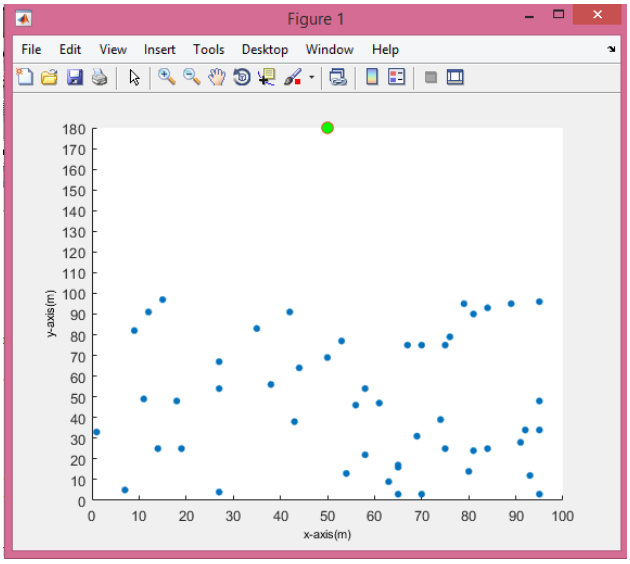

Fig.4. Network Initialization

Fig.5 shows the structure of the proposed form in the scenario which is given. CH of all clusters creates a tree for inter-cluster communication.

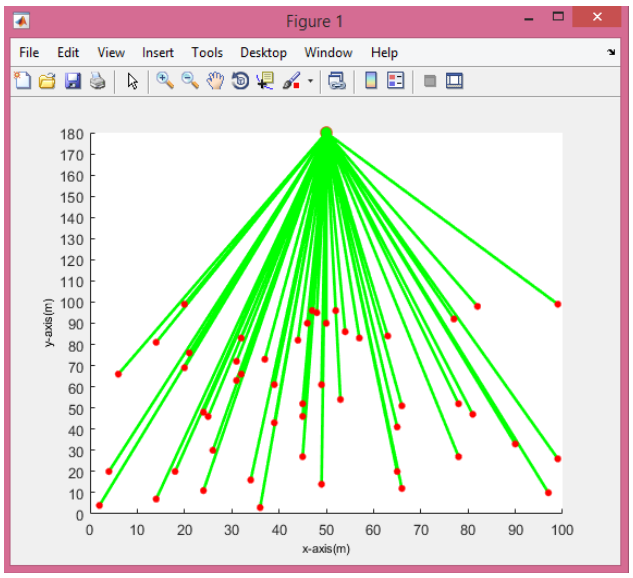

Fig.5. Inter-cluster communication

Fig.6 and fig.7 show the variance of alive nodes in different rounds and variance of energy in a different number of rounds correspondingly.

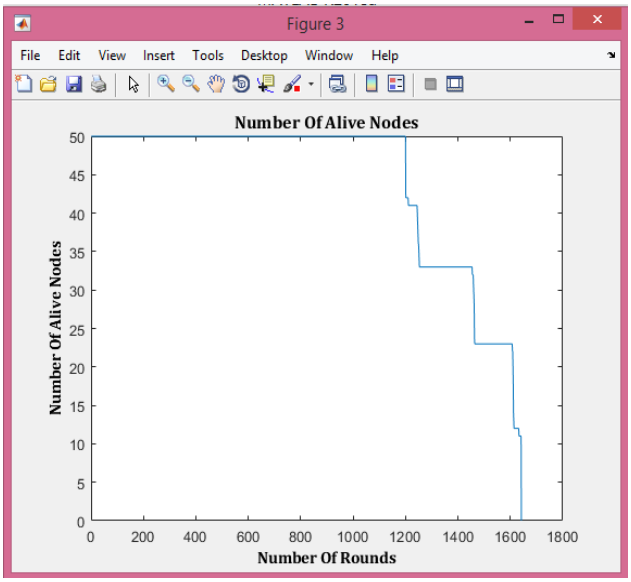

Fig.6. No. of alive nodes over no. of rounds

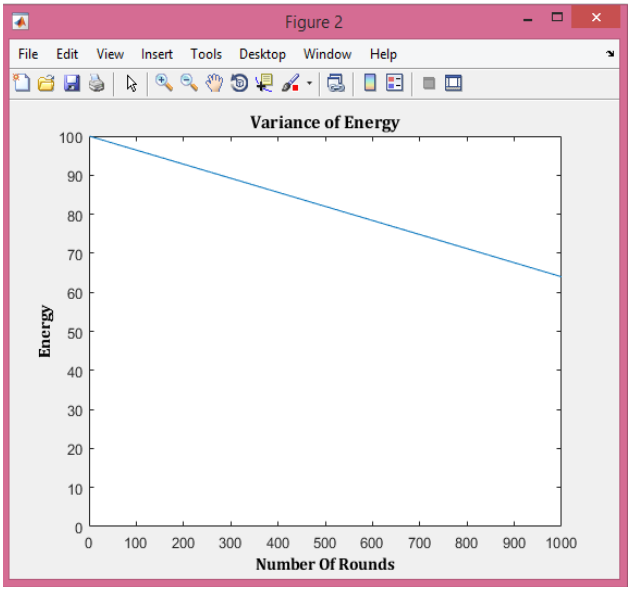

Fig.7.The variance of Energy level

B. Our Proposed:

Fig.8 shows the proposed structure of the given scenario. The structure is formed using a Dijkstra algorithm. CH of all clusters creates a tree for inter-cluster communication.

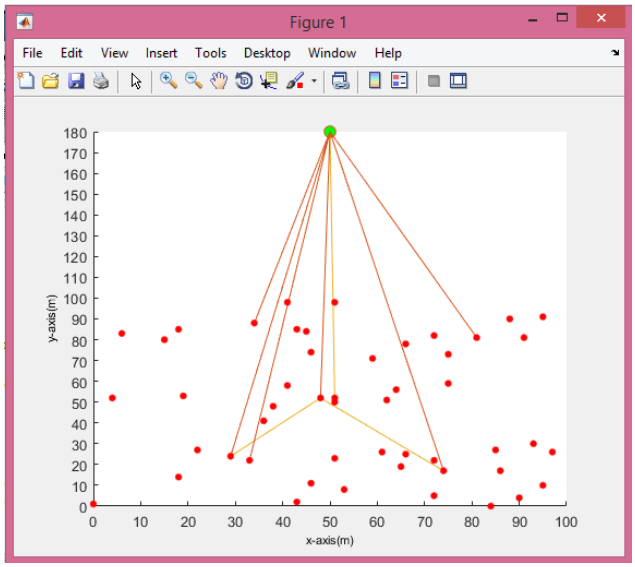

Fig.8. Inter-cluster communication

Fig.9 shows the variance of alive nodes in different rounds. And fig.10 shows energy variance in the various rounds of no.

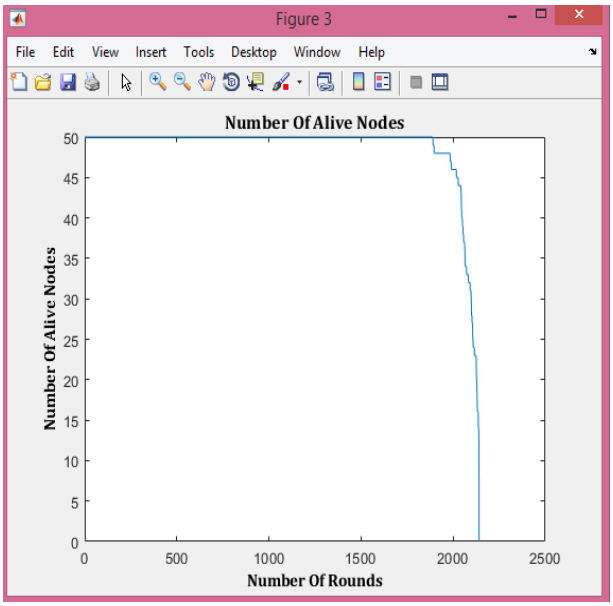

Fig.9. Comparison of no of nodes alive in every round

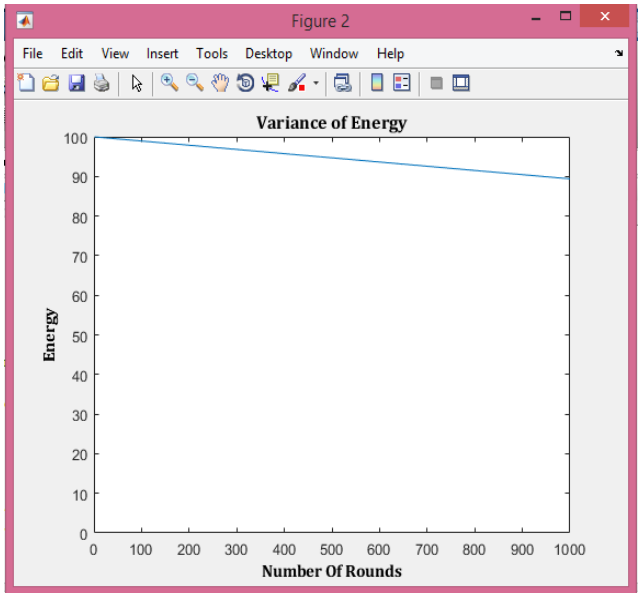

Fig.10. Comparison of energy in each round

In table 3 below table, we show the comparison among the three parameters such as Energy left, First Dead Round and Last Dead Round.

Table 3: Parameters with Base and Proposed Value

Parameters	Wang [5]	Our Proposed
Energy Left	63.973945	89.389925
First Dead Round	1201	1891
Last Dead Round	1642	2141

VI. CONCLUSION

The main problem with WSNs is the efficient network lifetime of reducing energy usage in network nodes. The K-medoids algorithm was used to create cluster heads for each cluster. These CHs use a load sharing system to rotate into an active cluster scale. It will protect the remaining energy of the nodes and extend network life. In addition, reducing the number of clustering again and the number of data packets sent during network operation significantly increased. It has detected suggested results to improve its existing clustering algorithms under its unique multi-cluster head media. In this paper, the shoretest path formed to transmit the data from source to BS. Dijikstra algorithm used to form the free and by this the performance of the network improved and it can be explained from the results in terms of energy and nodes lifetime.

REFERENCES

1. Sanjeev Kumar Gupta and, Poonam Sinha, "Overview of Wireless Sensor Network: A Survey", International Journal of Advanced Research in Computer and Communication Engineering Vol. 3, Issue 1, January 2014, pp. 5201-5207.
2. Chunjuan Wei, Junjie Yang, and YanjieGao,"Cluster-based routing protocols in wireless sensor networks: A survey", Published in "Computer Science and Network Technology (ICCSNT), 2011 International Conference on", Dec. 2011.
3. Dr.S. SelvaNidhyananthan, S. Surya and H. Basheer Ahamed

Batcha, "Fuzzy Logic-Based Clustering Algorithm to Improve Lifetime of Wireless Sensor Networks", International Journal of Pure and Applied Mathematics, Volume 119 No. 12 2018, 13601-13607.
4. Lehsaini, M., &Benmahdi, M. B. (2018). An improved K-means cluster-based routing scheme for wireless sensor networks. 2018 International Symposium on Programming and Systems (ISPS).doi:10.1109/isps.2018.8379004
5. Jin Wang, Kai Wang, JunmingNiu, Wei Liu, A K-medoids based Clustering Algorithm for Wireless Sensor Networks, International Workshop on Advanced Image Technology (IWAIT), 2018.
6. Mazinani, A., Mazinani, S. M., &Mirzaie, M. "FMCR-CT: An energy-efficient fuzzy multi cluster-based routing with a constant threshold in wireless sensor network", Alexandria Engineering Journal, Volume 58, Issue 1, March 2019, Pages 127-141.
7. Amri, S., Khelifi, F., Bradai, A., Rachedi, A., Kaddachi, M. L., &Atri, M. "A new fuzzy logic based node localization mechanism for Wireless Sensor Networks", Future Generation Computer Systems, Volume 93, April 2019, Pages 799-813.
8. Mehra, P. S., Doja, M. N., &Alam, B. (2018). Fuzzy based enhanced cluster head selection (FBECS) for WSN. Journal of King Saud University – Science, pp. 1-19.
9. Zhang Siqinga, Tao Yanga, and Yang Feiyuea, "Fuzzy Logic-Based Clustering Algorithm for Multi-hop Wireless Sensor Networks", Procedia Computer Science 131 (2018) 1095–1103.
10. M. Uddin Harun Al Rasyid1, Bih-Hwang Lee2, Iwan Syarif1, MokhammadMuqoffi Arkham1, LEACH Partition Topology for Wireless Sensor Network, 2018 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW).
11. Z. Fan, M. Cai, X. Li, "A New Dynamic Energy-Based Clustering Model for Wireless Sensor Networks," IEEE International Conference on Computational Science and Engineering (CSE) and Embedded and Ubiquitous Computing (EUC), 2017.
12. Senouci, Mustapha &Mellouk, Abdelhamid &Senouci, Mohamed &Oukhellou, Latifa. (2014). Belief Functions in Telecommunications and Network Technologies: An Overview. annals of telecommunications - annales des télécommunications. 69. 135-145. 10.1007/s12243-014-0428-5.
13. Gui, Tina & Ma, Christopher & Wang, Feng & Li, Jinyang& E. Wilkins, Dawn. (2016). A novel cluster-based routing protocol for wireless sensor networks using Spider Monkey Optimization. 10.1109/IECON.2016.7794106.

AUTHORS PROFILE

Garima Sharma currently M.techStudent from Madhav Institute of Technology & Science, Gwalior, Madhya Pradesh, India. She has obtained B.E. degree in 2016 from Rustamji Institute of Technology, BSF Academy, Tekanpur Gwalior, Madhya Pradesh, India. She has recently attended and successfully publishing paper in the series Algorithms for Intelligent System(AIS), Springer Nature. Her current research interest includes Wireless Sensor Networks.

Praveen Kumar currently working as an Assistant Professor in MITS , Gwalior, Madhya Pradesh, India. He has done M.tech from NIT Rourkela in 2016 and B.tech from A. K University, Patna, Bihar, India in 2013. He has attended two International Conference and also presented papers on IOT.

Dr. Laxmi Shrivastavacurrently working as an Associate Professor in MITS , Gwalior, Madhya Pradesh, India. She has received Ph.D Degree amdM.tech degree from RGPV, Bhopal, India. She is graduated from G.B. Pant University of Agriculture and echnology, Udham Singh Nagar, Pantnagar, Uttarakhand, India. She has published 56 Conference Papers and 25 Journal Papers. Her Research Interest includes Wireless and Ad-hoc Networking.