

# A Novel Data Gathering Algorithm For Wireless Sensor Networks Using Artificial Intelligence

Swapna Ch (✉ [swapnaeng@gmail.com](mailto:swapnaeng@gmail.com))

Sri Venkateshwara College of Engineering <https://orcid.org/0000-0001-5991-4526>

Vijayashree R Budyal

Sri Venkateshwara College of Engineering

---

## Research Article

**Keywords:** Wireless Sensor Network, Fuzzy logic, Clustering, Cluster Leader, Virtual Points, Mobile Element

**Posted Date:** June 2nd, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-463490/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# A NOVEL DATA GATHERING ALGORITHM FOR WIRELESS SENSOR NETWORKS USING ARTIFICIAL INTELLIGENCE

Swapna Ch<sup>1</sup> and Vijayashree R Budyal<sup>2</sup>

<sup>1,2</sup>Sri Venkateshwara College of Engineering, Bengaluru  
<sup>1,2</sup>Email: [swapnaeng@gmail.com](mailto:swapnaeng@gmail.com), [vijayashreeerb@gmail.com](mailto:vijayashreeerb@gmail.com)

**Abstract**—The most challenging task in wireless sensor network is energy efficiency, as energy is the major constraint in the wireless sensor network to improve the life time of the network. Hence developing algorithms to improve network life time is the major task. In wireless sensor network most of the energy is wasted while gathering the data, hence an efficient algorithm which conserves energy has to be designed. Thus our proposed work A Novel Data Gathering Algorithm for Wireless Sensor Networks using Artificial Intelligence (NDGAI) uses mobile element and deals with the conservation of energy while gathering the data. Appropriate clustering, cluster leader selection and proper path determination of mobile element helps to conserve energy and improve the over all network life time. In our proposed work initially the clusters are forged by using Amended Expectation Maximization(AEM) algorithm, which is the maximum likelihood estimate. It is used along with Gap statistic method to find the optimal number of clusters. AEM algorithm helps in obtaining the centres of the cluster with maximum number of nodes near the cluster centres. For each cluster, Cluster Leader (CL) is selected by using Fuzzy Logic. Fuzzy logic selects the node which is near to the cluster centre by using parameters such as Closeness of node to the Cluster Centroid, direction of node towards base station, number of Neighbouring Nodes. After the CL's are determined, to reduce the path length virtual points(VP) are selected so that mobile element reaches this virtual point and collects the data. These VP's are selected only when the CL has data in it. The mobile elements can reach these virtual points intelligently by using optimal path, that is obtained by using hybrid of Particle Swarm Optimization and Artificial Bee Colony algorithm. Thus the mobile element travels in the optimal path and gathers the data from the entire network intelligently and efficiently with less amount of energy. With this approach the performance and life time of the network is improved while gathering the data. The simulation results are compared with Scalable Grid-Based Data Gathering Algorithm for Environmental Monitoring Wireless Sensor Networks (SGBDN) and proved that the proposed method is better than SGBDN .

**Index Terms**—Wireless Sensor Network, Fuzzy logic, Clustering, Cluster Leader, Virtual Points, Mobile Element

## I. INTRODUCTION

Large number of sensors are deployed in the field, termed as Wireless Sensor Network (WSN) [1]. It consists of many sensors that helps in sensing the environment and communicates the same with base station and from base station to the user and helps the user to monitor and control the environment remotely. Thus the WSN has wide range of applications in military surveillance, agriculture, medical domain [2]- [5] etc and helps to monitor information such as temperature,

pressure, humidity, etc. While gathering the information from the sensor node lot of power is consumed to transmit the information from each sensor to the base station.

Designing Energy Efficient algorithms [6] is a major challenge in the WSN. Communication in the network has to happen in an efficient way as to conserve the energy and improve the network life. Because the sensors which are deployed in the field cannot be recharged once the battery is drained due to the absurd conditions of the network. Nodes can communicate directly to the base station but at the cost of the more power. Thus the entire network is partitioned and clusters are formed and then cluster heads. The LEACH [7] is one of the hierarchical clustering method where the network is partitioned or clustered and based on the probability one of the node is selected as cluster head (CH). The CH may not have sufficient energy and hence the battery power is not sufficient for effective communication and the connectivity in the network may be lost, thus leading to the decrease in the network life time. To increase the network life time Mobile Elements[ME]are introduced where these ME's [8] [9] travel to the CH and collects the data from the network. To improve the performance and reduce the energy consumptions various artificial intelligence algorithms, which is a sub division of Machine Learning [10] [11] [12] can be imbibed.

In our proposed work, A Novel Data Gathering Algorithm for Wireless Sensor Networks using Artificial Intelligence (NDGAI)the base station is considered to be at the center as shown in Fig. 1.

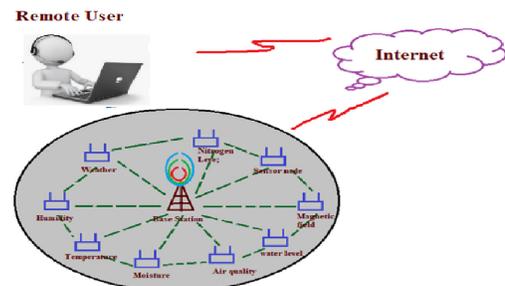


Fig. 1. WSN scenario

The network consists of hundred of nodes(also known as motes) and the entire network is sub divided into optimal number of clusters by using Gap statistics algorithm along with Amended Expected Maximization (AEM). The CL leader is selected based on the Fuzzy logic near the centroid of the each cluster. The ME has to reach all the clusters and gather the information. The stop points are the virtual points that are considered with in the communication range of the CL and the ME's reaches these virtual points and gather the data. The path of the ME is determined by Particle Swarm Optimization(PSO) algorithm along with Artificial Bee Colony (ABC) named as PSO-ABC algorithm is used, which solves the Travelling Salesman Problem. The residue of paper is categorized as Literature review in section II,section III presents the system model and assumptions, section IV consists of proposed work, section V consists of simulation results and finally section VI deals with the conclusions.

## II. LITERATURE REVIEW

Researchers have proposed many energy efficient data collection algorithms to prolong battery and network life by using ME's and these ME's uses multi hop communication.

DW. B. Heinzelman et. al., [13] proposed an hierarchical routing algorithm known as LEACH. Here the network is divided into clusters and CH's are selected based on the probability function, that selects the node with high energy as CH. However the performance of the network degrades as the CH sends the data directly to the base station. Amer O. Abu Salem et. al., [14] has proposed an enhanced version of LEACH protocol. The CH is selected based on the distance function in such a way that the distance is low from base station and this helps to reduce the energy consumption by increasing the network life time. In [15], an algorithm is designed in which CH's are selected by considering the node residual energy and degree of the node to its neighbours and achieves unvarying CH distribution in the entire network. However more energy is utilized when data is transmitted directly to the base station.

S. Lindsey et. al., [16], has proposed an algorithm, PEGASIS, in which chain among the sensor nodes are formed and the node transit its information to a nearest neighbour nodes and then finally to the base station. However the nodes near the base station dies faster and the network lifetime is reduced. Cluster head selection is also considered in [17] by which it achieves minimum latency by terminating the algorithm for  $O(\log\log N)$  when  $N$  nodes are given and also provides the path loss exponent of 2. In [18], hierarchical clustering method is employed. Each cluster consists of a CH, that collects the data from its members and aggregates the obtained data and transfers it to the base station by the nearest CH. The CH's in a cluster takes turns based on the time period. However as the time period is considered the residual energy might be less than the threshold energy thus breaking the chain of communication.

As the utilization of wireless sensor networks is increasing drastically there is lot of research happening and the authors

have designed algorithms based on mobile node [19]. That is a mobile node has been used so that the mobile node travels to the cluster head or the cluster and transmits data to the base station. S. Cho et. al., [20], has proposed an algorithm in which mobile sinks are used for collecting the data. In the designed P-LEACH algorithm, the clusters are formed in the shape of circle and the cluster is divided into four areas, all the four areas are independent, and could efficiently track mobile sink and save energy. However duplication of data is more in the proposed algorithm. In [21], the authors have used integer-linear-programming problem and proposed maximum amount shortest path problem and reduced energy consumption, by finding optimal mapping between its members and sub-sink. The authors have used genetic algorithm to solve the maximum amount shortest path problem and also adopted the problem related to dynamic topology changing. However mobile sink trajectory is not taken into consideration. [22] has employed a data gathering method based on fuzzy rule based CH selection by considering residual energy of nodes, degree of centrality, sink distance and number of neighbour nodes. This ensures uniform cluster head distribution. Along with CH election, intra cluster data aggregation using correlation function in fuzzy theory is adopted. However optimal path is not considered. V.Saranya et al., [23] has proposed a protocol in which, they have improved the throughput, network life time and waiting time. In the proposed work the authors have adopted clustering approach in which members of the cluster relays data to cluster head by N threshold method. The mobile sink travels to the cluster head and collects data by using dynamic polling mechanism.

An algorithm proposed in [24], have adopted fuzzy logic for CH selection along with particle swarm optimization to increase the range of membership function. The mobile element transit data to these cluster heads and gathers data in a single hop. The predetermined and optimal path is determined by Ant Colony Optimization algorithms. However during clustering optimal number of clusters are not formed.

In [25] has proposed LEACH-EA protocol which is based on LEACH algorithm. To increase the CH probability, along with residual energy they have added node distance. To find the optimal transit path ant colony optimization algorithm has been used. However the nodes near the base station dies fast. In [26], [27], [28] the authors have used Rendezvous points so that the mobile element reaches to these locations and collects the data from the corresponding nodes. These Rendezvous points are considered mainly in the delay tolerant applications. In [29], the authors have proposed an algorithm where Rendezvous points are selected for the mobile sinks. In the proposed work the authors have used multiple mobile sinks. A modified multihop layer model is used to find optimal clusters. However as multiple sinks are used the cost of the network is increased.

In the aforementioned work the authors have addressed problems related to clustering, CH selection, rendezvous point selection and data collection using mobile elements. However they didn't address the problem of finding optimal number of

clusters, appropriate CL selection and optimal path for the mobile element. Hence there is a requirement to address all the aforementioned problems to improve the network life time by consuming less energy.

### III. SYSTEM MODEL AND ASSUMPTIONS

#### A. System Model

The main motivation for our proposed work A Novel Data Gathering Algorithm for Wireless Sensor Networks using Artificial Intelligence (NDGAI) is to find optimal number of cluster, CL selection, VP selection and Path selection for mobile element using Artificial Intelligence to the existing algorithms. In our proposed work, S number of sensor nodes are deployed in the  $m^2$  network area. The location of the sensor nodes is known by using the localization technique. The base station will be aware of the location of all the sensor nodes. The starting knowledge about the network will be obtained by the *Hello* message packets by the base station. The reply message from the node consists of location information, node identity and their initial energy level. The network is based on the hierarchical network, i.e., it consists of Clusters with k cluster centroids denoted as  $k_1, k_2, k_3, \dots, k_n$ . Each cluster consists of Cluster Leaders, VP points between CL and base station, Mobile element to gather the information from Cluster Leaders and a center base station. Fig. 2 is the system model of our proposed algorithm

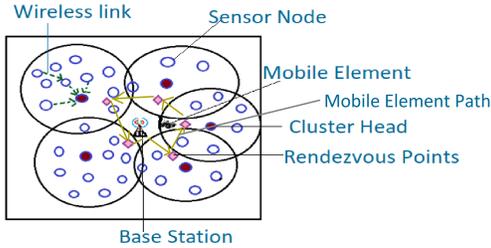


Fig. 2. System Model

#### B. Assumptions

The following are some of the assumptions considered. They are

- All the nodes are deployed according to Gaussian Distribution Function
- The communication range of all the sensor nodes including mobile element have the communication range of R and is able to communicate properly.
- The time taken to obtain the data from each CL is sufficient to gather the complete data.
- Base station is located at the center of the wireless sensor network area.
- Each cluster have a Cluster Leader.
- Mobile element starts from the base station and knows the location of sensor nodes.
- The space between the sensor nodes are considered to be free space.

### IV. PROPOSED WORK

The proposed work NDGAI deals with the data gathering from the sensor nodes in the WSN. Whilst sending and receiving the data packets, lot of energy is consumed and in general the energy consumed for sending the data is given by eq.1 and the amount of energy consumed for receiving the data packet is given by eq.2.

$$e_{tr} = \begin{cases} lE_{ele} + l\epsilon_{fs}d^2 & d < d_0 \\ lE_{ele} + l\epsilon_{ampli}d^4 & d \geq d_0 \end{cases} \quad (1)$$

$$e_{rt} = lE_{ele} \quad (2)$$

where energy required for transmission is given by  $e_{tr}$ ; energy required for receiving is denoted as  $e_{rt}$  transmitted bits is denoted by l; each bit transit energy is  $E_{ele}$ ; energy cost for transmission and reception of signal is given by  $\epsilon_{fs}d^2$  and  $\epsilon_{ampli}d^4$ ; transmission distance is denoted as  $d$ ,  $d_0$  is the error free receiving distance. Thus from eq.1 it is clear that the energy required for transmitting the data is directly proportion to the square of the communication distance. Hence the reduction of communication distance is utmost important parameter to conserve energy while gathering the information. In our proposed work data gathering is performed efficiently by following four different phases. The four different phases are

- 1) Clustering Phase
- 2) CL and VP election phase
- 3) Path Determination Phase

The amount of energy consumed for transmitting data packet given by  $e_{tr}$  and receiving data packet given by  $e_{rt}$  are calculated using eq.1 and eq.2 respectively. By applying appropriate Artificial algorithms, the required energy can be further reduced, prolonging the network life time by utilizing the energy efficiently during data gathering. The detailed analysis of all the phases are conveyed in the following sections.

#### A. Clustering Phase

Clustering the sensor network means dividing the entire network into sub networks. The benefits of clustering include tactical resource usage, scalability of network, elevated performance and to avoid longer distance multi hop communication. Because of these advantages clustering is considered to be one of the major step in data gathering in WSN. Hence various methods are existing to form clusters in the network. However in our present work, Amended Expected Maximization (AEM) algorithm is adopted, to reduce sum of squares of intra cluster communication distance. The AEM algorithm forms the clusters, with cluster center surrounded by 66 percent of the nodes. Along with the clustering it is also important to obtain optimal number of clusters so that the amount of energy required for communication can be reduced. if the number of clusters formed are more, then latency is increased and if less number of clusters are formed, then more data flooding happens. Hence to avoid these problems, an optimal number of clusters are to be

determined. To determine optimal number of clusters, in our proposed work along with AEM, Gap Statistics method is considered. Therefore it is essential to apply Gap statistics method to the nodes of the WSN, to find the optimal number of clusters.

1) *Gap Statistics Method to determine optimal number of clusters:* In gap statistics method the maximum number of cluster  $k_{max}$  has to be defined initially. Let the number of nodes to be  $1, 2, 3, \dots, N$  and the number of clusters be  $k_1, k_2, \dots, k_{max}$ . The optimal number of clusters can be evaluated by using the following steps.

Step.1: Calculate squared Euclidean distance  $D_{pq}^2$  between  $p^{th}$  node with the  $q^{th}$  node, using eq.3.

$$D_{pq}^2 = \sum_{i=1}^M (p_i - q_i)^2 \quad (3)$$

Let  $k_r$  denote the indices of observations in cluster  $r$  and  $N_r = |K_r|$ .

Step.2: Calculate the summation between the pairs of distances of all the nodes that belong to the cluster  $r$ , using eq.4

$$D_r = \sum_{p, q \in K_r} D_{pq} \quad (4)$$

Step.3: Measure dispersion within intra-cluster, denoted by  $W_k$ , which gives the pooled summation of squares within cluster means and is given by eq.5.

$$W_k = \sum_{r=1}^k \frac{1}{2N_r} D_r \quad (5)$$

Step.4: Generate  $D$  reference points using uniform random distribution. Cluster each of these data points by considering  $k_1, k_2, \dots, k_{max}$ . By considering all these data points calculate and the different sizes of clusters, enumerate dispersion within intra-cluster given by  $W_{kD}$

Step.5: Compute the deviation between  $W_k$  and  $W_{kD}$  using eq.6

$$Gap_n(k) = \frac{1}{D} \sum_{d=1}^D \log(W_{kd}) - \log(W_k) \quad (6)$$

and Standard deviation can be calculated using eq.7

$$S_k = \left[ \frac{1}{D} \sum_d \log(W_{kd}) - \log(W_k) \right] \left[ \sqrt{1 + \frac{1}{D}} \right] \quad (7)$$

Step.6: Finally the optimal number of cluster value is calculated in such a way that the smallest value of  $k$  provided the condition stated in eq.8 is satisfied.

$$Gap_n(k) \geq Gap_n(k+1) - S_{K+1} \quad (8)$$

Thus using Gap statistics, the optimal number of clusters are determine. When the optimal number of clusters are identified, Clusters are formed by using Amended Expectation Maximization algorithm which is a statistical method to determine Maximum likelihood function. This method is mainly

applied to the latent variable, where complete data is not observed or known. Based on the observed data and sufficient statistical values, it is possible to find the unobserved data using likelihood function, that can be maximized by taking the first derivative and equating it to zero, thus the name Maximum Likelihood Estimate(MLE) function. Thus using MLE function and EM algorithm, clusters can be formed.

2) *AEM algorithm:* EM algorithm is a statistical approach and belongs to unsupervised learning method and is used in the formation of clusters in the WSN. EM algorithm is one of the soft clustering algorithms, which means that the clusters may overlap, i.e., the node may belong to more than one cluster and the amount of relationship of the node with each cluster is identified by calculating the degree of responsibility of the node to the cluster. As the optimal number of clusters are determined, randomly choose  $k$  cluster centroids denoted as  $k_1, k_2, \dots, k_{max}$ . Initially the clusters are formed by using Euclidean Distance function. For example if  $p$  is the random cluster centroid as denoted by  $(x_p, y_p)$ , then the distance between the point  $p$  with each node  $(x_n, y_n)$ , where,  $n = 1, 2, \dots, N$  in the network is calculated and based on the distance the node  $x$  belongs to cluster with centroid  $p$  is determined, if the distance between the node and point  $p$  is less when compared with other cluster centroid. Hence to calculate the Euclidean distance the generalized formula adopted is given in eq.9.

$$D_{pn} = \sqrt{(x_p - x_n)^2 + (y_p - y_n)^2} \quad (9)$$

The node in a WSN belongs to a particular cluster if its distance to the centroid points is less when compared with the distance from other centroids. After initial clusters are formed the mean value for each cluster denoted by  $\mu_k$  is calculated by considering the average of all the nodes that belong to a particular cluster. Covariance,  $\xi_k$  is calculated using eq.10

$$\xi_k = \frac{1}{N} \sum_{n=1}^N (x_n - \mu_x)(y_n - \mu_y) \quad (10)$$

Thus the initial cluster parameters  $\mu_k$  and  $\xi_k$  are obtained and these are the position vectors of cluster centroids. For further improvement in the clustering and decreasing the communication distance the following steps are considered.

- Initialization Step
- Expectation Step
- Maximization Step
- Convergence of log likelihood function

During Initialization step, nodes in WSN are deployed according to the Gaussian distribution function with Probability function, given by eq.11

$$P(x_n) = \sum_k \pi_k N \left( \frac{x_n}{\mu_k}, \xi_k \right) \quad (11)$$

where  $x_n$  is the  $n_{th}$  node position vector,  $k$  are the number of clusters,  $\mu_k$  is the mixing coefficient and the value of  $\pi_k$  depends on the number of clusters and the condition on  $\pi_k$  is

given by  $\sum_{k=1}^K \pi_k = 1$  and  $0 \leq \pi_k \leq 1, \quad \forall k$ .

$\mu_k$  is the mean value, which is calculated by using Euclidean distance function and  $\xi_k$  indicates covariance matrix, calculated based on  $\mu_k$  and it is a  $2 \times 2$  diagonal matrix.  $N\left(\frac{x_n}{\mu_k}, \xi_k\right)$  is the Gaussian Distribution function given by eq.12.  $N\left(\frac{x_n}{\mu_k}, \xi_k\right) =$

$$\frac{1}{(2\pi)^{|\xi_k|} |\xi_k|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(x_n - \mu_k)^T \xi_k^{-1} (x_n - \mu_k)\right] \quad (12)$$

During Expectation step, the degree of responsibility, of each node to the cluster has to be calculated. Degree of responsibility, denoted as  $\Gamma_{kn}$  can be obtained using eq.13.

$$\Gamma_{kn} = \frac{\pi_k N\left(\frac{x_n}{\mu_k}, \xi_k\right)}{\sum_{j=1}^K \pi_j N\left(\frac{x_n}{\mu_j}, \xi_j\right)} \quad (13)$$

As a next step, Maximization step has to be executed. In this step using the degree of responsibility function, which is obtained in the Expectation Step has to be used to find the new parameters of the cluster centroids. This can be updated with the new values as shown from eq.14- eq.17. The new mean value is calculated as in eq.14

$$\mu_k^{new} = \frac{1}{N_k} \sum_{n=1}^N \Gamma_{kn} x_n \quad (14)$$

The total number of nodes that belong to cluster K is given by eq.15

$$N_k = \sum_{n=1}^N \Gamma_{kn} x_n \quad (15)$$

The mixing coefficient  $\pi_k$  can be updated using eq.16

$$\pi_k^{new} = \frac{N_k}{N} \quad (16)$$

Covariance can be updated by using eq.17

$$\xi_k^{new} = \frac{1}{N_k} \sum_{n=1}^N \Gamma_{kn} (x_n - \mu_k^{new})(x_n - \mu_k^{new})^T \quad (17)$$

. After updating the parameters of the centroid of the cluster, as a last step, the EM algorithm is evaluated by evaluating convergence of log likelihood function as given in eq.18

$$P = \ln p\left(\frac{x_n}{\mu_k}, \xi_k, \pi_k\right) = \sum_{n=1}^N \ln\left(\sum_{k=1}^K \pi_k N\left(\frac{x_n}{\mu_k}, \xi_k\right)\right) \quad (18)$$

Thereafter Expectation step and the Maximization step the log likelihood function given in eq.18 is evaluated and based on these values, EM algorithm is continuously executed until the log likelihood function converges to a small value, i.e., there is no or very small difference in the past calculated log likelihood function value and the present calculated log likelihood function value. When the log likelihood function converges the algorithm terminates and the final clusters are

determined. The main advantage of using EM algorithm with amendment is that it reduces the over all communication distance with in the cluster. This reduction in communication distance is as it provides the final cluster centroids where the density of the nodes are more.

### B. Cluster Leader and Virtual Point Selection

1) *CL election phase:* Following the clusters formation, the CL has to be determined near the centroid, so that all the nodes within the cluster can send their gathered information to the CL. Initially all the nodes near the centroid of the cluster, whose direction is towards the base station is eligible to be the CL. However once the CL residual energy is below threshold ( $t_h$ ), then a new CL has to be obtained, so that communication happens without any interruption. Fuzzy logic, an Artificial Intelligence technique can be employed to find proper CL. Fuzzy logic is used as the thinking and reasoning capability is similar to human reasoning and human thinking. Instead of only two binary values True or False, fuzzy logic involves intermediate values between True and False and hence fuzzy logic helps to deal with the vague problems. Thus fuzzy logic can be implemented using Fuzzy Logic Controller System (FLCS). It mainly consists of four parts. They are 1. Fuzzifier 2.Fuzzy rule base system 3.Intelligent Fuzzy Inference System (IIS) 4. Defuzzifier. as shown in Fig. 3

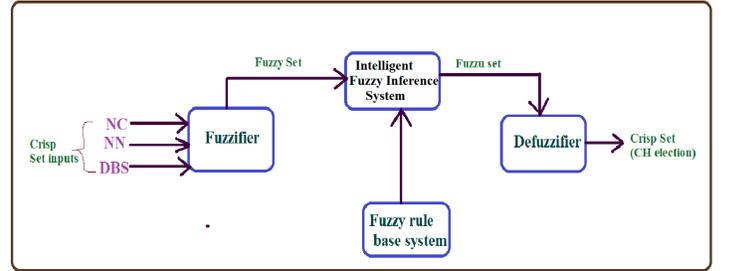


Fig. 3. Fuzzy Logic Controller System

**1. Fuzzifier:** Fuzzifier converts crisp numbers from the inputs to the fuzzy set of values. The inputs to this fuzzifier are i) Closeness of node to the Cluster Centroid (CCC): Node centrality conveys the information that how much near the node is to the centroid in each cluster. Thus CC can be calculated by using eq.19

$$CCC = \sqrt{D_{n_k, C_k}} \quad (19)$$

where  $n_k$  is the  $n^{th}$  node in  $k^{th}$  cluster and  $C_k$  is the centroid point of the  $k^{th}$  cluster. The lesser the NC value, the more the chances to become CL.

ii) Neighbour Nodes (NN): It tells us how many nodes are surrounded by the elected CL and the equation to calculate NN is given in eq.20.

$$NN = \frac{L}{M} \quad (20)$$

where L is the number of neighbours and M is the total number of nodes in the cluster. The more the value of NN the more

the chances to be the CL.

iii)Direction of node towards Base Station(DBS): It tells us that how much the node is near to the base station. It can be calculated using eq.21

$$DBS = \sqrt{D_{n_k,b}} \quad (21)$$

where  $n_k$  is the  $n^{th}$  node in  $k^{th}$  cluster and  $b$  is the location of base station. The less the DBS value, the more the chances to become CL.

For each parameter the linguistic values are given as  $CCC = [Near, Medium, Far]$

$NN = [More, Medium, Less]$

$DBS = [Far, Medium, Near]$

Based on the obtained crisp values of CCC,NN,DBS, the fuzzifier converts them to a fuzzy set, whose outcome is given by the value of Possibility (Pos) as.

$Pos = [VeryHigh, High, MediumHigh, MediumLow, Low, VeryLow]$

**2. Fuzzy rule base system:** It is the system that consists of if-then rules and is provided as an input to Intelligent FIS system. The framing of this if-then rules are very important and the if-then rules for our proposed work is as shown in Table I

TABLE I  
IF-THEN FUZZY SET OF RULES

No.	Closeness of node to Cluster Centroid	Neighbour Nodes	Direction of node towards Base Station	CL election
1	Near	More	Near	Very High
2	Near	More	Medium	High
3	Near	More	Far	Medium High
4	Near	Medium	Near	High
5	Near	Medium	Medium	Medium High
6	Near	Medium	Far	Medium
7	Near	less	Near	Medium Low
8	Near	less	Medium	Medium Low
9	Near	Less	Far	Low
10	Medium	More	Near	Very High
11	Medium	More	Medium	High
12	Medium	More	Far	Medium
13	Medium	Medium	Near	High
14	Medium	Medium	Medium	Medium High
15	Medium	Medium	Far	Medium Low
16	Medium	Less	Near	High
17	Medium	Less	Medium	Medium Low
18	Medium	Less	Far	Low
19	Far	More	Near	High
20	Far	More	Near	Medium High
21	Far	More	Far	Medium low
22	Far	Medium	Near	High
23	Far	Medium	Medium	Medium High
24	Far	Medium	Far	Low
25	Far	Less	Near	Medium Low
26	Far	Less	Medium	Low
27	Far	Less	Far	Very Low

**3. IIS:** It determines the match between fuzzy inputs from the fuzzifier and If-Then rules from the fuzzy rule base system. It combines the fired rules and forms the controlled action in the form of fuzzy output set.

**4. Defuzzifier:** Defuzzifier helps in Defuzzification.

Defuzzifier takes the fuzzy output set from the IIS and provides crisp output, that helps in determining the CL accurately. Thus to obtain the crisp output from defuzzifier, the Centroid method, which is also known as the Center of Area is used. In this method the center point is determined, where the area under the curve on both sides of this point is equal and the same is obtained by using eq. 22

$$\int_a^{x^*} \mu_{\bar{A}}(x)dx = \int_{x^*}^b \mu_{\bar{A}}(x)dx \quad (22)$$

where  $a = \min\{x|x \in X\}$  and  $b = \max\{x|x \in X\}$ .

Construction of membership function is a crucial step in the FLCS and they denote the degree of the truth in the FLCS. The graphical representation of input values mapped to membership values between 0 and 1 is determined by Membership function. The x-axis represents the universe of discourse and the y-axis represents the degree of membership in the interval  $[0, 1]$ . Many type of membership functions are available in literature, how ever the most commonly used membership function is a triangular membership function and provides best results. Thus in our proposed work, triangular membership function is selected for the inputs linguistic variables and the output variable. The membership function for the input variables to the fuzzifier is given in Fig. 4

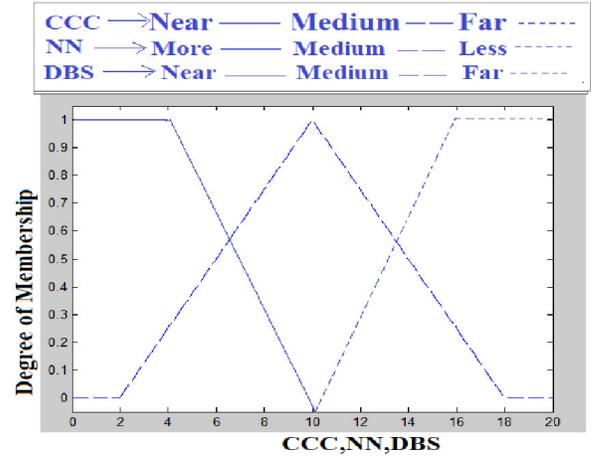


Fig. 4. Triangular Membership Function for CCC, NN, DBS

The triangular membership function for output variable Possible is given in Fig. 5

Hence fuzzy logic is adopted to select CL and the algorithm is shown in Algorithm 1.

Algorithm 1:

$N$ =Number of nodes in WSN

$n$ = a particular node in the network

$h$ =Neighbouring nodes

$t_h$ =The energy that is required for the node to be a CL

$DBS = D_{n_k,b}$ = The distance of  $n^{th}$  node in a cluster to a base station

$E_{res}$ = Nodes remaining energy

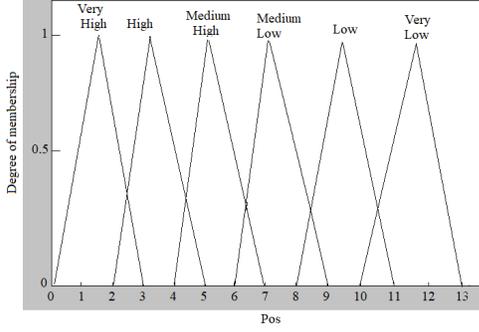


Fig. 5. Triangular Membership Function for Pos

$NN = L =$  Number of neighbours

$CCC = D_{n_k, C_k} =$  Distance of node to the Cluster Centroid

1. if  $(Val \{E_{res}\} > t_h$  and  $D_{n_k, C_k} \&\& D_{n_k, b}$ ) is less then

$CH(n_k) \leftarrow n_k$

2. if  $(E_{res} < t_h)$  then

ClusterLeaderlec  $\leftarrow$  fuzzyrule( $NN, DBS, CCC$ )

3. Send ClusterLeaderlec to the neighbours.

4. if  $(ClusterLeaderlec(n) > ClusterLeaderlec(h))$

CLected  $\leftarrow$  ClusterLeaderlec(n)

5. Broadcast(Halt Election)

6. else

round=round+1

7. Continue Step 2 to Step 6 for all the rounds until all the energy in the nodes are exhausted.

2) *VP election phase*: The selection of Virtual Point (VP) is very important to reduce the path length of the mobile element. Selecting the optimal number of VP points has to be achieved, so that the data is collected effectively. The CL communicates with the base station by sending HELLO message, along with its location, only if the CL has data in it or else it will not communicate HELLO message and If the CL has data in it then only the VP is selected or else the VP in that particular cluster does not exist. To determine the VP points, it is required to consider parameters such as Communication Range of CL, Distance towards BS. If the communication range of CL is R then the VP point should be selected with in this range and it should be very close to base station. The distance from the CL and the base station can be calculated by using Euclidean distance and then by subtracting the range of communication as shown in eq. 22

$$D_{C,B} = \sqrt{(c_i - b_1)^2 - (c_j - b_2)^2} \quad (23)$$

$$V_i = D_{C,B} - r \quad (24)$$

where  $i$  is the  $i_{th}$  CL,  $(c_i, c_j)$  are the CL ordinates and  $(b_1, b_2)$  are the base station coordinates. Thus by using eq.23 the exact point of VP can be obtained.

### C. Path Determination Phase

The mobile element has to reach each and every VP formed and collect data from the CL. Hence finding an optimal route

is very important. To find the optimal path, Particle Swarm Optimization (PSO) algorithm along with Artificial Bee Colony (ABC) named as PSO-ABC algorithm is used, which solves the Travelling Salesman Problem. ABC is used along with PSO, as ABC helps to eliminate the defects of PSO and helps PSO to obtain its local optimum quickly.

ABC algorithm is a meta heuristic algorithm, and it is based on the honey bees searching for their food. The model consists of three essential components: employed and unemployed foraging bees, and food sources. The first two components, employed and unemployed foraging bees, search for rich food sources, which is the third component, close to their hive. In order to implement ABC algorithm, it is very important to find the objective function. The search is made in such a way that the objective function value is minimized. The objective function is fixed by considering the parameters such as  $x_m = (D_{ij}), (e_{CL}), (t_m)$ .

where  $D_{ij}$  is the distance between the nodes,  $e_{CL}$  is the energy in CL,  $t_m$  is the time taken by the mobile element to reach the VP. Thus the objective function is given as in eq. 23

$$f_i = \min(\alpha_1 D_{ij} + \alpha_2 e_{CL} + \alpha_3 t_m) \quad (25)$$

Once the objective function is fixed the artificial bees that are present in the search space randomly selects a initial vector and later the best route is found from the initial population by using greedy solution. This is achieved by iterative method and employing the strategy by moving towards the better solution and by abandoning the poor solution. To determine whether the solution is best or poor, the objective function as given in eq.23 and the fitness function as given in eq.24 are employed.

$$fit_m = \begin{cases} \frac{1}{1+fit_i} & \text{if } fit_i \geq 0 \\ 1 + abs fit_i & \text{if } fit_i < 0 \end{cases} \quad (26)$$

The flow chart of PSO-ABC for path determination helps to solve TSP problems and finds the best and optimal route is as shown in Fig. 6

## V. SIMULATION RESULTS

This section helps to analyse the simulation results of the proposed method. The performance of wireless sensor networks is evaluated by implementing the proposed methodology using MATLAB. To evaluate the performance, 1000m x 1000m sensing area is considered. The initial energy of each node is assumed to be 50J and the energy required to transmit or receive each data packet is considered to be 50nJ. The Table II gives the list of parameters considered.

Fig. 7 represents the number of clusters that can be considered by using Gap statistics algorithm. The number of clusters here are determined by considering the gap between  $Log(W_k)$  and  $ELog(W_k)$ . The lesser the gap the better is the reduction in intra cluster distance and more the distance in the inter cluster distance, which further reduces the variance and provides the optimal number of clusters. The optimal number of clusters are obtained and by using the optimal number of clusters, the clusters are formed by using the

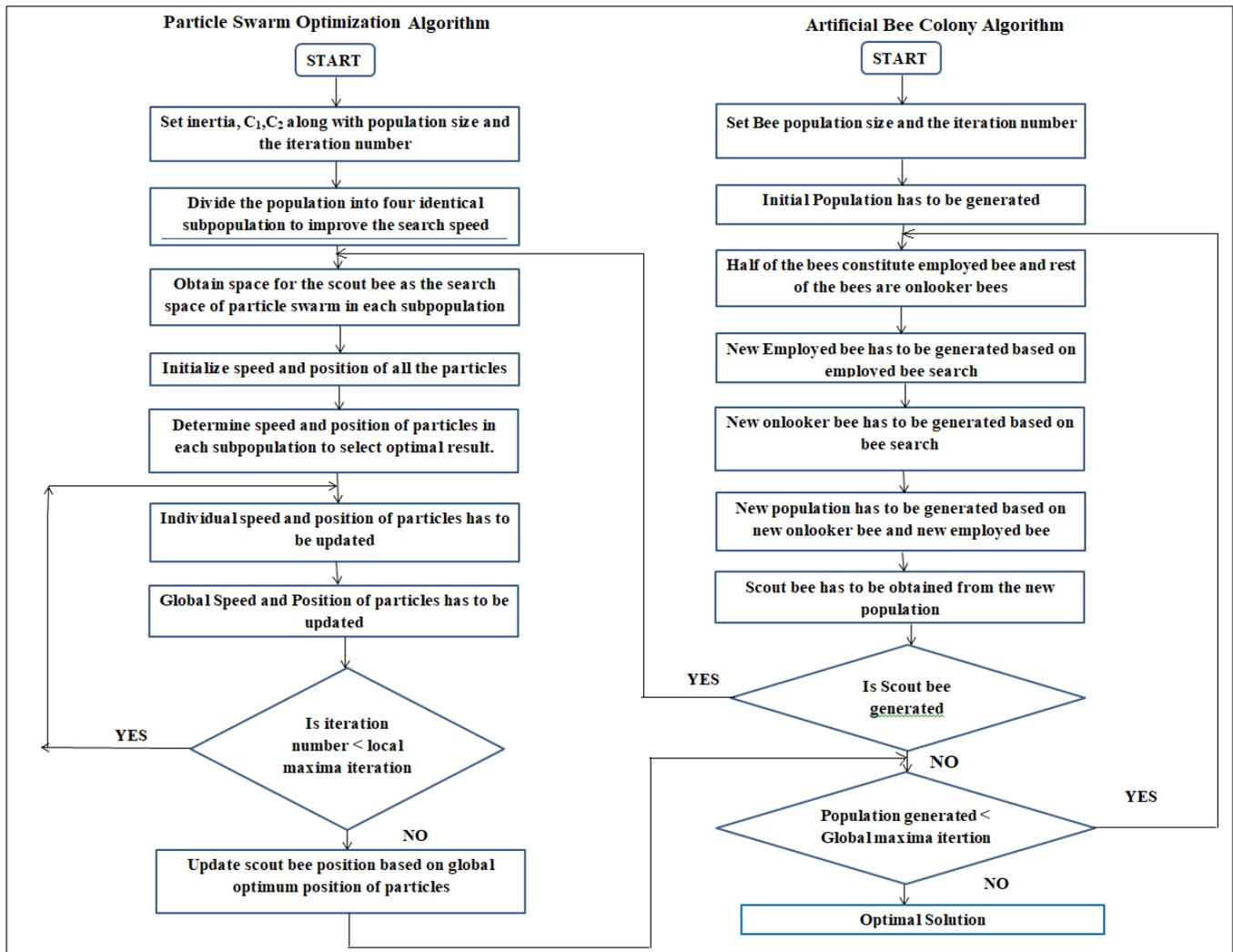


Fig. 6. PSO-ABC for Path Determination

TABLE II  
SIMULATION PARAMETERS

Simulation Parameters	Values
Area for sensing,	1000 X 1000m
Total Nodes, S	100-200
Nodes Initial Energy, E	50J
Transmission Energy required, $e_{tr}$	50nJ
Data Receiving energy required, $e_{rt}$	50nJ
Packet length of data, P	512 bits
Threshold energy of model, $D_0$	50J
Each bits transmission energy, $E_e$	50nJ/bit
Amplifier Coefficient, $\epsilon_s$	5PJ/bit/m <sup>2</sup>
Value of convergence, $\epsilon$	0.01
Transmission Euclidean distance, $D$	80m
Threshold, $T_H$	25J

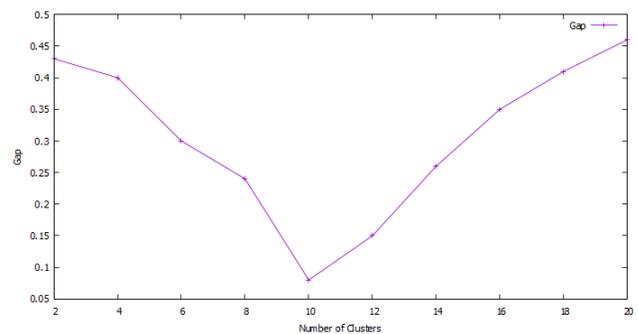


Fig. 7. Determination of optimal number of clusters

AEM algorithm where the AEM algorithm helps to obtain the centroid where the density of nodes are more and CL is selected by using the Fuzzy logic and based on the parameters considered the CL is taken near the centroid. The selection

of CL near the centroid reduces the communication distance between the nodes in the cluster and the cluster centroid. Since the communication distance is reduced, the energy consumed while transmitting the information from the nodes in the

cluster to the centroid is reduced drastically. Fig. 8 shows how the average energy is consumed within the cluster and our proposed algorithm requires less energy when compared with SGBDN. The network lifetime of the proposed network as the nodes varied from 10 to 200 is as shown in Fig. 9

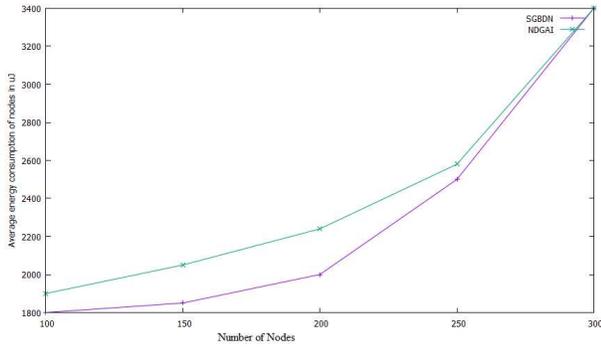


Fig. 8. Average Energy consumed within a cluster

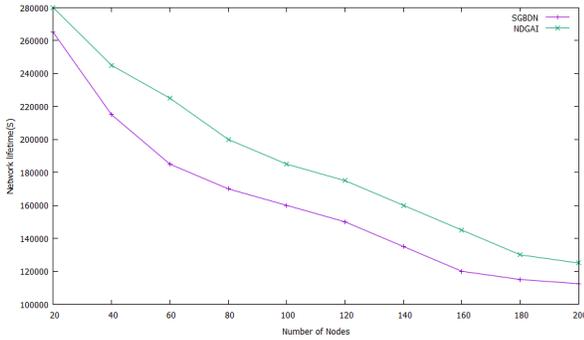


Fig. 9. Network life time using SGBDN and NDGAI

Fig. 10 shows that the number of nodes that are alive after each each round is given in the figure and it shows that our proposed algorithm is efficient as the nodes that that dyeing while collecting the data is less as the intra cluster distance has been reduced and the communication distance is efficiently considered. Fig. 11 also shows the number nodes that are alive is reducing with the increase in time. The proposed algorithm better performs as the path determined is the shortest path.

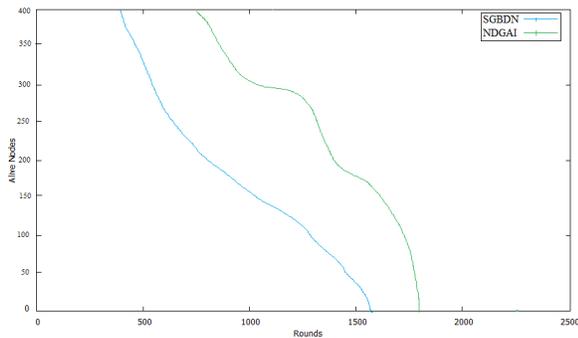


Fig. 10. Number of nodes alive

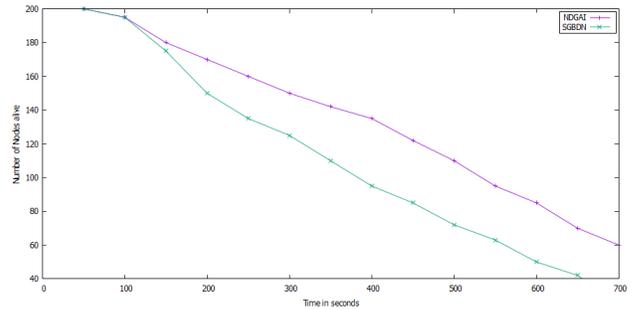


Fig. 11. Number of nodes alive as a function of time

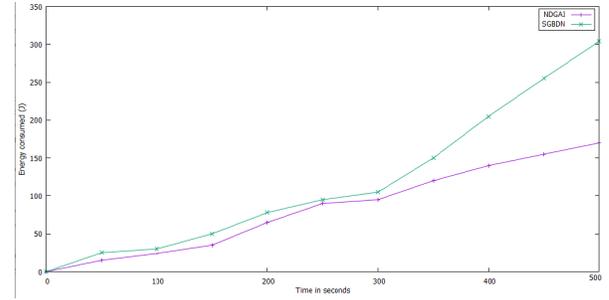


Fig. 12. Total Energy Consumed

Fig. 12 shows the total energy consumed with respect to the time. As the VP as selected based on the information present in the CL, the energy consumed will be less in the proposed algorithm. And also the optimal path is determined by using the PSO-ABC algorithm which is used to solve the TSP problem. Hence the Energy consumed has been reduced drastically.

## VI. CONCLUSIONS

The proposed methodology, A Novel Data Gathering Algorithm for Wireless Sensor networks using Artificial Intelligence (NDGAI) is a hybrid approach various various Artificial Intelligence algorithms are used along with the Mobile Element. The proposed methodology balances the power consumption and the energy is efficiently used while data is gathered in the WSN. NDGAI uses Gap statistics algorithm along with distance function for the selecting the optimal number of clusters. The clusters which are formed are fine tuned by using the AEM algorithm which forms the final clusters, where the cluster centroids are selected in such a way that the node density is more. This form of selection reduces the communication distance between the nodes to the centroid within the cluster. In each cluster the Cluster Leaders are selected near the centroid by using Fuzzy logic. Each clusters may have the Virtual Points based on the information present in the CL and these VP's are formed within the communication range of CL. And if the CL doesn't have any information, then the VP is not formed for that particular cluster. Once the VP's are formed the Mobile Element travels to this VP and collects the data from the CL. The optimal path for mobile element is determined by using PSO-ABC algorithm

which solves the TSP problem. Thus the entire methodology of considering ME along with the intelligent algorithms, helps in data gathering the information in the most energy efficient way and conserves lot of energy. The simulation results also shows that the proposed methodology is very efficient and improves network lifetime.

## VII. DECLARATIONS

**Funding:** Not Applicable

**Conflicts of interest/Competing interests :**Not Applicable

**Availability of data and material :**Available

**Code availability :** Available

**Authors' contributions :** Complete work has been carried out by us.

## REFERENCES

- [1] I. Akyildiz, W. Su, Y. Sankarasubramaniam and E. Cayirci, "A survey on sensor networks", IEEE Commun. Mag., vol. 40, no. 8, pp. 102-114, Aug. 2002.
- [2] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam and E. Cayirci, "Wireless sensor networks: A survey", Comput. Netw., vol. 38, no. 4, pp. 393-422, Mar. 2002.
- [3] Diamond and M. Ceruti, "Application of wireless sensor network to military information integration", Proc. 5th IEEE Int. Conf. Ind. Inform., vol. 1, pp. 317-322, Jun. 2007.
- [4] I. Bekmezci and F. Alagz, "Energy efficient delay sensitive fault tolerant wireless sensor network for military monitoring", Int. J. Distrib. Sens. Netw., vol. 5, no. 6, pp. 729-747, 2009.
- [5] L. Yu, N. Wang and X. Meng, "Real-time forest fire detection with wireless sensor networks", Proc. Int. Conf. Wireless Commun. Netw. Mobile Comput., vol. 2, pp. 1214-1217, Sep. 2005.
- [6] Akkaya, K, Younis M, (2005). A survey on routing protocols for wireless sensor networks. Ad Hoc Networks, 3(3), 325-349.
- [7] W. R. Heinzelman, A. Chandrakasan and H. Balakrishnan, "Energy-efficient communication protocol for wireless microsensor networks", Proc. 33rd Annu. Hawaii Int. Conf. Syst. Sci., vol. 2, pp. 10, Jan. 2000.
- [8] Swapna ch, Vijayashree R Budyal, "Expectation Maximization and Fuzzy Logic Based Energy Efficient Data Collection in Wireless Sensor Networks with Mobile Elements," 2020 7th IEEE International Conference on Signal Processing and Integrated Networks (SPIN), Noida, India, pp. 21-26, 20 April 2020 ,doi: 10.1109/SPIN48934.2020.9071055
- [9] Ruonan Zhang, Jianping Pan, Di Xie, Fubao Wang, "NDCMC: A Hybrid Data Collection Approach for Large-Scale WSNs Using Mobile Element and Hierarchical Clustering", IEEE Internet of Things Journal, Vol.3, No.4, pp. 533-543, AUG 2016.
- [10] S. B. Kotsiantis, "Supervised machine learning: A review of classification techniques," Informatica, vol. 31, no. 3, pp. 249-268, 2007.
- [11] C. Zhang, P. Patras and H. Haddadi, "Deep Learning in Mobile and Wireless Networking: A Survey," in IEEE Communications Surveys and Tutorials, vol. 21, no. 3, pp. 2224-2287
- [12] X. Wang, X. Li and V. C. M. Leung, "Artificial Intelligence-Based Techniques for Emerging Heterogeneous Network: State of the Arts, Opportunities, and Challenges," in IEEE Access, vol. 3, pp. 1379-1391, 2015
- [13] DW. 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
- [14] Abu Salem, A.O., Shudifat, N. Enhanced LEACH protocol for increasing a lifetime of WSNs. Pers Ubiquit Comput 23, 901-907 (2019). <https://doi.org/10.1007/s00779-019-01205-4>.
- [15] O. Younis and S. Fahmy, "HEED: a hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks," in IEEE Transactions on Mobile Computing, vol. 3, no. 4, pp. 366-379, Oct.-Dec. 2004
- [16] S. Lindsey and C. S. Raghavendra, "PEGASIS: Power-efficient gathering in sensor information systems," Proceedings, IEEE Aerospace Conference, Big Sky, MT, USA, 2002
- [17] Yan Jin, Ling Wang, Yoohwan Kim, and Xiaozong Yang "EEMC: An energy-efficient multi-level clustering algorithm for large-scale wireless sensor networks", Computer Networks, Volume 52, Issue 3,pp.542-562,22 February 2008.
- [18] A. Manjeshwar and D. P. Agrawal,"TEEN: A routing protocol for enhanced efficiency in wireless sensor networks", 15th International Parallel Distribution Process Symposium, pp.2009-2015, Apr. 2000.
- [19] Yinying Yang, Mirela I Fonoage, Mahaela Cardei, "Improving network lifetime with mobile wireless sensor networks", Computer Communications, Volume-33, No.4, March 2010.
- [20] S. Cho, L. Han, B. Joo, and S. Han,"P-LEACH: An efficient cluster-based technique to track mobile sinks in wireless sensor networks",International Journal of Distributed Sensor Networks, vol. 2014, pp. 1-10, Sep. 2014.
- [21] S. Gao, H. Zhang and S. K. Das, "Efficient Data Collection in Wireless Sensor Networks with Path-Constrained Mobile Sinks," in IEEE Transactions on Mobile Computing, vol. 10, no. 4, pp. 592-608, April 2011
- [22] Wen si, Hui-yuan Shi and Pan Liu, "A Collaborative Data Gathering Mechanism Based on Fuzzy Decision for Wireless Sensor Networks", International Journal of Distributed Sensor Networks, Volue 2015, Article ID 952428, pp.1-10, 2015
- [23] Saranya, V., Shankar, S. Kanagachidambaresan, G.R. "Energy Efficient Data Collection Algorithm for Mobile Wireless Sensor Network" Wireless Pers Commun 105, pp.219-232, December 2018.
- [24] Ghosh, N., Banerjee, I Sherratt, R.S. On-demand fuzzy clustering and ant-colony optimisation based mobile data collection in wireless sensor network. Wireless Netw 25, 1829-1845 (2019).
- [25] W. Wang and G. Tong, "Multi-path unequal clustering protocol based on ant colony algorithm in wireless sensor networks," in IET Networks, vol. 9, no. 2, pp. 56-63, 3 2020
- [26] C. Konstantopoulos, G. Pantziou, D. Gavalas, A. Mpitziopoulos and B. Mamalis, "A Rendezvous-Based Approach Enabling Energy-Efficient Sensory Data Collection with Mobile Sinks," in IEEE Transactions on Parallel and Distributed Systems, vol. 23, no. 5, pp. 809-817, May 2012
- [27] Anwit, Raj and Jana, Prasanta,"An Approximation Algorithm to Find Optimal Rendezvous Points in Wireless Sensor Networkx",International Conference on Advanced Computing Networking and Informatics, pp.193-204, 2018
- [28] Konstantopoulos, C., Pantziou, G., Gavalas, D., Mpitziopoulos, A. and Mamalis, B., "A Rendezvous-Based Approach Enabling Energy-Efficient Sensory Data Collection with Mobile Sinks", IEEE Transactions on Parallel and Distributed Systems. [Online] 23 (5), pp.809-817,2012
- [29] Jian Zhang, Jian Tang, Tianbao Wang and Fei Chen, "Energy-efficient data-gathering rendezvous algorithms with mobile sinks for wireless sensor networks" International Journal of Sensor Networks ,Volume 23, No.4, 12 April 2017.

# Figures

**Remote User**

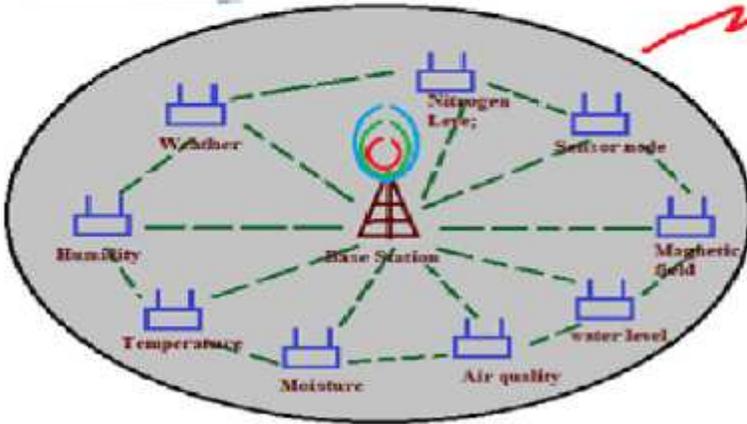


Figure 1

WSN scenario

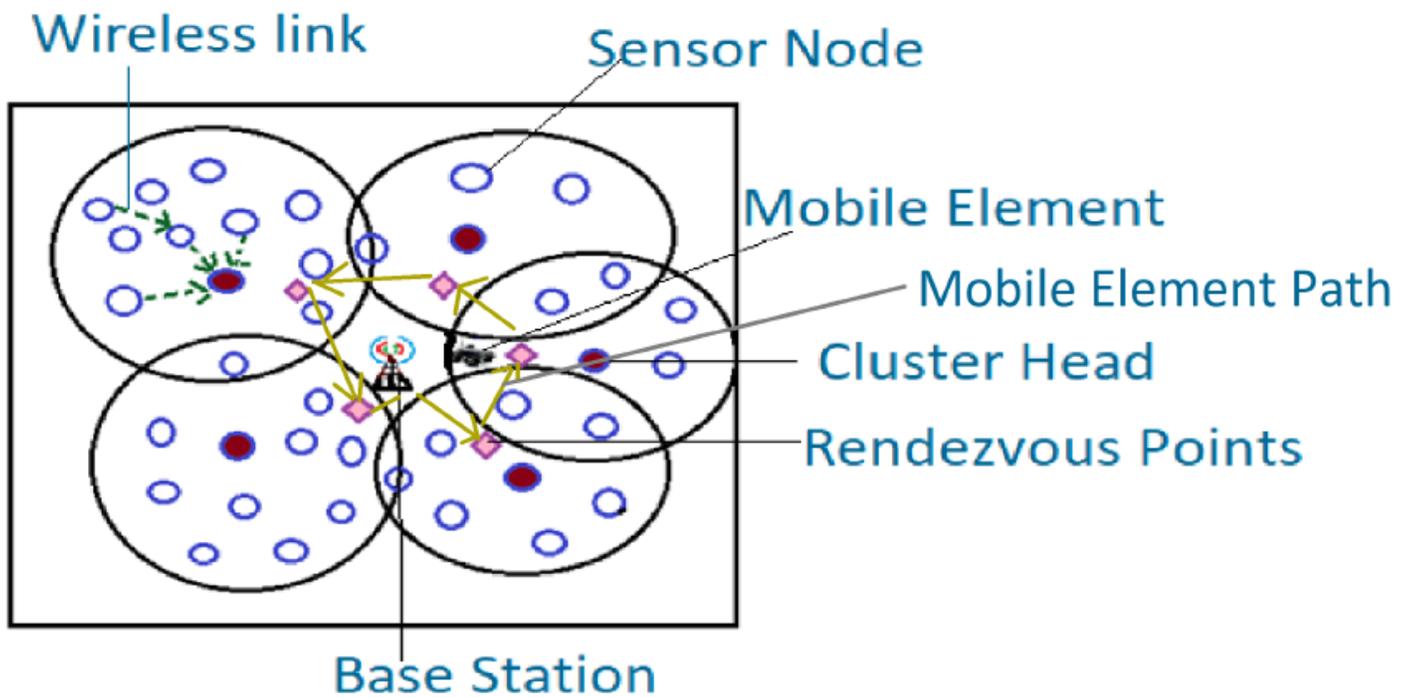


Figure 2

System Model

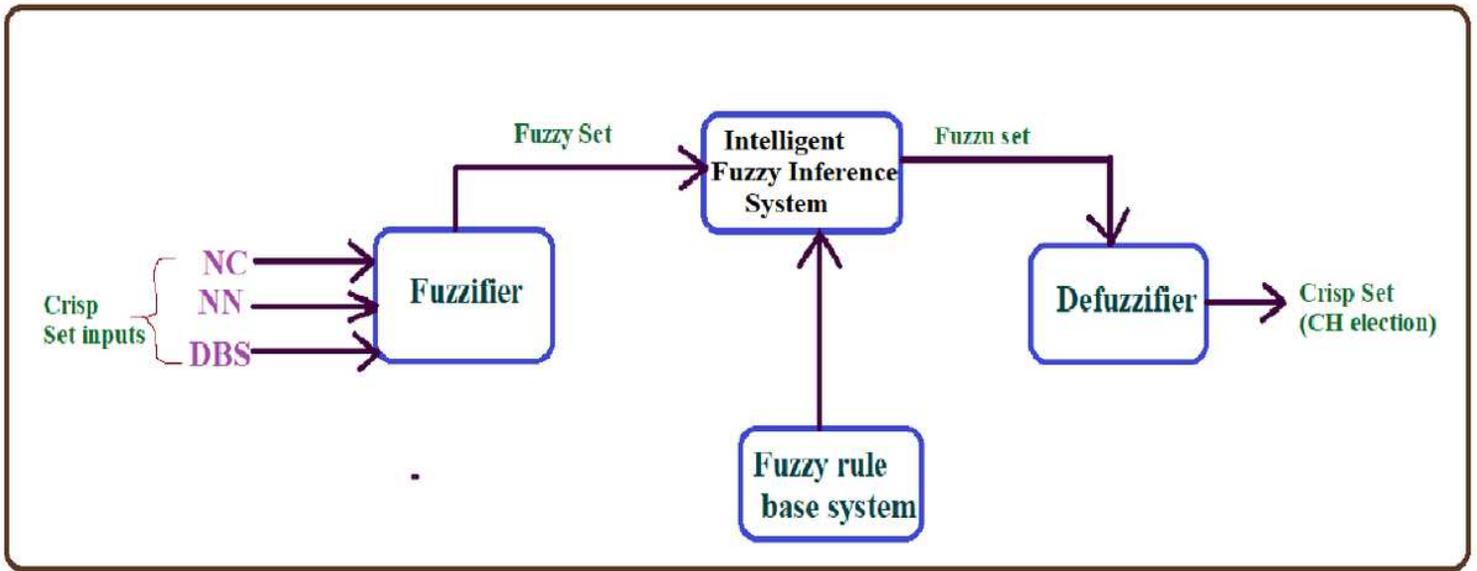


Figure 3

Fuzzy Logic Controller System

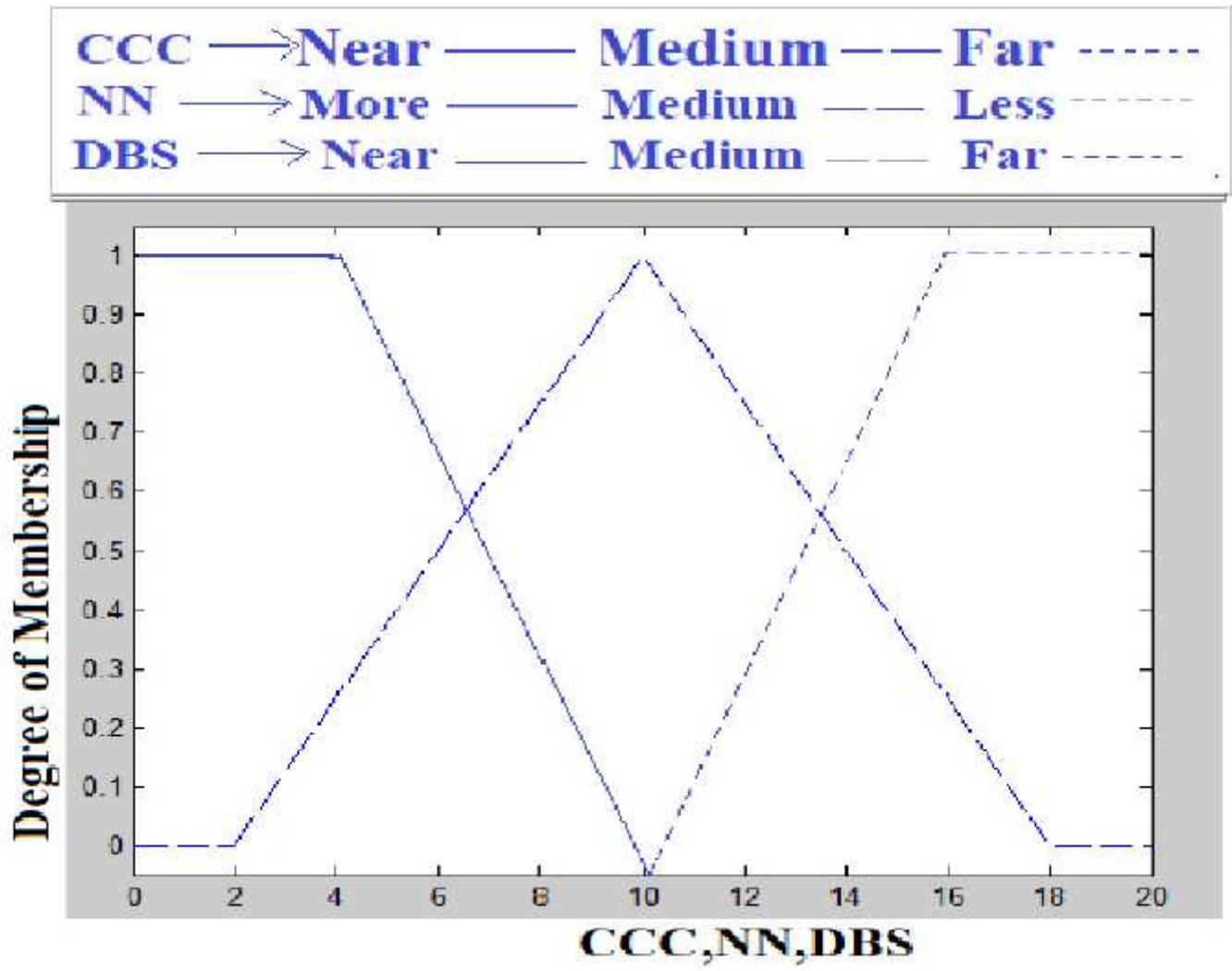


Figure 4

Triangular Membership Function for CCC, NN, DBS

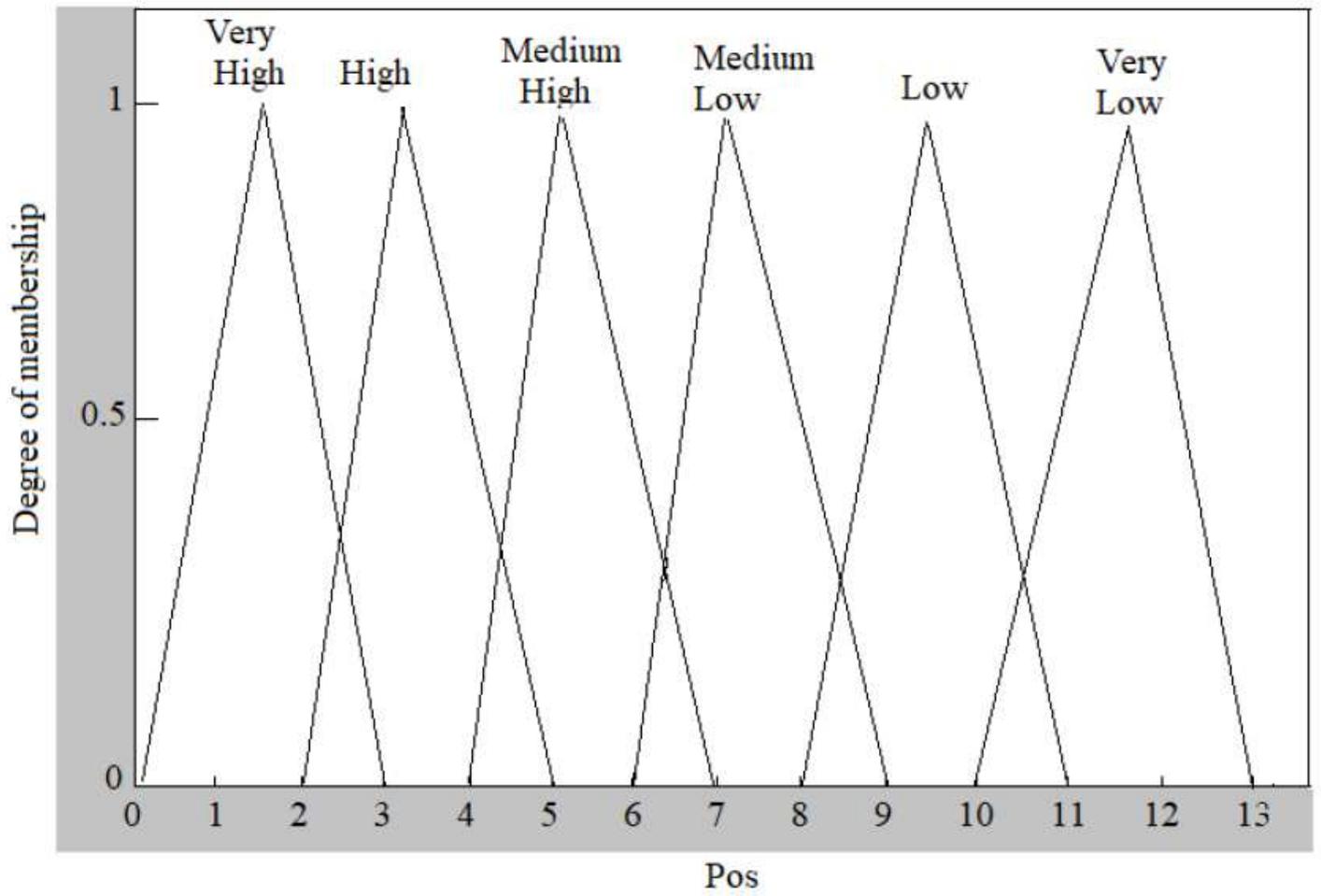


Figure 5

Triangular Membership Function for Pos

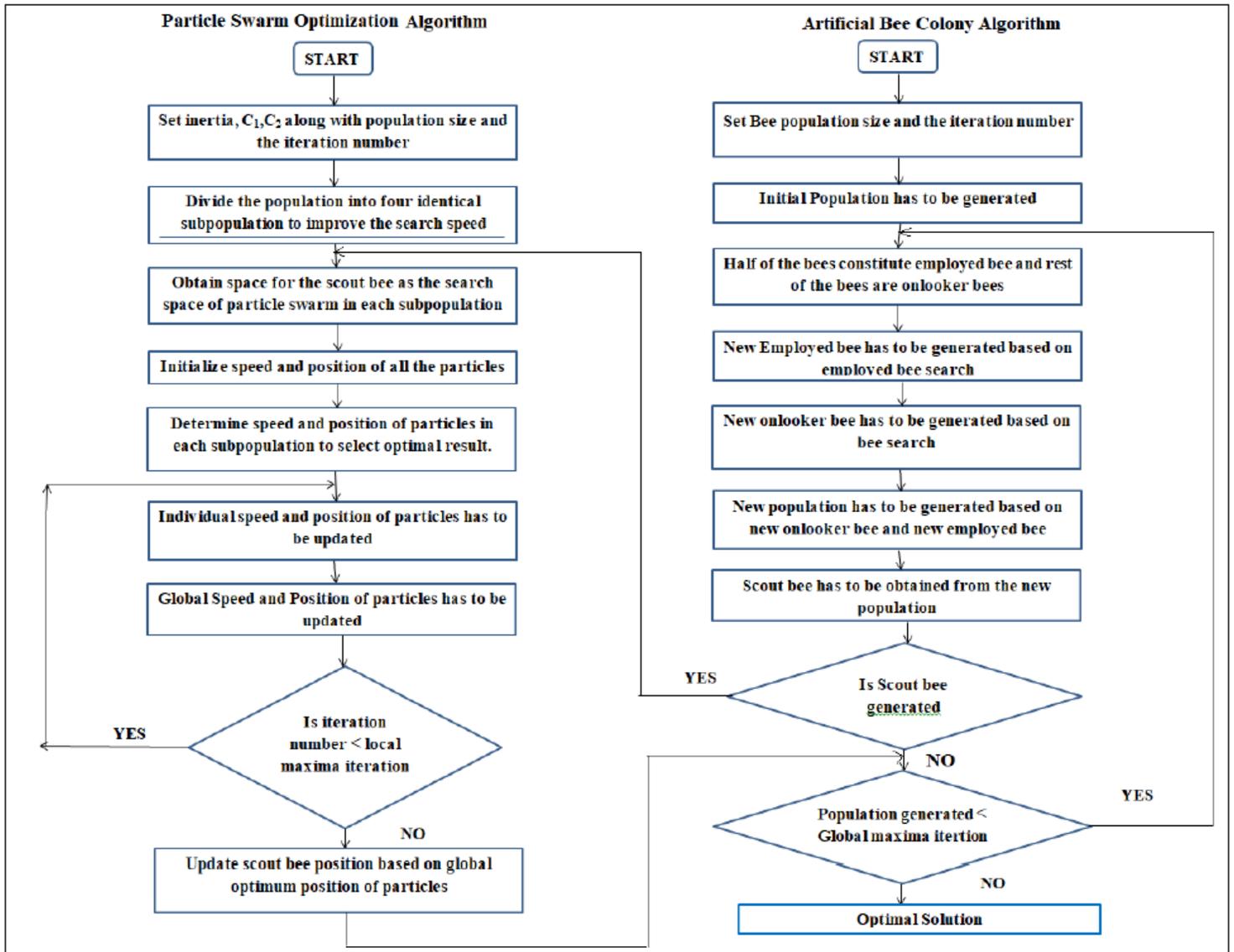


Figure 6

PSO-ABC for Path Determination

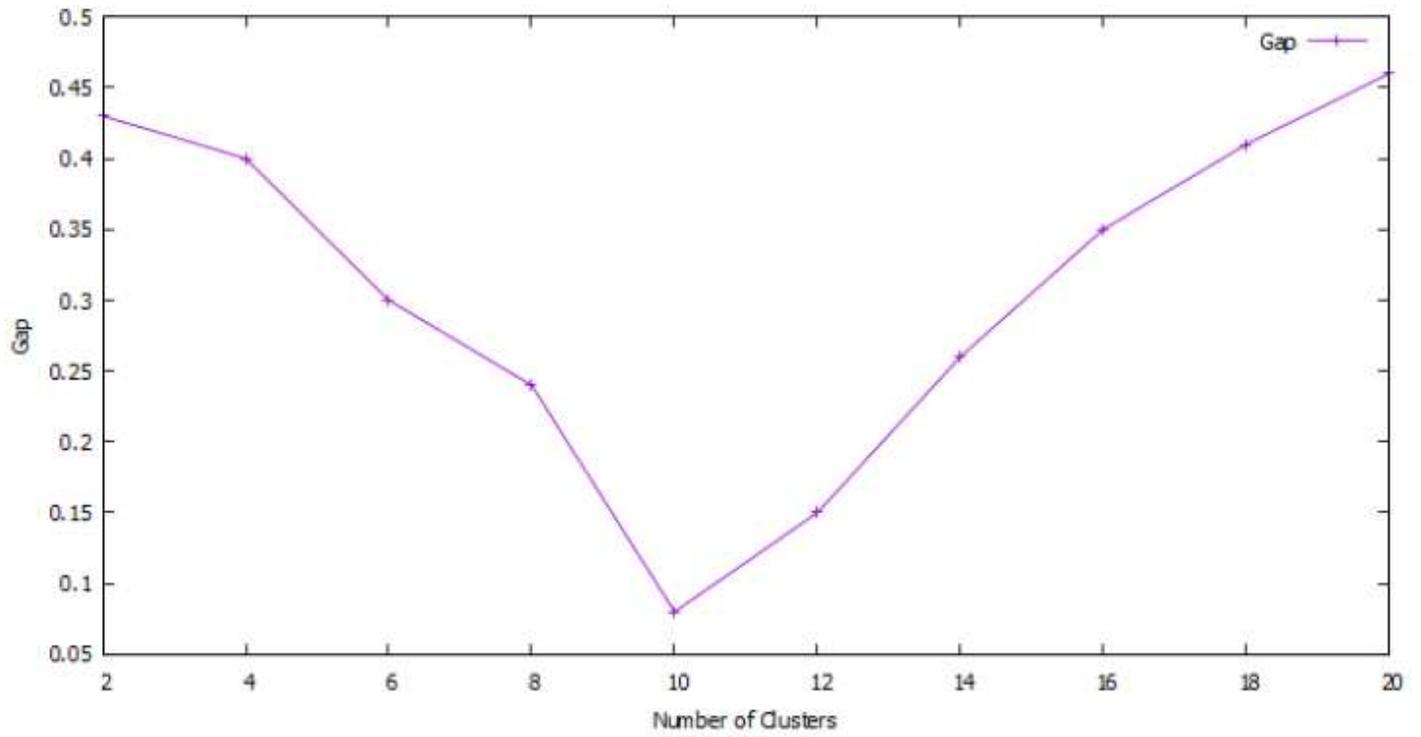


Figure 7

Determination of optimal number of clusters

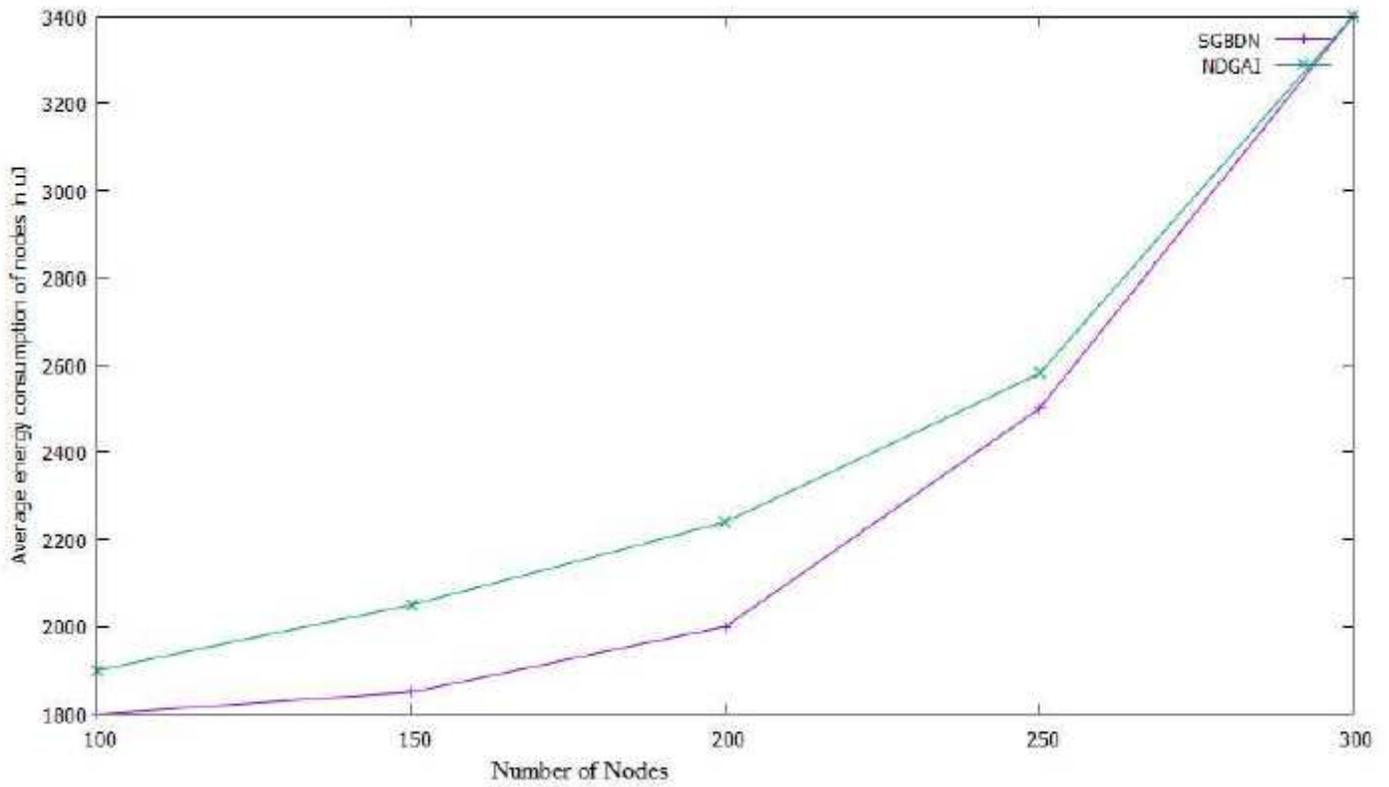


Figure 8

Average Energy consumed within a cluster

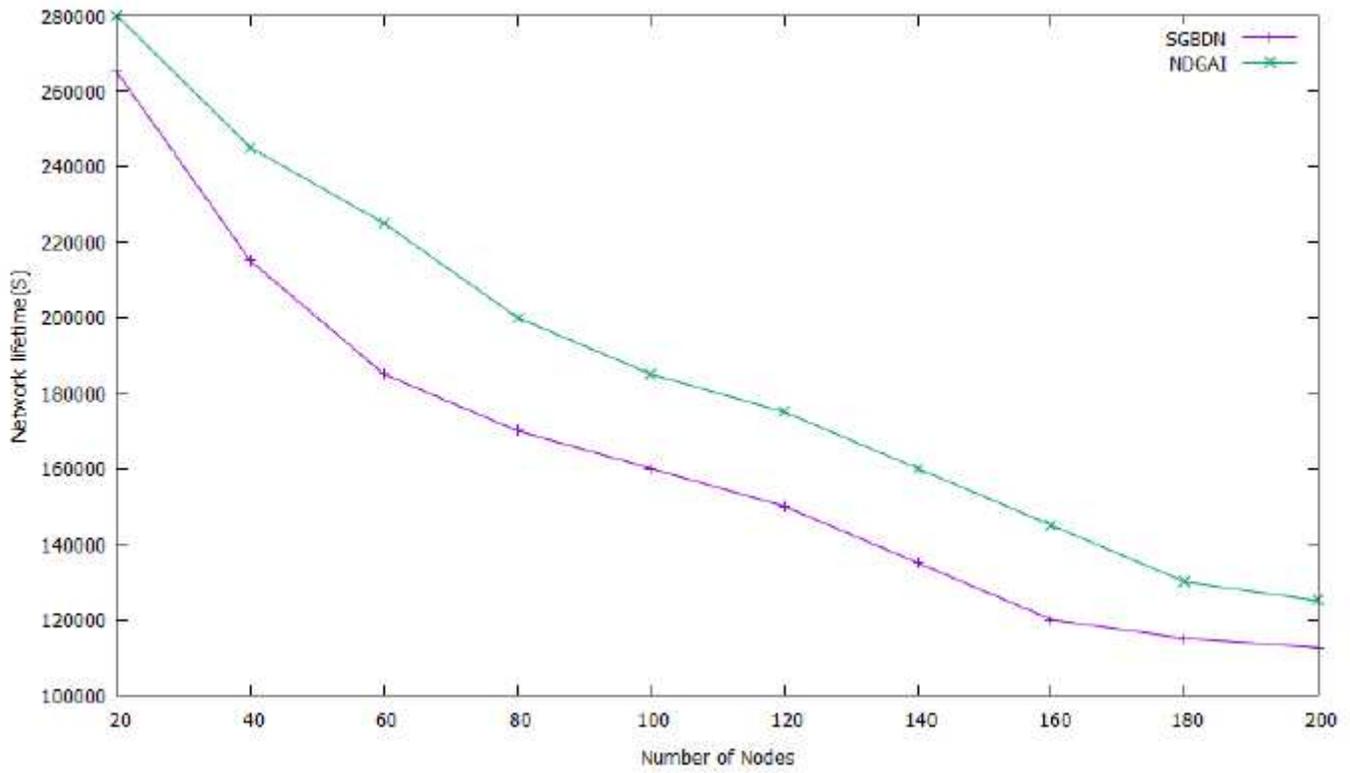


Figure 9

Network life time using SGBDN and NDGAI

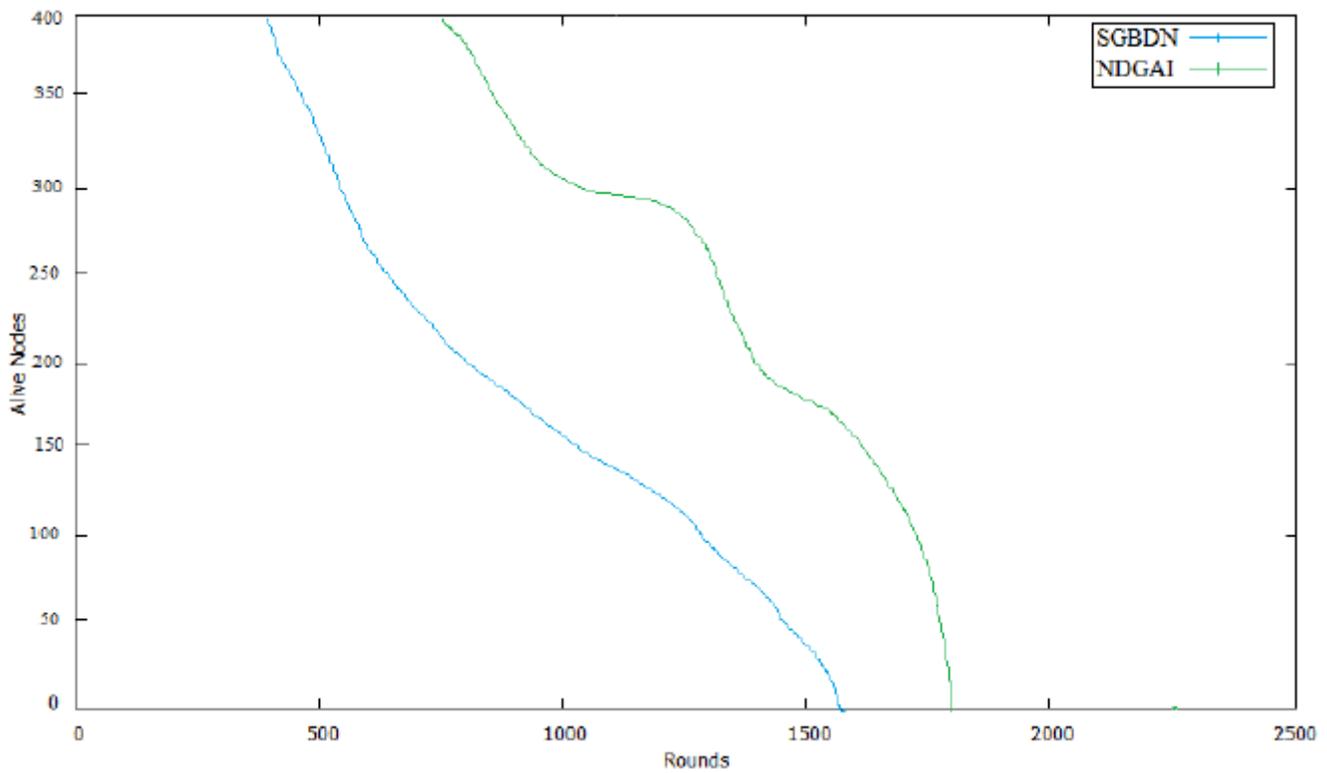


Figure 10

Number of nodes alive

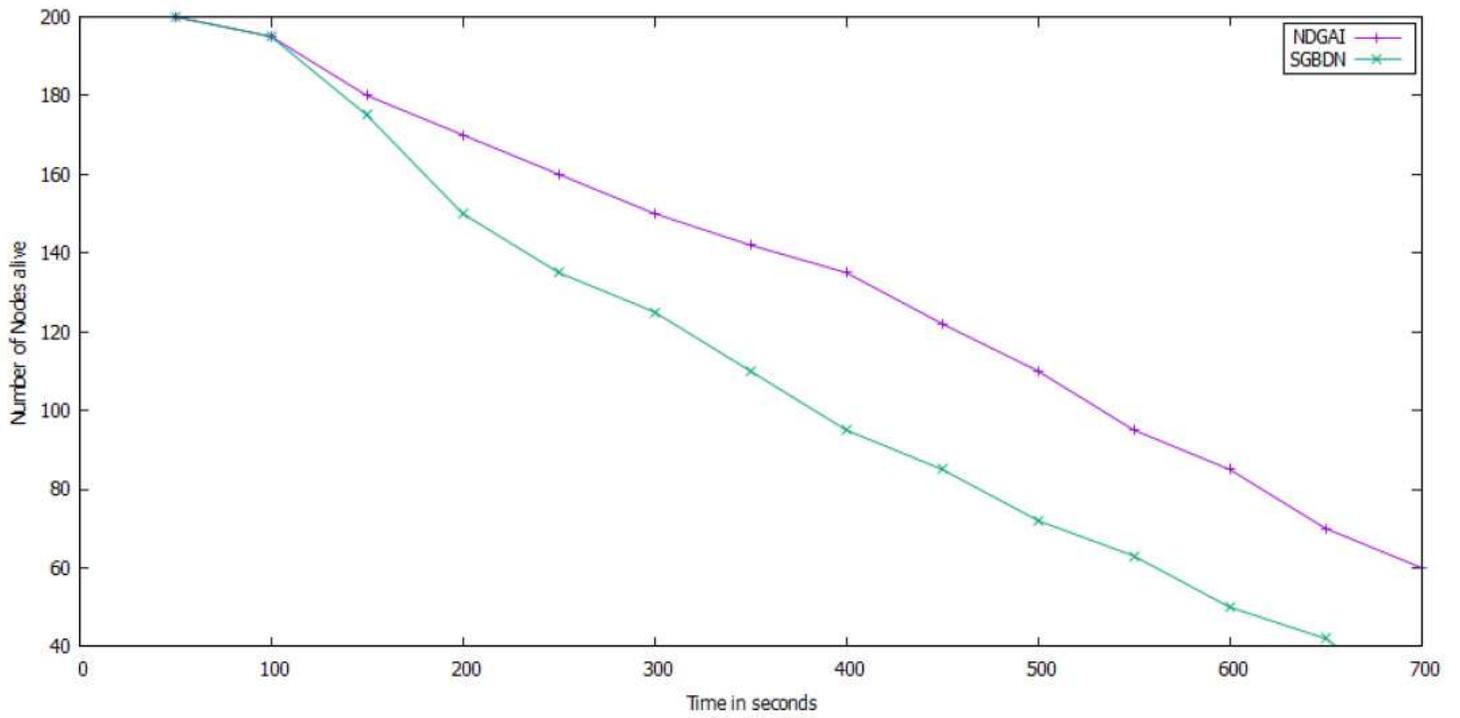


Figure 11

Number of nodes alive as a function of time

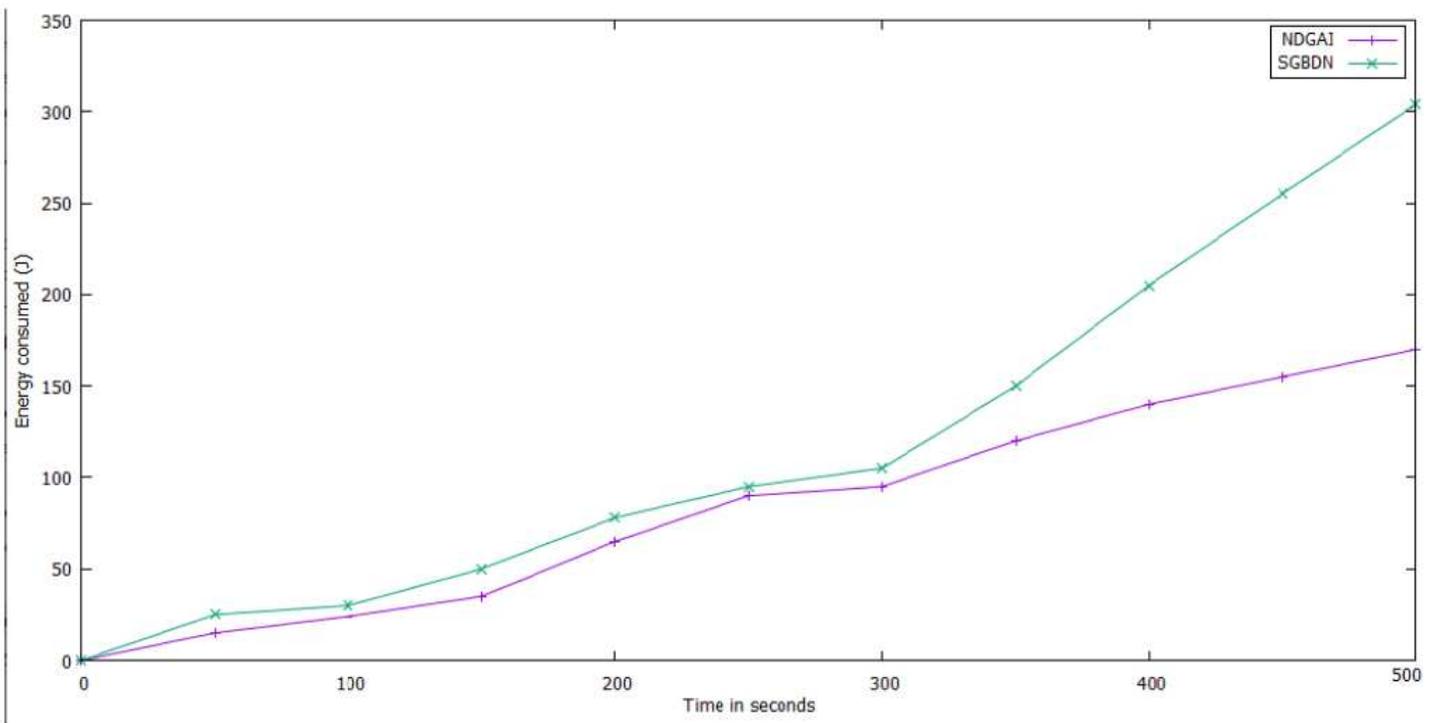


Figure 12

Total Energy Consumed