

Optimization Algorithm for Wireless Sensor Networks using K-means and Evolutionary Algorithm

R. M. El-Awady¹, Rehab F. Abdel-Kader², and M. F. Tobal²

ABSTRACT

Recently, many new types of wireless networks have emerged for both civil and military, applications, such as wireless sensor networks, ad hoc networks, among others. In this paper, we mainly focus on static sensor networks. In these networks, every node is capable of sensing, data processing, and communication, and operates on its limited amount of battery energy consumed mostly transmission and reception and its radio transceiver. We propose a merged algorithm between Evolutionary Algorithm (EA) and K-means clustering. K-means divide the sensor network into K-clusters (K-known); EA finds the optimal Fitness of the network. The proposed algorithm determines the optimal number of clusters, cluster elements, and the head of each cluster. Thus, the total distance will be reduced and this will lead to energy saving and long life time network.

Keywords: sensor networks; clustering; K-means; Evolutionary Algorithm; Network life time

1. INTRODUCTION

A wireless sensor network (WSN) is a network of wireless sensor nodes, which consists of spatially distributed devices using sensor nodes to monitor physical or environmental conditions, such as sound, temperature, and motion, at different locations. Figure 1 shows a typical sensor network structure, in which, sensor nodes are gathering the information from the environment, as well as transmitting sensed data back to the base station [1].

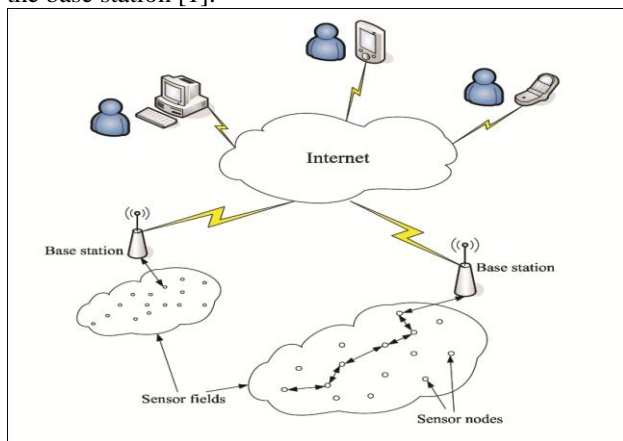


Figure 1: Typical structure of a WSN

The energy consumption during communication is a major energy depletion factor, because the battery is not changeable. Therefore, sensor nodes should reduce the number of transmissions or communicate within short distances to achieve extended battery life [2].

Network lifetime can be defined as the time elapsed from the network operation starts until the first node (or the last node) in the network depletes its energy (dies).

In applications using direct transmission, sensor nodes transmit their sensed data directly to a base station (BS). Thus, the nodes located far from the BS will die quickly since they dissipate much energy in transmitting data.

Clustering is Grouping sensor nodes into clusters has been widely pursued by the research community in order to achieve the network scalability objective. Every cluster has a leader, often referred to as the cluster-head (CH). CH is data aggregation for information flows coming from the local cluster, and forwarding (or relaying) the aggregated information to the base station as shown in Figure 2 [3].

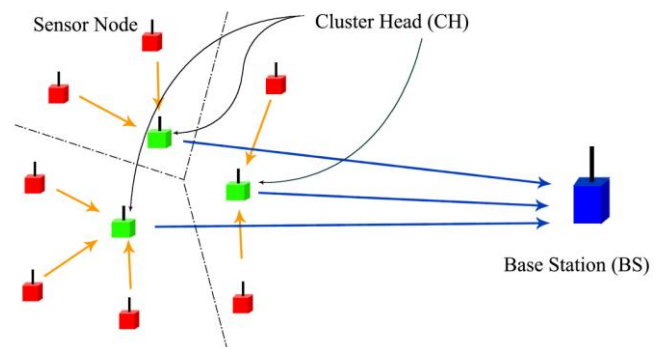


Figure 2: A cluster-based sensor network

In our study, we propose an optimal algorithm for clustering the sensor nodes. The algorithm creates clusters of uniform size such that the distance between the sensor nodes and the CHs is minimized to reduce the consumption of transmission energy.

A new merged algorithm based on k-means clustering algorithm and Evolutionary Algorithm (EA) is proposed. K-means is the simplest and most popular among iterative partitional clustering algorithms generate a single partition, with a specified or estimated number of non entanglement clusters, we use it to determine the cluster members of known clusters. Evolution Strategies is type of EA's used to obtain the optimal number of

¹ Faculty of Engineering, Mansoura University, Mansoura, Egypt
(actt_egypt@yahoo.com)

² Faculty of Engineering, Port-Said University, Port-Said, Egypt
(r.abdelkader@eng.psu.edu.eg - mftobalsh@gmail.com)

clusters and the cluster heads. This provides energy efficient solutions to the optimizer.

The rest of the paper is organized as follows. Section 2 presents a brief survey of related work in the literature. In section 3 a description of the k-Means algorithm is presented. Section 4 briefly describes Evolutionary Algorithms. Section 5 introduces the proposed algorithm. The experimental evaluation, results and discussions are reported in Section 6. Finally, we summarize the paper with some concluding remarks and future research directions in Section 7.

2. RELATED WORK

Heinzelman et al. [4] described the low-Energy adaptive clustering hierarchy (LEACH) protocol, a hierarchical, self-organized cluster-based approach for monitoring applications. He divided the network into several clusters. Each cluster has a special node as a cluster head and several other nodes. Nodes gather data and transmit them to the cluster head, cluster heads deliver gathered data to the sink using time division multiple access (TDMA). This idea had been used by T.Voigt et al. [5] and developed as Solar LEACH clustering algorithm.

O. Moussaoui, et al. [6] assumed clustering algorithm considers nodes number that a cluster head can handle aiming at ideally partitioning the network. In fact, the nodes number constituting the cluster is limited around a pre-defined threshold, facilitating the optimal operation of the medium access control (MAC) protocol. The clustering algorithm is not periodic, and is aimed to decrease the computation and communication costs.

Rahul Khanna, et al. [7] presented a new approach to design a self-organizing network based on genetic algorithms. Sensors that are randomly placed are assigned functions (sensing node, cluster-head, router, or inactive-node) based upon the results of Genetic Algorithm (GA). The GA approach optimizes the network to maximize energy usage along with battery conservation with route optimization. They use multi-objective genetic algorithms (MOGA) which is used to develop a set of solutions (the population) towards the Pareto-optimal front where trade off analysis can be performed to select a suitable solution.

Noritaka Shigei, et al. [8] proposed two types of clustering methods for wireless sensor networks. The first type, which is based on centralized management, uses vector quantization (VQ) for effective clustering. The second type, which is performed in a distribute, takes into account remaining battery level and node density. The proposed methods are compared with conventional methods LEACH, HEED and ANTCLUST. The effectiveness of the proposed methods are demonstrated by numerical simulation.

Ehsan Heidari, et al. [9] assumed an effective method based on genetic algorithms (GAs) to find a solution for

a sensor network optimization problem. Long communication distances between sensors and a sink in a sensor network can greatly drain the energy of sensors and decrease the lifetime of a network. By clustering a sensor network into a number of independent clusters using a GA, They greatly reduce the total communication distance, thus increase the network lifetime.

Mohaned Al. Obaidy, et al. [10] presented the implementation two of the main Evolutionary Computation techniques which are; Genetic Algorithms (GAs) and Particle Swarm Optimization (PSO) to optimize the energy dissipation in a dynamic Wireless Sensor Network (WSN). The GAs in the first stage are used to divide the sensor network into K-clusters (K-unknown). The output of the first stage will be used as an initial population for the PSO which represents the dynamic Sensor Network. GAs proved to be used effectively in the optimization of static Sensor Networks. In this paper, we compare with the first stage because we study static Sensor Network life time.

3. K-MEANS CLUSTERING

K-means clustering is an algorithm to classify or to group given objects according to attributes or variables (e.g. in our paper by distance) into K- known number of groups. K is a positive integer number. The K-Means clustering algorithm was developed by J. Mac-Queen (1967) and then by J. A. Hartigan and M. A. Wong around 1975. The grouping is done by minimizing the sum of squares of distances between data and the corresponding cluster centroid. Given k, these centroids should be put in a skilful way because different location causes different result. The k-means algorithm is performed [11] [12] in 4 steps:

1. Partition objects into k nonempty subsets and Compute seed points as the centroids $z_1(1), z_2(1), z_3(1), \dots, z_K(1)$ of the clusters of the current partition.
2. At the k-th iterative step, distribute the object $\{x\}$ among the K clusters using the relation,

$$x \in C_j(k) \text{ if } \|x - z_j(k)\| < \|x - z_i(k)\| \quad (1)$$

For all $i = 1, 2, \dots, K; i \neq j$; where $C_j(k)$ denotes the set of samples whose centre is $z_j(k)$.

3. Compute the new cluster centers $z_j(k+1), j = 1, 2, \dots, K$ such that the sum of the squared distances from all points in $C_j(k)$ to the new cluster centre is minimized. The measure which minimizes this is simply the sample mean of $C_j(k)$. Therefore, the new cluster centre is given by

$$Z_{j(k+1)} = \frac{1}{N_j} \sum_{x \in C_j(k)} x \quad (2)$$

where $j = 1, 2, 3 \dots K$ and N_j is the number of samples in $C_j(k)$.

4. Go back to Step 2, stop if $z_j(k+1) = z_j(k)$ for $j = 1, 2, \dots, K$ then the algorithm has converged and the procedure is terminated. Otherwise go to Step 2.

The K-means has sensitivity to initial condition in different initial condition may produce different result of cluster. The algorithm may be trapped in the local optimum.

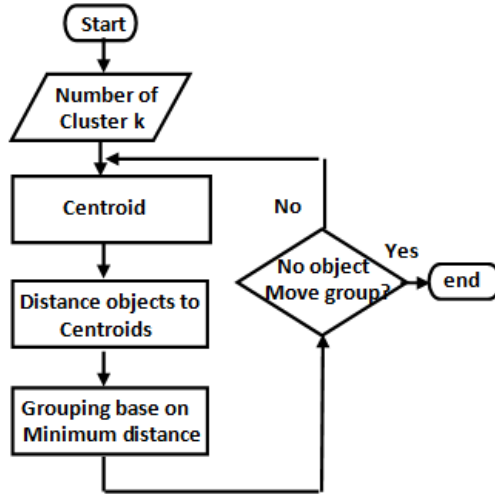


Figure 3: Flow chart of k-means

Although it can be proved that the procedure will always terminate, the K-means algorithm does not necessarily find the most optimal configuration, corresponding to the global objective function minimum. The algorithm is also importantly sensitive to the initial randomly selected cluster centroids. The K-means algorithm can be run multiple times to reduce this effect. K-means is a simple and fast algorithm that has been adjusted to many problem applications.

4. EVOLUTIONARY ALGORITHMS

Evolutionary algorithms (EAs) are population-based metaheuristic optimization algorithms that use biology inspired mechanisms like mutation, crossover, natural selection, and survival of the fitness test in order to develop a set of solution candidates repeatedly [13]. Evolutionary algorithms have advantage compared to other optimization methods. EAs make only few assumptions about the underlying objective functions. The basic cycle of evolutionary algorithms as shown in Figure 4 [14]

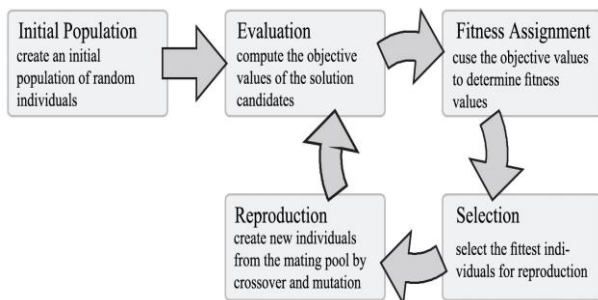


Figure 4: The basic cycle of evolutionary algorithms.

Evolutionary Algorithms are classified into five categories: Genetic Algorithm, Evolution Strategies, Genetic Programming, Learning Classifier Systems and Evolutionary Programming [13].

4.1. Genetic algorithm (GA)

GA is an iterative process that maintains a community of structures that represent candidate solutions. Each solution candidate called chromosome. The chromosome can be a clone of itself, or an adjusted version of its parent. The best chromosome in the population is selected considering a fitness function which is closely correlated to the problem's objective [15]. GA is based on three basic operators:

1. Reproduction produced through genetic algorithms breeding new generations of solutions modified by selecting the parents which have high estimates of efficiency or by giving such parents more likely to be a contribution in the production of new generations.

2. Crossover uses several algorithms genetic sequences of a binary code (1,0) to represent the solutions and transit by choosing a site randomly on the chain (e.g. after double-digit), and exchange sectors, either from the right or to the north of this point with the other series is divided in the same way.

3. Mutation is a random change in the transformation of attitudes and this procedure is required in some cases to save the algorithm from being trapped at local optimum. Mutation randomly changes some bits of the chromosome from 1 to 0, or vice versa through a property transfer.

4.2. Evolution Strategies (ES)

ES introduced by Rechenberg is a heuristic optimization technique based on the ideas of adaptation and evolution, a special form of evolutionary algorithms. Mutation and selection are the primary operators and recombination is less common. In this paper we used ES when a population only consists of a single individual which is reproduced from the elder and the offspring, the better individual will survive and form the next population.

4.3. Genetic Programming (GP)

It is often used to subsume all evolutionary algorithms that have tree data structures as genotypes. In GP, generally some inputs or situations and corresponding output data samples are known or can be produced or simulated. The aim is to find a program that links them, or that shows the kind of behavior required in harmony with the specific cases.

4.4. Learning Classifier Systems (LCS)

These systems are a special case of production systems and consist of four major parts:

1. A set of production rules of interaction, called classifiers.
2. Performance of algorithm that directs the actions of the system in the environment.
3. Learning algorithm which maintains the success of each work and the distribution of rewards.
4. creation genetic algorithm which modifies a set of classified so that the good works of parameters still exist and new ones that are likely to be best in an efficient way.

4.5. Evolutionary Programming (EP)

EP is an evolutionary approach that manages the cases of the genome as different kinds rather than as individuals. Over the decades, it has more or less merged into Genetic Programming and the other evolutionary algorithms. Mutation and selection are the only operators used in EP and recombination is usually not applied.

5. PROPOSED ALGORITHM

In this work, a new algorithm is proposed to save the dissipation and increase the life time network. Before we enumerate the proposed algorithm we should explain the important factors of each sensor.

5.1. Energy Model for Optimisation

We studied the influence of the transmission range of sensor nodes and positioning of the sink in minimising the communication energy in a sensor network. The important elements of each sensor are the data, control processing unit and the radio for communication. The microprocessor used in the processing unit should be energy efficient with less energy consumption. The energy dissipation in the radio depends on the different traits of the radio. The energy model used in this work is adopted from [16] [10] and is briefed here. The energy dissipation for transmitting b bits to d distance is shown in Equation 3.

$$E_{tx}(b, d) = E_{elec} \times b + E_{amp} \times b \times d^2 \quad (3)$$

The energy dissipation in a node to receive b bits of data is shown in Equation 4.

$$E_{rx}(b) = E_{elec} \times b \quad (4)$$

Where E_{elec} is the radio energy dissipation and E_{amp} is the transmitting amplifier energy dissipation. Energy consumption of a wireless sensor node transmitting and receiving data from another node at a distance d can be splitted into two main elements: Energy used to transmit, receive and amplify data and energy used for processing the data, mainly by the microprocessor. Leakage current can be as large as a few mA for the microprocessor, and leakage current can be omitted for higher frequencies and lower supply voltage. We assume the leakage current is negligible, the total energy loss for the wireless

sensor network in our system due to the distance E_{dd} can be computed according to Figure 2 using the following equation:

$$E_{dd} = \sum_{j=1}^k \sum_{i=1}^{n_j} (d_{ij}^2 + \frac{D_j^2}{n_j}) \quad (5)$$

For more details about the derivation and proof refer to [17] [10].

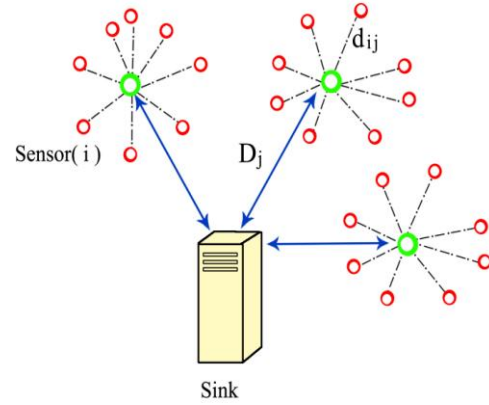


Figure 5: Energy Model for distance based Sensor Network

Where D_j is the distance between cluster-heads and the sink-point, d_{ij} is the distance between the sensor-nodes and its related cluster-heads. k represents the number of clusters and n is the total number of sensors in the network.

5.2. Proposed Algorithm

We suppose the fitness function can be calculated using the following equation:

$$Fitness = \min \left(\sum_{j=1}^k \min \left(\sum_{i=1}^{n_j} (d_{ij}^2 + \frac{D_j^2}{n_j}) \right) \right) \quad (6)$$

This fitness depends on the distances d_{ij} and D_j for each cluster. The main objective is to minimize the fitness function by our algorithm where the distance calculated by using the following equation:

$$Euclidian = \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2} \quad (7)$$

We use K-means algorithm to divide to k -known clusters of n population and Evolutionary Algorithm to obtain the minimum fitness of each cluster and sum this fitness with other clusters. Repeat this process at a different number of clusters and compare with pervious results. Finally, we know the optimal number of clusters and its head by EA and also the members of each cluster by k-means.

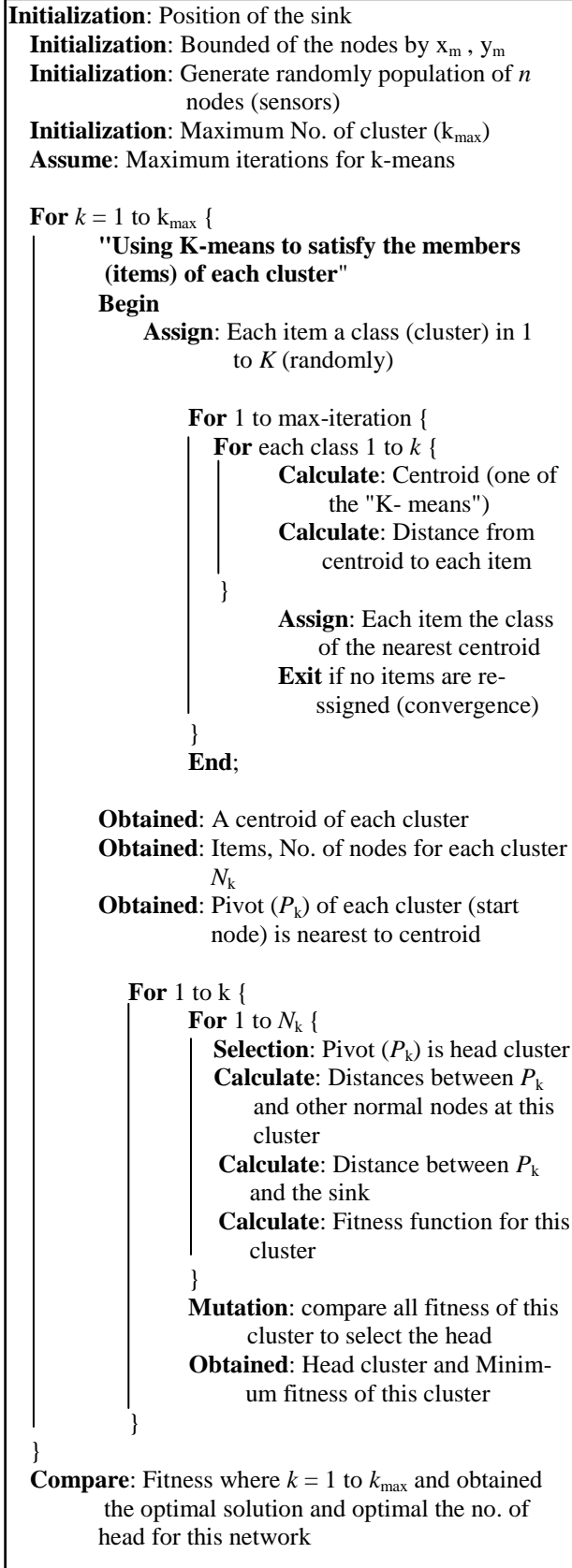


Figure 6: Proposed Algorithm

In the proposed algorithm K-means operator is used as a search operator instead of crossover. Mutation and selection are the primary operators and recombination is less common

6. RESULTS AND DISCUSSION

Experimental results are performed in order to evaluate the performance of the proposed algorithm. The Proposed method is implemented using MATLAB version 7.2 on Intel i3 Core Processor Laptop with 2 MB of RAM memory to simulate an experiment with 80 generated random nodes in a simulated 2-D environment at bounded area 200 by 200 m with two different sink positions located at (0,0) and (100,100).

The system is evaluated using different size sensor networks with 80, 160,320,640 and 1280 nodes.

6.1. Experiments

Case1: For an 80-node network with the sink point located at (0,0) (i.e. the down left corner) as shown in Figure 7. This network distribution is suitable when the application environment is inflexible like some military applications.

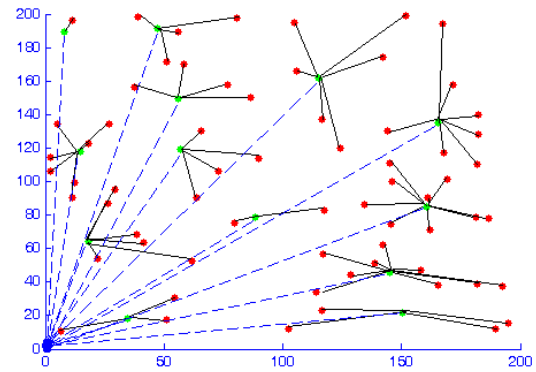


Figure 7: Clustered network when sink point at (0, 0).

The proposed approach was able to find the optimal solutions quickly. For 80-node problem, a good solution can be achieved by 13 clusters as shown in Figure 8

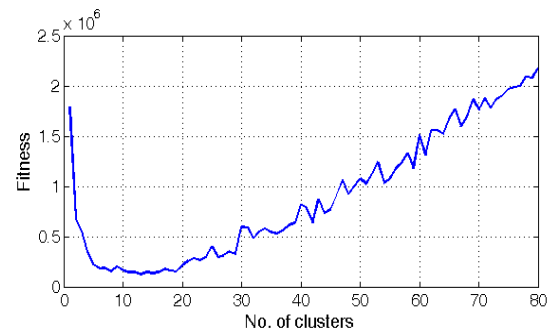


Figure 8: Fitness values over No. of cluster

It is hoped that the fitness is minimised as possible. In the beginning, the number of clusters is few and the fitness is much. With the increase of number of clusters, the fitness decreases. The fitness is optimal when the number of clusters is 12. And then with the increase of the number of clusters, the fitness will increase.

Case2: For an 80-node network with the sink point located at (100,100) as shown in Figure 9. This network distribution is suitable when the sensor nodes are

distributed around a centralized safe area. For example the Mobile networks.

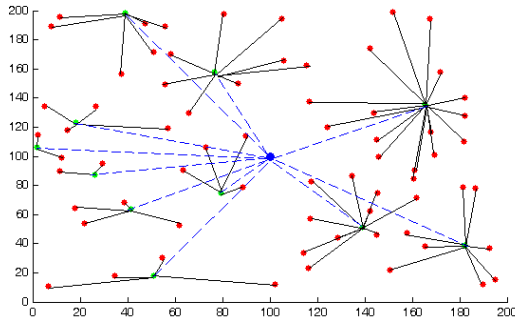


Figure 9: Clustered network when sink at (100, 100).
In this case a good solution can be achieved by 11 clusters and the fitness reduces as shown in Figure 10

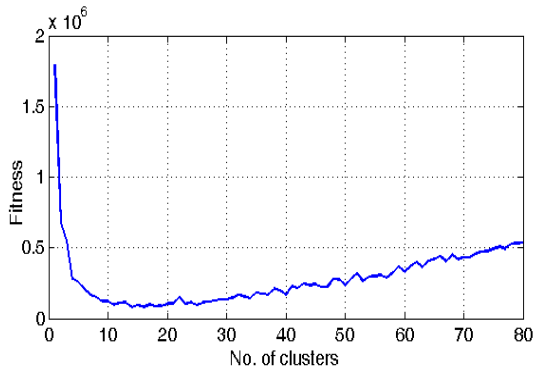


Figure 10: Fitness values over No. of cluster

When the sink is at the centre of network, the fitness will be achieved by a smaller number of clusters compared to the network with the sink node located at (0, 0). Therefore, the distances are reduced.

Table 1 presents some initial test results when the sink-point is assumed to be at point (0,0). As an average value, the distance is reduced by 81%.

Table 1: Test results for different problem size

No. Nodes	Optimal No. clusters	Total distance without clustering	Total distance with clustering	Distance decreased
80	13	5639	1616	71.34%
160	15	23901	5206	78.22%
320	32	54208	9277	82.89%
640	39	95473	14542	84.77%
1280	56	197813	23502	88.12%

Compared to the distance when direct transmission is used, the percentage reduction in the total distance will slightly increase as the number of nodes increases. This is due to the fact that as more nodes will be deployed in the network with denser distribution, this will result in more efficient cluster optimization.

6.2. Comparison between proposed algorithm and genetic algorithm

The comparison between our results as shown in Figure 8 and number of cluster heads over generations as shown in Figure 11 by using GA [10].

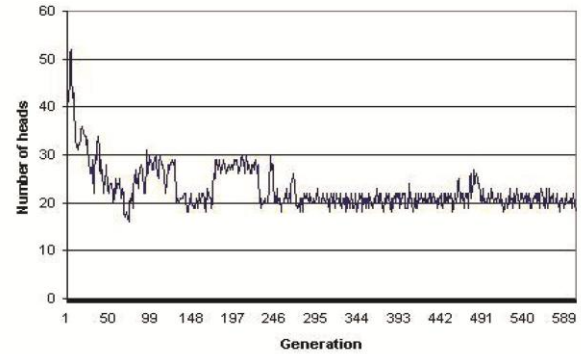


Figure 11: Number of cluster heads over generations

In our approach, its results are more relative than genetic algorithm. GA achieved the optimal fitness value with 20 clusters after about 600 iterations [10]. But our proposed algorithm obtained the optimal No. of head clusters is 12 easily and speedily. Hence the proposed algorithm decreased time and cost.

7. CONCLUSIONS AND FUTURE WORK

In this paper, a merged algorithm between Evolutionary Algorithm (EA) and K-means clustering is proposed to minimize the communication distance in a sensor network. This is achieved by dividing it into clusters by using K-mean algorithm to divide to k -known clusters of n population and Evolutionary Algorithm to obtain the minimum fitness of each cluster. Repeat this process with a different number of clusters and compare with pervious results. Finally, we know the optimal number of clusters and member of each cluster.

Experimental results show that the suggested algorithm is an efficient solution for solving the problem of clustering easily and speedily. It means that the suitable number of clusters determined a suitable fitness.

We plan to apply the algorithm on dynamic Sensor Network by using hybrid algorithm between our proposed algorithm and Particle Swarm Optimization (PSO) or assuming a hierarchical structure where a cluster-head can have a super cluster-head which sends data directly to the sink.

References

- [1] Y. Sankarasubramaniam I. Akyildiz, W. Su and E. Cayirci, "A survey on sensor networks", IEEE Communications Magazine, 102–114, (August 2002).
- [2] V. Mhatre, C. Rosenberg, D. Kofman, R. Mazumdar, and N. Shroff, "A minimum cost heterogeneous sensor network with a lifetime

- constraint," IEEE Transactions on Mobile Computing (TMC) , vol. 4, no. 1, pp. 4 –15, 2005.
- [3] A. A. Abbasi, and M. Younis. "A survey on clustering algorithms for wireless sensor networks", Computer Communications, Volume 30, Issues 14-15, 2007.
- [4] W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-efficient communication protocol for wireless microsensor networks," in Proceedings of the Hawaii International Conference on System Sciences , January 2000.
- [5] T Voigt, A Dunkels, J Alonso, H Ritter, J Schiller, " Computers and Communications, 2004" ,Proceedings. ISCC 2004.
- [6] O. Moussaoui, A. Ksentini, M Naimithe, and M. Gueroui, " A novel clustering algorithm for efficient energy saving in Wireless Sensor Networks", IEEE International Symposium on Computer Networks (ISCN' 06), 66-72, 2006.
- [7] Rahul Khanna, Huaping Liu, and Hsiao-Hwa Chen, "Self-organization of sensor networks using genetic algorithms", IEEE ICC, 3379-3382 2006.
- [8] Noritaka Shigei, Hiromi Miyajima, Hiroki Morishita, and Michiharu Maeda, "Centralized and distributed clustering methods for energy efficient wireless sensor networks" ,International Multi Conference of Engineers and Computer Scientists 2009 Vol I, IMECS, March 18-20-2009, Hong Kong.
- [9] Ehsan Heidari, and Ali Movaghar, "An efficient method based on genetic algorithms to solve sensor network optimization problem", international journal on applications of graph theory in wireless ad hoc networks and sensor networks (GRAPH-HOC) Vol.3, No.1, 18-33, March 2011.
- [10] M. Obaidy, A. Ayesh, and A. Sheta., "Optimizing the communication distance of an ad hoc wireless sensor networks by genetic algorithms", Artificial Intelligence Review, Springer, 29(3):183-194, November 2009.
- [11] S. Ray and R. H. Turi, "Determination of Number of Clusters in K-Means Clustering and Application in Colour Image Segmentation"; Image, In: N. R. Pal, A. K. De and J. Das, Eds., Proceedings of the 4th International Conference on Advances in Pattern Recognition and Digital Techniques, Calcutta, 27-29 December, 1999.
- [12] Asif Khan, Israfil Tamim, Emdad Ahmed, and Muhammad Abdul Awal, " Multiple parameter based clustering (MPC): prospective analysis for effective clustering in wireless sensor network (WSN) using k-means algorithm", doi:10.4236/wsn.2012.41003 Published Online January 2012,(<http://www.SciRP.org/journal/wsn>)
- [13] Thomas Back, Ulrich Hammel, and Hans-Paul Schwefel, "Evolutionary computation: comments on the history and current state", IEEE Transactions on Evolutionary Computation, 1(1):3–17, April 1997. Online available at <http://sci2s.ugr.es/docencia/doctobio/EC-History-IEEEETEC-1-1-1997.pdf> and <http://citeseer.ist.psu.edu/601414.html>.
- [14] Thomas Weise, "global optimization algorithms theory and application", <http://www.it-weise.de/>, 2009.
- [15] Sajid Hussain, Abdul Wasey Matin, and Obidul Islam, "Genetic algorithm for hierarchical wireless sensor networks", journal of networks, vol. 2, no. 5, September 2007.
- [16] A.Wang and A. Chandrakasan, "Energy-efficient dsps for wireless sensor networks", Signal Processing Magazine, IEEE, 19, No.4, 68–78, (2002).
- [17] S. Halgamuge S.M. Guru, A. Hsu and S. Fernando, "An extended growing self-organising map for selection of clustering in sensor networks", International Journal of Distributed Sensor Networks, 1, No.2, (2005).