IJCSC VOLUME 5 • NUMBER 2 JULY-SEPT 2014 PP. 116-123 ISSN-0973-7391

Performance Comparison of Various Clustering Techniques in Wireless Sensor Networks

Gaurang Raval[1], Madhuri Bhavsar[2], Nitin Patel[3]
CSE Department, Institute of Technology, Nirma University, Ahmedabad, Gujarat, India
[1]gaurang.raval@nirmauni.ac.in, [2]madhuri.bhavsar@nirmauni.ac.in, [3]12mict18@nirmauni.ac.in

Abstract

This paper presents the comparison of various Wireless Sensor Network (WSN) clustering protocols like LEACH-Centralized, KMeans based clustering, Fuzzy C-Means clustering and Harmony Search Algorithm based clustering. The protocols have been compared with respect to network lifetime, energy consumption and effectiveness of clustering. It was observed that HSA based method of clustering outperformed LEACH-C, K-Means and FCM based protocols in almost every performance measurement aspects. However, few shortfalls of HSA method were also identified. All these protocols only consider intra cluster distance of members while clustering. Due to that the data delivery of these protocols are not uniform and clusters are not covering the deployment field in uniform way. The clustering problem was also observed for these protocols, wherein large variation was found in the cluster sizes during a particular round of the clustering algorithms

Keywords - LEACH-C, FCM, Harmony Search, Clustering.

1 Introduction

Communication Technology foster the fast growing world with quick and reliable data transfer and information sharing. With the advent of various generations of networking technologies, applications have been developed for automatic data acquisition and information compilation and computation. Sensor networks is one such application where hundreds to thousands of sensor nodes are deployed for gathering information related to environment conditions, performing necessary simple computations and sharing the information with peer nodes or with an external Base Station. A sensor network can be created by manual as well as random deployment. This results into a fault tolerant and maintenance free network for data gathering. Since these sensor networks employ nodes working on battery, it is also important to conserve energy of the nodes so that the life time of network is enhanced. The research of sensor network primarily focuses on reducing the energy drain of the nodes, thereby increasing network life. Data sensing and reporting in sensor networks is dependent on the application and time criticality of data. So, a sensor network can be categorized into time-driven or event driven networks. Time driven network is suitable for applications that require periodic data monitoring, wherein nodes periodically switch on their sensors and transmitters, sense the environment and perform data transmission at constant periodic time intervals. Hence, through this type of network, snapshot of relevant attributes are provided at regular intervals. In the event driven network, a monitored external event triggers the data acquisition and information is transmitted to the base station or peers.

There are many challenging factors which highly influence the design of WSN routing protocols [1]. Deployment of sensor network is random and hence ad hoc. This mandates the system to be able to cope up with the distribution of nodes and establish connections between nodes. Due to limited energy supply, computation and transmission as per the application requirements in a wireless environment with reliability is highly difficult. Limited computation capabilities of sensor nodes does not allow porting of sophisticated network protocols, as is the case with wired networks wherein the nodes are rich in terms of resources. Since communication range of sensor nodes is limited, route will generally consist of multiple wireless hops for communicating with peers or Base Station. A sensor node consists of many hardware components, integrated in less than a cubic centimetre. Such components consume extremely low power and operate in an unsupervised mode. They must also adapt to the environment of sensor network precisely. Collisions generally adds more trouble in to the network like retransmissions, which in turn triggers latency problem and increases energy usage as well. The overhead of control packet in this scenario linearly related to the density of nodes. At times, data is time-critical within a particular duration of time. If failed to deliver during this deadline, the data becomes useless. For time-constrained applications, bounded latency is yet another constraint. To deal with these constraints, sensor networks need highly customized routing protocols.

Routing protocols specifically designed for sensor networks highly depends on application requirements. Major categories as mentioned in [2][3], are Flat which are data centric, Hierarchical and Location based. These protocols are further classified into multipath based, query based, negotiation based, QoS based and coherence based protocols. However, some hybrid protocols exist that fit under more than one categories. D J Dechene et al [4] explored various routing algorithms based on clustering for WSNs and classified them into Heuristic, Weighted Schemes, Hierarchical Schemes and Grid based approaches. Specifically they examine performance of these algorithms in terms of the power and quality aspects of these schemes. G Santosh Kumar et al [5] discussed four routing protocols for wireless sensor networks viz. Flooding, Gossiping, Gradient Based Routing (GBR) and LEACH. They simulated these protocols using TinyOS. As sensor network deployments for critical applications are generally large scale in nature, specific routing techniques are required to handle such large networks. Scalability and efficiency of communication is also achieved using Hierarchical routing technique. Here, higher energy nodes can be used to process and send information, while low energy nodes can be used to perform sensing in the proximity of target. Many such routing protocols have been proposed in the literature such as LEACH [6], LEACH-C[7], K-Means[8][9], FCM[11][13][14][15], HSA[16][17][18]. LEACH is considered as the most popular routing protocol that use cluster based routing in order to minimize the energy consumption. A variant of LEACH, LEACH-C is a centralized version of LEACH which exploits base station for clustering purpose. In this paper, we analyse the performance of various centralized clustering protocols. In section II, we discuss various clustering techniques in detail. Section III present the simulation results and analysis of these methods. Section IV and V conclude the paper.

2 Clustering Techniques

Clustering approaches have become an emerging technology for building scalable, robust and energy efficient applications in order to ensure efficient network operations. Conserving energy resources in such networks is considered as the most important research avenue. As per the review of literature, many contributions have been made for minimizing energy usage during communication. Data is routed form all nodes to a sink node using the process called convegecast. Moreover in these applications it is observed that, energy is consumed more in data transmissions instead of data processing. Hence, an energy efficient technique called data aggregation is used which collects local data at intermediate nodes and forwards the aggregated data. Data aggregation methods are generally synchronized with the packet flow in the network. The protocols have been designed by researchers which allows in-network processing along-with routing of packets. These type of protocols fall in to two different categories, namely tree based aggregation and cluster based aggregation. Tree based aggregation may suffer from latency when number of nodes are large. To control the latency, clustering techniques may be applied. Every cluster reports aggregated data to the base station. Within a network, all sensor nodes group themselves in to several regions using short messages in cluster based data aggregation technique. Here, at any instance one node acts as a cluster-head and sends short messages to other remaining nodes. Depending upon the signal strength the sensors decide to join the groups.

LEACH [6] is very popular protocol used by researchers in hierarchical routing related work. LEACH attempts to minimize energy dissipation. It is based on a simple clustering mechanism by which energy can be conserved since cluster heads are selected for data transmission instead of other nodes. LEACH operates iteratively in rounds till all nodes are not expired. Each round has a setup-phase and steady-state-phase. In setup phase clusters are organized and in steady-state phase data is transferred to sink. To reduce overhead setup phase is shorter compared to steady-state phase. Tyagi and Kumar [10] present detailed review of LEACH related protocols. The literature review suggest that LEACH protocol suffers from several problems, one of them is non-uniform clustering. Better clusters may be produced using centrally controlled algorithm called LEACH Centralized (LEACH-C) [7]. It has similar phases like LEACH. The steady-state phase of LEACH-C is exactly similar to LEACH. During set-up phase of LEACH-C, all nodes send their geographic details and present energy level to base station. BS calculates the average network energy, and assigns the role of cluster-heads (CH) to the members who are having energy above the threshold for the current round. BS finds clusters using the simulated annealing algorithm [12]. LEACHC controls energy usage of non-cluster head nodes, by creating clusters wherein the cluster-head and members are as close as possible. The algorithm assigns the members to clusters in an optimized way, by minimizing sum total of squared distances between all member nodes and the candidate cluster-head. After the cluster-heads and associated cluster members are finalized, BS broadcasts message containing cluster-head ID for each node. All receiving nodes match their own ID with the received cluster-head ID, to determine whether they are cluster-heads or not. If a node is not designated as cluster-head, it determines its TDMA slot for data transmission and changes the operating mode to sleep state until the desired TDMA slot is available to transmit data. The energy model used is as follows. Energy required to transmit L bits at a distance d,

$$E_{TX}(L,d) = \begin{cases} L(E_{elec} + \varepsilon_{fs} d^2 & d < d_0 \\ L(E_{elec} + \varepsilon_{mp} d^4 & d \geq d_0 \end{cases} \quad (1)$$

Where d_0 is the deciding factor whether to use free space propagation model or multi-path radio propagation model. $\square_{fs}$and $\square_{mp}$ are the amplification components depending on the propagation model in use. Energy required to receive L bits,

$$E_{RX}(L) = LE_{elec}$$

In [9], authors have proposed K-Means based clustering where, they minimize the objective function,

$$F = \sum_{j=1}^{K} \sum_{i=1}^{N} \left|X_i^{(j)} - C_j\right|^2 \quad (2)$$

Here $\left|X_i^{(j)} - C_j\right|^2$ is a chosen distance measure between a data point $X^{i(j)}$and the cluster centreC_j. It is an indicator of the distance of the n data points from their respective cluster centres. K-Means Algorithm [9] creates 'k' centroids randomly and then computes the Euclidian Distance of each node with every centroid. It iteratively computes the location of centroid in each cluster and looks for change in the position with respect to previous iteration. If change is found in the centroid position, the algorithm continues, otherwise the clusters are considered final and the algorithm converges.In [8], authors have proposed BPK-Means clustering algorithm. They refer the algorithm as balanced parallel K-Means algorithm. Authors balance the cluster size by keeping the number of members exactly same in every cluster.In [11] and [15], authors have applied Fuzzy C-Means clustering algorithm for centralized clustering. The objective function to be optimized is,

$$J_m = \sum_{i=1}^{N} \sum_{j=1}^{C} \delta_{ij}^m \left\|X_i - C_j\right\|^2, where\ 1 \le m < \infty (3)$$

where

$$\delta_{ij} = \frac{1}{\sum_{k=1}^{C} \left(\frac{\|X_i - C_j\|}{\|X_i - C_k\|} \right)^{\frac{2}{m-1}}} (4)$$

The norm, $\left\|X_i - C_j\right\|$measures the similarity (or closeness) of the data point X_i to the centre vector C_j of cluster j. Cluster centers are computed with,

$$C_j = \frac{\sum_{i=1}^{N} \delta_{ij}^m . X_i}{\sum_{i=1}^{N} \delta_{ij}^m} (5)$$

The degree of membership for data point i to cluster j is initialised with a random valueδ_{ij}, $0 \le \delta_{ij} \le 1$, such that $\sum_j^C \delta_{ij} = 1$.Initially the degree of membership is assigned randomly. As the algorithm converges, clusters are evolved with termination criteria. The termination conditions are the least overlapping membership of members with neighboring clusters and the fuzziness coefficient of the membership. The FCM algorithm terminates when the desired accuracy is achieved for the degree of membership, till then the algorithm iteratively executes and compares the previous and current degree of membership. The largest difference between the current and previous degree of membership for all data points with all clusters is the termination acts as termination criteria here. Once the clusters are created with the centroids, the members near the centroid are labelled as cluster-head depending on members' available energy.

Harmony Search Algorithm was initially proposed by Geem and Kim[19]. Later applied for clustering process in WSN by Hoang, Kumar and Panda[16]. Authors find the best clusters using intra cluster distance and residual average energy. They recommend nodes with higher residual energy to become cluster head along-with consideration of intra cluster distance of cluster members from cluster heads. They define Harmony Memory as matrix HM (eq. 6) having randomly selected member elements in every row. Each vector in the matrix HM indicates that elements in the row may be cluster head after the algorithm converges. These vectors are partially altered or replaced altogether with harmony improvisation logic.

$$HM = \begin{bmatrix} V_1^1 & V_1^2 & \cdots & V_1^k \\ V_2^1 & V_2^2 & \cdots & V_2^k \\ \cdots & \cdots & \cdots & \cdots \\ V_{HMS}^1 & V_{HMS}^2 & \cdots & V_{HMS}^k \end{bmatrix} \begin{bmatrix} F^1 \\ F^2 \\ \cdots \\ F^{HMS} \end{bmatrix} (6)$$

Once the HM is derived, a new harmony memory is devised. The new Harmony from HM is derived using Harmony Memory Consideration Rate (HMCR). A new harmony may be directly selected from the candidates present in the

HM with probability HMCR or they may be generated fresh from the pool of all the sensor nodes which are alive using the probability 1 - HMCR.

$$V_j' \leftarrow \begin{cases} V_j' \in HM(j) & \text{with probability } HMCR \\ V_j' \in V_{candidate} & \text{with probability } (1 - HMCR) \end{cases} \quad (7)$$

If the new harmony has been derived from HM itself then it is further fine-tuned with Pitch Adjustment Rate (PAR). With probability PAR, member elements in the new harmony may be mutated with other members in the same position in the Harmony Memory having residual energy higher than the average energy. Otherwise the new harmony will be left as it is with the probability 1-PAR. Pitch Adjustment for selected $V_j^{`}$ is,

$$V_j' \leftarrow \begin{cases} V_j^n \in HM & \text{with probability } PAR \\ V_j' & \text{with probability } (1 - PAR) \end{cases} \quad (8)$$

The goal is to minimize the objective function:

$$f_{obj_{min}} = \alpha \times f_1 + (1 - \alpha) \times f_2 \quad (9)$$

where

$$f_1 = max_{j \in (1,k)} \left\{ \frac{\sum_{\forall node_i \in C_j} d(node_i, CH_j)}{|C_j|} \right\} \quad (10)$$

The function f_1 finds the maximum of the average member distance with the cluster head of every cluster.

$$f_2 = \sum_{j=1}^{k} \left\{ \frac{\sum_{\forall node_i \in C_j} V_i^{res}}{V_{CH_j}^{res}} \right\} \quad (11)$$

The function f_2 finds the effect of residual energy of CH on to the sum of total energy of every member in every cluster. The objective function is designed in such a way that this term's contribution is minimized, thereby ensuring that members with higher residual energy become cluster-heads.

3 Simulation Results and Analysis

Protocols discussed in previous section were simulated in NS-2 with C++ and TCL. Table 1 shows the network specific simulation parameters used. Table 2 shows the energy dissipation parameters used during simulation. The simulations were performed for different network parameters settings. The deployment was varied from 50 to 200 meters. The number of nodes were changed from 50 to 200 nodes. The base station location was also changed between center position and 50,175 in the field. Figure 1 shows the network performance in terms of number of nodes alive during the simulation. It may be observed that Harmony Search Algorithm preforms better compared to other protocols, as it looks for the best possible clustering using evolutionary technique.

Table 1: Simulation Parameters

Network Parameter	**Value**
Node distribution	(0,0) to (200,200)
BS location	Center and (50, 175)
No. of Nodes	50,100,150,200
Initial Node Energy	2J
Simulation Time	3600s
Desired No. of cluster-heads	5% (Optimal)
Bandwidth of the channel	1 Mbps
Packet header size	25 Bytes
Message size	500 Bytes

Table 2: Radio parameters

Operation	**Symbol**	**Energy dissipated**
Energy consumed for transmitting and receiving	E_{elec}	50nJ/bit
Energy consumed by amplifier to transmit at shorter distance $(d_{toBS} \leq d_0)$	$\square_{fs}$	10pJ/bit/m^2
Energy consumed by amplifier to transmit at longer distance $(d_{toBS} \geq d_0)$	$\square_{mp}$	0.0013pJ/bit/m^4
Energy consumed while data aggregation	E_{DA}	5nJ/bit

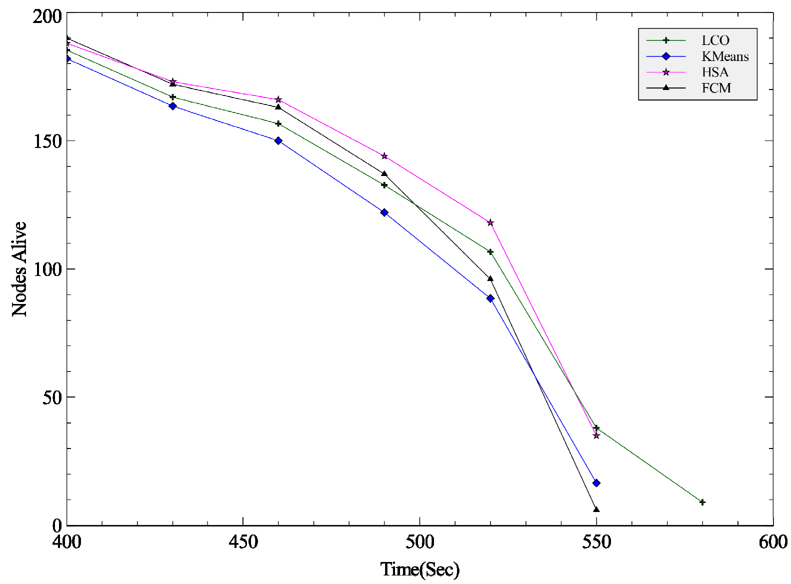

Figure 1: Alive nodes for LCO, KMeans, FCM and HSA - 200 nodes deployment

Figure 2 shows the data reported by various protocols during simulation to the base station via cluster heads. The data reporting of HSA and FCM based methods are almost similar. But carefully examining Figure 1 shows that number of alive nodes quickly falls in FCM whereas, in HSA it is gradually falling, but it is still performing better compared to other protocols when number of alive nodes is the comparison criteria.

Figure 3 shows the clustering done by KMeans algorithm for 100 nodes deployment. It was captured at a particular round that which nodes are elected as cluster-heads and how many members are assigned to them. It may be seen that one cluster-head is assigned only 2 member nodes while another cluster-head is assigned close to 50 nodes. So clustering method is not distributing members uniformly across all clusters. Similar problem was also observed for FCM based clustering. It is presented with Figure 4. Here also, one cluster-head is assigned less than 10 nodes whereas the other cluster-head is assigned more than 30 nodes, out of total 100 nodes.

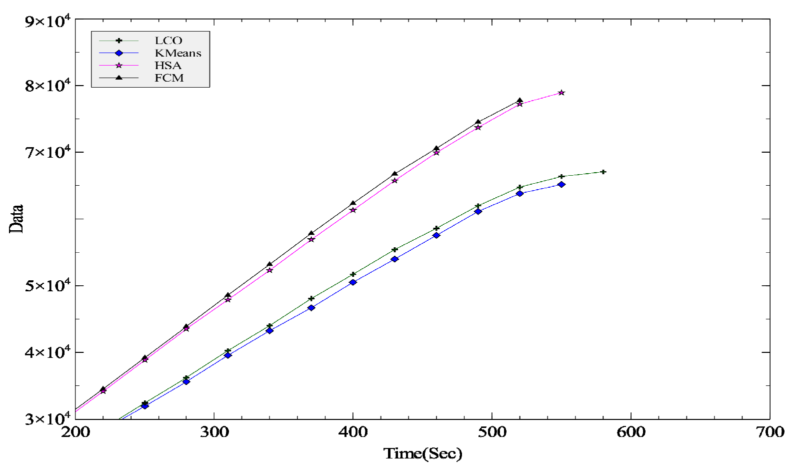

Figure 2: Data Delivered for LCO, KMeans, FCM and HSA - 150 nodes deployment

HSA [16] algorithm outperforms LEACHC, KMeans, FCM and several other evolutionary computing methods, but it sometimes creates worst clustering. The problem behind this issue is the selection of the function optimization parameters. It does not consider inter-cluster distance while doing clustering. Due to this, sometimes, the cluster-heads selected may be very close to each other resulting in to very less separation from neighboring clusters. The clustering effect of HSA is shown in Figure 5. The clustering process snapshot is shown for 50 nodes deployment in the field.

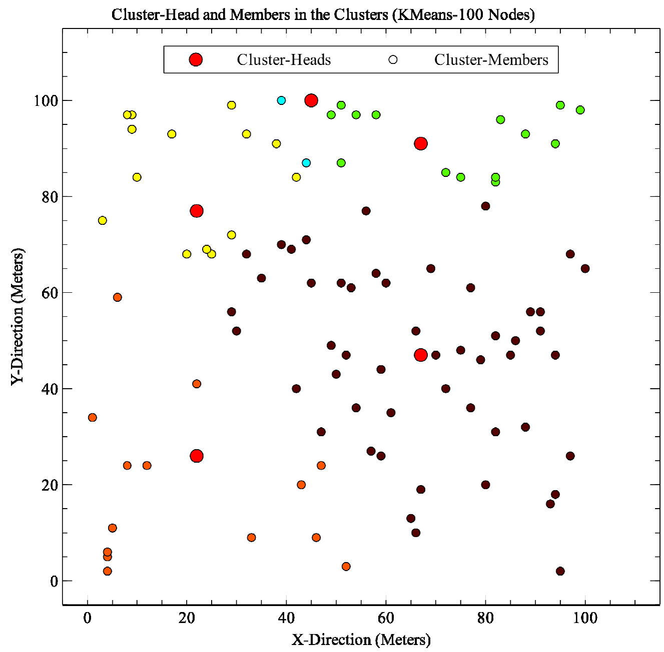

Figure 3: Clustering process with KMeans - 100 nodes deployed

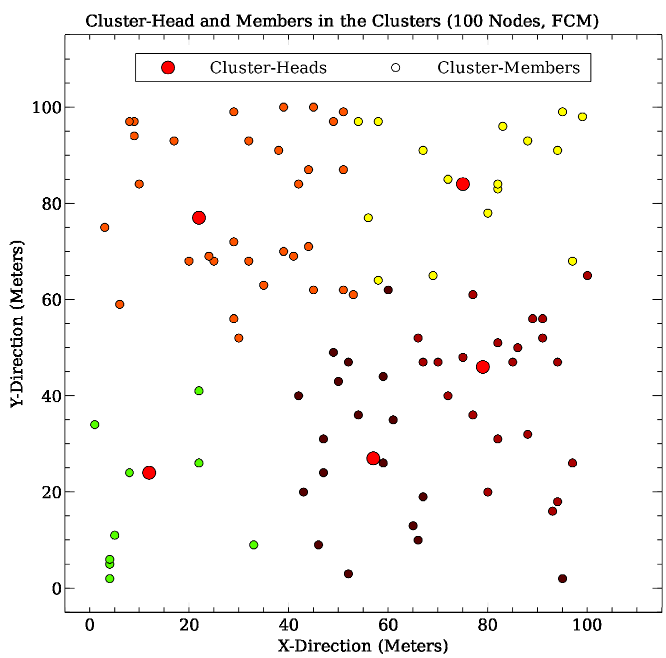

Figure 4: Clustering process with FCM - 100 nodes deployed

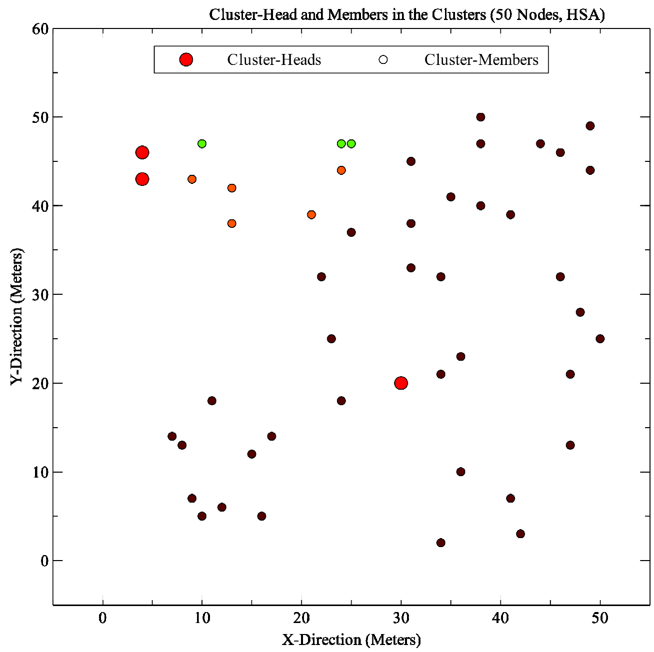

Figure 5: Clustering process with HSA - 50 nodes deployed

4 Conclusion

Various clustering protocols where compared especially for the clustering process. Harmony Search Algorithm was found performing best when compared with LEACHC, FCM and KMeans based clustering methods. All these methods were applied with centralized clustering approach. HSA based method exhibited superior performance when data delivery and lifetime of the network was compared with other protocols. However all of these algorithms suffer from the problem of unbalanced clustering including HSA based method.

5 Future Work

The problems of the clustering techniques can be eliminated up to some extent with some kind of density control approach. The uniformity of data delivery process may also be analyzed further, to check the network coverage during the data arrival at the base station.

References

[1] Mohammad Ilyas and Imad Mahgoub."Handbook of sensor networks : compact wireless and wired sensing systems", Book, Chapter 1 & 6 CRC Press, 2005.

[2] Kemal Akkaya, Mohamed Younis, "A survey on routing protocols for wireless sensor networks",in the Elsevier Ad Hoc Network Journal, Vol. 3/3 pp. 325-349, July, 2005.

[3] Jamal N. Al-Karaki,Ahmed E. Kamal,. "Routing Techniques In Wireless Sensor Networks: A Survey",IEEE Wireless Communications, December 2004.

[4] D. J. Dechene, A. El Jardali, M. Luccini, and A. Sauer, "A Survey of Clus- tering Algorithms for Wireless Sensor Networks",Computer Communications, Butterworth-Heinemann Newton, MA USA, October 2007.

[5] G.Santhosh Kumar, Lino Abraham Varghese, Jose Mathew K and K. Poulose Jacob, "Evaluation of the Power Consumption of Routing Protocols for Wireless Sensor Networks", Ad Hoc and Ubiquitous Computing, ISAUHC '06. Interna- tional Symposium IEEE, 2006.

[6] W. Heinzelman, "Application-specific protocol architectures for wireless networks", Ph.D. dissertstion, Mass. Inst. Technol., Cambridge, 2000.

[7] Wendi B. Heinzelman,Anantha P. Chandrakasan,Hari Balakrishnan, "An Application-Specific Protocol Architecture for Wireless Microsensor Networks",2002 .

[8] Liansheng Tan, Yanlin Gorg, Gong Chen, "A Balanced Parallel Clustering Protocol for Wireless Sensor Networks Using K-means techniques" , IEEE Second International Conference on Sensor Technologies and Applications, 2008.

[9] P. Sasikumar, Sibaram Khara,"K-Means Clustering In Wireless Sensor Networks",IEEE Fourth International Conference on Computational Intelligence and Communication Networks, 2012.

[10] Sudhanshu Tyagi, NeerajKumar,"A systematic review on clustering and routing techniques based upon LEACH protocol for wireless sensor networks", Journal of Network and Computer Applications,2012.

[11] A. S. Raghuvanshi, S Tiwari, R Tripathi and N. Kishor. "Optimal Number of Clusters in Wireless Sensor Networks: An FCM Approach",IEEE International Conference on Computer and communication Technology, 2010.

[12] T. Murata and H. Ishibuchi, "Performance evaluation of genetic algorithms for flowshop scheduling problems", Proc. 1st IEEE Conf. Evolutionary Computation, vol. 2, pp. 812-817, June 1994.

[13] Raghuvanshi, Ajay Singh and Tiwari, Sudarshan and Tripathi, Rajeev and Kishor, Nand, "Optimal number of clusters in wireless sensor networks: a FCM approach", IJSNet, 2012

[14] Medhat, Fady and Ramadan, Rabie A and Talkhan, Ihab, LM, "Smart Clustering for Multimodal WSNs", IEEE Seventh International Conference on Broadband, Wireless Computing, Communication and Applications, 2012

[15] Hoang, DC and Kumar, R and Panda, SK, "Fuzzy C-means clustering protocol for wireless sensor networks", IEEE International Symposium on Industrial Electronics, 2010

[16] Hoang, DC and Kumar, R and Panda , SK , "A Robust Harmony Search Algorithm based Clustering Protocol for Wireless Sensor Networks", IEEE International Conference on Communications Workshops (ICC), 2010

[17] Osama Mohd Alia, Zyad Shaaban, Ahmad Basheer, Alaa Al-Ajouri and Ahmed Alsswey, "Musicians'-Inspired Clustering Protocol for Efficient Energy Wireless Sensor Networks", IEEE International Conference Communications and Networking (ComNet), 2014

[18] Hoang, DC and Kumar, R and Yadav, P and Panda, SK ,"Real-Time Implementation of a Harmony Search Algorithm-Based Clustering Protocol for Energy-Efficient Wireless Sensor Networks", IEEE Transactions on Industrial Informatics, Vol. 10, No. 1, February 2014

[19] Geem, Zong Woo, Joong Hoon Kim, and G. V. Loganathan. "A new heuristic optimization algorithm: harmony search." Simulation 76.2 (2001): 60-68.