



The Third Information Systems International Conference

An Energy-Aware Routing Protocol for Wireless Sensor Networks Based on new combination of Genetic Algorithm & k-means

Behrang Barekattain^a, Shahrzad Dehghani^b, Mohammad Pourzaferani^c

^a*Faculty of Computer Engineering, Najafabad Branch, Islamic Azad University, Najafabad, Iran*

^b*Department of Software Engineering, Faculty of Computer Engineering, Higher Education Institute of Allameh Naeini, Naein, Iran*

^c*Department of Software Engineering, Faculty of Computer Engineering, University of Isfahan, Isfahan, Iran*

Abstract

Wireless Sensor Networks (WSNs) consist of large number of sensors which having capabilities such as sensing, computing, and communicating. Beside these features, sensors have limited computational and communication power. Therefore, energy is a challenging issue in WSN networks. Clusters-based routing protocols are used to maximize network lifetime. In this paper, we propose a new combination of K-means and improved GAs to reduce energy consumption and extend network lifetime. The proposed method, reduce energy consumption by finding the optimum number of cluster head (CHs) nodes using improved Genetic Algorithm (GA). To balance energy distribution, a k-means-based algorithm, dynamically cluster the network. The simulations in NS-2 show the proposed algorithm has longer network lifetime than famous algorithms like LEACH, GAEEP and GABEEC protocols.

Keywords: wireless sensor networks, clustering, genetic algorithm, k-means, energy-efficient;

1. INTRODUCTION

Wireless Sensor Networks (WSNs) consist of large number of sensors with limited battery power, works in dangerous environment which power recharge or replacing is not cost effective or in some cases is not possible [1, 2]. Therefore, maintain energy is so important in WSN networks. Routing played a significant part in energy consumption [2, 3]. The most important parameters in the design of routing protocols are energy consumption and lifetime of the network. In cluster-base routing methods, nodes with more energy can be used to process and transmit data while nodes with less energy are used to sense variables environment [4-6].

In cluster-based methods, cluster arrangement and assigning specific tasks to cluster head will be guaranteed the scalability and manageability of these networks. In cluster-based routing algorithms each cluster has a cluster-head (CH) that collects data from all member nodes in its cluster and then aggregates

data and sends them to Base Station (BS) [7, 8]. Therefore, energy consumption will be efficient by aggregating data to reduce the number of messages sent to the base station. Clustering method faces challenges which should be considered such as: Forming clusters and selects the appropriate cluster heads such that have a maximum lifetime in the network [7, 9].

However, some algorithms do not consider residual energy of the sensor nodes and in some algorithms CHs are not distributed appropriately in WSN. This selection technique leads to inappropriate distribution of CHs and the disregard of the optimal number of cluster head causes reducing the network lifetime [8, 9].

So in this paper, we proposed a new hybrid Genetic Algorithm and K-means that is developed to efficiently maximize the network lifetime. This method uses improved Genetic Algorithm (GA) and dynamic clustering the network environment that is done by k-means algorithm. The remainder of this paper is organized as follows. Section II presents some related work about various clustering protocols, Section III describes the proposed method, the simulation results and discussion are given in Section IV. Finally, Section V describes conclusion.

2. RELATED WORK

Clustering is the most effective approach to reduce energy consumption in WSNs and have been used by many researchers recently [3]. Many clustering algorithms have been proposed to optimize energy consumption and LEACH is the first of them which proposed by Heinzelman et al [10]. This algorithm played an important role for development of new clustering algorithms. In this algorithm cluster heads are selected randomly and after the determination of cluster heads, each node which is not CH, choose one of its neighboring CHs and assigned to it based on its minimum distances to the CHs.

This method has some weaknesses such as, non-uniform distribution of cluster heads and choosing cluster-head randomly without considering the remaining energy of nodes. LEACH-C was a new version of leach algorithm that proposed by [11]. In this algorithm formation of clusters is performed in centralized manner and done by the base station. In LEACH-C, base station should guarantee uniform distribution of energy among all the clusters. Finally, this algorithm defines a threshold energy and each node which has more energy than the threshold will be candidate for being clusters head.

Authors in [12] suggested GABBEC method as a hierarchical cluster-based routing protocol. They uses a Genetic Algorithm to determine the number of clusters, the cluster heads, the cluster members, and the transmission schedules to optimize the lifetime of WSN. This method is cluster-based approach such as LEACH and consists of two phases which are set-up phase and steady-state phase:

- Set-up phase:

The Set-up phase execute clustering at first and clustering not changed in network lifetime. So in each round, there are static clusters with dynamically changing CHs.

- Steady-state phase:

In this phase selection of CH is done. The selection of the new CH is based on the residual energy of the current CH and its member nodes.

This method uses a binary representation of the network. CH nodes represented as “1” and non-CH nodes are represented as “0”. The size of a chromosome is equal to the size of the nodes in the network. The fitness function is:

$$F = \sum i (fi \times wi) \forall fi \in (R_{fnd}, R_{lnd}, -C)$$

In this equation C is the cluster distance. R_{FND} is the round number which the first node dies and R_{LND} is the round number which last node dies, finally W is the weight for fitness parameter.

The advantages of this method are:

- Facilitate clustering operations.
- The proposed method has performed well in the number of live nodes

Disadvantages of this method are:

- Protocol does not consider residual energy of nodes to balanced nodes power.
- Protocol does not consider mobility of sensor nodes.
- Protocols is not scalable.

In [13] a genetic algorithm proposed for clustering network. In this method the main objective is finding suitable cluster head nodes and then forming clustering using Genetic Algorithm based on the nearest neighbor distance. In each chromosome, "0" represents the nodes of members and "1" represents a cluster head node.

The fitness function is:

$$F = \frac{\bar{E}}{NE(0)} + \frac{\bar{E}}{\Delta E} + \frac{1}{\sum_i d(S_i, B)}$$

N: Total nodes in the network.

\bar{E} : Residual energy.

ΔE : Distance between the cluster head and the base station.

$\sum_i d(S_i, B)$: Total distance between sink and nodes.

The advantage of this method is:

- This algorithm has improved network lifetime.

Disadvantages of this method are:

- Non-uniformed Cluster-Head Distribution.
- Slow convergence.

A new hybrid method of genetic algorithms and fuzzy logic presented in[14] to balance energy consumption among CHs. In this method calculating the fitness function depends on the energy difference between the current and the previous round of chromosomes and BS selects chromosome that has minimum difference.

The fitness function is:

$$F = |E_{NETWORK}^k - E_{NETWORK}^{k-1}|$$

$E_{NETWORK}^k$ = Energy in the round k (Energy flow in the network)

$E_{NETWORK}^{k-1}$ = Energy in the round k-1 in the network

The algorithm consists of the following steps:

1. Initialize network (specifying the number of sensors).
2. Each node knows its position and sending this information to its neighbors.
3. Done fuzzy definitions (such as energy, density and centrality that are used to calculate the "probability" for fuzzy).
4. Nodes that have a higher probability of fuzzy parameters will be selected as a candidate for cluster head.
5. Now applies genetic algorithms to select the cluster head.
6. Cluster head nodes are introduced to all nodes.
7. Each sensor node joins the nearest cluster head in adjacent.
8. Each sensor node uses TDMA method to transmit data to the cluster head.
9. After all the data are received, the data aggregation is done in each cluster heads and cluster head sends the information received under the package.

The advantage of this method is:

- In this algorithm improved network lifetime which have been clustered using the proposed fuzzy clustering method.

Disadvantages of this method are:

- In fitness function only considered the minimum energy difference compared to the last round.
- The only criterion to calculate the fitness is energy and many influential parameters such as

distance are not considered.

K-means algorithm was presented to clustering networks [15, 16]. K-means is a commonly used partitioning based clustering technique that tries to find a user specified number of clusters (k), which are represented by their centroids, by minimizing the square error function developed for low dimensional data, often do not work well for high dimensional data and the result may not be accurate most of the time due to outliers.

The advantages of using this method are:

- In the large scale of networks with a large number of sensor nodes, this method has very high speed to performed clustering.
- The density of each cluster in the algorithm is very high.

Disadvantages of this method are:

- Different results (data represented in the form of Cartesian co-ordinates and polar co-ordinates will give different results).
- Euclidean distance measures can unequally weight, underlying factors.
- The learning algorithm provides the local optima of the squared error function.
- Randomly choosing of the cluster center cannot lead us to the fruitful result.
- Applicable only when mean is defined, i.e. fails for categorical data.
- Unable to handle noisy data and outliers.
- Algorithm fails for the non-linear data set.

The following Table 1 shows the advantages and disadvantages of the methods have been studied.

Table 1: Advantage and disadvantage of clustering algorithm

Method	Advantages	Disadvantages
GABEEC	<ul style="list-style-type: none"> • Facilitate clustering operations. • The proposed method has been performing well in the number of live nodes 	<ul style="list-style-type: none"> • Don't consider remaining energy in fitness functions • Slow Convergence in large scale network • Non-uniform CH • Do not stability • Do not mobility • Constant number of clusters • Constant Cluster members
ESNAGA	<ul style="list-style-type: none"> • Clustering with genetic algorithm has improved lifetime of the network. 	<ul style="list-style-type: none"> • Constant and of chromosomes and equal than the size of nodes that caused slow convergency • Non-uniform CH • Do not stability • Slow convergence in large scale network
GA&Fuzzy	<ul style="list-style-type: none"> • The proposed method uses a clustering algorithm based on genetic algorithms and fuzzy algorithm has improved lifetime of the network. • Improved density in cluster with fuzzy logic 	<ul style="list-style-type: none"> • In fitness function only considered the minimum energy difference compared to the last round. • The only criterion to calculate the fitness is energy and many influential parameters such as distance is not considered • Constant and of chromosomes and equal than the size of nodes that caused slow convergency
K-means	<ul style="list-style-type: none"> • The proposed method has been performing well in the number of live nodes • High speeds in clustering • High density in the cluster 	<ul style="list-style-type: none"> • Constant number of clusters • Unable to handle noisy • Unable to recognize nonspherical shape • Possible Stay in Local optimization • Different results dependent initial random points • Non-uniform CH

3. PROPOSED APPROACH

This paper presents a new hybrid method which is based on genetic algorithm and k-means for providing a cluster-based routing algorithm to reduce energy consumption in wireless sensor networks.

In some related researches Genetic Algorithm is used to calculate the optimum number of CH and minimize the long distance communications between the sensors and sink which leads to reducing the energy consumption in WSN. However, some genetic algorithms do not consider residual energy of the sensor nodes. Also, some algorithms consider constant number of CH and don't find optimum number of them. In this research we tried to calculate the optimum number of CH and clustering them with efficiency.

✓ GA methodology

1) Problem Representation

In proposed approach, after placement nodes in the environment, sends their remaining energy to base station. Afterward, BS segments the WSN coverage into nonequivalent cells. Cells near the sink are set smaller than farther cells. This segmentation leads to appropriate distribution of CHs. Since, difference in cell size makes more CHs near sink which play traffic relays role in network and decrease hotspot phenomenon. In Figure 1 shown this segmentation.

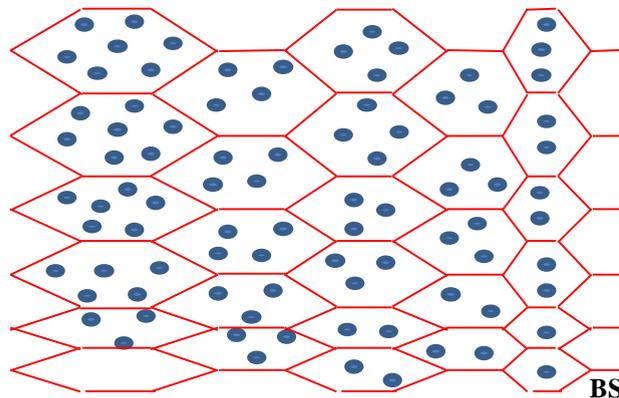


Figure 1: nonequivalent cells segmentation

Afterwards, BS calculate residual energy of nodes for each cell and any node which has residual energy more than the average of its cell could be candidate for being CH. As the next step, BS compose a chromosome based on number of candidates (to prevent all nodes being candidate, especially at first stage when network starts, we set a maximum threshold).

The chromosome is represented by a list of bits also called (genes) and each bit corresponds to one sensor node in set N. If the gene value is 1, then the corresponding node is a cluster head and 0 means that it's a regular node.

The proposed approach, shortened chromosome-length which accelerate convergence to optimum solution.

2) GA Operations

The initial population is generated randomly and the algorithm consist of three steps:

- Selection which equates to the survival of the fittest.

- Crossover which represents mating between individuals.
- Mutation which introduces random modifications.

Selection Operator

- The selection process gives preference to better individuals, allowing them to pass on their genes to the next generation.
- The goodness of each individual depends on its fitness.
- Fitness may be determined by an objective function or by a subjective judgment.

In our proposed algorithm Roulette wheel selection is used. In this method the CHs chromosomes with higher fitness values are more likely to be selected as the chromosomes of the population in the next generation.

Crossover Operator

The crossover is a binary genetic operator that used to create new solutions from the existing solutions available in the mating pool after applying a selection operator. Crossover selects any two parent CHs chromosomes are selected randomly to exchange information after that point to produce two children. Therefore, two offspring are generated as shown in Figure 2.



Figure 2: Two-point crossover

Mutation Operator

The mutation operator is applied to each gene of CHs chromosome with a probability of mutation m . Mutation is the occasional introduction of new features into the solution strength of the population pool to maintain diversity in the population. Mutation operator changes a 1 to 0 or vice versa, with a mutation probability of m . The mutation probability is generally kept low for steady convergence. But we tried to increase the speed of convergence in achieve optimum solution. In that case, if nodes which are selected for the mutation acts, are in the radio range of each other mutations be done, otherwise do not mutate. This method would cause only the neighboring nodes are join together.

3) Fitness function

A fitness function value quantifies the optimality of a solution. The value is used to rank a particular solution against all the other solutions. It is calculated on distance parameter. The fitness function establishes the basis for selecting chromosomes that will be mutated. The fitness function used as follows:

$F(x) = W (\text{Total energy consumption} / \text{Total energy residual}) + W - 1 (\text{Number of CH} / \text{Number of nodes alive})$

4) Stopping Criterion

The stopping criterion achieves when the objective function does not change for a certain number of generations or when the number of generations exceeds the specified maximum generations.

✓ K-Means methodology

Using genetic algorithms have many problems such as it is very slowly converge for large data. Also, it could be stuck in local optimum solution, not global optimum. Therefore, we decided to use it with K-means algorithm. In this method at first we used genetic algorithm with limited number of repetitions and achieve to near optimum solution, then sent the points that achieved by genetic algorithm to K-means algorithm as an initial points and performed node clustering. Then each datum, based on similarity, is attributed to one of the clusters. In this algorithm, the following relation is used to determine the distance between points and the center of each cluster:

$$j = \sum_{j=1}^k \sum_{i=1}^n \|x_i^j - c_j\|^2$$

In which C_j is the center of j -cluster.

If we use only K-means for clustering cannot achieve the best answer, because of sensitivity of the algorithm to the data center, and inability to detect noise and inability to detect non-spherical clusters and finally maybe get stuck in local optimum.

4. SIMULATION RESULTS

Our purposed algorithm developed using NS2 simulator [17] on Fedora10. The machine computer used for this simulation is equipped with the core i7 processor, 8 GB RAM. The cluster formation is performed on wireless sensor network consisted of 50,100,150, 200,250 nodes.

In, Figure 3 and Figure 4 shown simulation result of purposed method and compare with other clustering algorithm.

We consider the following parameters:

- The Area of region is $100 \times 100 \text{ m}^2$,
- Crossover probability (P_c) is 0.9,
- Mutation probability(P_m) is 0.1,
- W is 0.7,
- Selection operator: Roulette-Wheel,
- Crossover type: One point.

The radio model used in our work is the same in [11] that are free space radio model which stated in the equations described below:

given $E_{elec} = 50 \text{ nJ/bit}$ as the energy being dissipated to run the transmitter or receiver circuitry and $\epsilon_{amp} = 10 \text{ pJ/(bit} \times \text{m}^2)$ as the energy dissipation of the transmit amplifier. K is a bit of data, d is the distance between N_1, N_2 Transmission (E_{Tx}) and receiving costs (E_{Rx}) are calculated by following equation:

$$\begin{cases} E_{elec} * k + E_{fs} * k * d_{N1,N2}^2 & \text{if } d \leq d_0 \\ E_{elec} * k + E_{mp} * k * d_{N1,N2}^4 & \text{if } d > d_0 \end{cases}$$

$$E_{Rx(k)N1,N2} = E_{elec} * k$$

$$D_0 = \sqrt{\frac{E_{fs}}{E_{mp}}}$$

In **Error! Reference source not found.** shows network Lifetime per seconds with a various number of nodes. It is clear that the proposed protocol (GA-Kmeans) improves the network lifetime compared to the LEACH, GABEEC and GAEEP. Also in Figure 4 shows the enhancement of the proposed method throughput compared to the LEACH, GABEEC and GAEEP.

Throughput is defined as the time taken for a packet to be transmitted across a network from source to destination. The throughput script of the network in Ns2 can be estimated as follows:

$$\text{RecvdSize} / (\text{stopTime} - \text{startTime}) * (8/1000)$$

That in this formula recvdSize computed as follows:

$$\text{hdr_size} = \text{pkt_size} \% 512$$

$$\text{pkt_size} -= \text{hdr_size}$$

$$\text{RecvdSize} += \text{pkt_size}$$

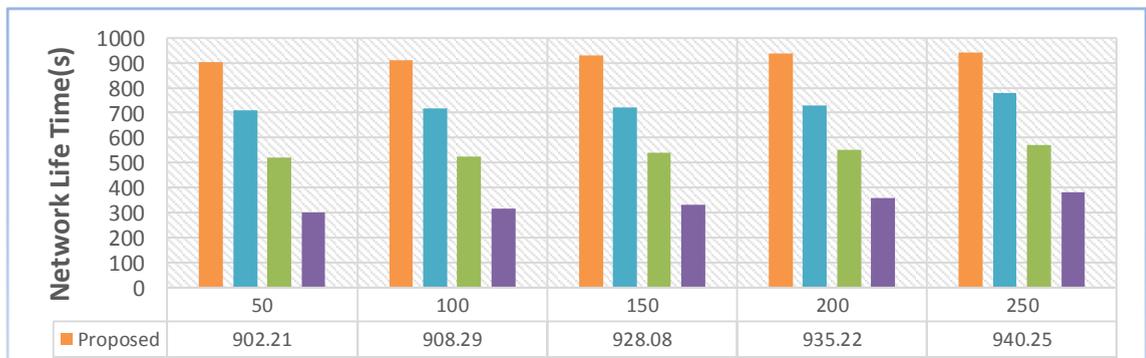


Figure 3: Compare network Lifetime Obtained using GA-Kmeans, GAEEP, GABEEC, LEACH Protocols

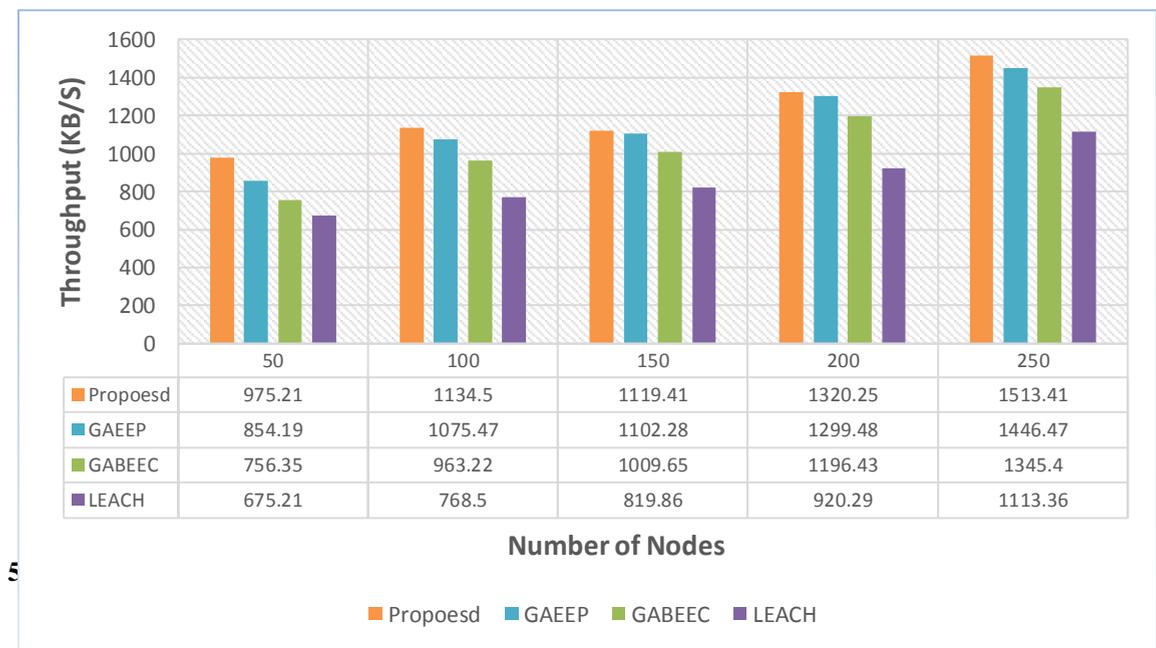


Figure 4: Comparison Throughput Obtained using GA-K-means, GAEEP, GABEEC, LEACH Protocols

clustering using genetic algorithm is an applicable and scalable approach where it can be implemented in the large number of nodes with different base station locations and nodes deployment styles. NS2 simulation results showed that the proposed protocol is more energy efficient and more reliable in clustering process as compared to LEACH, GAEEP and GABEEC.

References

- [1] M. Sabet and H. R. Naji, "A decentralized energy efficient hierarchical cluster-based routing algorithm for wireless sensor networks," *AEU - International Journal of Electronics and Communications*, 2015.
- [2] M. Hammoudeh and R. Newman, "Adaptive routing in wireless sensor networks: QoS optimisation for enhanced application performance," *Information Fusion*, vol. 22, pp. 3-15, 2015.
- [3] V. Kochher and R. K. Tyagi, "A Review of Enhanced Cluster Based Routing Protocol for Mobile Nodes in Wireless Sensor Network," *Advance in Electronic and Electric Engineering*, vol. 4, pp. 629-636 2014.
- [4] J. Changjiang, X. Min, and S. Weiren, "Overview of cluster-based routing protocols in wireless sensor networks," in *Electric Information and Control Engineering (ICEICE), 2011 International Conference on*, 2011, pp. 3414-3417.
- [5] S. P. Barfunga, P. Rai, and H. K. D. Sarma, "Energy efficient cluster based routing protocol for Wireless Sensor Networks," in *Computer and Communication Engineering (ICCCCE), 2012 International Conference on*, 2012, pp. 603-607.
- [6] N. Pantazis, S. Nikolidakis, and D. Vergados, "Energy-efficient routing protocols in wireless sensor networks," *IEEE Commun. Surv. Tutorials*, pp. 551-591, 2013.
- [7] Kaicheng. Yin and Chaosheng. Zhong, "Data collection in wireless sensor networks," *Cloud Computing and Intelligence Systems (CCIS), 2011 IEEE International Conference on*, pp. 98-102, 15-17 Sept. 2011 2011.
- [8] L. Jianhua, Z. Baili, and X. Li, "A data correlation-based wireless sensor network clustering algorithm," in *Computer Application and System Modeling (ICCSM), 2010 International Conference on*, 2010, pp. V8-61-V8-65.
- [9] L. Weifa, L. Jun, and X. Xu, "Prolonging Network Lifetime via a Controlled Mobile Sink in Wireless Sensor Networks," in *Global Telecommunications Conference (GLOBECOM 2010), 2010 IEEE*, 2010, pp. 1-6.
- [10] W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-Efficient Communication Protocol for Wireless Microsensor Networks," in *Proc. of 33rd Hawaii International Conference on System Sciences*, pp. 1–10, 2000.
- [11] W. B. Heinzelman, a. P. Chandrakasan, and H. Balakrishnan, "An application-specific protocol architecture for wireless microsensor networks," *IEEE Transactions on Wireless Communications*, vol. 1, pp. 660–670, 2002.
- [12] S. Bayraklı and S. Z. Erdogan, "Genetic Algorithm Based Energy Efficient Clusters (GABEEC) in Wireless Sensor Networks," *Procedia Computer Science*, vol. 10, pp. 247-254, // 2012.
- [13] M. Elhoseny, Y. Xiaohui, H. K. El-Minir, and A. M. Riad, "Extending self-organizing network availability using genetic algorithm," in *Computing, Communication and Networking Technologies (ICCCNT), 2014 International Conference on*, 2014, pp. 1-6.
- [14] E. Saeedian, M. N. Torshiz, M. Jalali, G. Tadayon, and M. M. Tajari, "CFGA: Clustering Wireless Sensor Network Using Fuzzy Logic and Genetic Algorithm," in *Wireless Communications, Networking and Mobile Computing (WiCOM), 2011 7th International Conference on*, 2011, pp. 1-4.
- [15] P.Sasikumar and S. Khara, "k-MEANS CLUSTERING IN WIRELESS SENSOR NETWORKS," *Fourth International Conference on Computational Intelligence and Communication Networks*, 2012.
- [16] P. Wei and D. J. Edwards, "K-Means Like Minimum Mean Distance Algorithm for wireless sensor networks," in *Computer Engineering and Technology (ICCT), 2010 2nd International Conference on*, 2010, pp. V1-120-V1-124.
- [17] NS2 Simulator, <http://www.NS2.org/>.