THE DEVELOPMENT OF AN ENHANCED WIRELESS SENSOR NETWORK LIFESPAN USING HYBRIDIZED CLUSTER AND K-MEANS ALGORITHM

[1]J.O. Amoo, [2]O. Osunade, [3]O.E. Makinde, [4]D.N. Ajobiewe & [5]E.O. Adebayo

[1,4,5]Department of Computer Science,
School of Secondary Education (Science Programmes)
Federal College of Education (Special), Oyo.
Email: lekanamoo247@gmail.com

[2]Department of Computer Science,
University of Ibadan, Nigeria.
Email: seyiosunade@gmail.com

[3]Department of Computer Science,
Ajayi Crowther University, Oyo.
Email: eo.makinde@acu.edu.ng

Abstract

Wireless sensor networks (WSN) are useful in situations where a low-cost network needs to be set up quickly and no fixed network infrastructure exists. Typically, its applications are for monitoring system such as (environmental, air and water quality etc), natural disaster prevention, forest fire prevention, military exercises and emergency rescue operations. Wireless sensor network consists of large number of a group of small sensor nodes which utilizes battery and are with limited energy and life span that makes these sensors node inefficient, non-scalable, in- mobile and non-load balance. The nodes close to each other have more overlap; they sense the same data from environment and cause a waste of energy by generating repetitive data. In this study, a hybridized clustering and k-means algorithm are introduced, in each round, a certain number of nodes are specified; the nodes which have at least one neighboring node at a distance less than the threshold. Then, the wireless sensor network nodes were partitioned into different clusters and one of the nodes elected from each cluster to act as a cluster head with the use of a hybridized cluster and k mean algorithm. Among them, the nodes with less energy and greater overlap with their neighbors have been chosen to go to sleep mode. Also, the energy imbalance among sensor nodes is reduced by integrating the distance of the nodes from the base station into clustering policies. A simulator was developed with python programming language to evaluate the performance of this system. On the basis of the simulation results, the proposed scheme elongates the network life span, then, balances the load among sensor nodes further.

Key words: *Sleep mode, network infrastructure, life span, overlap, cluster head, elongate.*

Introduction

Since the need for data transfer from one place to another is part of our everyday events, the invention of wireless sensor networks is a great tool which helps to facilitate data transfer as well as helps to save resources and energy. Wireless sensors are small devices that make local measurements such as environmental conditions like temperature or pressure and contain the hardware necessary to transmit this information to other devices. They may be dropped or scattered over an area to provide information on current conditions. The sensors are typically reliant on internal battery power and therefore have a limited lifespan (Abbasi and Younis, 2007). The sensors may form a network known as a wireless sensor network (WSN) in order to communicate with one another and also send information to a special node on the network called a base-station or sink node. The sink node transmits information to other systems where it is used for analysis (Akyildiz, Su, Sankarasubramaniam, & Cayirci 2002). Wireless Sensor Networks have no pre-determined infrastructure and therefore the network must be set up in an ad hoc manner. WSN nodes may have limited mobility such as sensors floating in the ocean but generally are in fixed positions.

According to Rault, Bouabdallah, & Challal (2014). The design of these sustainable wireless sensor networks (WSNs) is a very challenging issue. On the one hand, energy-constrained sensors are expected to run autonomously for long periods. However, it may be cost-prohibitive to replace exhausted batteries or even impossible in hostile environments. On the other hand, unlike other networks, WSNs are designed for specific applications which range from small-size healthcare surveillance systems to large-scale environmental monitoring. Thus, any WSN deployment has to satisfy a set of requirements that differs from one application to another. There are many other situations where WSNs are useful including military applications, emergency response, or natural disasters (Abbasi and Younis, 2007). Unlike traditional wired networks which typically have dedicated hardware to route network traffic, each node in a WSN acts as a router. Each WSN node either communicates directly with other nodes within its transmission range or uses other nodes to relay messages to nodes outside its range (Zhou and Haas, 1999).

These same features that make WSNs attractive for the uses noted above also create additional resource and security issues as compared to standard wired and wireless networks (Blum, Eskandarian, & Hoffman 2004). With no central authority or management, each node in a WSN, along with acting as a router, must also run its own intrusion detection (Safa, Artail & Tabet, 2010). An additional challenge with WSNs is the limited resources available. The mobile devices used as nodes of a WSN typically have limited battery power and therefore any protocols devised for WSNs should strive to limit increases in network overhead and CPU load (Abbasi and Younis, 2007). There are many attempts in the literature to improve WSN resource consumption by implementing a clustering algorithm, either separately (Cheng, 2012; Abbasi and Younis, 2007) or as part of a routing protocol (Hajami, Oudidi and ElKoutbi, 2010; Safa *et* *al.*, 2010). The idea behind these efforts is that groups of nodes can be clustered together and a cluster head can be chosen to act on behalf of the cluster.

The author implemented a clustering scheme called “Hybrid Energy-Efficient Distributed” (*HEED)* Clustering that was based on the residual energy of the nodes along with a “secondary parameter such as proximity to its neighbors or node degree” (Younis & Fahmy, 2004). The clusters were configured such that each node was a member of exactly one cluster and cluster heads were located such that all nodes were within transmission range of at least one cluster head. *HEED* (Younis & Fahmy, 2004) was compared to an existing algorithm from the literature known as “Low Energy Adaptive Clustering Hierarchy” (*LEACH)* (Heinzelman, Chandrakasan & Balakrishnan, 2002). *LEACH* had been shown to improve network lifetime significantly over static network clustering (Heinzelman, Chandrakasan & Balakrishnan, 2002).

Clustering was also noted to increase network lifespan through “reducing the number of nodes contending for channel access” (Younis and Fahmy, 2004) and “routing through an overlay among cluster heads, which has a relatively small network diameter” (Younis and Fahmy, 2004). Therefore, the problem of clustering WSN nodes can be viewed as a graph-partitioning problem.

Statement of Problem

It was established that wireless sensor network consist of large number of a group of small sensors node that uses battery and are with limited energy and lifespan hence makes these sensors node to be inefficient, non-scalable and non-load balance. Also, it has been widely established that the transmission and reception per bit is much larger than sensing and processing energy per bit.

In general, the raw data of a node’s measurement is of large volume. Transferring raw measured data not only consumes large amount of energy but also increases network traffic which poses high bandwidth demand.

Aim and Objectives

The aim of this research work is to enhance the lifespan of wireless sensors network through a designed hybridized cluster and k-means algorithm model while the following are objectives of the study:

1. To use the developed model to calculate the distance between nodes for an appropriate clustering and selection of a cluster head
2. To compare the performance of this clustered WSN with dispersed nodes type in term of lifespan elongation

Scope of the Study

The researchers opt to develop a clustering model depending on the energy of the nodes in order to elongate the lifespan of wireless sensor network.

Significance of the Study

Charging and recharging battery of sensor nodes becomes quite difficult especially in hostile environment. Even though, it is possible to replace the battery, this step will interrupt the continuous operation of WSNs which in turn causes lost of some of the packets. Meanwhile, losing some of these packets is not good for the application where continuous surveillance is required. All systems, processes and communication protocols for sensors and sensor networks must minimize power consumption by using the technology to control the amount of node that will be involved in the process of data transfer and also maintain the aggregate transfer of data

Methodology

Since clustering has been proved to be an effective method to increase the life span of wireless sensor network, then, to make use this effectiveness of clustering, this researcher tried to present a new technique for clustering the sensor nodes.

K-Means Algorithm

K-means is the simplest algorithm used for clustering which is unsupervised clustering algorithm. This algorithm partitions the data set into k clusters using the cluster mean value so that the resulting clusters intra cluster similarity is high and inter cluster similarity is low.

The following steps were taken:

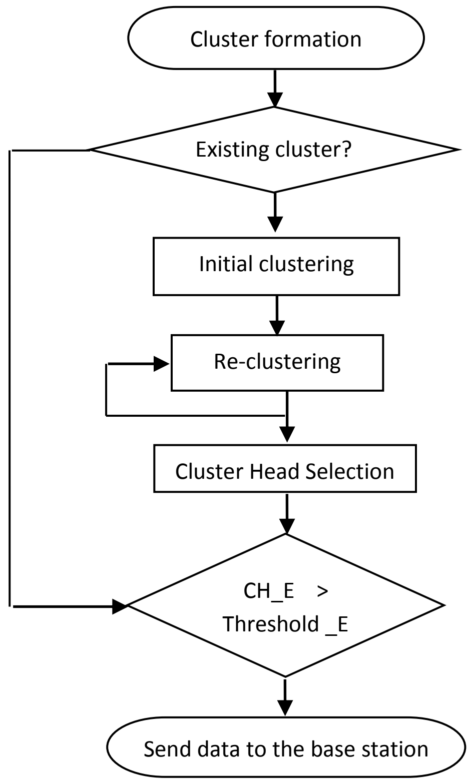
1. Arbitrarily generated k points (cluster centres), k being the number of clusters desired.

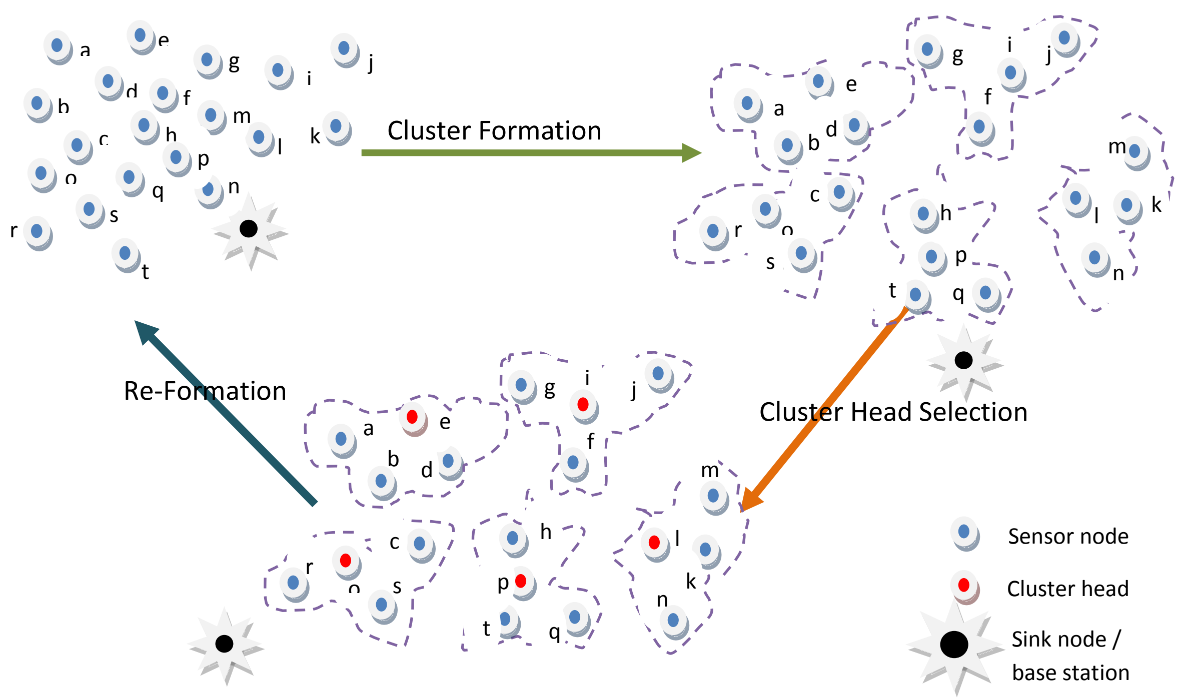
2. The distance between each of the data points to each centre were calculated, and assigned each point to the closest centre.

3. Determined the new cluster centre by calculating the mean value of all data points in the respective cluster.

4. With the new centres, step 2 is repeated. If the assignment of cluster for the data points changes, repeat step 3 else stop the process.

Flowchart for the scheme





The distance between the data points is calculated using Euclidean distance as follows. The Euclidean distance between two points or tuples,

Table showing the distance between pair nodes

Pairs of nodes	Distance features
a, b	7.5
b, c	5.5
c, d	6.5
d, e	5.6
e, f	5.8
f, g	4.3
g, h	3.8
h, i	4.1
i, j	4.1
j, k	5.2

Intra-cluster Distance

In intra-cluster, the distance between the cluster nodes to its cluster centres to determine whether the clusters are compact is calculated using the below formula.

$$\boldsymbol{intra} = \frac{\mathbf{1}}{\boldsymbol{N}} \sum_{\boldsymbol{i=1}}^{\boldsymbol{K}} \sum_{\boldsymbol{x \in C_i}} \|\boldsymbol{x} - \boldsymbol{Z_i}\|$$

Where N is the number of nodes in the network, K is the number of clusters, and Z_i is the cluster centre of cluster C_i

Inter-Cluster Distance

This is the distance between clusters. We calculate this as the distance between cluster centres, and take the minimum of this value, defined as

Using norm space

If $\mathcal{E} \in R^n$ and $\eta \in R^n$

$\mathcal{E} = (x_1, x_2, x_3, \ldots\ldots\ldots\ldots x_n)$

$\eta = (y_1, y_2, y_3, \ldots\ldots\ldots y_n)$

Now,

$$d\ (\varepsilon, \eta) = \sqrt{(x_1 - y_1)^{\,2} + (x_2 - y_2) + (x_3 - y_3)} \ldots\ldots\ldots\ldots (x_n - y_n)$$

If n = 2

$$d\ (x, y) = \sqrt{(x_1 - y)^2 + (x_2 - y_2)^2}$$

Now, if distance of nodes from the origin, have

$$d\ (\varepsilon, 0) = \sqrt{(x_1 - 0)^2 + (x_2 - 0)^2 + \ldots\ldots\ldots\ldots}$$

$$\| \varepsilon \| = \sqrt{(\sum_{i=1}^{n} xi)^2)}$$

$$= ((\sum_{i=1}^{n} x_i)^2)^{1/2}$$

Now let a_{11}, a_{12} ……… indicate the nodes in the clusters, then fulfilling the conditions stated, we can have the two models.

$$\sum_{i=1}^{n} (a1i \ - \ a2j)^2 n)\, i = 1)^{1/2}$$

OR

$$d\ (X, Y) = ((\sum_{i=1}^{n} (x_{1i} - y_{2i})^2)^{1/2}$$

Where X = (a_{11}, a_{12}, …………. a_{1n})

Y = (a_{21}, a_{22}, …………. a_{2n})

between the two nodes in the cluster

Simulation Results and System Implementation

The simulation parameters are taken in accordance with the designed network scenario and used energy model. 100 static sensor nodes and one sink node (base station) are distributed uniformly in the square region of observation O with M = 500m and W = 500 m. In the simulations, all nodes have initial battery energy of 0.5W. The Data packet size is taken to be 512 bits. The simulation time can be varied to study the effect of the proposed scheme under different scenarios as per the requirements. In this study, the whole results were obtained by taking 300 seconds simulation time. Two ray ground was used for the radio propagation model.

In the simulation, Energy Consumption of the entire network was taken as the main performance evaluation parameter. The overall energy consumption of network includes the energy consumed by all the nodes in sensing, processing and transmitting the data. Also, the network life span was used as the second parameter for evaluating our model. This is computed by considering the time when first node of the network dies out. It is the difference of total energy of the network and the summation of average used energy of nodes and their energy deviation.

It is inferred that the slope of curve in our graph gives the degree of optimization of the network life span and, a better algorithm will have the greater slope.

First, the researcher evaluates the performance of the proposed algorithm and compares LEACH. In the simulation results as reported in table below, shows the performance in term of network throughput, energy consumption and energy utilization efficiency of the respective algorithms with different simulation time. Our proposed algorithm energy consumption is low and therefore, this makes energy depletion in each node to be less and results in longer network life span.

In our simulation we have clustered the network in same number of clusters. As we have mentioned that the cluster heads can be placed randomly or separated with some minimum distance. Results show that if the cluster heads are separated with some minimum distance it gives the better performance.

The Performance Analysis

The proposed scheme has been validated through simulation using Table 4.0 parameters and comparing its performance with the Low Energy Adaptive Clustering Method-Mobile (LEACH-M), Augmented Lagrangian Multiplier Method (ALM), and Triangle Closure Method (TCM) algorithms. The proposed scheme aims to preserve as less energy as possible by selecting secure CH and consumes less energy. The result comparison among proposed schemes and LEACH-M, ALM, and TCM has been carried out using the above simulation parameters.

Table 1: Showing comparative analysis of different schemes.

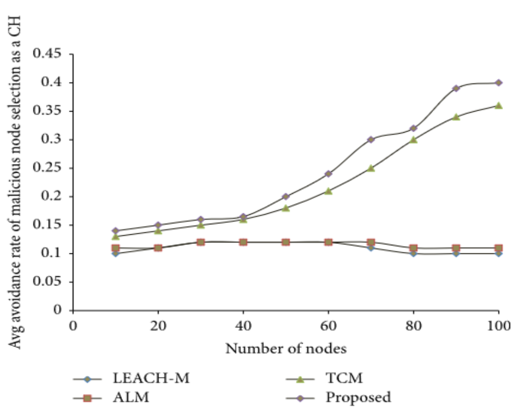
Scheme	CH Selection Parameter	Trust	Mobility	CH security
K-Means	Battery awareness	Yes	Yes	No
LEACH-M	Random number and remoteness	No	Yes	No
ALM	Weight	No	Yes	No

TCM	Weight	Yes	Yes	No

The main objective of the proposed scheme is to secure the Cluster Head selection process with minimum energy consumption, so that to avoid malicious nodes selection as a CH. Because CH carries the whole member data, the selection of malicious node as a CH will definitely waste the network resources and data. The following figure shows the avoidance rate of malicious nodes selection as a CH. We have deployed 10 nodes as malicious nodes in the whole network to see the avoidance rate of malicious nodes selection. It is clear from the above table that the proposed scheme avoids the malicious node selection as a CH more efficiently because of node behavior measurement through trust management and also avoidance of unauthorized node to join any cluster.

Table 2: Showing the Average avoidance rate of malicious node selection as Cluster Head versus number of nodes

Number of nodes	**Proposed**	**TCM**	**ALM**	**LEACH**
20	0.15	0.14	0.12	0.1
40	0.15	0.14	0.11	0.1
60	0.25	0.22	0.12	0.12
80	0.3	0.28	0.12	0.13
100	0.4	0.35	0.125	0.11

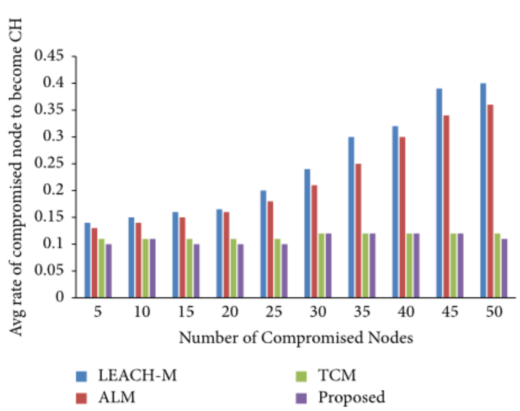


Average avoidance rate of malicious node selection as CH versus number of nodes

Average rate that a compromised node turns into a CH is figured by excluding the CH from the entire cluster which was a compromised node and independent from anyone else. This metric represents how well a cluster formation scheme expels compromised nodes. LEACH-M and ALM have no defense mechanism against compromised node and that is why the average rate of compromised node to become CH is too high.

Table 3: Average rate of compromised node become to Cluster Head versus number of compromised node

Average rate of nodes	**LEACH-M**	**ALM**	**TCM**	**Proposed**
5	0.14	0.13	0.12	0.11
10	0.15	0.14	0.12	0.12
15	0.16	0.15	0.11	0.1
20	0.17	0.16	0.11	0.1
25	0.2	0.18	0.11	0.1
30	0.24	0.21	0.11	0.11
35	0.3	0.25	0.11	0.11
40	0.34	0.31	0.11	0.11
45	0.4	0.35	0.11	0.11
50	0.42	0.39	0.11	0.1

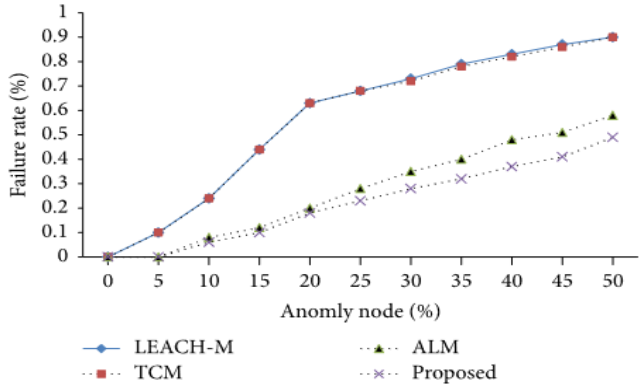


Average rate of compromised node become CH versus number of compromised nodes.

We define one failure as an anomaly node selected to be CH; failure rate is to compute coordinate influence of one malicious node, likewise called unsuccessful anomaly detection rate. Generally, when anomaly node rate is low, failure rate is also low. As rate goes up, failure rate additionally goes higher. LEACH-M and TCM are schemes with no trust and authentication mechanism, so it performs worse than ALM and the proposed schemes. In contrast, the proposed scheme is a converged model (trust supervision) with a strong defense to anomaly nodes, so it shows the highest robustness.

Table 4: Percentage of failure rate versus anomaly node in percentage

Percentage (%) of nodes	**TCM**	**LEACH-M**	**ALM**	**Proposed**
0	0	0		
5	0.1	0.1	0	0
10	0.25	0.25	0.05	0.05
15	0.45	0.45	0.11	0.1
20	0.65	0.65	0.2	0.19
25	0.7	0.7	0.3	0.25
30	0.75	0.75	0.35	0.3
35	0.8	0.8	0.4	0.35
40	0.85	0.85	0.45	0.4
45	0.9	0.9	0.45	0.4
50	0.95	0.95	0.5	0.43



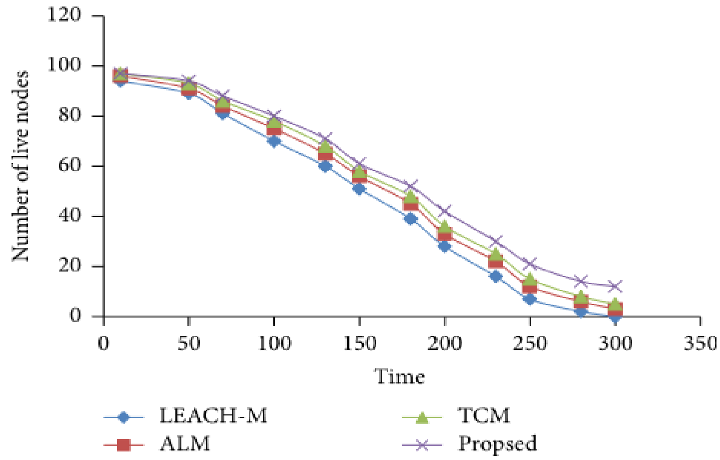
Percentage of failure rate versus anomaly node in percentage

The network lifetime is the time interval from initial deployment of the network until the death of all the live nodes. It can be, for instance, the moment when the remaining sensors die, a percentage of sensors die, the network is partitioned, or the loss of coverage occurs. In the simulation, the proposed scheme measured the time span of the network in which none of the nodes can perform the designated tasks and compared these results with the other approach.

Table 4.4: Alive number of nodes versus time

Time	**Proposed**	**TCM**	**ALM**	**LEACH-M**
5	98	97	96	95
50	96	94	93	91
70	90	88	86	84
100	85	83	80	76
130	74	70	68	65
150	64	61	59	55
170	60	57	55	50
200	43	40	38	35
230	38	35	33	30
250	33	29	28	25
280	20	18	16	14

300	19	15	14	10

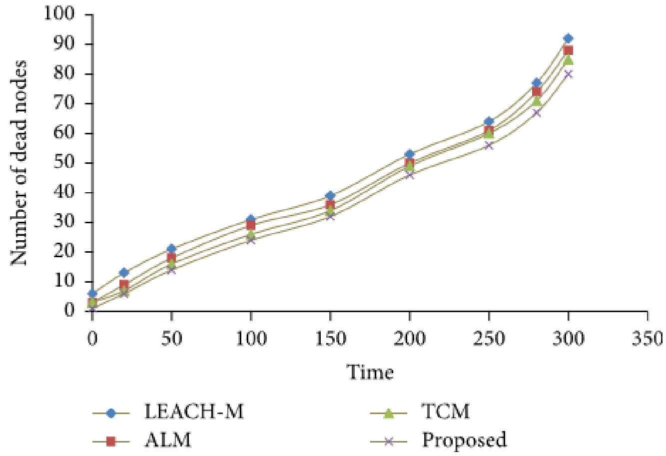


Alive number of nodes versus time

Figure below shows that nodes died more slowly in the proposed scheme because of stable and trusted nodes selection as a CH. The proposed scheme extends the stability period by selecting suitable CH on the basis of calculated weight using relative trust, less energy consumption ratio, and high success factor. The slow node death rate of the proposed scheme reflected in figure below is to secure efficient and high stability among member and CH.

Table 4.5: Number of dead nodes and time

TIME	PROPOSED	TCM	ALM	LEACH-M
0	2	4	4	6
50	6	6.2	8	12
150	15	17	19	21
200	25	26.5	29	30.5
250	31	32	36	40
300	48	50	51	53
350	58	60	61	65
400	80	85	90	95



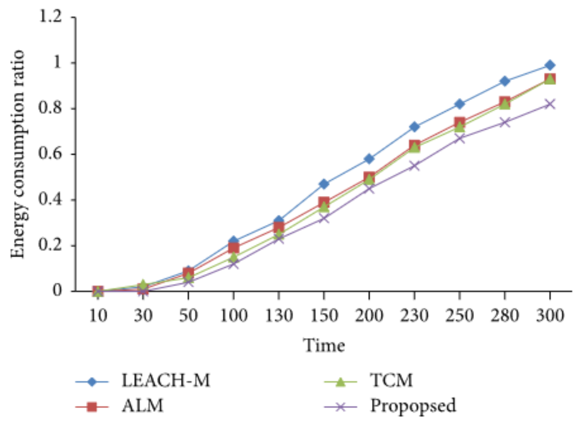
Number of dead nodes and time

The average energy consumption ratio of the entire topology is the average distinction between the initial energy and the final level of remaining energy of network. This metric is important because the energy level of the network used is proportional to the networks life span. The lower the energy consumption ratio, the longer the networks life span. Figure 4.6 shows energy consumption ratio comparison of the proposed scheme with other schemes. From this chart, it can be seen that the average energy consumption ratio of the proposed scheme is less than LEACH-M, ALM, and TCM, because the proposed scheme first selects most stable and secure CH in cluster.

Table 4.6: Average energy consumption ratio and time

Time	**LEACH-M**	**ALM**	**TCM**	**Proposed**
10	0	0	0	0
30	0.03	0.02	0.01	0
50	0.09	0.08	0.06	0.05
100	0.22	0.2	0.18	0.15
130	0.32	0.28	0.25	0.23
150	0.37	0.35	0.33	0.3
200	0.52	0.43	0.42	0.39
230	0.7	0.62	0.6	0.54

250	0.8	0.73	0.71	0.67
280	0.9	0.81	0.8	0.75
300	1	0.91	0.91	0.85



Average energy consumption ratio and time

Conclusion

This work has been able to achieve some far reaching objectives by optimizing the lifespan of wireless sensor network to become more mobile, scalable, and efficient and balance through the use clustering method by the K-means algorithm.

The process has a model that consists of two major phases i.e. the cluster formation and the cluster head selection. The base station executes the algorithm to cluster the sensor nodes and assign the proper roles to them (cluster head). The k-means algorithm helps to setup clusters, it will take the set of node input with their location and then evaluate the distance between them to form the clusters and then use the mean value of their distance because K-means uses clustering means. It evaluates the new population with fitness function and selects the best population with the best nodes for cluster head in each given cluster; the fitness function selection is based on energy level and the distance from the sink. It follows that if this system is applied in network data transfer, it will be noted that wireless sensor network nodes will last longer than they use to and more energy will be gained and there will be no data lost and the quality of information that will be gotten will still remain the same if not better than the former.

Recommendation

From the achievement listed above, the researchers recommend that the system be used under stringent conditions; that is all protocols should be followed and necessary rules be stringently followed, adequate resources and favourable environment that will enhance smooth running of the simulation so as to give good optimization of the wireless sensor network in order to increase the life span of its batteries.

References

1. Abbasi, A. A., & Younis, M. (2007). A survey on clustering algorithms for wireless sensor networks. *Computer communications*, *30*(14), 2826-2841.
2. Akyildiz, I. F., Su, W., Sankarasubramaniam, Y., & Cayirci, E. (2002). Wireless sensornetworks: a survey. *Computer networks*, *38*(4), 393-422.
3. Blum, J. J., Eskandarian, A., & Hoffman, L. (2004). Challenges of intervehicle ad hoc networks. *Intelligent Transportation Systems, IEEE Transactions on*,*5*(4), 347-351.
4. Cheng, H. (2012, May). Genetic algorithms with hyper-mutation for dynamic load balanced clustering problem in mobile ad hoc networks. In *Natural Computation (ICNC), 2012 Eighth International Conference on* (pp. 1171-1176). IEEE.
5. Hajami, A., Oudidi, K., & ElKoutbi, M. (2010, May). An enhanced algorithm for MANET clustering based on multi hops and network density. In *New Technologies of Distributed Systems (NOTERE), 2010 10th Annual International Conference on* (pp. 181-188). IEEE.
6.Heinzelman, W. B., Chandrakasan, A. P., & Balakrishnan, H. (2002). An application-specific protocol architecture for wireless microsensor networks. *IEEE Transactions on wireless communications*, *1*(4), 660-670.
7. Rault, T., Bouabdallah, A., & Challal, Y. (2014). Energy efficiency in wireless sensor networks: A top down survey. *Computer Networks*, *67*, 104-122.
8. Safa, H., Artail, H., & Tabet, D. (2010). A cluster-based trust-aware routing protocol for mobile ad hoc networks. *Wireless Networks*, *16*(4), 969-984.
9. Younis, O., & Fahmy, S. (2004). HEED: a hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks. IEEE Transactions on mobile computing, 3(4), 366-379.
10.Zhou, L., & Haas, Z. J. (1999). Securing ad hoc networks. *Network*, *IEEE*, *13*(6), 24-30.