RESEARCH ARTICLE OPEN ACCESS

Load Balanced Clustering Approach in Wireless Sensor Network using Genetic Algorithm

Saziya Tabbassum[1], Sanjeev Bangarh[2]
[1]*Department of Computer Science, OPJS University Churu, India*
[2]*Department of Computer Science, OPJS University Churu, India*

ABSTRACT
Extending network lifetime in Wireless Sensor Network (WSN) has become one of the major challenges as sensor nodes have limited battery source. All of these sensor nodes have the ability to communicate either with each other or directly to an external base station. Studies show that Low Energy Adaptive Clustering Hierarchy (LEACH) is one of the protocols of WSNs which is a clustering based protocol where randomized rotation of local cluster base stations is used to evenly distribute load among the sensor nodes. The major challenge in this clustering technique is balancing the load among different clusters in the network so that early death of some nodes does not affect network lifetime. In this paper we propose a genetic algorithm based load balanced clustering approach for optimizing the problem of using energy efficiently. Simulation on our proposed work, LEACH, GA-LEACH and DE-LEACH have been done. Result shows that the proposed algorithm converges earlier than traditional DE and GA but it performs better in terms of network lifetime, energy consumption and number of dead sensor nodes.

Keywords- Wireless Sensor Network, Genetic Algorithm, Sensor Node, Cluster Head, Network Lifetime.

Date of Submission: 05-10-2020 Date of Acceptance: 19-10-2020

I. INTRODUCTION

Wireless Sensor Network (WSN) is a wireless network which consists of spatially distributed autonomous devices using sensors to monitor the physical or environmental condition. Sensors measure conditions in the environment surrounding them and then transform these measurements into signals which can be processed to find out characteristics ofthe phenomenon located in the area around these sensors. Then the signal is transferred from these sensor nodes to the base station through the gateway where the distance between the place where sensor nodes are deployed and base station depends on application of the network. The data to be sent from the sensor node to the base station may go directly to the base station or it may follow a certain path using multiple hops. The nodes transmit the sensed to the sink via relay nodes and from sink to the base station [1-5]. The concept of sensor network was initially used in applications of inaccessible, disaster scene and military environments but now there is special interest in commercial applications. With recent technological advances in micro-electro mechanical systems (MEMS), [6]wireless communication and digital electronics have proved low-cost, low-power, multi-functional sensors with capabilities of sensing, data processing and wireless communication within short range. The intrinsic properties of individual sensor nodes pose additional challenges to the communication protocols in terms of energy consumption. These sensor nodes consist of a sensing, communication, processing, power unit which helps to execute all the functionality of the sensor nodes. Location information can easily be provided by GPS, which provides accuracyup to 10m through the recent GPS unit developed for WSNs. However, the cost of these units is significantly higher than a single sensor node so instead, a limited number of nodes, which use GPS or other means to identify their location, are used to help the other nodes determine their locations. Due to the short transmission ranges, large numbers of sensor nodes are densely deployed and neighboring nodes may be very close to each other. But for far transmission more power is required which leads to dead sensor nodes very early. Hence, multi-hop communication is used in communications between nodes since it leads to less power consumption than the traditional single-hop communication. In this paper, we propose a protocol which works on clustering for energy efficiency and balancing the load in the cluster, as sensor nodes have limited energy. Our protocol uses two parameters: energy of network and cluster distance between clusters to prolong the network lifetime. In a cluster-based WSN, CH consumes energy more because of the extra workload of receiving the sensed data, aggregation of sensed data and transmitting that aggregated data to the base station. If the formation of clusters is not properly done then it may result in the quick death of CHs which leads to the partition of network and degrades the overall performance of the WSN due to CH overload.We use genetic algorithm for optimizing the energy and cluster distance of network and find out amore efficient way to maximize the network lifetime by delaying the initial death of sensor nodes. Here, we are going to propose a clustering algorithm for WSNs to increase the lifetime of the network by controlling the faster death of highly loaded CHs.

II. RELATED WORK

In recent years, numerous studies have been done on LEACH protocol for clustering and routing in WSNs. Cluster creation and assigning of cluster-heads greatly contribute to overall system scalability, life, and energy efficiency. Heinzelman and et al [7] develop communication protocols which can have a significant impact on the overall energy dissipation of the networks. In this work, the author describes LEACH as a clustering based routing protocol where the whole load of the network is distributed to all nodes at a different point in time resulting in minimization of global energy usage. In LEACH clusters are formed by selecting the cluster head (CH) with a probability in each cluster, then data collection has been done by all the nodes in a cluster; later all the collected data is sent to CH of the cluster, this process is known as data aggregation. Once data aggregation has done then CH sends these aggregated data to the base station either directly or by multiple hops. However, the performance in heterogeneous networks is not very well because it selects CH without considering residual energy of the nodes. To solve this problem, researchers improved LEACH and proposed many algorithms [7-11].

LEACH [7], TEEN [12], APTEEN [13], PEGASIS[14] are prominent routing techniques for WSNs. Main procedure of cluster head selection was given by LEACH and that is further enhanced by SEP. TEEN introduced the concept of thresholds which gives good results in network lifetime by showing reactive nature whereas, in APTEEN nodes react on time critical situations as well as gives an overall scenario of the network at periodic intervals in a very energy efficient manner.

The authors in [15] highlight the challenges in clustering a WSN, discussing the several key issues that affect the practical deployment of clustering techniques in sensor network applications. To reduce the communication overhead and exploit data aggregation in sensor network, node clustering is a useful topology management approach. Here, the focus is made on a distributed clustering approach which is suitable for the large-scale sensor network. The most compelling challenges are to schedule concurrent intra-cluster and inter-cluster transmissions, to compute the optimal cluster size, and to determine the optimal frequency for CH rotation in order to maximize network lifetime.

A new application specific low power routing protocol named ASLPR is introduced that takes into account some concepts from sensor nodes to elect the optimal cluster heads by Jalali and Shokouhifar in [16]. In this work, a hybrid algorithm based on genetic algorithm and simulated annealing is applied to optimize ASLPR in order to prolong the network lifetime, based on the application specifications. On the other hand, the proposed ASLPR protocol can maximize the defined lifetime scheme (e.g., FND, HND, etc.), based on the application. Simulation results show that the proposed hybrid optimization algorithm can efficiently balance the energy consumption of nodes and maximize the network lifetime.

All the above mentioned algorithms are based on the assumption that all the nodes in networks are distributed uniformly. In networks with non-uniform distribution, considering the network coverage problem, [17] proposed some unique cluster head selection technique to evenly distribute the clusters and prolong the network lifetime. In [18] P. K. Jana, et al has proposed a novel differential evolution (DE) based clustering algorithm for WSN to prolong the network lifetime by preventing the faster death of highly loaded CH. The main objective of the proposed algorithm is to prolong the network lifetime by preventing the initial death of the gateways.

III. METHODOLOGIES

1 Energy Model

The proposed model incorporates radio model discussed by Heinzelman et al. [11] for energy consumed by sensor nodes in transmission of data, in which a free space (fs) model is used when the distance is less than a threshold value ; otherwise, multi-path (mp) model is used which is shown in figure 1.

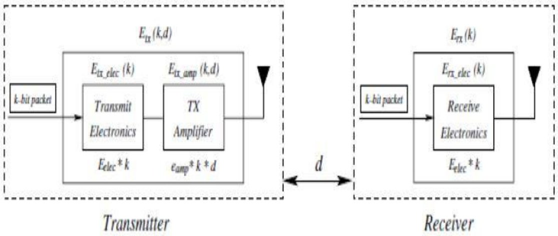

Figure1: Energy Model

Let E_{elec}, Ɛfsand Ɛmp be the energy required by the electronics circuit and by the amplifier in free space and multipath respectively. Then energy consumed by the sensor node radio to transmit a k-bit data over a distance d is given as follows:

$$E_{Tx}(k,d) = \begin{cases} \left(E_{elec} + \varepsilon_{fs} \times d^2\right) \times k, & d < d_0 \\ \left(E_{elec} E + \varepsilon_{mp} \times d^4\right) \times k, & d \geq d_0 \end{cases} \quad (1)$$

The energy required by the radio to receive a k-bit message is given by

$$E_{Rx}(k, d) = E_{elec} \times k \quad (2)$$

The energy consumed in generation of k-bit messages is given by E_{elec} which depends on various factors such as digital encoding, modulation, filtering, and spreading of the signal. The amplifier energy $\varepsilon_{fs}\, d^2$ /$\varepsilon_{mp}\, d^4$, depends on the distance between transmitter and the receiver and also on the acceptable bit error.

2 Network Model

A wireless sensor network model is assumed in which sensor nodes are randomly deployed in the region, where each sensor node is assigned unique ids. From these sensor nodes, some CHs are selected using a threshold value and after deployment they become stationary. All sensor nodes and CHs can be heterogeneous but they cannot be recharged once deployed. A sensor node can be assigned to any CH which is in the communication range of the node. Sensor nodes have a list of CHs near them and they can be assigned to any one of the CH. Now the data collection is divided into rounds, in each round sensor nodes collect data from their environment and send sensed data to their respective CH. Upon receiving the data from sensor nodes CH first aggregates the data and then sends the aggregated data to the base station through the next hop relay node.

All communications are over a wireless link which is established between two nodes only if they are within the communication range of each other. Network lifetime can be calculated in various ways based on the residual energy of sensor nodes as the time at which the first node is dead, time at which half of the nodes are dead, and time at which the last node dies.

3 Proposed Work

In LEACH protocol, there exist two phases: a setup phase and steady state phase. During the setup phase clusters are formed by choosing the node which satisfies the condition to become a CH. Now in the steady state phase, data transmission is done between the sensor nodes to CH and from CH to the base station. This method has a drawback that if data has to be sent from a CH which is far from the base station then more energy is required to transfer the data from so far and the node will die soon. So, to overcome this problem we are going to use a protocol based on LEACH and then using a genetic algorithm to optimize our problem.

In this paper, we are going to use a genetic algorithm for optimizing the problem of using energy efficiently in single hop clustering. Initially, network setup is done in which two steps are followed: the first step is to give unique IDs to all the sensor nodes, then from those sensor nodes CHs are selected based on their energy and the threshold value in equation (1) as same like LEACH. Once the CHs are chosen then clusters are formed by sensor nodes. CH broadcasts their IDs using CSMA MAC protocol. The CH can collect the information about sensor nodes and other CHs in its communication range. Now the sensor nodes decide to join one of the CH to which communication range they belong and then clusters are formed which is the second step. Sensor nodes send their data to CH according to the TDMA scheduleassigned to it. Now we present the methodology used in our algorithm for initialization of population vector, calculation of fitness function, mutation crossover, and selection in the following subsections.

3.1 Initialization of Population Vector The population vector is initialized by using a random generator which generates uniformly distributed number Rand (0, 1). It is generated independently of each component. The vector represents the complete assignment of sensor nodes to the CH in a cluster which can be represented as

$$X_{i,G} = [x_{1,i,G}, x_{2,i,G}, x_{3,i,G}, \ldots, x_{N,i,G}] \quad (3)$$

Where N is the dimension of the vector and $x_{1,i,G}$ represents the assignment of sensor nodes to the CHs is the set of sensor nodes deployed in the network where, $S = \{s_1, s_2, \ldots, s_n\}$ and s_{n+1} is the base station. C is the set of CHs selected from sensor nodes where, $C = \{c_1, c_2, \ldots, c_m\}$. In general, each sensor node is assigned here a unique ID and based on the communication range of nodes to the CH they form a cluster. The population vector represents the information about the sensor nodes in the cluster. This type of vector representation is a part of a clustering algorithm and addition or deletion of any node in the network requires re-clustering.

3.2 Fitness Function: Now we construct the fitness function which evaluates the individual particle in the population. The main objective of our proposed work is to maximize the lifetime of the network by delaying the initial and final death of the nodes. It is achieved by minimizing the energy consumption and minimizing the cluster distance, i.e., minimizing the distance between sensor nodes and corresponding CHs.

3.2.1 Energy of CH: the main principle of the maximization of the energy of CH is that the CH with lower residual energy should have a lower rate of energy consumption per round as compared to the CH with higher residual energy. Therefore, energy consumption of CH with number of member sensor node due to inter cluster activity in a single round is given as:

$$Egate\ (c_i) = n_i * E_R + n_i * E_{DA} + E_T (c_i , BS) \tag{4}$$

Where, CH c_i has n_i number of sensor nodes assigned. E_R is the energy consumption due to data receiving from other sensor nodes. E_{DA} is the energy consumption due to data aggregation of data collected from different sensor nodes. E_T is the energy consumption due to data transmission to the base station. *dist* (c_i , BS) is the Euclidean distance between the CH and the base station. Let *Eres* denote the residual energy of CH or remaining energy of that node. Therefore, lifetime of the CH defined as

$$L(i) = \frac{Eres\,(ci)}{Egate\,(ci)} \tag{5}$$

Our objective is to maximize the lifetime, therefore larger the value of *L (i)* higher is the fitness value i.e.

$$Fitness \propto L \tag{6}$$

3.2.2 Average Cluster Distance: Some sensor nodes are forced to join the CH which is far from them. These sensor nodes consume extra energy for transmitting data to the CH resulting in the faster death of the node. Therefore, we should minimize the average cluster distance so that this problem does not occur. Thus, average cluster distance is given by

$$AvgDist=\sqrt{\frac{1}{m}\sum_{j=1}^{m}\{\mu_D - AvgClusDist(ci)\ \}^2} \tag{7}$$

Where, *m* is the number of CH, and

$$\mu_D = \frac{1}{m}\sum_{j=1}^{m} AvgClusDist(ci) \tag{8}$$

$$AvgClusDist(ci) = \frac{1}{n_j}\sum_{i=1}^{n}\{dist(s_i , c_j) * \alpha_{i,j}\} \tag{9}$$

Where, $\alpha_{i,j}$ is a Boolean Variable such that $\alpha_{i,j}$ is 1 if sensor node s_i is assigned to CH c_j, otherwise it is 0.

The shorter the *AvgDist* higher is the fitness value. Therefore, fitness function is

$$Fitness \propto \frac{1}{AvgDist} \tag{10}$$

Equation (6) and (10) combined imply that

$$Fitness \propto \frac{L}{AvgDist}$$

$$Fitness = K * \frac{L}{AvgDist} \tag{11}$$

Where *K* is a proportionally constant and we assume *K* = 1.

$$\text{Therefore, Fitness} = \frac{L}{AvgDist} \tag{12}$$

Vectors are evaluated by the fitness function. Higher the fitness value better is the chromosome.

3.3 Crossover, Mutation & Selection: To produce a new population from a selected population we have randomly used uniform crossover or *k*-point crossover where, *k* = 1, 2, 3, or selected randomly for crossover operation.

To improve the fitness value of an individual, mutation is applied after the process of crossover. To perform the mutation over an individual instead of randomly selecting a gene, as in standard GA, we select the node, say node i, in the chromosome which dissipates the maximum energy due to receiving and transmitting its data. The main purpose of selecting node i as the critical node for mutation is to reduce the total energy dissipation by the node and hence to increase the network lifetime. It can be done either by replacing edge i to j from critical node i to edge i to k where k is randomly chosen from nodes which are closer to node i than j and also closer to the base station than i, or by diverting some incoming flow away from node i by randomly deleting an existing edge u to i, and adding an edge u to v to some node v which is closer to the base station than u. In our work, we have considered the mutation rate as 0.05, and during this process the load of CH in a highly loaded cluster is reduced.

Selection for an individual is carried out by using Roulette-Wheel Selection method, where the probability of being selected increases with the fitness value of the individual chromosome [42, 43].

IV. SIMULATION RESULT

We performed an extensive experiment on our proposed work and it was taken on a diverse number of sensor nodes between 100 to 500 and cluster heads ranging between 15 and 50 in an area of 200m×200m. All the sensor nodes are assigned initial energy 2J. To execute our proposed algorithm using GA we considered an initial population of 100 chromosomes, crossover rate, scaling factor, and mutation rate is taken as 0.7, 0.5 and 0.05 respectively. We have compared our proposed work with protocols like LEACH, GA-LEACH, and DE-LEACH. Experiments were performed on two scenarios and in both of them the sensing region is 200m×200m.

First Scenario: we choose the network in which default deployment of nodes is done and number of

CH is fixed to 30.

Case I: 300 sensor nodes were taken and the base station is considered in the center of the region. Figure 2 shows a comparison of proposed protocol with other protocols in terms of dead sensor nodes. Here, our derived fitness function takes care of residual energy and energy consumption of nodes as well as the distance between clusters. Here we find out that as compared to other protocols initial death is somewhat early than that of DE. But after half of the nodes are dead the energy of DE gets deplete earlier as compared to that of our proposed model. Our approach takes care of the load of CHs in the cluster and it reduces the load in setup phase only. In the second case of the first scenario, we have considered

Case II: 500 sensor nodes were taken and the base station is at the center of the network which is shown in figure 3. Here, we find that if the region is small and a number of nodes are more after a certain point the node's energy starts to deplete faster.

Second Scenario:we choose node deployment to be done randomly and the number of CH changes every time range between 27 and 50 and selected randomly.

Case I:we have taken 300 sensor nodes and a base station in the center of the region. In this case, we have compared different approaches with our approach and it is very similar to the first case of the first scenario as shown in figure 4. It is clear that from comparing the two scenarios of 300 sensor nodes, the first scenario has a longer lifetime as compared to that of the second scenario, this is because in the second scenario number of CH is not fixed and load distribution of cluster is uneven.

Case II:we have deployed 500 sensor nodes in the environment and base station, is at the center of the region. As shown in figure 5the lifetime of our approach is better than any other because we have chosen CH on the basis of residual energy of the node. If a node has energy less than determined than it will not take that node as a CH. The rate of being dead of a node increases faster after half of the sensor nodes are dead because there is a large number of sensor nodes in a small region.

FIRST SCENARIO:

Case I

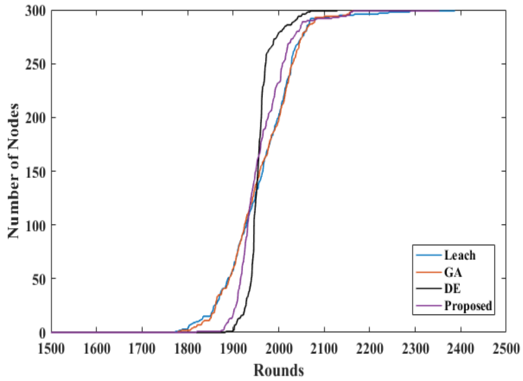

Figure 2: Comparisonin terms of dead sensor node

Case II

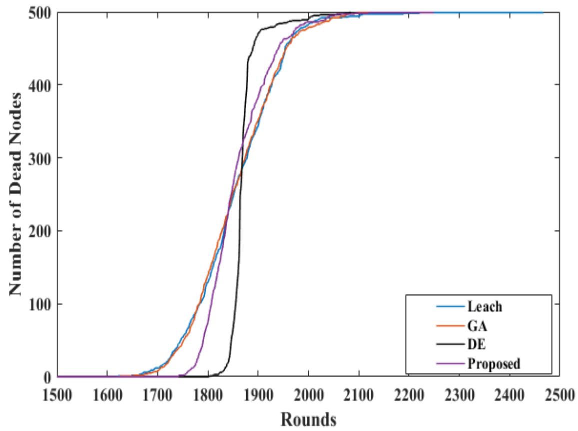

Figure 3: Comparison in terms of dead sensor node

SECOND SCENARIO:

Case I

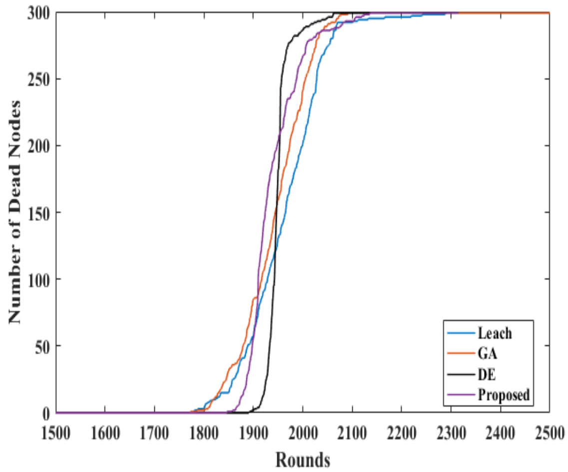

Figure 4: Comparison in terms of dead sensor nodes

Case II

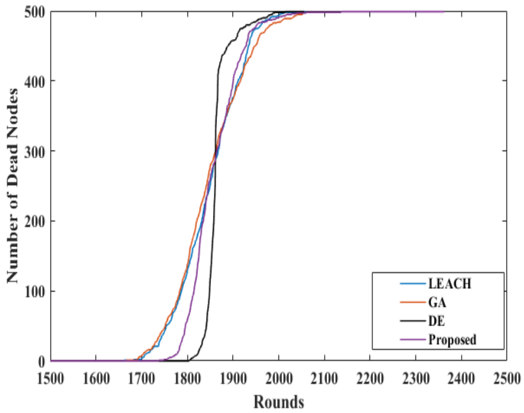

Figure 5: Comparison in terms of dead sensor nodes

V. CONCLUSION

In this work, we have presented a genetic algorithm based clustering approach for wireless sensor networks. The proposed approach has introduced an efficient vector encoding scheme to improve the performance of the clustering algorithm. We have also derived an efficient fitness function for enhancing network lifetime significantly. The fitness function takes care of energy consumption of CH as well as sensor nodes. The experimental result has shown that proposed algorithm converges faster than traditional DE and GA, but it performs better than the existing algorithm in terms of network lifetime, energy consumption, and a number of dead sensor nodes. Our objective is here achieved not only by minimizing energy consumption of network, but also maximizing the lifetime by taking care of lifetime of CH. In our work, we have introduced a threshold point for the selection of CH so that the CH with lower residual energy cannot become a CH. This is the main reason behind maximizing the lifetime of the network because nodes with higher residual energy are only selected as CH. Here, we have used standard deviation for minimizing the average distance between clusters. Standard deviation takes care of the distant nodes from the base station and nodes which are near to the base station, find out the variance between each cluster. Hence, our second objective is also fulfilled hence, a delay in nodes early death. The node deployment in the first case has the best lifetime among all four cases because in this case the nodes having lower residual energy have a lower rate of energy consumption as compared with nodes with higher residual energy. In this case node distribution is evenly done which is not present in both cases of second scenario.

REFERENCES

[1]. Wireless Sensor Network Perspective by JUN ZHENG.

[2]. Mts420 environment sensor board. Available at http://www.xbow.com.

[3]. D. Trossen and D. Pavel, “Sensor networks, wearable computing, and healthcare Applications”, IEEE Pervasive Computing, vol. 6, no. 2, Apr. – June 2007, pp. 58 – 61.

[4]. F. Koushanfar, M. Potkonjak, and A. Sangiovanni - Vincentelli,“Fault – tolerance techniques for ad hoc sensor networks”, Proceedings of IEEE Sensors, vol. 2, June 2002, pp. 1491 – 1496.

[5]. Wireless Sensor Networks Ian F. Akyildiz and Mehmet Can Vuran c 2010 John Wiley & Sons, Ltd.

[6]. “An introduction to MEMS”, Prime Faraday Technology Watch – January 2002.

[7]. Wendi Rabiner Heinzelman, Anantha Chandrakasan, and Hari Balakrishnan, “Energy Efficient Communication Protocol for Wireless Micro-Sensor Networks”, Proceedings of the 33rd Hawaii International Conference on System Sciences – 2000.

[8]. J. N. Al - Karaki and A. E. Kamal, “Routing techniques in wireless sensor networks: A survey”, IEEE Wireless Communications, vol. 11, no. 6, Dec. 2004, pp. 6 – 28.

[9]. Siva D. Muruganathan, Daniel C. F. Ma, Rolly I. Bhasin, and Abraham O. Fapojuwo “A Centralized Energy-Efficient Routing Protocol for Wireless Sensor Networks”, IEEE Radio Communications March 2005.

[10]. Md. Solaiman Ali, Tanay Dey, and Rahul Biswas, “Advanced LEACH Routing Protocol for Wireless Microsensor Networks”, Department of Computer Science & Engineering Khulna-9203, 5th International Conference on Electrical and Computer Engineering ICECE 20-22 December 2008.

[11]. 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.

[12]. Arati Manjeshwar and Dharma P. Agrawal,“A Routing Protocol for Enhanced Efficiency in Wireless Sensor Networks”, Center for Distributed and Mobile Computing OH 45221-0030, 0-7695-0990

[13]. Arati Manjeshwar and Dharma P. Agrawaly“A Hybrid Protocol for Efficient Routing and Comprehensive Information Retrieval in Wireless Sensor Networks Center for Distributed and Mobile Computing”, ECECS Department, University of Cincinnati, Cincinnati, OH 45221-0030.

[14]. Stephanie Lmdsey and Cauligi S. Raghavendra, “Power-Efficient GAthering in

Sensor Information Systems", Computer Systems Research Department, the Aerospace Corporation, CA 90009-2957.

[15]. Ossama Younis, Marwan Krunz, and Srinivasan Ramasubramanian, "Node Clustering in Wireless Sensor Networks: Recent Developments and Deployment Challenges", 0890-8044 IEEE Network 2006.

[16]. Mohammad Shokouhifar Ali Jalali, "A new evolutionary based application specific routing protocol for clustered wireless sensor networks", 10.1016/j.aeue.2014.10.023.

[17]. Jiguo Yua, Yingying Qia, Guanghui Wangb, and Xin Gua, "A cluster-based routing protocol for wireless sensor networks with non-uniform node distribution", nt. J. Electron. Commun. (AEÜ) 66 (2012) 54– 61.

[18]. Pratyay Kuila, Prasanta K. Jana: "A novel differential evolution based clustering algorithm for wireless sensor networks", Applied Soft Computing, 2014 ElsevierB.V.

Saziya Tabbassum, et. al. "Load Balanced Clustering Approach in Wireless Sensor Network using Genetic Algorithm." *International Journal of Engineering Research and Applications (IJERA)*, vol.10 (10), 2020, pp 28-34.