International Journal of Latest Trends in Engineering and Technology
Special Issue SACAIM 2017, pp. 040-047
e-ISSN:2278-621X

DGC-SOM CLUSTERING ALGORITHM FOR EFFICIENT BIG DATA GATHERING IN DENSELY DISTRIBUTED WIRELESS SENSOR NETWORK

Doreswamy[1] & Kunal.G.S[2]

Abstract- Wireless sensor networks (WSN) have emerged as an important research topic due to advancement in internet technologies and communication. WSN composed of sensor nodes which are enabled with wireless technology and precise energy. The key challenges in WSN are efficient continuous data gathering which is generated by each sensor and save energy. In order to accomplish these key challenges, various clustering methods have been applied to densely distributed WSNs. There is a need for an effective method for the selection of clusters and utilize the mobility of sink nodes to address the big data gathering. In this paper to address continuous data gathering and energy efficiency of densely distributed WSNs a Dynamically Growing Cellular Self Organizing Map (DGC-SOM) data clustering algorithm is proposed. The proposed algorithm reduces the energy consumption by dynamic selection of clusters with mobile sink strategy.
Keywords – Wireless Sensor Network, Self Organizing Map, Data Gathering, Mobile Sink, Clustering

1. INTRODUCTION

Advancement in internet, wireless communication, global position systems (GPS), played an important role human daily life. In this context, wireless sensor networks (WSN) are widely used in monitoring physical circumstances such as temperature, vibration, pollutants, humidity, habitat monitoring, security surveillance, target tracking, and medical applications. A typical WSN device is composed of the transceiver, an analog to digital converter module, power supply unit and data processing unit as shown in figure 1.

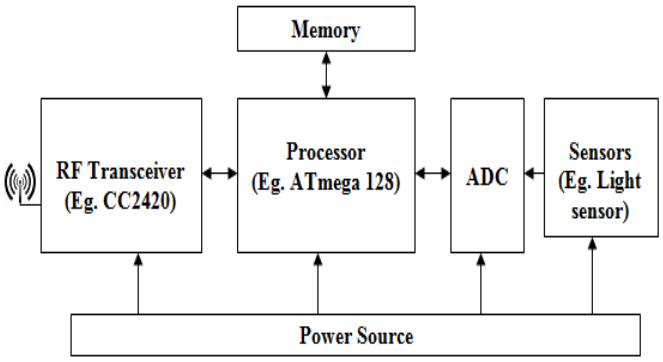

Figure 1. Typical WSN Device

The sensing unit may have multiple sensors. Relevant data required from the region of interest is processed by data processing unit. The memory unit is used for storing data and programs. ADC task is covert analogy signal of sensors to digital signals for information processing. The limited power supply is provided by the battery or power generators. WSN architecture mainly contains four entities namely sensor node, monitored events, base station and users. Sensors are fit for observing, measuring and responding to occasions as well as wonders in a predetermined domain. The base station or sink nodes is the system hubs connecting clients to the checked or detected occasions, including by means of different systems, for example, web, satellite, and LAN. The monitored events of interest are detected by the sensor nodes. The users, sensor nodes, and sinks can be mobile or fixed depending on requirements of application [1][2]. Figure 2 shows a typical WSN architecture

[1] Department of Computer Science, Mangalore University, Mangalore, Karnataka, India
[2] Department of Computer Science, Mangalore University, Mangalore, Karnataka, India

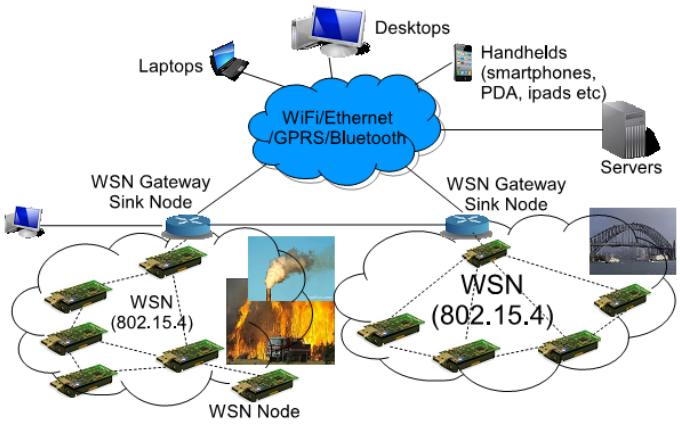

Figure 2. WSN Architecture

In sensors, the major problem is in energy capacity of each sensor nodes. Many research works have been carried out in order to save energy of each sensor and improve the network lifetime. To overcome saving of energy problem routing protocols should consider energy utilization of nodes in implementation. There are many routing protocols have been proposed and evaluated to minimize this problem in WSN using clustering, data aggregation, data-centric methods, bio-inspired algorithms and so on [3]. Few of the approaches use clustering techniques to achieve a saving of energy in WSN. Clustering technique contains cluster head (CH) which is responsible for gathering data and its cluster members and send it to the base station (BS). Clustering process is carried out in a number of iterations. In each iteration, the CH gets exchange based on the balance of energy capacity in given interval. Figure 3 shows the data gathering by the CH on the event of interest.

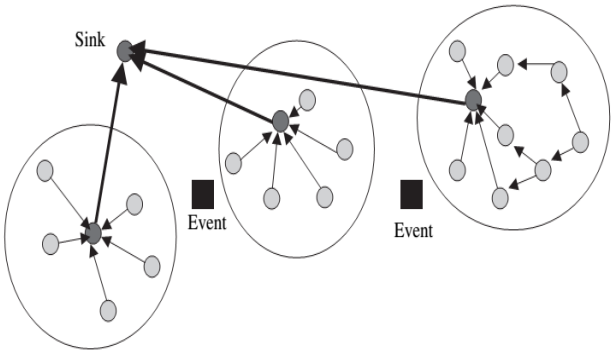

Figure 3. Data gathering at Cluster Heads and direct reporting to sink.

A sensor node with less energy may be utilized for performing sensing work and transmission of sensed data to its CH in short distance from its members of the cluster. Transmit the data gathered to sink nodes. This flow cannot minimize the energy loss for communication, balance the traffic load is difficult and leads to scalability issues when the network area grows in size. The major issue in clustering is deciding how to elect the CH and organize each cluster. Data aggregation may be performed at CH to minimize the required data transmission to sink and improve energy efficiency. In this concern, many clustering techniques have been explored and it is discussed related work session.

This paper is formulated as follows, section II discusses related work on clustering methods for WSN, section III gives the proposed approach for efficient big data gathering in WSN, section IV presents the implementation details and results obtained V concluding the paper by presenting future work.

2. RELATED WORK

Several clustering algorithms for efficient data gathering were presented and evaluated in the decade. Among that LEACH [4][5] is most popular hierarchical routing protocol in which CH is selected in a rotation manner based on probability model and all sensor nodes get enough chance to become a CH. LEACH consists of two steps. Step 1 is a set up where cluster formation is computed and done. Next step is a steady state, actual data transmission takes place. Each node selects a random number range from 0 and 1 to become a CH. Sagiroglu et al. [7] recognized various wellsprings of enormous information, for example, online exchanges, messages, sounds, recordings, pictures, click-streams, logs, posts, look inquiries, wellbeing records, long-range informal communication collaborations, cell phones furthermore, applications, logical gear, and sensors. Likewise, it was called attention to, in their work, that the enormous information is troublesome to catch, shape, store, oversee,

share, investigate, and imagine by means of traditional database apparatuses. PEGASIS [8] and KAT mobility [9] are the algorithms for centralized clustering. PEGASIS clustering algorithms form clusters of nodes chain positioned on location and loop for CH election. The limitation of the transmission range is considered in PEGASIS clustering algorithm and uniform energy utilization is achieved. The authors of [10] determined the limitations based on the sensor nodes buffer size. These limitations are sensible assumptions, data request energy utilization is not considered.

To decrease the correspondence cost, grouping on person hubs was examined in DEM [11]. DEM accept that the information records have a similar number of normal (obscure) appropriations on each hub. The learned shows parameters are passed to the following hub in light of the pre-specified arrange. This correspondence is essential due to the presumption of similar dispersions on all figuring hubs. The cluster head is selected based on the range of converging. The mobile sink node will broadcast data request message to all member node. After that, all member nodes send the sensing data through the direct or indirect path to cluster head [12]. After each round of social affair and transmission of information, bunch cluster head lost more vitality than the part node, so after some time it winds up noticeably terminated. A Complete framework is produced with a convention that continues detecting the cluster head node in each round of transmission. In the event of any bunch head disappointment, the framework chooses the most a founding vitality node the group head. The cluster head imparts the data to rest of the cluster head in arrange and investigate the sink node. Just the data sharing group heads partake in a choice of re-modification courses process. The main advantage this technique is it has dynamic routing capability (i.e.) it will change the transmission path if any cluster head fails and also in case of the base station is changed [13]. The pitfall in this technique is it consumes more energy and also it takes more time for data transmission.

BEE-C [14] designed based on bio-inspired algorithm view of the conduct of honey bees for grouping the system. BEE-C works in two stages. In the principal stage, called amass arrangement, the convention utilizes a bio-inspired algorithm and a target capacity to design (grouping) the system. The target work depends on the vitality utilization model of the system. In this stage, every hub sends to the Base Station the data of its topographical area and the measure of vitality. The Base Station utilizes this data to compute the normal vitality of the sensor nodes. Just nodes with vitality higher than normal are qualified to the group head. Along these lines, the Base Station runs the bio-enlivened calculation in view of the honey bees called BEE-C Mating Optimization (HBMO) [15] to frame the best bunches. The target work functions as a model to locate the best groups in the system. In the second stage, called information transmission, the cluster head gets the information parcels from its group, totals it into the bundle and sends to the Base Station.

Mobility Based Metric for Clustering [16] proposes a nearby mobility metric to such an extent those portable sensor nodes with low speed in respect to their neighbors have the opportunity to wind up bunch heads. By figuring the difference of a portable hub's speed in respect to each of its neighbors, the total nearby speed of a versatile hub is assessed. Low change esteem shows that this portable node is generally less versatile to its neighbors. Thus, versatile hubs with low change esteem in their neighborhoods are picked as CH. Along these lines, a chose CH can typically guarantee the low portability regarding its part nodes. Notwithstanding, if portable nodes move haphazardly the execution may lessen.

Clustering for energy conservation [17] accepts two nodes as master and slave. A slave node must be associated with one master node just and there is no immediate association between slave hubs. Each master node can set up a group in view of associations with slave sensor nodes. The downside of this plan is paging process before each round of correspondence expends a lot of vitality. WSNs may not be ideal as far as vitality Master node race isn't versatile and the strategy for choosing the master node isn't determined. On vote based grouping algorithm [18], the authors have considered neighbor's number and remaining battery time of every versatile node. The essential idea is the Hello message, which is transmitted on the mutual channel. Making utilization of node area and power data, this work proposes voting idea which is the weighted entirety of a number of legitimate neighbors

3. PROPOSED ALGORITHM

To design Clustering algorithms few important parameters need to keep in mind for serving best solutions. These parameters are as follows:

- The quantity of clusters in given area: CH selection and the required number of clusters is the deciding factor in the efficiency of the WSN routing protocol.
- Nodes and mobility of CH: dynamic nature of changing the members of cluster nodes to CH.
- Node roles and types: Capability of computation of some sensor nodes or not.
- Cluster head election: Growing cellular probabilistic approach.
- The complexity of the GCPSOM algorithm: Primary goal is to achieve faster execution.
- Overlapping: Reduce the overlapping of sensor nodes of various clusters.

The following subsection explains the proposed approach implementation based on an assumption:

- A fixed number of sensor nodes are deployed in a rectangular area.
- Distributions of nodes are uniform in given geographical area.
- Every sensor nodes will know their neighbor sensor nodes and location not known.
- Uniform transmission range.

- The connectivity between sensor nodes is bi-directional.

3.1 *Energy Model*

The energy calculation formula is adopted from [10][3].We have utilized normal radio model to compute the energy dissipation to the communication of data from the sensor nodes. For the transfer of b no of bits for a distance d the energy dissipation is computed using the equation given below:

$$E_{diss} = E_{diss}(b) + E_{dissampp}(b,d)$$

The dissipation of energy for receiving *b* bits of collected data is given in

$$E_{receivediss} = E_{enloss} \times b$$

Energy loss per bit transmission or receiver is calculated in E_{enloss} = 50 nJ/bit and constant amplifier □ amp = 0.1nJ / bit / m^2. Ep has mainly two factors for losses of energy at each sensor node are a loss for switching and loss because of leakage. We can show that function Ep to supply voltage V and the frequency f related to the sensor node. The average capacitance Davg shifter per cycle and the average number of clock cycles required for the execution of task Scyc for a given microprocessor device [11]. Therefore, total capacitance switched is Scyc * Davg assuming I0 as the leakage current. The total energy loss can be computed using the equation given below.

$$E_{loss} = (100 + 0.1d^2)b + S_{cyc}D_{avg}W^2 + VI_0e^{v/nv_t}S_{cyc}/f$$

3.2 Mobile Sink Strategy

Design and developing of data gathering using dynamic routing protocol with one or more mobile sinks reduces the overall route length and can minimize the total energy utilization. The flowchart given in figure 4 shows the mobile sink strategy used in the proposed approach.

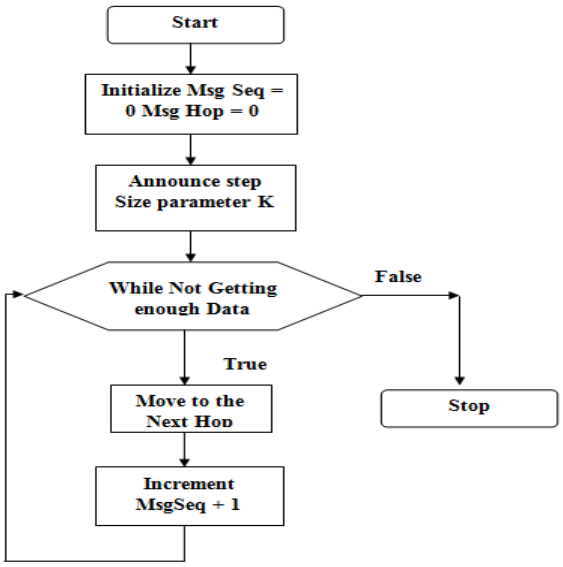

Figure 4. Flow Chart for Mobile sink's strategy

3.3 *Growing Cellular Probabilistic Self Organizing Map*

The dynamically growing cellular algorithm is based on SOM [19][20], the cells selection starts at random positions with n parameters. Parameters can be energy, node density, and a number of clusters. The network formed by DGCSOM defines the nodes of in form of triangles. The DGC consists of phases such as initialization, growing and reconcile. The algorithm outcomes will the network graph with a set of nodes and its connectivity. In the initialization phase, the node vector with weights are selected with random number and growth threshold (GT) is computed according to the requirements of the end users. The pictorial representation of the initial GSOM is shown in figure 4. Next phase is growing phase, in this phase find the weight node vector which is closest to the initial mode and map with the current sensor nodes using Euclidean distance. The finding of the new node is as shown in figure 5.

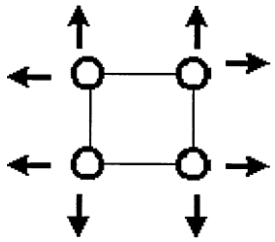

Figure 5. Initial Growing SOM

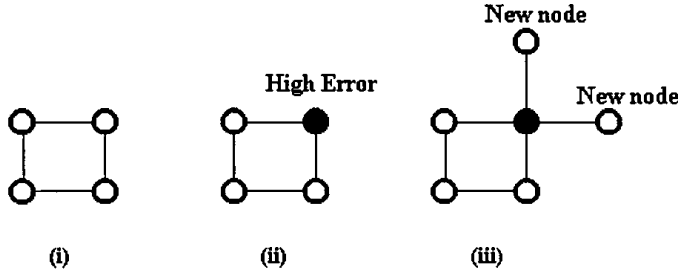

Figure 6. New sensor node formation from the clusters in WSN

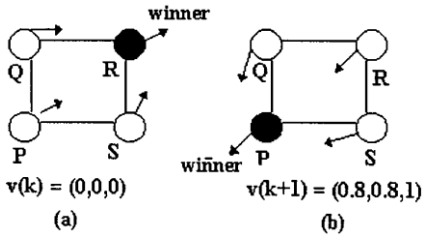

Figure 7. Weigh calculation at the new nodes

Figure 6 shows the weight node vector calculation integrated at the neighborhood to find the winner. Here winner is the next new node for CH elected. In DGC-SOM, CH selection will be based on smaller weights from the initial GSOM so the energy consumption can be reduced gradually at each iteration. The proposed DGC-SOM algorithm for energy efficient data gathering in WSN is given below.

Algorithm: Proposed Dynamically Growing Cellular Self Organizing Map Algorithm
Input: N quantity of wireless sensor nodes
Output: Selected Cluster head and data gathered by sensors in different regions.
Step 1: Deploy 'n' number of sensors in densely distributed area m x n
Step 2: Load SOM with feature hash maps for each cellular topology
Computer the distance using the formula:

$$\sigma_X = |\partial x - P|$$

Where ∂x is the distance from the node x and pre-allocated degree P of a node in the cluster.
Step 3: Initially, assign the Cluster Head (CH) based on problem range
Step 4: Compute Growing threshold (GT) using node density and location of each sensor.
Step 5: Hash map the sensor nodes by using communication energy
Step 6: Compare the GT with total energy consumed for communication. If exceeds GT threshold then dynamically select the node with highest energy dissipation as CH.
Else share the neighbors of CHs
Step 7: Apply the mobile sink strategy to gather data.
Step 8. Repeat the 1 to 5 for the sensor nodes which has not been selected as a cluster head in the given number of iterations.

The growing factor GT is computed form the spreading factor (SF) which is defined by the end user requirement. The SF controls the mapping of the number of clusters. The GT is computed using the formula it is defined in [21]. In our work, we have set D = 2 because we have assumed for 2D input space.

$$GT = -D \times \ln(SF)$$

In reconcile phase of DCG-SOM works based on Kohonen's SOM [22] algorithm smoothing phase, it tunes the required topology by mapping with the output from growing phase of DCG-SOM without modifying the vector size of the hash map.

4. SIMULATION AND RESULTS

In this work is simulated in NS2, the energy utilized and energy required for transmission is calculated, and the efficiency of our proposed DGC-SOM clustering algorithm by varying the number of nodes. A varying number of sensors are deployed uniformly in a 4000 x 4000 square meters area. The transmission range is initialized to 412.45 meters and we can compute the and efficiency of DGC-SOM approach.

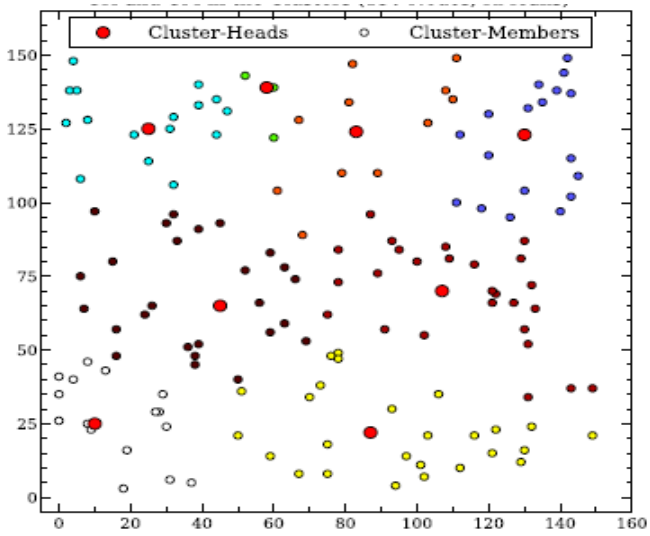

Figure 8. Clusters and cluster head distribution

Figure 8, shows the sensor nodes deployment and cluster members with CH.

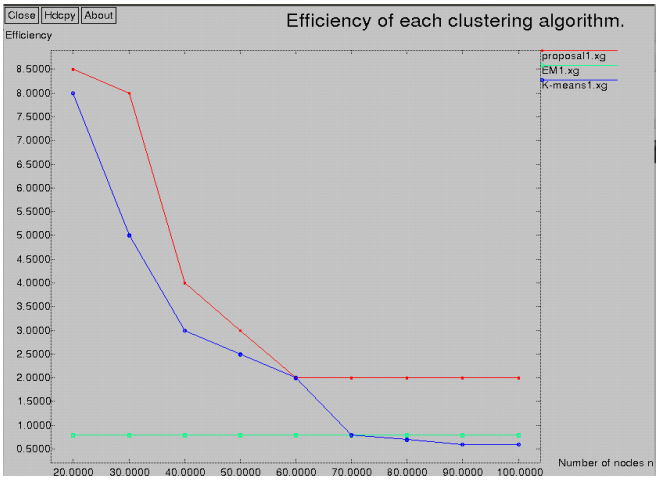

Figure 9. Energy Efficiency of DGC-SOM clustering approach

Figure 9, shows the proposed approach DGC-SOM minimizes the energy gradually. The reason for this minimization is a dynamic selection of CH and distribution of new clusters. The improvement occurred in terms of connectivity and distance of the transmission range. Mobile sink strategy applied gives the significant improvement in the communication of sensor nodes.

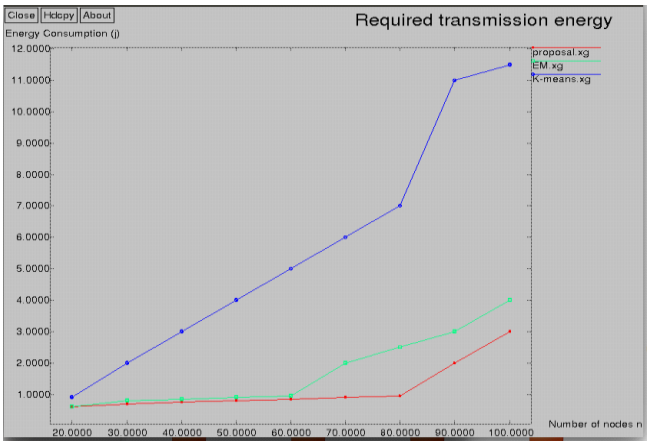

Figure 10. Transmission energy

Figure 10, represents the energy transmission of the proposed DGC-SOM (marked in light green color). The reason for this is DGC-SOM computes the CH nodes using the GT, node location and saves the energy consumption in densely distributed WSN. The number of intermediate nodes is reduced with help of mobile sinks.

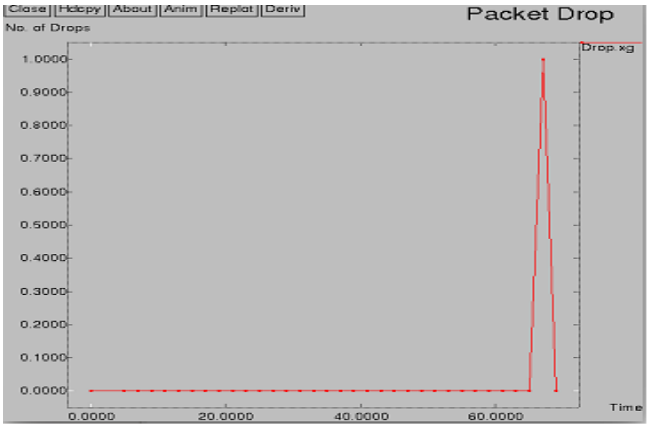

Figure 11. Packet drop

Figure 11, shows the total number of packet drops in communication. The reasons may be due to dead sensor nodes. Some of the sensors may die due to energy loss.

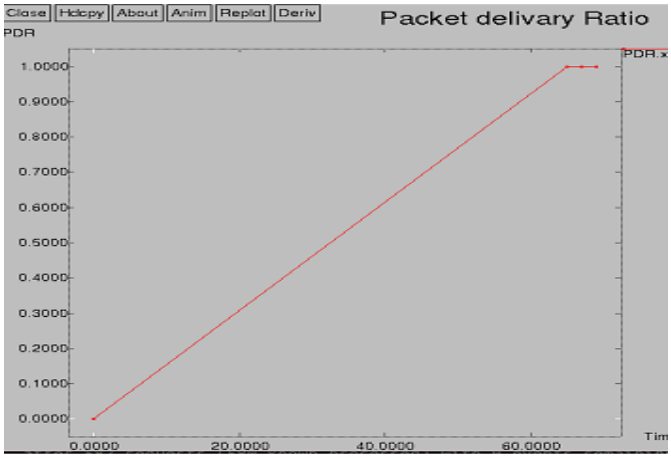

Figure 12. Packet delivery ratio

Figure 12, shows the significant improvement in the packet delivery ratio this is because of the efficient mobile sink used at low efficiency. Data successfully delivered to the base station in densely distributed WSN.

5. CONCLUSION

In view of reducing the energy consumption, we described energy model based on communication distance and node density. A DGC-SOM algorithm is proposed to select the cluster head in WSN. Simulated the algorithm and cluster formation. DGC-SOM give better performance in terms of throughput, packet ratio, packet delivery and transmission. overall cluster achieves better energy. Proposed algorithm DGC-SOM helps in introducing new nodes to the network when energy loss crosses the threshold. DGC-SOM is quite simple to design and develop also gave better QoS. It can be used for monitoring high-speed data communication. Further, this work will be extended with different clustering algorithm with different transmission ranges for the large problem domain.

6. REFERENCES

[1] M.younis, M. Eltoweissy and A.Wadaa "On handling Qos traffic in wireless sensor network" in proceedings of the 37th Annual hawai international conference, IEEE, 2004, pp. 1-10.

[2] S.Tilak, N.B Abu-Ghazaleh and W.Heinzelman "A Taxonomy of wireless micro sensor network models" ACM Mobile Computing and Communications Review, Vol. 6,No 2 ,2002, pp. 28-36.

[3] J. N. Al-Karaki and A. E. Kamal, "Routing techniques in wireless sensor networks: a survey," *Wireless communications, IEEE*, vol. 11, no. 6, pp. 6–28, 2004.

[4] W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energyefficient communication protocol for wireless microsensor networks," in *Proc. 33rd Hawaii Int. Conf. Syst. Sci. (HICSS)*, Washington, DC, USA, Jan. 2000, pp. 1–10.

[5] W. B. Heinzelman, A. P. Chandrakasan, and H. Balakrishnan, "An application-specific protocol architecture for wireless microsensor networks," *IEEE Trans. Wireless Commun.*, vol. 1, no. 4, pp. 660–670, Oct. 2002.

[6] S. Lindsey and C. S. Raghabendra, "PEGASIS: Power efficient gathering in sensor information systems," in *Proc. IEEE Aerosp. Conf.*, Mar. 2002, pp. 3-1125–3-1130.

[7] S. Sagiroglu and D. Sinanc, "Big data: A review," in *International Conference on Collaboration Technologies and Systems (CTS)*, 2013.

[8] S. Lindsey and C. Raghavendra, "PEGASIS: Power-efficient gathering in sensor information systems," in *IEEE Aerospace Conference*, 2002.

[9] H. Nakayama, N. Ansari, A. Jamalipour, and N. Kato, "Fault-resilient sensing in wireless sensor networks," *Computer Communications*, vol. 30, no. 11-12, pp. 2375–2384, Sep. 2007.

[10] L. He, Z. Yang, J. Pan, L. Cai, J. Xu, and Y. Gu, "Evaluating service disciplines for on-demand mobile data collection in sensor networks," vol. PP, no. 99, 2013.

[11] B. Gong, L. Li, S. Wang, and X. Zhou, "Multihop routing protocol with unequal clustering for wireless sensor networks," in 2008 ISECS International Colloquium on Computing, Communication, Control, and Management, vol. 2. IEEE, 2008, pp. 552–556.

[12] [12] J. H. B. Neto, A. d. S. Rego, A. R. Cardoso, and J. Celestino Jr, "Mhleach: A distributed algorithm for multi-hop communication in wireless sensor networks," ICN 2014, pp. 55–61, 2014.

[13] T. H. Cormen, Introduction to algorithms. MIT press, 2009.

[14] [14] A. da Silva Rego, J. Celestino, A. dos Santos, E. C. Cerqueira, A. Patel, and M. Taghavi, "Bee-c: a bio-inspired energy efficient cluster-based algorithm for data continuous dissemination in wireless sensor networks," in Networks (ICON), 2012 18th IEEE International Conference on. IEEE, 2012, pp. 405–410.

[15] O. B. Haddad, A. Afshar, and M. A. Mari˜no, "Honey-bees mating optimization (hbmo) algorithm: a new heuristic approach for water resources optimization," water resources management, vol. 20, no. 5, pp. 661–680, 2006.

[16] Basu, Khan, "A Mobility Based Metric for Clustering in Mobile Adhoc Networks", IEEE, 2001, pp. 413-418.

[17] Ryu, Song, Cho, "New Clustering Schemes for Energy Conservation in two tiered Mobile Adhoc Networks", IEEE, vo1. 3, 2001, pp. 862-866.

[18] Li, Zhang, Wang, "Vote Based Clustering Algorithm in Mobile Adhoc Networks", International Conference on Networking Technologies, 2004.

[19] L. D. Alahakoon, S. K. Halgamuge, and B. Srinivasan, "A structure adapting feature map for optimal cluster representation," in *Proc. Int.* Conf. Neural Information Processing, 1998, pp. 809–812.

[20] "A self growing cluster development approach to data mining," in Proc. IEEE Conf. Systems, Man, and Cybernetics, 1998, pp. 2901–2906.

[21] A. Nuernberger, "Interactive text retrival supported by growing self-organizing maps," presented at Proceedings of the International Workshop on Information Retrival (IR' 2001), (Infotech, Oulu, Finland, 2001).

[22] T. Kohonen, Self-Organizing Maps, Berlin, Spinger, 1997