GMM-DT: A Novel Model of Gaussian Mixture Model-based Clustering using Dijkstra Algorithm in WSN

Karishma Verma
Dept. of EC
Maharana Pratap College of Technology
Gwalior, India
verma.karishma07@gmail.com

Pankaj Shrivastav
Associate Professor
Dept. of EC
Maharana Pratap College of Technology
Gwalior, India
pankajsamyak@gmail.com

***Abstract*— Wireless sensors network as an energy consumption technique has been widely discussed in several studies. However, factors such as cluster formation or Cluster Head (CH) assignment node approach have a dramatic effect on network performance as reducing energy efficiency involves restricted resources or network data traffic. For this purpose, we have developed a new model which consists of the Gaussian mixture model and the Dijkstra tree. The Gaussian Mixture model is used for clustering to generate clusters. The key purpose of this approach is to reduce sensor node communication by using clustering techniques. In this paper, Gaussian Mixture is used to measure a log-likelihood for sensor nodes to obtain an optimal result of a cluster. The related cluster heads may then be selected. The cluster head selection is done based on distance from a base station and the energy level of nodes. Dijkstra shortest path tree algorithm is used later for establishing the communication among the CHs hence making a secure connection for nodes. Simulation findings show the efficiency in terms of energy consumption or cellular network life of our proposed algorithm over the existing popular protocol. The results are compared with the R-LEACH protocol in MATLAB and parameters like nodes dead rounds, energy consumption has been calculated. The balancing of network demands among clusters has increase energy efficiency or effectively prolong network life.**

***Keywords*— WSN, IoT, Clustering, CH Selection, Residual Energy (RE), Gaussian Mixture Model, Dijkstra Tree.**

I. INTRODUCTION

IoT is considered pattern innovation's multifaceted region. In areas with a wide variety of businesses, IoT's innovation is attracting further recognition. The fundamental invention of IoT is commonly known for companies, where related processes engage with each other and are strongly connected to merchant compliance schemes, client aid schemes, corporate brainpower supplication & corporate logic. IoT consists of Wireless Sensor Networks (WSNs), where sensors link to the Internet without any human intervention and contact each other. WSN and IoT security is a serious concern, rather than a simple one. [1,2].

WSNs should be considered the core components of IoT and it can support users (humans or machines) communicate and respond to actual events. These WSNs consist of nodes that combine microelectromechanical, wireless and digital systems and can perceive their computing and communications world. This technique is costly and leads as new applications are envisaged to be redundant. It is not impossible to predict that potential WSN implementations would concurrently serve many applications with the advent of the IoT [3, 4, 5].

With the advent of WSN and its utilitarian useful applications, nowadays it is speedily introduced in researches to investigate the environmental conditions of remote and esoteric regions. WSN is the collection of wireless sensor nodes consists of small powered batteries with limited resources. These sensor nodes are used to capture the data from the environment, process it & sends data packets to a destination. While doing a process of reception and transmission large amount of energy has been consumed. Due to limited resources of power, energy consumption becomes a major issue for WSN design. At the same time number of energy-efficient routing algorithms have been established to minimize energy consumption & improve the lifetime of the sensor network [6].

Clustering is a division of data and related object groups. Active research in many areas, like analytics, machine learning algorithms, is carried out in clustering. Clustering is a crucial technology to extend sensor network life by reducing energy. Cluster formation makes a scalable network sensor. Cluster member is also called CH. Sensor in a cluster may elect a CH or the network designer can pre-assign it. Various clustering algorithms for scalability & effective communication were developed specifically for WSNs. the idea of cluster-based routing is often applied for effective WSN routing. The key strategy for energy conservation is clustering. [7].

The number of algorithms for clustering was suggested to advance the lifespan of the WSN. The WSN is divided into groups in clustering algorithms, which are called clusters, as well as CH is selected from the one sensor knot from each cluster. All data aggregation actions within the cluster were done, preceded by CH (cluster head) to BS (base station), also a sinks node, to relay information from a specific cluster. Periodic CH selection within clusters is recommended for stability energy usage in each cluster. [8]. the uniform distributed CH will optimize energy usage between Sensor Nodes (SNs) or expand the life of the network. The system may not always be efficient by the non-uniform distribution of sensor nodes, methods used to regulate power use, or extend network lifetime. Clusters have consistent cluster areas and can even out resource uses throughout clusters or nodes by uniformly spaced CH cluster heads. However, owing to the non-uniform node distribution of sensors, the imbalanced power utilization proceeds among CHs.

The most effective technique for energy efficiency is clustering. SNs are arranged in groups called clusters in these techniques. The cluster's regular nodes are related to as cluster members & a CH) is selected. The WSN clustering architecture is shown in Fig. 1. There are two kinds of traffic in clustered WSN: intra-cluster data transmission & data transmission among clusters called inter-cluster traffic. The members of the cluster since the variables of the real world to transfer the value to their CH. The Chamber of Commerce collects and adds data to delete obsolete data and to directly or via intermediate chamber transmissions. The members of the cluster cannot transfer the data to BS and they send it to CH only. Clustering has the benefit of reducing energy use by enhancing bandwidth use, reducing overhead, enhanced connectivity, stabilizing the topology of the network, decreasing delays, effective load balancing & reduced the size of routing. Close BS the CHs absorb more energy as well as drain faster than the CHs away from BS. CHs closest to BS is charged with heavy traffic through traffic intra-clusters, data collection, and other CH inter-cluster traffic for data relay to BS. This causes network access disruption and coverage problems in clusters closer to BS. Figure 1 shows the architecture of clustering in WSN. It shows the general concept of clustering and cluster head for data transmission towards BS [6].

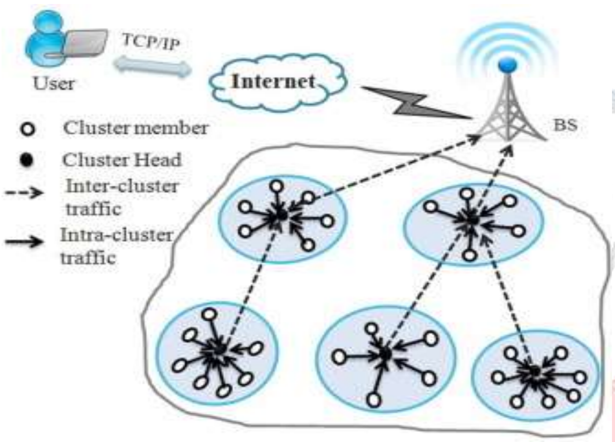

Fig. 1. The architecture of clustering in WSN

The remaining paper is structured as according to section II discusses literature survey. Section III tells about the problems that exist in the previous R-Leach model that introduce a GMM-DT model. It also shows the flowchart for this proposed model. Then section IV simulates the results for described simulation network parameters. At last, this paper is concluded in section V.

II. LITERATURE SURVEY

Altakhayneh, W. A. [9] This paper evaluates a genetic LEACH(G-LEACH) for a total number of live nodes, energy utilization, CH's, and the packet distribution to CHs in 100 uniformly distributed WSNs. This paper includes: CH selection is accomplished by the application of a genetic optimization technique that can evaluate effective and usable cluster heads. G-LEACH has some benefits over LEACH as G-stability LEACH's zone is larger by 358 rounds than the stability zone since the first node dies following the 1544 round of G-LEACH. The results are seen by the results in this study. That's the mean; 61.7 percent longer stays on the network with G-LEACH rather than LEACH. At the same time, G-LEACH increases CHs by 10% over LEACH.

Malshetty, G., & Mathapati, B. [10] implemented a clustering method called LBSO, which helps to balance the node. The key objective of their algorithm is to achieve a better clustering and the dynamic clustering method, dividing the technique into three distinct stages, and selecting CH for the first phase, the second stage consists of building the cluster. The third stage comprises the rotational process in which the CH is chosen. A large-scale simulation has been conducted to test the algorithm including comparative analysis is carried out and we equate our methods with the current LEACH protocol and our algorithm evaluates the traditional protocol.

Islam, S. et al. [11] The suggested protocol was tested and contrasted with LEACH-C & the selection of the efficient CH (ECHS). They implemented their protocol and observed significant efficiency improvements for 1st Node Death (1ND), End Node Death (END), and power vs. round consumption.

Dhivya, K. [12] To advance critical stinting of the finder sensor hubs, efficient plan as appropriate CH assurance plans are provided. In this paper, the vibrant DCHSM protocol is mostly used CHs are obtained in 2 steps. This calculation improves mandatory stinting on a broad scale and is hence also used for IoT implementations. Initially, the QB Cluster model is used for the segmentation of insights of molded groups of points. In 2 steps, CH Assurance is carried out.

Aalavandhar, A., & Arjunan, A. [13] Focusing on various security parameters that are built and applied using the Markov process. Markov is a pioneering and rapidly evolving mathematical model used in the selection of the CHs in WSN. In the analysis of the credibility of nodes, rather than another mathematical process in sensor networks, the Markov model offers a high level of security.

Manzoor, K. et. al. [14] the routing of the sensor data to BS from the SN is a difficult task in a WSN framework. There are some limitations on SNs, like low energy, low memory, computing capacity, etc. These limitations must be considered when designing the RP to ensure effective maintenance of network life and data robustness overall.

III. RESEARCH METHODOLOGY

A. Problem Formulation

Typically, WSNs contain thousands of resource-limited sensors to monitor their setting, gather data, and upload it for further processing to remote servers. Although WSN is considered extremely flexible ad hoc networks, given their deployment scale and their related efficiency issues, like resource management or reliability, Network management was an important challenge for these forms of networks. Because of scarce funds in WSN, direct sensor node communications to BS or BS multi-hop sensor node communication are not feasible.

B. Proposed Approach

1) GMM

GMM [15] is a probability model of a mixture of Gaussians. It seems to be closer to the natural distribution and an easy model to do mathematical manipulation for having Gaussian function. If the distribution is not Gaussian in nature then different methods can be used to form the Gaussian-like clusters having Gaussian distribution and this analysis is based on GMM. The univariate Gaussian distribution (or "normal distribution," or "bell curve") is the distribution in which the result is the average of events that occurs again and again.

$$G(x|\,\mu,\sigma) = \frac{1}{\sqrt{(2\pi\sigma 2)}}\, e^{\frac{(x-\mu)2}{2a^2}} \quad (1)$$

Where G is the Gaussian function, μ is mean & σ 2 is variance. The mean (μ) indicates the maximum likelihood and variance (σ^2) is the deviation from the maximum likelihood within a univariate Gaussian distribution field. If Gaussian is multivariable, the univariate normal is generalized with two or more variables [16][17]. It has the mean vector μ & covariance matrix Σ to be parameterized.

$$N(x|\mu,\sum) = \frac{1}{\sqrt{(2\pi|\sum|)}} e^{[=\frac{1}{2}(x-\mu)^T \sum^{1}(x-\mu)]} \quad (2)$$

The mean (μ) indicates the maximum likelihood and Σ is the covariance between different Gaussian distribution fields [18].

$$\ln p(x|\mu,\sum) = \frac{1}{\sqrt{(2\pi|\sum|)}} = -\frac{1}{2} In(2\pi) - \frac{1}{2} in|\sum| - \frac{1}{2}(x - \mu)T\sum - 1(x - \mu) \quad (3)$$

It is simply a single Gaussian distribution where then we can set the derivative of ln p(x|μ,∑) to zero

$$\frac{\delta In\, p(x|\mu,\sum)}{\delta\mu} = 0 \quad (4)$$

$$\frac{\delta In\, p(x|\mu,\sum)}{\delta\sum} = 0 \quad (5)$$

Solve directly for μ and ∑.

$$\mu = \frac{1}{N}(\sum_{N=1}^{N} x_n) \quad (6)$$

$$\sum = \frac{1}{N}\sum_{n=1}^{N}(x_n - \mu)(x_n - \mu)^T \quad (7)$$

2) Dijkstra Tree

Dijkstra Algorithm (DA) [19] for the minimum spanning tree is very related to the Prim algorithm. Including Prim's MST, we create the SPT (shortest path tree). We create 2 sets, one set contains a vertical in the shortest path tree, another set contains vertices not included in the SPT yet. Each stage of the algorithm contains a vertex (not yet included) that is in other sets with a minimum distance from the source.

Algorithm:

The following stages are being applied to find the SP from one source vertex to all other vertices in the specified graph in a precise DA.

1) Construct set sptSet, which tracks vertices that are involved in SPT, i.e. which calculates & finalizes minimum distance from the source of sptSet. This set is currently void.
2) Allocate the distance value in the input graph to all vertices. Create as INFINITE all distance values. Allocate distance value 0 to the root vertex to be selected first.
3) Since sptSet does not represent all vertices:

    a) Choose a vertex u that does not have a minimum distance value in sptSet.
    b) Enter u for sptSet
    c) Modify the adjacent vertices of u's distance value. Use all adjacent vertices to modify distance values. If the sum of the distance of u & weight of edge u-v for every adjacent vertex v is less than the value of v, then update the value of v to the value of distance v.

Example: Let us be clear about the below:

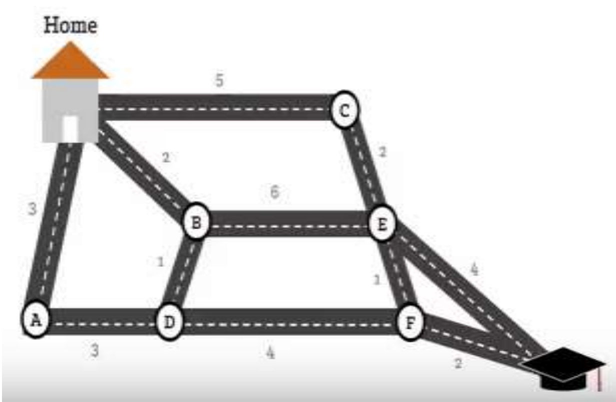

Fig. 2. Flow diagram of Dijkstra's

SP that can be identified using the algorithm in Dijkstra is Home→*B*→*D*→*F*→School.

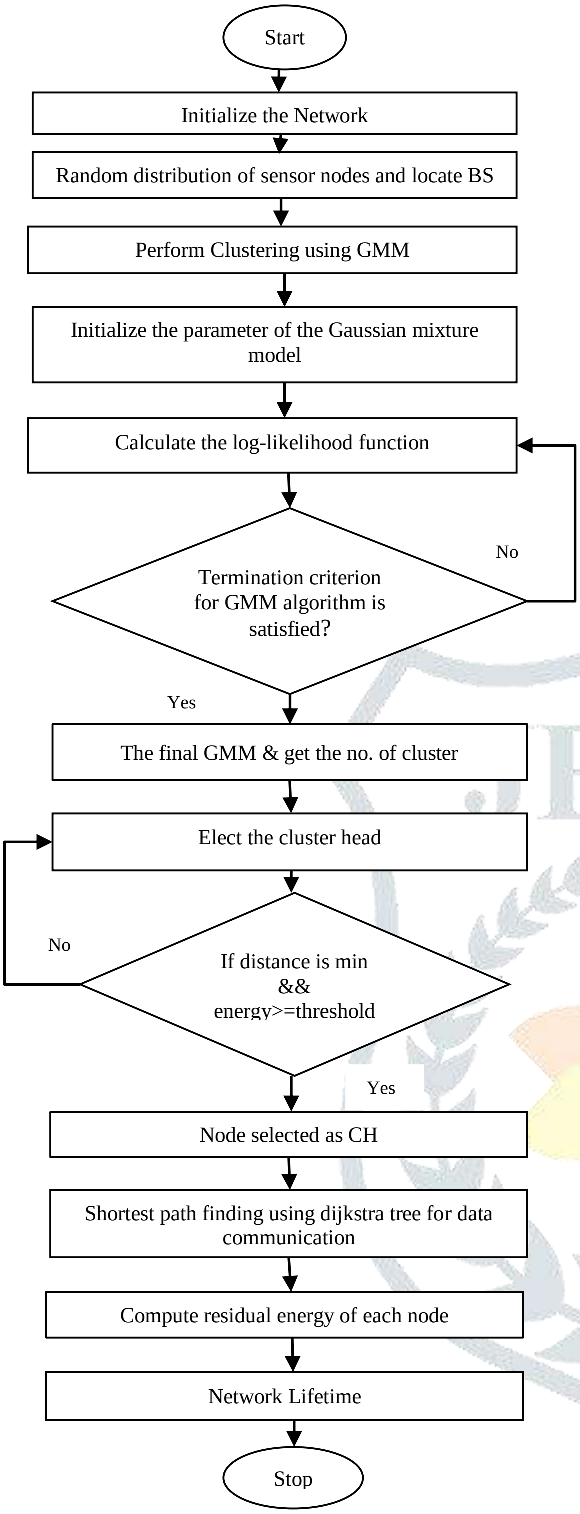

Fig. 3. Flow diagram of proposed System

IV. RESULTS AND DISCUSSION

Table 1 includes a description of network parameters assumed for MATLAB model simulation. The size of a packet is 4000-bit. As seen in Figure 4, 100 nodes are distributed randomly of BS in a network area. Network model and Network Lifetime are two main metrics for consistently delivering better network output in an IoT system. Stability of the network is the time from initiation of the network, before FND (death of the first node), & time from FND & LND (death of the last node) in the network, decide a time that has elapsed. We evaluate parameters like network stability FND, network life LND, or network HND (Half node death) for our analysis of the behavior of the proposed Gaussian mixture model for clustering model. For network analysis, initial energy E_0 of 0.25, 0.5 & 1J is considered.

Table I. Simulation parameters

Parameters	Value
Network diameter	100 meters^2
Total no. of nodes (n)	100 nodes
Total network energy (E_0)	0.25J, 0.5J, 1J
Energy dissipation receiving (E_{amp})	0.0013 pJ/bit/m^4
Energy dissipation free space model (E_{fs})	10 pJ/bit/m^2
Energy dissipation power amplifier (E_{amp})	100 pJ/bit/m^2
Energy dissipation aggregation (EDA)	5 nJ/bit

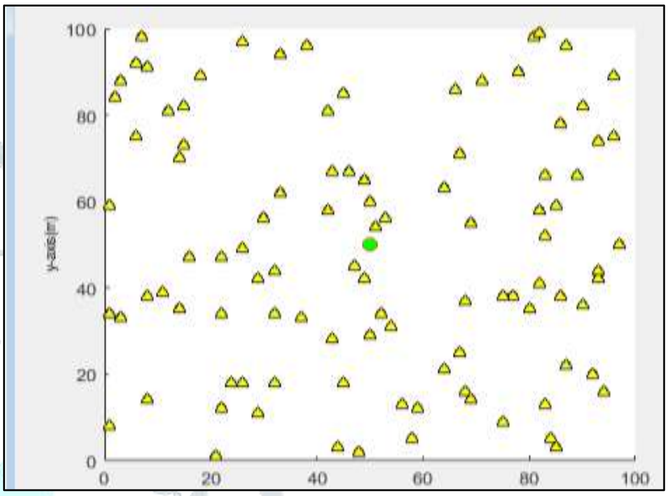

Fig. 4. Node Deployment with total dead node rounds by R-LEACH (.25 J)

Figures 4 and 5 display the resultant at 0.25J as initial energy. The figures clearly show values of all the dead rounds found in R-LEACH, which are further improved by the Gaussian mixture model for clustering where the life of dead rounds is increased hence the Gaussian mixture model for clustering model is capable to enhance the network lifespan of the nodes working in an area.

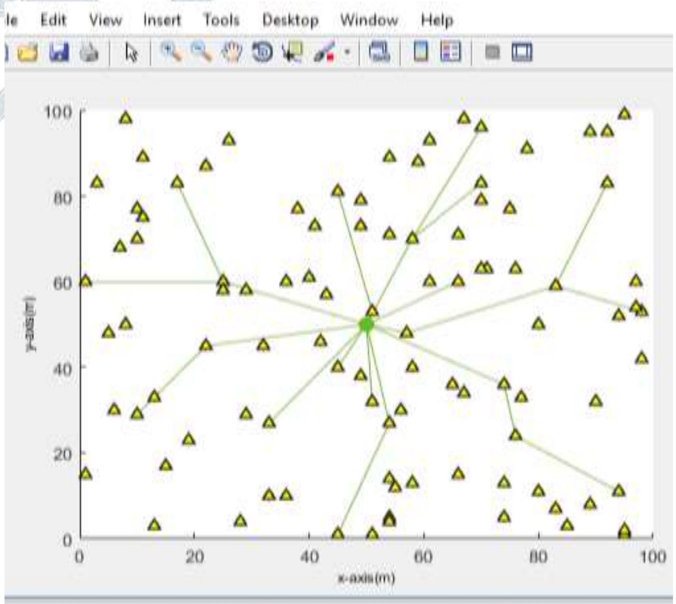

Fig. 5. Node Deployment with total dead node rounds by GMM-DT model for clustering initial Energy (0.25 J)

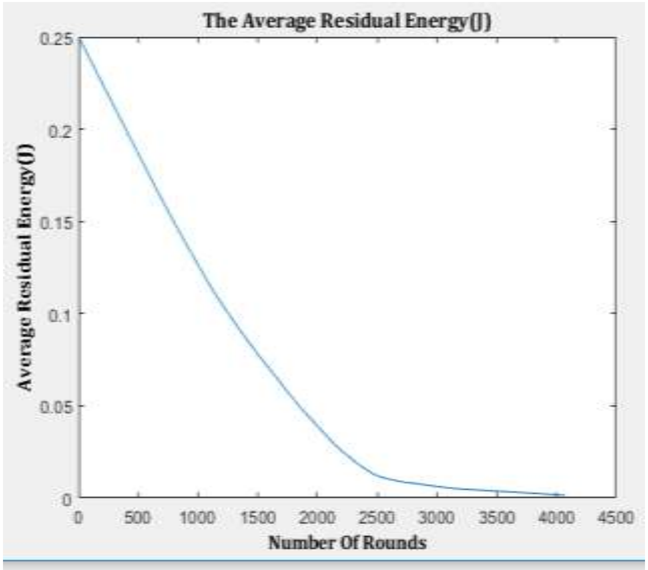

Fig. 6. GMM-DT model average RE (.25J)

Figure 6 demonstrates the average residual energy of the proposed GMM-DT at a 0.25-joule energy level. The Gaussian mixture model for clustering has a slight variation because the residual energy in the Gaussian mixture model for clustering the energy has decreased rapidly at 2500 rounds then it decreases constantly. Hence, we are capable to run our model till 4000 rounds having no remaining energy.

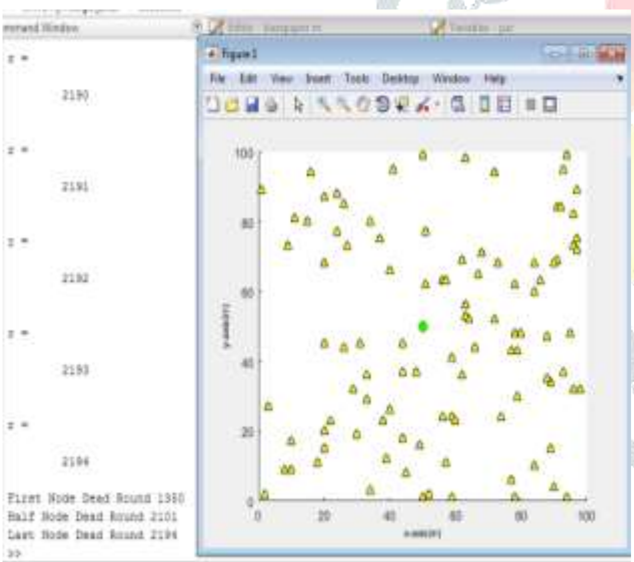

Fig. 7. Node Deployment with total dead node rounds by R-LEACH (0.5J)

Fig. 7 displays node deployment of R-LEACH protocol with initial energy (0.5 Joule) for network life. The simulation outcome demonstrates that at 1350 rounds first R-LEACH node dies, while it is at 1326 rounds for the proposed algorithm i.e. Gaussian mixture model for clustering. Likewise, R-LEACH's last node dies at 2194 rounds, while the Gaussian Mixture model for clustering is still alive as shown in fig 8. The last node that died for Proposed GMM-DT is 6846 rounds. R-LEACH protocol assumes that for each round CHs dispense the same energy, which results in an inefficient selection of CHs or affects the durability of the network. The proposed protocol, the Gaussian Mixture Model for clustering, chooses CHs to maximize network life by considering the remaining node energy and the optimum number of clusters.

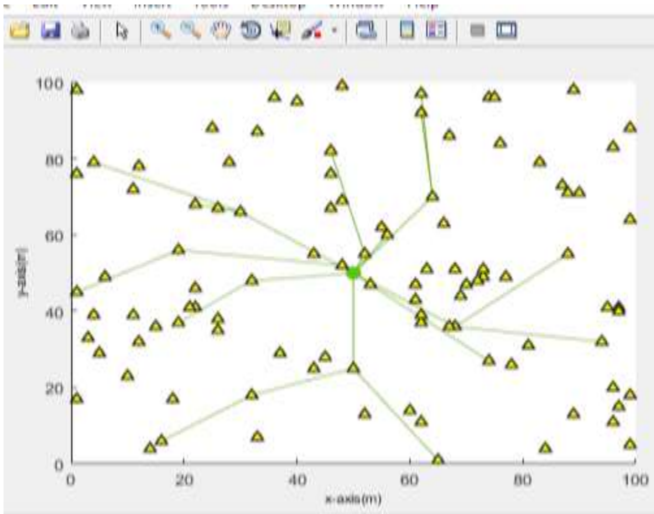

Fig. 8. Node Deployment with total dead node rounds by GMM-DT initial energy (0.5 J)

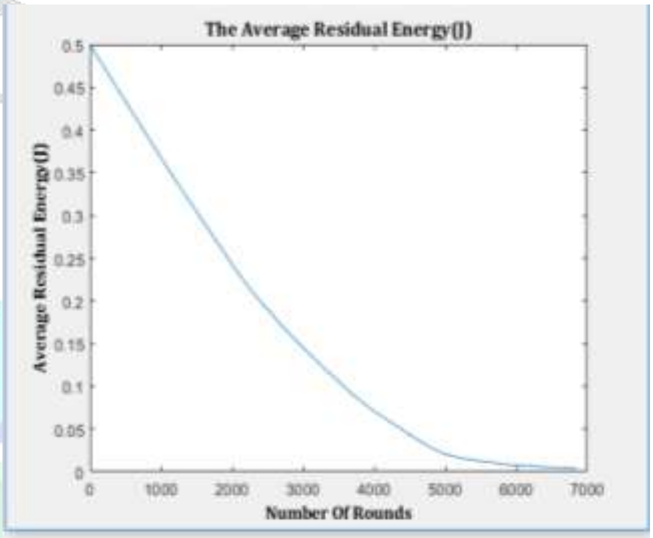

Fig. 9. GMM-DT model average RE (0.5J)

Figure 9 shows the average residual energy of the proposed GMM-DT at a 0.5-joule energy level. The Gaussian mixture model for clustering has a slight variation because the residual energy in the Gaussian mixture model for clustering the energy has decreased rapidly at 5000 rounds then it decreases constantly. Hence, we are capable to run our model till 7000 rounds having no remaining energy.

```
r =
          4029
r =
          4030
r =
          4031
r =
          4032
r =
          4033
r =
          4034
r =
          4035
First Node Dead Round 1350
Half Node Dead Round 3979
Last Node Dead Round 4035
>>
```

Fig. 10. Node Deployment with total dead node rounds by R-LEACH initial energy (1 Joule)

Figures 10 and 11 visualize the dead node rounds of R-LEACH and the proposed GMM-DT model. From the above figures, we noticed that FND, HND, and LND of R-LEACH is at 1350, 3979, 4035 respectively while FND HND and LND of the proposed model are not found as the nodes are not dead even at 5000 rounds hence we can say that the network lifetime of Gaussian mixture model for clustering is more efficient than that of the R-LEACH.

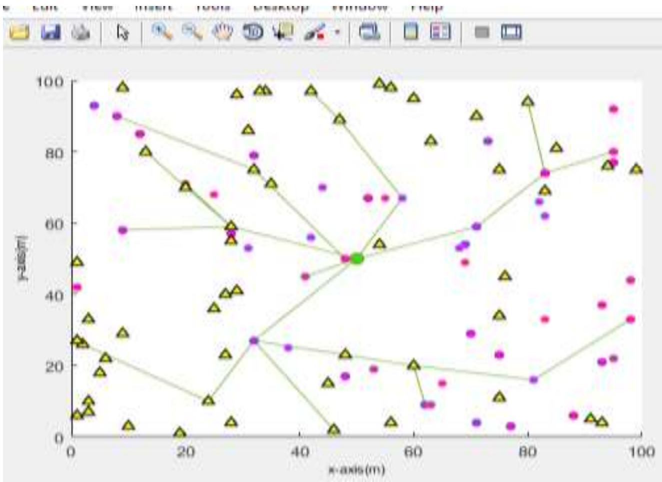

Fig. 11. Node Deployment with total dead node rounds by GMM-DT model for clustering initial energy (1 Joule)

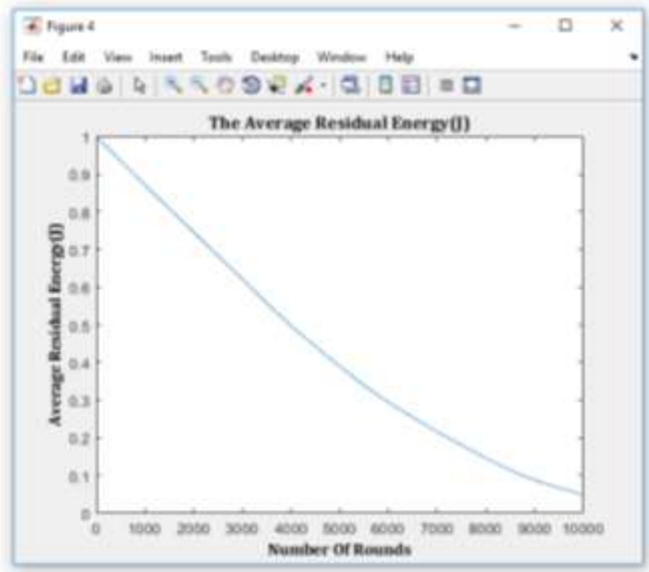

Fig. 12. GMM-DT model average RE (1J)

Figure 12 demonstrations average residual energy of proposed GMM-DT at a 1-joule energy level. The Gaussian mixture model for clustering has a lot of variation as shown in figure 12 because the residual energy in the Gaussian mixture model for clustering the energy has remained even at 5000 rounds. Hence, we are capable to run our model till 10000 rounds having remaining energy of 0.99J.

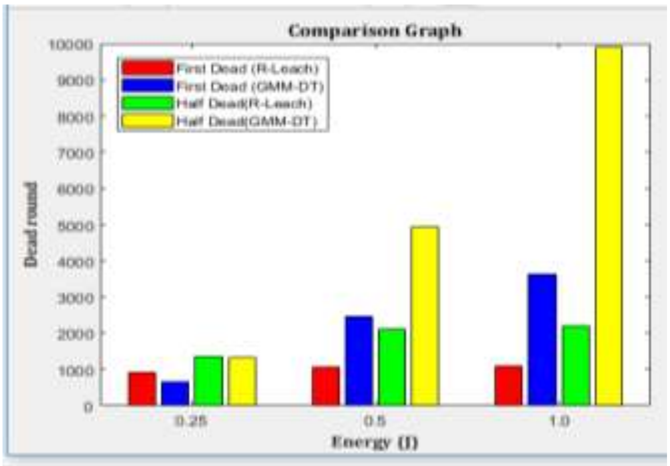

Fig. 13. Dead node comparison between existing R-Leach and proposed GMM-DT

Figure 13 shows a Dead node comparison between the existing R-Leach and proposed GMM-DT. A gaussian mixture model for clustering model aims at increasing the life of the network by covering further rounds of HND & FND for all energy levels, so it selects stable nodes for CH.

V. CONCLUSION

As the energy and lifespan of each routing protocol are two major constraints for WSN, there has been much study into achieving the aim. It is a difficult process to select an effective routing technique, which evenly distributes the load in the network. The R-LEACH protocol allows for an adaptive model but has some drawbacks. This study proposes a new model GMM-DT for performing clustering and CH selection to enhance network life by controlling the energy discharge in the network. In situations like environmental monitoring with IoT, the improved routing mechanism can be effectively used as a protocol that offers better results for a heterogeneous system based on R-LEACH. The proposed model is based on GMM and Dijkstra shortest path tree algorithm. The approach described tested for a WSN based IoT method in various practical scenarios. It performs on different energy levels are 0.25J, 0.5J, and 1J. The results of the simulation show improved network efficiency for measures like residual energy, dead nodes, and network lifetime. More variables for CH selection in a mobile network, which constantly changes its position, can be applied for the latest work. Optimal CH can be determined using optimization techniques like GA, PSO, etc., and can be used in different clustering techniques in the future.

References

[1] C. Worlu, A. A. Jamal and N. A. Mahiddin, "Wireless Sensor Networks, Internet of Things, and Their Challenges," *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, Vol. 8 Issue 12S2, October 2019, pp. 556-566.

[2] M. Mamdouh, M. A. I. Elrukhsi and A. Khattab, "Securing the Internet of Things and Wireless Sensor Networks via Machine Learning: A Survey," *2018 International Conference on Computer and Applications (ICCA)*, Beirut, 2018, pp. 215-218, DOI: 10.1109/COMAPP.2018.8460440.

[3] M. A. Feki, F. Kawsar, M. Boussard, and L. Trappeniers, "The Internet of Things: The Next Technological Revolution," Computer, vol. 46, no. 2, pp. 24–25, 2013.

[4] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: a survey," Computer Networks, vol. 38, no. 4, pp. 393–422, Mar. 2002.

[5] N. M. M. K. Chowdhury and R. Boutaba, "Network virtualization: state of the art and research challenges," IEEE Communications Magazine, vol. 47, no. 7, pp. 20–26, Jul. 2009.

[6] A. Nikolaos Pantazis, A. Stefanos Nikolidakis, and D. Dimitrios Vergados, "Energy-Efficient Routing Protocols in Wireless Sensor Networks: A Survey" IEEE communication survey & tutorials, VOL. 15, No.2, secondquarter2013.

[7] V. Katiyar, "A Survey on Clustering Algorithms for Heterogeneous WSNs," *Int. J. Advanced Networking and Applications* 745, Vol. 02, Issue 04, 2011, pp. 745-754.

[8] P. Kumarawadu, D. J. Dechene, M. Luccini, and A. Sauer, "Algorithms for Node Clustering in Wireless Sensor Networks: a Survey," Information and Automation For Sustainability, ICIAFS, 4th International Conference on, pp. 295–300, Dec. 2008

[9] Altakhayneh, W. A., Ismail, M., Altahrawi, M. A., & AbuFoul, M. K. (2019). Cluster Head Selection Using Genetic Algorithm in Wireless Network. 2019 IEEE 14th Malaysia International Conference on Communication (MICC).

[10] Malshetty, G., & Mathapati, B. (2019). Efficient clustering in WSN-Cloud using LBSO (Load Based Self-Organized) technique. 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI).

[11] Islam, S., Khan, M. N. I., Islam, S. J., & Akhtar, M. J. (2019). Cluster Head Selection Technique Using Four Parameters of Wireless Sensor Networks. 2019 International Conference on Computer Communication and Informatics (ICCCI). doi:10.1109/iccci.2019.8822137.

[12] Dhivya, K., Premalatha, G., & Sindhuparkavi, B. (2019). Energy Efficient Power Management Scheme for WSN Using Network Encoding Protocol. 2019 International Conference on Intelligent Computing and Control Systems (ICCS). doi:10.1109/iccs45141.2019.9065334.

[13] Aalavandhar, A., & Arjunan, A. (2019). Analysis of Markov Model through reliable Cluster Head Selection In WSN. 2019 3rd International Conference on Electronics, Communication, and Aerospace Technology (ICECA). doi:10.1109/iceca.2019.8821799.

[14] Manzoor, K., Jokhio, S. H., Khanzada, T. J. S., & Jokhio, I. A. (2019). Enhanced TL-LEACH routing protocol for large-scale WSN applications. 2019 Cybersecurity and Cyberforensics Conference (CCC). doi:10.1109/ccc.2019.00-12.

[15] Shaveta Gupta and Vinay Bhatia, "GMMC: Gaussian Mixture Model-Based Clustering Hierarchy Protocol in Wireless Sensor Network", International Journal of Scientific Engineering and Research (IJSER), Volume 3 Issue 7, July 2015, pp. 44-49.

[16] Reynolds, D.A.: "A Gaussian Mixture Modeling Approach to Text-Independent Speaker Identification". Ph.D. thesis, Georgia Institute of Technology (1992).

[17] Jeff A. Bilmes," A Gentle Tutorial of the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models" International Computer Science Institute.

[18] Gaussian mixture models and the EM algorithm Ramesh Sridharan.

[19] *Skiena, S. Dijkstra's Algorithm Animation. Retrieved April 22, 2016,* from http://www3.cs.stonybrook.edu/~skiena/combinatorica/animations/anim/dijkstra.gif