International Journal of Electrical and Computer Engineering (IJECE)
Vol. 9, No. 2, April 2019, pp. 1168~1183
ISSN: 2088-8708, DOI: 10.11591/ijece.v9i2.pp1168-1183

Genetic-fuzzy based load balanced protocol for WSNs

Pankaj Kumar Kashyap, Sushil kumar
School of Computer and System Sciences, Jawaharlal Nehru University, New Delhi, India

Article Info

Article history:

Received Jul 16, 2018
Revised Oct 16, 2018
Accepted Nov 6, 2018

Keywords:

A*(A-star routing)
Fitness function
Fuzzy logic inference system
Genetic algorithm
Inter-clustering routing
Load balancing

ABSTRACT

Recent advancement in wireless sensor networks primarily depends upon energy constraint. Clustering is the most effective energy-efficient technique to provide robust, fault-tolerant and also enhance network lifetime and coverage. Selection of optimal number of cluster heads and balancing the load of cluster heads are most challenging issues. Evolutionary based approach and soft computing approach are best suitable for counter the above problems rather than mathematical approach. In this paper we propose hybrid technique where Genetic algorithm is used for the selection of optimal number of cluster heads and their fitness value of chromosome to give optimal number of cluster head and minimizing the energy consumption is provided with the help of fuzzy logic approach. Finally cluster heads uses multi-hop routing based on A*(A-star) algorithm to send aggregated data to base station which additionally balance the load. Comparative study among LEACH, CHEF, LEACH-ERE, GAEEP shows that our proposed algorithm outperform in the area of total energy consumption with various rounds and network lifetime, number of node alive versus rounds and packet delivery or packet drop ratio over the rounds, also able to balances the load at cluster head.

Corresponding Author:

Pankaj Kumar Kashyap,
School of Computer and System Sciences,
Jawaharlal Nehru University, New Delhi, India.
Email: pankaj76_scs@jnu.ac.in

1. INTRODUCTION

Recent advancement in VLSI architecture, wireless communication with the help of sensors technology and recently evolved a new architecture Internet of things (IoT) gives upper hand to creation of wireless sensors networks for different applications such as battlefield surveillance, health care, disaster detection (forest fire, flood detection), remote habitat monitoring, home automation, environment data monitoring (temperature, humidity, pressure, seismic vibration) etc. [1]-[3]. Wireless sensor network (WSNs) consist of thousands of tiny sensors networks spatially dispersed over monitoring area to sense specific parameter and these sense data periodically deliver to base station by sensors node. Sensor node powered by limited capacity inexpensive small battery. Battery of sensor node is almost irreplaceable in nature when it runs out of power because of generally sensor nodes are dispersed on the area where human intervention is very difficult [4]. Therefore efficient use of available energy is most prominent issue in wireless sensors networks for running the network for long time.

Clustering is the top ranked approach for data gathering in WSNs. Clustering process divide the network in the form of small clusters. Division of networks in optimal number of clusters is NP- hard problem. Every cluster have a leader called as cluster head (CH), which perform the operation of data gathering from its member nodes, data aggregation and transfer the aggregated packet to base station (BS). Clustering has many significant advantages such as it reduce the number of transmission as only one leader per cluster, conserves the bandwidth of network because of nodes are communicated with their cluster head only. Also it is easily managed and fault-tolerance so it improves the scalability of networks [4]. Therefore balancing the load of leader is backbone of the clustering protocol that opens the door for efficient selection of cluster heads among nodes. Also clustering provide multi-hop routing for large scale network, so finding load balanced route from cluster head to base station is another issue.

To address the problem of selection of optimal number of cluster heads and to balance the load, we proposed Genetic based approach with the fusion of Fuzzy Logic technique named as Genetic Fuzzy Logic Based Energy-Efficient Load Balanced Clustering Algorithm (GFELC), which works in three rounds. Set-up phase, cluster binding phase and inter-clustering (A*(A-star) based algorithm) phase. The rest of the paper is organized as follows. Section 2 describes the previous algorithms related to Genetic algorithm as well as fuzzy logic. In Section 3 system and energy model is described. Section 4 described the proposed algorithm. Section 5 shows the extensive simulation work. Finally, we conclude our paper with brief discussion on conclusion and future scope.

2. RELATED WORK

LEACH [5-6] have certain limitations such as low-energy sensor node would be selected as CH, it does not uses multi-hop routing between inter cluster heads, leads to unevenly distribution of cluster formation. In PEGASIS [7] whereas nodes are placed in chain where at a time only one node (selected as leader without considering residual energy) transfer the data directly to BS and the role of leader rotated in chain. HEED [8] which eliminate the problem of unevenly distribution of CH by including the parameter residual energy and node density. In TEEN [9] sensor node sends only sensitive data to BS that is controlled in nature, Whereas APTEEN [10], [11] improve the TEEN and objective is to capture both the periodical data as well as sensitive data by implementing both proactive and reactive scheme, but it require more complexity in formation of clusters and its performance lies between LEACH and TEEN. In [12] proposed algorithm EELBCA creates min-heap of cluster heads on the basis of number of sensor nodes are join to respective clusters of cluster heads.

In [13] proposed algorithm is based on genetic technique, which creates 3-level hierarchical cluster. Initially it selects the optimal number of CH by genetic algorithm and finally routing is done based on criteria minimizing the transmission distance. But the redundancy may occur at level 2 clusters also energy of modes not conserve as much needed. In [14] proposed genetic based routing algorithm in which data is routed through relay nodes in two-tier sensor network architecture, genetic algorithm determine suitable route for the upper–tier relay nodes in sensor networks. It is centralized approach and sensor nodes (GPS enabled made algorithm costlier) are static after deployment as well as it require extra relay node.

In [15], proposed genetic algorithm is load balanced clustering approach which transfers the data from node to BS via gateways having high powered sensors than ordinary sensor nodes. Traffic load of each sensor nodes are determined prior the cluster formation. The nodes are assigning to only gateways to represent the valid chromosome. The fitness function of gateway is depending upon standard deviation of the load of the gateway which causes evenly distribution of traffic load among gateways. Lower the standard deviation higher the fitness value of gateways. The mutation is replacement of gateway having higher load in terms more number of nodes transferring or receiving the message with another lower loaded gateway.

The improved version of LEACH is proposed by J.L liu and C.V. Ravishankar in [16] (LEACH-GA) uses genetic approach. This algorithm include preparation phase before the set-up phase and steady state phase once for first round, where selection probability of node to become CH is evaluated. The fitness function is relying on the basis of minimizing the total energy consumption required for each clustering round. Selection of CH does not include residual energy. In GAEEP [17] fitness function is relying on minimizing the overall dissipation energy by considering optimal number of cluster heads. It does not include node density and distance factor to base station in the evaluation of fitness function. Multi hop routing and selection procedure of cluster head is also not explained in the GAEEP.

In [18] Gupta has proposed improved version of LEACH based on Fuzzy logic using variables; energy level, concentration level (node density) and centrality. The BS is responsible for collecting the energy level and location of each node. There is no use of multi-hop routing, so network consumes more energy. IN CHEF [19] which overcomes the problem of Gupta protocol by including local distance with residual energy. Distance to base station metric is excluded, which have major impact on network lifetime in the case of mobile nodes. In [20] proposed F-MCHEL improved version of CHEF where selection of cluster head depend upon fuzzy logic variables residual energy and proximity distance. The cluster head having maximum residual energy among the elected cluster heads play the role of Master cluster head. The Master cluster head is only responsible for transfer of the aggregated data to base station. F -MCHEL provides more stable network and energy efficient compare to LEACH and CHEF.

In [21] proposed algorithm (FM-SCHM) which show improvement over F-MCHEL by taking mobility as third parameter with residual energy and distance to base station. As they consider BS is mobile, but when mobility of the BS is decreases or increases then there is no effect on lifetime it remains constant. In [22] proposed algorithm by same author to overcome the problem of mobility by taking centrality parameter instead of distance to base station as third parameter and all the other assumption are same, which show the improvement of lifetime over FM-SCHM [21]. In LEACH-FL [23] is also yet another improved version, where selection of CH depend on three fuzzy variables, energy level, node density and distance between cluster head and base station. It differs from the Gupta protocol only in one parameter centrality versus distance to base station. But LEACH-FL also suffers from same problem as Gupta protocol such as multi-hop routing. In LEACH-ERE [24] fuzzy Logic based protocol which consider the expected residual energy (as prediction) as well as residual energy of node as fuzzy variables. But it does not consider the distance between cluster head and base station, node density and centrality as well, which leads to unbalanced load and energy consumption.

In [25] Alshawi et al. proposed an algorithm to balance the traffic load using fuzzy and A-star approach which give the least burden path with forwarding nodes having higher energy and minimum hop count. In CFGA [26] BS creates a balanced cluster based on Genetic-fuzzy based algorithm. As the BS situated in central of the network and all nodes check their validity for CH in fuzzy module consumes more energy of nodes and lifetime of network decreases and they does not show how multi-hop routing works [27], [28]. In [29] A-star based algorithm (ASSER), BS find optimal route and broadcast in the two-tier network, by which CH send data to BS using optimal route, which leads to balance the traffic load.

From the above proposed algorithms, it appears that no better algorithm developed which uses both evolutionary (Genetic algorithm) and soft computing (Fuzzy logic based) approaches to overcome the limitation of each other and provide optimal number of clusters in network and balanced the load among CHs along with multi-path routing. Our proposed algorithm uses genetic algorithm for selection of optimal number of cluster heads which uses fuzzy logic based inference system to evaluate the fitness function of sensor nodes, and balancing the traffic load between cluster heads done through A*(A-star) algorithm.

3. SYSTEM AND ENERGY MODEL

We consider that all the sensors nodes are homogenous in terms of sensing, computation and transmissions capability, they all have equal initial energy and unique identification number (ID). Sensors nodes adjust their radio power to transmit data. To transmit m-bit of message by any sensor node [5-6] in either free space (d^2 power loss) or multipath (d^4 power loss) model dissipates energy E_{elec} radio electronics and $\mathcal{E}_{fs}$ is free space amplifier energy or $\mathcal{E}_{mp}$ is multipath fading amplifier energy over distance d required Energy $E_{TX}(m,d)$ is the summation of both electronic $E_{TX-elec}$ and amplifying energy E_{TX-amp} as follow.

$$E_{TX}(m,d) = E_{TX-elec} + E_{TX-amp} = \begin{cases} m\,E_{elec} + \mathrm{m}\,\mathcal{E}_{fs}\,d^2, if\ d < d_0 \\ m\,E_{elec} + \mathrm{m}\,\mathcal{E}_{mp}\,d^4, if\ d \geq d_0 \end{cases} \quad (1)$$

For a node to receive a message of m-bit dissipates energy $E_{RX}(m)$ in radio electronics as,

$$E_{RX}(m) = E_{RX-elec}(m) = m\,E_{elec} \quad (2)$$

Whereas d_0 is reference distance $d_0 = \sqrt{\mathcal{E}_{fs}/\mathcal{E}_{mp}}$ to use differentiate between free space model and multipath model. E_{elec} is the required unit electronic energy to process one bit of message, which depends upon several factors such as modulation, digital coding, signal, acceptable bit-rate etc.

If there are N nodes and C clusters and each cluster have on average N/C nodes per clusters (one cluster head and remaining ((N/C)-1) are member nodes). The expected consumed energy of cluster head $EC_{expcons}^{CH}$ [4] is the sum of consumed energy in receiving packet from its member nodes and aggregating them into single packet fixed size require energy E_{DA} and transferring to base station directly (d_{CH-BS}^4 power loss) or to another cluster head (multi-hop routing) require (d_{CH-CH}^4) multipath power loss.

$$E_{expcons}^{CH}(m,d) = N_{frame}[E_{TX}(m,d) + \frac{N}{C}(E_{RX}(m) + mE_{DA})]$$
$$E_{expcons}^{CH}(m,d) = \begin{cases} N_{frame}(mE_{elec} + m\mathcal{E}_{amp}\,d_{CH-BS}^4 + \frac{N}{C}\,mE_{DA} + mE_{elec}), d_{CH-BS} < d_0 \\ N_{frame}(mE_{elec} + m\mathcal{E}_{amp}\,d_{CH-CH}^4 + \frac{N}{C}\,mE_{DA} + mE_{elec}), d_{CH-CH} > d_0 \end{cases} \quad (3)$$

We assume that all the sensor nodes receive and transmit same size of data packet m-bit. Number of frame N_{frame} transmitted by cluster head in one round (time duration of a node act as CH) calculated as.

$$N_{frame} = \frac{t_{SSS}}{k * t_{slot} + t_{CH-BS}} \quad (4)$$

Where $k = \left(\frac{N}{C} - 1\right)$ is the number of member node, t_{SSS} is the time period (steady-state phase) of a node to be CH, t_{slot} is slot duration in which node send their packet into frame and t_{CH-BS} is time taken by CH to transfer the packet to base station. Now, expected residual energy of cluster head $E_{expres}^{CH}(m, d)$ after steady state phase is difference between residual energy of node (defined as residual energy of node before cluster head selection) and expected consumed energy of node given by Equation (5).

$$E_{expres}^{CH}(m, d) = E_{res}(m, d) - E_{expcons}^{CH}(m, d) \quad (5)$$

On the contrary, expected consumed energy of non-cluster head node is depend upon number of frame transmitted in one round. As member node of clusters are close to its cluster head so it follow free space path model ($d_{to\,CH}^{2}\ power\ loss$).

$$E_{expcons}^{NCH}(m, d) = N_{frame} \times (E_{TX}(m, d)) = N_{frame}(mE_{elec} + m\mathcal{E}_{fs}\, d_{to\,CH}^{2}) \quad (6)$$

Expected Residual Energy of non-cluster head is

$$E_{expres}^{NCH}(m, d) = E_{res} - E_{expcons}^{NCH}(m, d) \quad (7)$$

4. GENETIC FUZZY LOGIC BASED ENERGY-EFFICIENT LOAD BALANCED CLUSTERING ALGORITHM (GFELC)

GFELC works in three phases. Set-up phase provide selection of optimal number of cluster head based on Genetic fuzzy logic inference system. Cluster binding phase relate to calculation of cluster area done by CH. The sensor nodes inside the area occupied by cluster head send their data directly to cluster head and cluster head integrate several packets into single packet of fixed size using data aggregation. In Routing phase, CH forward the aggregated packet to base station through multi-hop inter cluster routing based on A* algorithm which minimizes the hop count, balance the traffic load on CH in terms of limit the number of the packet transmission. The block diagram and flow chart of algorithm is shown in Figure 1.

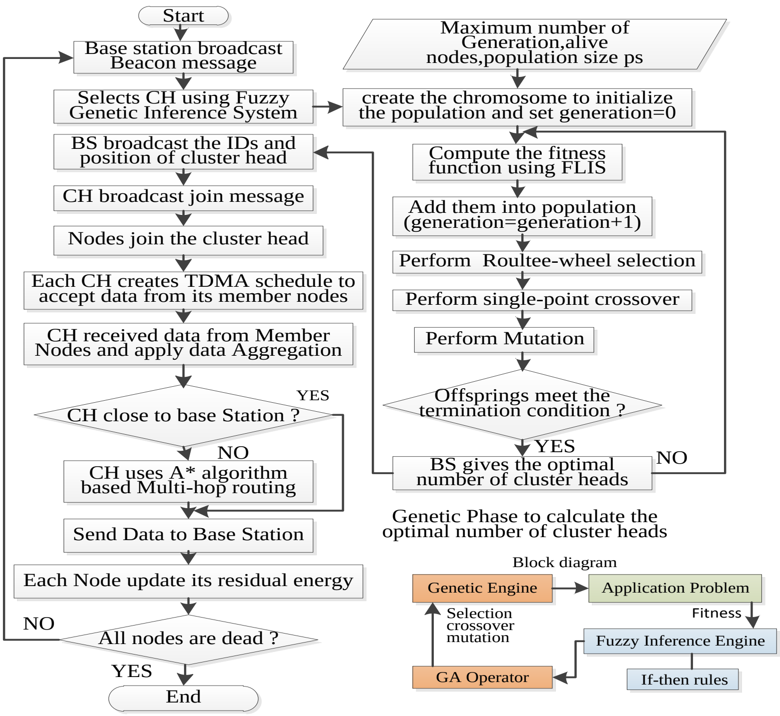

Figure 1. Flow chart and block diagram of GFELC algorithm

4.1. Set-up phase

In this phase initially base station broadcast a beacon message for collecting the information about nodes such as IDs, distance to base station, node density, expected residual energy and residual energy. The sensor nodes after listening beacon message they calculate distance to base (location) station through Received Signal Strength Indicator (RSSI). The received power P_r (in dBm) is defined as by [31].

$$P_r = P_o + 10\,\alpha\,\log\left(\frac{d_r}{d_o}\right) + S \quad (8)$$

Where P_o (in dBm) is the received power at a reference distance d_o and α ($2 \le \alpha \le 4$) is the path loss exponent. S is the Gaussian random variable represent medium-scale channel fading with zero mean and variance σ^2 (in dBm, $4 \le \sigma \le 12$). The measured distance d_r from base station is calculated as;

$$d_r = d_o\, 10^{\frac{RSSI-A}{10\,\alpha}} \quad (9)$$

Where A is the received signal strength meter in one meter distance from base station with no obstacle.

4.1.1. Chromosome representation and initial population

The base station creates a chromosomes (having equal length) as a string of sensors nodes having residual energy greater than or equal to average energy of all live nodes. This ensures that optimal number of cluster heads is selected by base station to reduce the energy consumption through restrict the number of message exchange also it's reduce the length of chromosome that makes faster convergence rate of GFELC The chromosome resembles to binary string of 1 or 0, where 1 represent the sensor node as cluster head and 0 represent the ordinary node shown in Figure 2.

Sensor Set (S_{Live})	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11
Cluster heads Genes	0	1	0	0	1	0	1	0	0	0	1

Figure 2. Binary representation of cluster heads chromosome

4.1.2. Fitness function

The fitness function for cluster head chromosome is defined as function of average residual energy level (ARE_n), average node density (AD_n), average distance to base station of node (ADB_n) and average expected residual energy ($AERE_n$) of chromosome.

$$ARE_n = \frac{1}{l}\left(\sum_{i=1}^{l} RE_i\right), AD_n = \frac{1}{l}\left(\sum_{i=1}^{l} D_i\right), ADB_n = \frac{1}{l}\left(\sum_{i=1}^{l} DB_i\right) \text{ and } AERE_n = \frac{1}{l}\left(\sum_{i=1}^{l} ERE_i\right) \quad (10)$$

Where l is the number of cluster heads in chromosome. RE_i Is the residual energy, D_i is the density of node, DB_i is the distance to base station and ERE_i is the expected residual energy of cluster head in the chromosome. The fitness function is as follow:

$$F = \alpha_i * f_i, \text{ where } f_i \epsilon \{ARE_n, AD_n, ADB_n, AERE_n\} \quad (11)$$

Where α_i is the weight of fitness function and updated according to the formula $\alpha_i = \alpha_i' + \beta_i \delta f_i$ .where $\delta f_i = f_i - f_i'$ with f_i and f_i' are fitness value for the current and previous generation chromosome and the coefficient β_i is calculated by formula $\beta_i = {}^{1}/_{(1 + e^{-} f_i')}$ which improve the further weight value for current chromosome. Initially weight value is chosen according to the simulation. Where function f_i is evaluated by fuzzy logic inference system.so basically fitness of chromosome defined as:

$$Fitness(Chromosome) \propto \{FLIS(Chromosome)\}. \quad (12)$$

4.1.3. Fuzzy logic inference system (FLIS)

The input for FLIS for selection of chromosome depends on four different metrics: average residual energy level (ARE_n)={low(L), medium(M), high(H)}, average node density (AD_n)={sparse(S), abundant(A), dense(D)}, average distance to base station of node (ADB_n)={close(C), near(N), remote(R)} and average expected residual energy ($AERE_n$)={often(O), valid(V), extreme(E)} of chromosome. Trapezoidal and triangular member ship function is used for linguistic variables shown in Figure 3. The chance of the choosing chromosome gives optimal number of cluster head PCH_n divided into seven linguistic variables={poor (P), tiny (T), fair (F), good (G), well (W), best (B), superb(S)} shown in Figure 4. Each parameters divide into three levels, so it require 3^4 =81 knowledge base rule shown in Table 1.

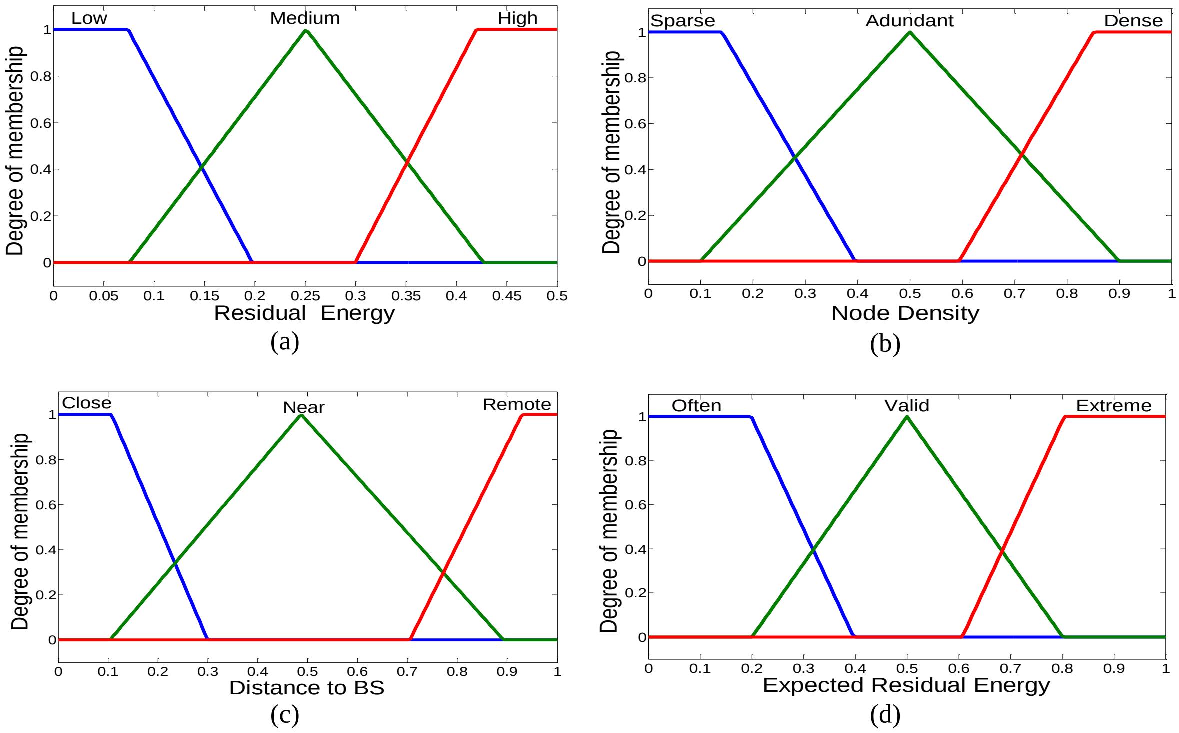

Figure 3. Membership function of (a) ADB_n , (b) AD_n ,(c) ADB_n, (d) $AERE_n$

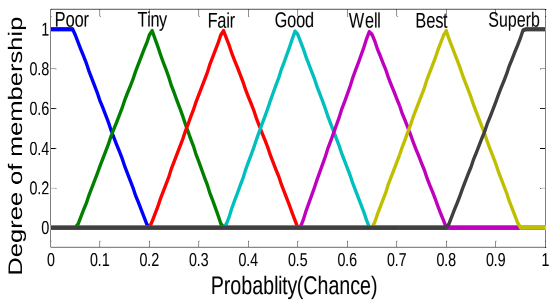

Figure 4. Membership function of Chance (PCH_n)

The two extreme cases are, if residual energy of a chromosome is high, node density is dense in nature, base station is close to node and expected residual energy is extreme then chromosome have superb chance to give optimal number of cluster head and second one is, if residual energy of node is low, it is sparse in nature, distance to base station is remote and expected residual energy is often then there is poor chance of chromosome to be selected by base station to produce optimal number of cluster head.

Table 1. Fuzzy Logic If-Then Rules

SL No.	1	2	3	4	5	6	.	.	.	.	.	.	76	77	78	79	80	81
ARE_n	L	L	L	L	L	L	.	.	.	.	.	.	H	H	H	H	H	H
AD_n	S	S	S	S	S	S	.	.	.	.	.	.	D	D	D	D	D	D
ADB_n	C	C	C	N	N	N	.	.	.	.	.	.	N	N	N	R	R	R
$AERE_n$	O	V	E	O	V	E	.	.	.	.	.	.	O	V	E	O	V	E
PCH_n	T	F	F	P	T	T	.	.	.	.	.	.	G	W	B	G	G	W

The fuzzy logic inference system works into four steps as follows:

a. Crisp value input and fuzzification- fuzzifier decides the value of inputs based upon triangular membership function which is the intersection point and creates fuzzy sets or simply it converts the numerical value into graph membership function.
b. Fuzzy Rule Base-It consists of series of 81 IF-THEN rules which runs parallel on fuzzy sets inputs in any order. As IF-THEN have multiples inputs, so minimum selection fuzzy AND operator is applied to select minimum of four membership value to get one single value to output set.
c. Aggregation of all output value- To aggregate multiple output value single fuzzy sets used fuzzy union operator OR, which selects maximum of our fuzzy rule base output to create fuzzy output set.
d. Defuzzificaton-Selection of chromosome depends on single crisp value not as collection of value (output fuzzy set consist of linguistic variable), so we apply centroid defuzzification method given as:

$$Node_{\text{chance}} = \frac{\int \mu_A(y) * y dy}{\int \mu_A(y) dy}, \quad (13)$$

Where $\mu_A(y)$ define degree of membership function of object y in fuzzy set A , which is defined as in terms of ordered pairs: $= \{(y, \mu_A(y)) / \, y \in U\}$, where U is the universe of discourse.

4.1.4. Selection

It is used to determine probability of chromosome in proportion with fitness value, higher the fitness value higher is the chance of selection. All the chromosomes of the population obtain a segment on virtual Roulette-wheel based on their fitness value, higher the fitness value bigger size of segment allotted to them, after then wheel is spinned. The chromosome corresponding to segment on which virtual Roulette-wheel stops, selected for crossover operation.

The average fitness value of the population $F_{p,k}$ for k^{th}generation is defined as follow:

$$F_{p,k} = \frac{\sum_{k=1}^{pop\_size} F_k}{pop\_size}, \text{ where p is} 1 \le p \le n \text{ and k is } 1 \le k \le pop\_size \quad (14)$$

Hence, the probability of selecting the K^{th} string

$$P_k(String\ K\ is\ chosen) = \frac{1}{F_K} = \frac{F_k}{\sum_{k=1}^{pop\_size} F_k}, \text{ where } F_K \text{is the fitness value of string k} \quad (15)$$

4.1.5. Crossover

In our proposed algorithm single-point crossover or uniform crossover (with swapping probability 0.7) is used. Point is chosen randomly based upon crossover rate after which parent chromosome exchanged their pattern shown in Figure 5.

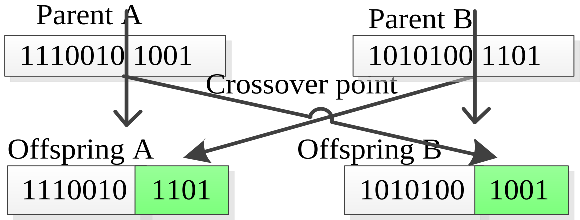

Figure 5. Single point crossover

4.1.6. Mutation

The process enables the search for optimal gene to convert ordinary node (bit ‘0’) into cluster head (bit ‘1’) and cluster head (bit ‘1’) into ordinary node (bit ‘0’) shown in Figure 6. The opposite case prevent from abnormal increases in the number of cluster heads in the network which fulfill our primary goal to selection of optimal number of cluster heads. After doing crossover and mutation, the position of cluster heads may be shifted.

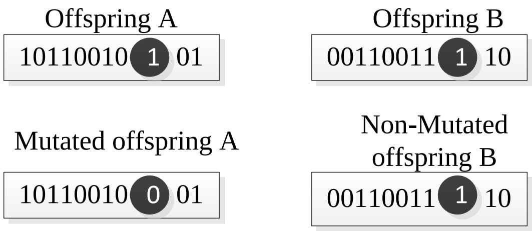

Figure 6. Mutation in offspring

4.1.7. Termination condition

In this step initially base station check the number of generation in the population and if the number of generation is more than maximum generation or fitness value of the offspring is uniform or converged for certain generation then genetic algorithm terminated and base station release the number of optimal cluster heads and their position based on highest fitness value chromosome.

4.2. Steady state phase/cluster binding phase

In this phase, each CHs broadcast join message in the network with by using CDMA mac protocol to reduce inter-cluster interference. Each node belong to only one CH, the node does not receive any join message declare itself as CH and send data directly to BS. The average radius [16] of cluster is evaluated as follow:

$$R_{min} = \sqrt{\frac{K*K}{\pi*n*c}} \tag{16}$$

There among, wireless sensor area of deployed nodes is represented by K*K, n is the total number of nodes and n*c is the number of cluster (c) formed. Generally clusters have larger radius than R_{min}. Cluster heads creates a TDMA schedule to avoid intra-cluster collision. This TDMA schedule is broadcast by each cluster heads, according to which member nodes turn on (wakeup mode) or off (sleep mode) their radio. The member nodes send their data directly to cluster head into allocated time slot only in wake up mode. The cluster heads gathers data from its member nodes according to TDMA schedule and applies data aggregation function to compress the data into single packet of fixed size.

4.3. Inter-cluster routing phase

If cluster head distance to base station is less than threshold distance TD_{max}, it transmits its data directly to base station otherwise cluster head select relay node (multi hop routing) from its candidate set. The candidate set S_{CH_i} of cluster head CH_i is the set of next forwarding neighboring node (CH) in the transmission range define as follow;

$$S_{CH_i} = \left\{CH_j \middle| d\left(CH_i, CH_j\right) \le \eta CH_i.TD_{max}, d(CH_i, BS) < d(CH_j, BS)\right\} \tag{17}$$

Whereas, $d\left(CH_i, CH_j\right)$ = distance between cluster head, $d(CH_i, BS)$ = distance between cluster head to base station and η is the minimum integer that $S_{CH}.CH_i$ has at least one node in the set to forward the data, if the value of η is ‘0’ that means candidate set have null value and CH_i send its packet directly to base station. At the start of process each CH broadcast a message (cluster IDs, Residual Energy, Traffic load and distance to base station). Initially all cluster heads have only one packet to transmit to BS but when it act as relay node for other CHs, they do not aggregate the other incoming packet with own packet into single packet because of data correlation between sensed data by different clusters are comparatively low. Thus the packet load on relay node is increases known as Traffic load.

Algorithm 1. Inter-cluster routing (A* (A-star) Algorithm)

1. Begin
2. Input: Graph $G(V, L)$, Source node CH_i, OPEN list (priority queue) say OL and CLOSE list say CL and predecessor of node V in set P.
3. Set P. CH_i = NIL // source node has no predecessor.
4. Set OL = ∅ and CL = ∅
5. ENQUEUE (OL , CH_i with $f(CH_i)$ attached)
6. *While (*OL≠ ∅) // Find the node with maximum $f(CH_i)$ in the OL, say it U.
7. U = DEQUEUE (OL)
8. Add (CL, U) and explore candidate set S_Q with $f(CH_i)$ attached to each node.
9. *For* each V ∈ G. S_{CH}[Q]
10. P.V = Q
11. *If* (U = destination node)
12. *Then* search is over, exits.
13. *Else*
14. ENQUEUE (OL , V with $f(v_i)$ attached)
15. RELAX (U, V, $f(CH_i)$,CL,OL)
16. Go to Step 6
17. Output the nodes from close list CL as shortest path from source to destination.
18. *END.*

Algorithm 2. RELAX (U, V, $f(CH_i)$)

1. Begin
2. *If* V = P.U // Explore node is predecessor of current node
3. *If* V = P.U // Explore node is predecessor of current node
4. *Else If* V. $f(v_i)$ > $U.f(v_i)$ + d (U, V)
   // choose the short-path based on $f(v_i)$
5. *Then* V. $f(v_i)$ = $U.f(v_i)$ + d (U, V)
6. V.P = U

4.3.1. Inter-cluster routing using A* (A-Star) algorithm

A-star search algorithm (tree-structure route) is used to find an optimal route from CH to BS applies on each cluster heads. Now CHs in network is modeled as a directed graph $G(V, L)$. Where V ($CH_1, CH_{2,} CH_3 \ldots\ldots\ldots CH_i \ldots\ldots CH_j$) is the set of cluster heads and L ($L \subseteq V^2$) is the set of links between cluster heads.The function for relay (tree) node selection is as follows;

$$f(CH_{Node}) = \frac{\alpha\ Residual\ Energy(RE_{Node})}{\beta\ (TL_{node})} + \frac{1}{\gamma(Distance\ to\ base\ station\ (DBs_{node}))} \quad (18)$$

Where α$(0 < \alpha \leq 1)$, β$(0 < \beta \leq 1)$, γ$(0 \leq \gamma \leq 1)$ is energy, traffic load and hop count coefficient respectively. If the value of α is '0' that means node have very less residual energy close to dead node, but in GEFLC selection of CH based upon the criteria that residual energy is more than average energy of all live nodes, thus dead node never selected for CH, otherwise the node have very high residual energy then value of α is '1'. As every CH have at least one packet to transmit to BS so value of β never touches to zero (i.e. β > 0) otherwise if the traffic load on the node is very high (i.e. 5) then value of β is '1'. Whereas value of γ is '0' refer that node either isolated node or very close to BS that send packet directly to BS otherwise if the value of γ is '1' indicates that node reaches to maximum hop count value (i.e. 5). The CH having more residual energy, less traffic load and minimum distance to base station selected as relay node from their neighboring set or candidate set. As a result node having largest value of $f(RL_{Node})$ is selected as relay node.

As A* (A-star) algorithms runs by each CH in the network creates tree-structure route from source to base station. An example is shown in Figure 7 where cluster head V_a runs A-star algorithm to find optimal route for transfer of data to base station. Inside the cluster head represent the residual energy and outside refer to traffic load or packet inside the buffer. Source node V_a Creates a route ($V_a - V_c\ V_g{-}V_n - BS$) with maximum residual energy 9, minimum traffic load (packet = 6) and minimum hop count 3 for transfer of data to base station. Other routes ($V_a - V_b - V_f - -V_n - BS$) have residual energy 9 but traffic load is 11 also hop count is increases up to 4. The Route ($V_a - V_d - V_k - -V_n - BS$) having residual energy 8 less than optimal path whereas traffic load and hop count is 10 and 3 respectively, which is not optimal so this path is also rejected.

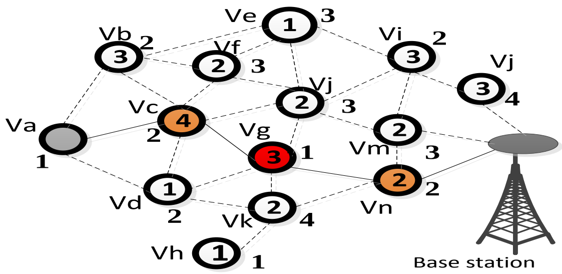

Figure 7. Multi-hop routing

5. SIMULATION AND RESULTS

The performance of our proposed algorithm GFELC evaluated with the help of MATLAB simulation tool. A number of experiment have done against both the approach Genetic based algorithm and Fuzzy logic based algorithm. Also, some experiment done towards inter clustering routing to show how much GFELC more efficient to deliver a packet over number of rounds. The simulation parameter is shown in Table 2.

Table 2. Simulation Parameter

Parameter	Value	Parameter	Value
Network size	100×100 m^2	TDMA frames per round	6
Number of nodes	100	Data packet size (m)	500 bytes
Transmission range of node	25 m	Header size	25 bytes
BS location	(50,175)	Bandwidth	1 mbps
Initial energy	0.5 J	Competition radius	25 m
$\mathcal{E}_{fs}$	10pJ/bit/m^2	Mutation rate	0.001
$\mathcal{E}_{mp}$	0.0013pJ/bit/m^2	Crossover rate	0.7
E_{elec}	50nJ/bit	Maximum generation	200
E_{DA}	5nJ/bit/message		

5.1. Comparison of number of alive nodes over rounds

In LEACH all sensor nodes dies around 800 rounds but in our algorithm lifetime in terms of rounds extended up to1000 rounds and also it does not decreases rapidly as LEACH, GFELC is more stable towards death of sensor nodes and decreases linearly until last node dies. Figure 8 shows that GFELC perform better with respect to CHEF, LEACH-ERE (Fuzzy logic based), GAEEP (genetic algorithm based) about 14%, 10%, 5% respectively in terms of rounds over number of alive sensor nodes.

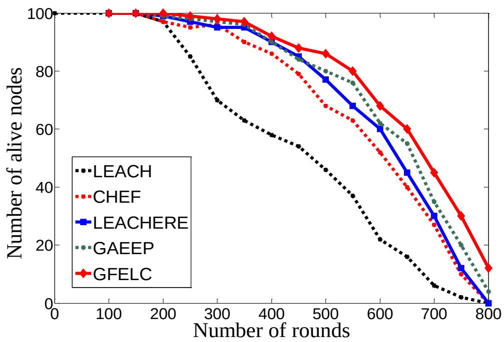

Figure 8. Network lifetime

5.2. Comparison of average residual energy of all sensor nodes over rounds

From the Figure 9, the average residual energy of all sensor nodes of the approach GFELC is more than LEACH, CHEF, LEACH-ERE and GAEEP. LEACH algorithm is poorest one having residual energy in unbalancing nature and it goes up to 800 rounds to death of last node. Whereas LEACH-ERE and GAEEP follow almost the same nature of distribution of residual energy. As a result, we can say that from figure that our algorithm GFELC performance in consuming energy among the nodes in uniform way that show it helps in the balancing the loads among nodes.

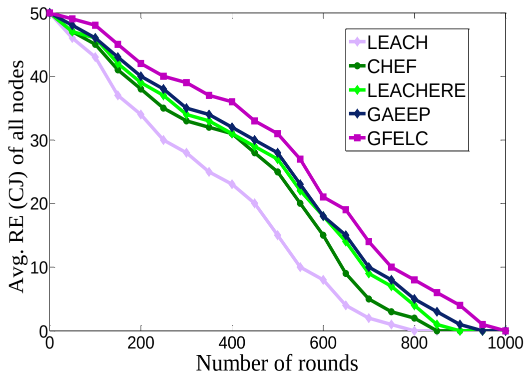

Figure 9. Avg. RE of all nodes per round

5.3. Node death percentage over rounds

First node death (FND), Half node death (HND) and Last node death (LND) are defined as number of rounds at which first node is dead, half of the nodes are died and last node died respectively. Figure 10 shows that FND for LEACH is occur in 114 round whereas for GFELC goes up to 220 rounds which shows twice the improvement in lifetime of network. The number of rounds at which HND for GFELC is also improves around 70% more than LEACH and finally when LND for GFELC reaches up to 940 rounds whereas LND for LEACH is takes place in 740th round. Figure 11 and Table 3 shows the general view of the node death percentage over increasing number of rounds. We can see from figure in algorithm GFELC node are died very slowly rate and runs for the long time (rounds) until last node died against LEACH, CHEF, LEACH-ERE, GAEEP shows lifetime of network increase.

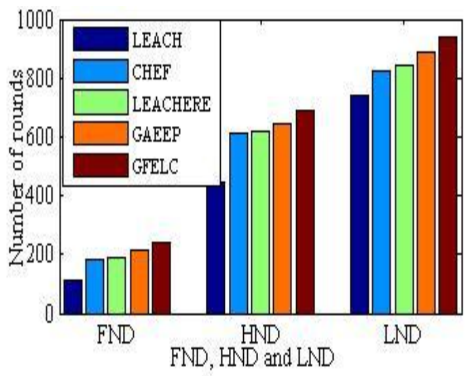

Figure 10. FND, HND, LND

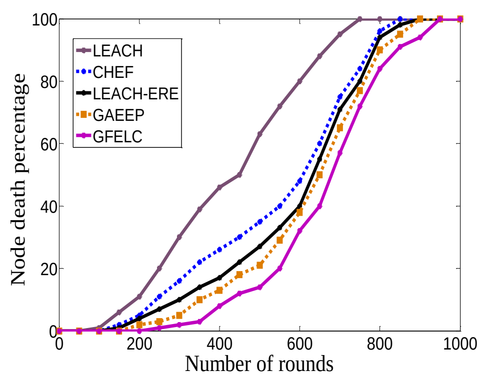

Figure 11. Node death percentage

Table 3. Node Death Percentage Up To 1000 Rounds

Death percentage	LEACH	CHEF	LEACH-ERE	GAEEP	GFELC
FND	114	184	206	214	220
20	272	358	478	514	598
40	386	556	598	605	678
HND (50)	447	610	620	645	690
60	512	648	667	710	766
80	690	710	733	798	830
LND (100)	740	825	840	890	940

5.4. Comparison of standard deviation of residual energy over rounds

As we know that standard deviation graph gives more precise view of how sensors nodes are spread towards mean. It visualizes us how much sensors nodes residual energy deviates from average residual energy.

As energy consumed in each round r by node i is given by

$$E_r i(m,d) = \left(\frac{N}{C} - 1\right) mE_{elec} + m\mathcal{E}_{amp}\, d^4_{CH-BS} + \frac{N}{C}\, mE_{DA} + mE_{elec}\ , \text{i} \in \text{CH (20)}$$
$$= mE_{elec} + m\mathcal{E}_{fs}\, d^4_{to\ CH}\ , \text{i} \in \text{non-CH} \quad (19)$$

The average consumed energy for round r:

$$\mu(E_{consumed}) = \frac{\sum_{i\,\epsilon\,\text{n}} E_r i(consumed)}{N} \quad 20)$$

The residual energy of (r+1)th round:

$$E_{r+1}(i)res = E_r(i)res - E_r i(consumed) \quad (21)$$

The average residual energy of rth round:

$$\mu\ (E_{res}) = \frac{1}{N}\sum_{i\,\epsilon\,\text{n}} E_r i(res) \quad (22)$$

The standard deviation of residual energy (square root of variance):

$$\sigma(E_{resdiual}) = \sqrt{\frac{1}{N}\sum_{i\,\epsilon\,\text{n}} [\mu(E_{residual}) - E_{res}(i)]^2} \quad (23)$$

From the Figure 12, we can see that standard deviation curve of LEACH is much narrower than other protocols, this show less number of nodes are close to mean on either side right or left. Whereas CHEF has significant amount of increase in standard deviation but our algorithm GFELC outperforms than CHEF, LEACH-ERE and GAEEP algorithms, where standard deviation of residual energy is bigger than all of them because of wider curve, which tell us that relatively more number of sensors nodes lies between one standard deviation. That show it balanced the load among nodes, and it increases the lifetime of network.

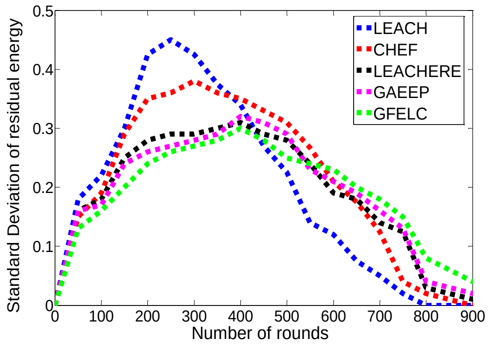

Figure 12. Standard deviation of RE

5.5. Distribution of clusters over rounds

Figure 13 show the variation of cluster head over number of rounds for each algorithm LEACH, CHEF, LEACH-ERE, GAEEP and GFELC. LEACH algorithm uses deterministic approach to select the node as cluster head so number of cluster head over rounds is vary in inconsistent manner over 730 rounds. GFELC algorithm produces the number of clusters more uniform way (on average 5.1) than other algorithms, the average number of clusters up to round 800 for LEACH, CHEF, LEACHERE and GAEEP are 3.8, 4.2, 4.3 and 4.2 respectively. Clearly, the GFELC provide higher network lifetime and increases stability period of network.

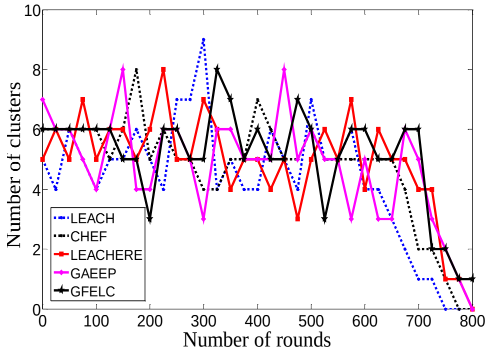

Figure 13. Number of clusters per round

5.6. Network throughput over rounds

The number of packets received per rounds by the base station is considered as network throughput. From the Figure 14 it is clear that GFELC received more number of packets through its lifetime maximum 11×10^4 packets over 800 rounds whereas LEACH, CHEF, LEACH-ERE, GAEEP receives maximum 5×10^4, 7×10^4, 9×10^4 and 9.7×10^4 packets respectively over 800 rounds. As our algorithm uses the best-first search, A-star algorithm for transfer the packets to base station using multi-hop routing between cluster heads, that systematic approach helps up in increasing the throughput of lifetime.

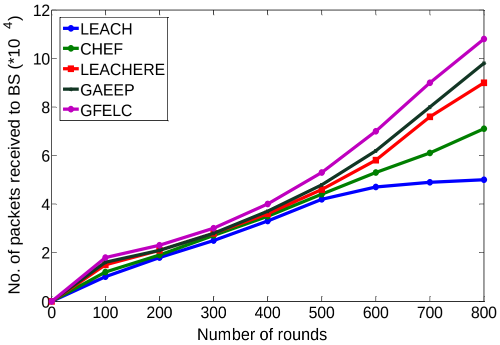

Figure 14. Network throughput

5.7. Average network residual energy over transmitted packets

Figure 15 shows the effectiveness of A-star routing on robustness of network. LEACH performs poor in energy consumption over transmissions in packets per rounds to base station. LEACH loses its energy completely over transmission of $10*10^4$ packets. whereas GFELC having maximum residual energy 27 cent joule over 10×10^4 packets are transmitted over base station.

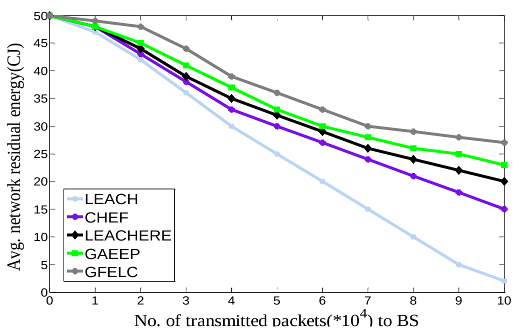

Figure 15. Avg. RE over transmitted packets

5.8. Packet drop ratio and Packet delivery ratio

Packet drop ratio defined as the ratio of total number of lost packet to total number of transmitted packets. Whereas Packet delivery ratio measures the success rate of algorithm to successfully deliver the number of packets. For a better algorithm, higher the packet delivery rate and lower the packet drop rate. Now, we calculate the packet drop ratio by using random uniformed model [30]; the average probability of packet drop (P_{dr}) updated dynamically based on the distance (d) between sensor nodes or sensor node to base station is calculated by Equation (20). If the probability of link is lower than P_{dr} then there is chance of packet loss; otherwise it will successfully receive by sensor nodes or base station.

$$P_{dr} = \begin{cases} 0\ if\ 0 \leq d < 25 \\ \left(\frac{1}{25}\right) * (d - 75), if\ 25 \leq d \leq 100 \\ 1\ if\ d > 100 \end{cases} \quad (26)$$

Figure 16 show the packet loss rate of LEACH, CHEF, LEACHERE, GAEEP and GFELC is 13.99, 5.76, 0.57, 0.32, 0.25 and 0.17 % respectively. Whereas Figure 17 shows the packet success rate of LEACH, CHEF, LEACHERE, GAEEP and GFELC is 86.01, 94.65, 96.82, 98.23 and 99.33 %. The simulation results proved that our algorithm GFELC have higher packet delivery ratio and lower packet drop ratio among others protocol. The decreasing rate of packet drop because of using A-star algorithm which balanced the congestion among clusters heads and also provide higher success rate in the network.

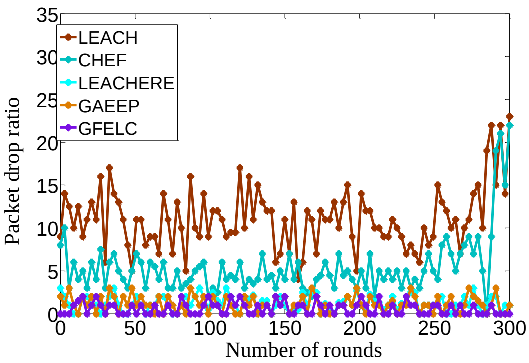

Figure 16. Packet drop ratio

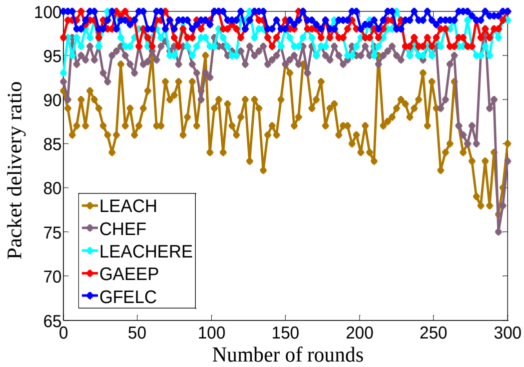

Figure 17. Packet delivery ratio

5 CONCLUSION

In this paper, we presented a load balanced algorithm (GFELC) using the hybrid technique of both genetic algorithm and fuzzy logic by selecting optimal number of CHs. Experimental result showed that proposed algorithm is outperforms in the area of Network lifetime over rounds, number of alive nodes over rounds, node death percentage, residual energy of nodes over rounds and network throughput in terms of number of packets deliver over rounds to the base station than other algorithms like LEACH, CHEF, LEACH-ERE and GAEEP. Moreover, our protocol GFELC utilizes A-star algorithm to deliver the packet to base station which improve the packet delivery ratio and provide more robustness as it reduces the packet drop ratio. Thus, overall GEFLC provided much better performance in the area, whether it is network lifetime, network throughput or other metrics also it well suited for both uniform and non-uniform environment.

REFERENCES

[1] Akyildiz I.F, Su W, Sankarasubramaniam Y, Cayirci E. "Wireless Sensor Networks: A Survey," *Computer Networks*, 38; 393– 422, 2002.
[2] Abbasi A. H, Mohamad Y.A. "Survey on Clustering Algorithms for Wireless Sensor Networks," *Computer Communications*,.30; 2826–284, 2007.
[3] Yick J, Mukherjee B, Ghosal D. "Wireless Sensor Network Survey" *Computer Networks, Elsevier*, 52; 2292-2330, 2008.
[4] Nguyen T T, Shieh C S, Horng M F, Ngo T G, Dao T D. "Unequal Clustering Formation Based on Bat Algorithm for Wireless Sensor Networks," *Knowledge and Systems Engineering, Springer*, 326; 667-678, 2015.

[5] Heinzelman W B, Chandrakasan A P, Balakrishnan H. "Energy-efficient Communication Protocol for Wireless Micro sensor Networks," in *Proceedings of the Annual Hawaii international conference on system sciences* 2000.
[6] Heinzelman W.B, Chandrakasan A P, Balakrishnan H. "An Application-Specific Protocol Architecture for Wireless Micro Sensor Networks," *IEEE Transactions on Wireless Communications*, 1(4); 660–670, 2002.
[7] Lindsey S, Raghavendra C S. "PEGASIS: Power Efficient Gathering in Sensor Information Systems", in: *Proceedings of the IEEE Aerospace Conference*, 3; 1125–1130, 2003.
[8] Younis O, Fahmy S. "HEED: Hybrid Energy-Efficient Distributed Clustering" Approach for Ad Hoc Sensor Networks," *IEEE Transaction on Mobile Computing*, 3; 366–379, 2004.
[9] Manjeshwar A, Agrawal D P. "TEEN: a Routing Protocol for Enhanced Efficiency in Wireless Sensor Networks," *Proceedings 15th International Parallel and Distributed Processing Symposium. IPDPS 2001*;. 2009–2015.
[10] Manjeshwar A, Agrawal D P. "APTEEN: a Hybrid Protocol for Efficient Routing and Comprehensive Information Retrieval in Wireless," *Proceedings 16th International Parallel and Distributed Processing Symposium*, Ft. Lauderdale, FL, 8, 2002.
[11] Manjeshwar A, Zeng A Q, Agrawal D P. "An Analytical Model for Information Retrieval in Wireless Sensor Networks Using Enhanced APTEEN Protocol," in *IEEE Transactions on Parallel and Distributed Systems*, 13(12); 1290-1302, 2002.
[12] Kuila P, Jana P.K. "Energy Efficient Load-Balanced Clustering Algorithm for Wireless Sensor Network," *ICCCS Procedia Technology*, 6; 771–777, 2012.
[13] Huruială P C, Urzică A, Gheorghe L. "Hierarchical Routing Protocol Based on Evolutionary Algorithms for Wireless Sensor Networks," *9th RoEduNet IEEE International Conference*, Sibiu, Romania, 387-392, 2010.
[14] Bari A, Wazwd S, Jaekel A, Bandyopadhyay. "A Genetic Algorithm Based Approach for Energy Efficient Routing in Two-Tiered Sensor Networks," *Ad Hoc Networks Elsevier*, 7(4); 665–676, 2009.
[15] Kuila P, Gupta, S K., Jana P K. "A Novel Evolutionary Approach for Load Balanced Clustering Problem for Wireless Sensor Networks", *Swarm and Evolutionary Computation*, Elsevier, 12; 48-56, 2013.
[16] Liu J L, Ravishankar C V. "LEACH-GA: Genetic Algorithm-Based Energy-Efficient Adaptive Clustering Protocol for Wireless Sensor Networks", *Int. J. Mach. Learn. Computation*, 1; 79–85, 2011.
[17] Zahhad M A, Ahmed S M., Sabor N, Sasaki S. "A New Energy Efficient Adaptive Clustering Protocol Based on Genetic Algorithm for Improving the Lifetime and the Stable Period of Wireless Sensor Networks," *International Journal of Energy, Information and Communications*, 5; 47-72, 2014.
[18] Indranil G, Riordan D, Sampalli S. "Cluster Head Election Using Fuzzy Logic for Wireless Sensor Networks," *Communication Networks and Services Research IEEE Conference, Proceedings*, 255–260, 2005.
[19] Kim J, Park S, Han Y, Chung T. "CHEF: Cluster Head Election mechanism using Fuzzy logic in Wireless Sensor Networks," *10th International Conference on Advanced Communication Technology, ICACT*, 654 – 659, 2008.
[20] Sharma T, Kumar B. "F-MCHEL: Fuzzy based Master Cluster Head Election Leach Protocol in Wireless Sensor Network," *Int. J. Comput. Sci. Telecommunication*, 3(10); 8–13, 2012.
[21] Nayak P, Anurag D, Bhargavi V V N A. "Fuzzy method based super cluster head election for wireless sensor network with mobile base station (FM-SCHM)," in *Proc. 2nd Int. Conf. Adv. Comput. Methodol.*, 422–427, 2013.
[22] Nayak P, Devulapalli A. "A Fuzzy Logic-Based Clustering Algorithm for WSN to Extend the Network Lifetime," in *IEEE Sensors Journal*, 16(1); 137-144, 2016.
[23] Ge R, Zhang H, Gong S. "Improving on LEACH Protocol of Wireless Sensor Networks Using Fuzzy Logic," *Journal of Information & Computational Science*, 3(7); 767-775, 2010.
[24] Shyan L J, Cheng W L. "Fuzzy-logic-based Clustering Approach for Wireless Sensor Networks Using Energy Predication," *Sensors Journal, IEEE*, 9(12); 2891-2897, 2012.
[25] Alshawi I S, Yan L, Pan W, Luo B. "Lifetime Enhancement in Wireless Sensor Networks Using Fuzzy Approach and A-Star Algorithm," in *IEEE Sensors Journal*, 10(12); 3010-3018, 2012.
[26] Saeedian E, Torshiz M N, Jalali M, Tadayon G, Tajari M M. "CFGA: Clustering Wireless Sensor Network Using Fuzzy Logic and Genetic Algorithm," in *IEEE Wireless Communications, Networking and Mobile Computing (WiCOM),7th International Conference on*, Wuhan, 1-4, 2011.
[27] X Yan, Q Wu, C Zhang, W Li, W Chen, W Luo, "An Improved Genetic Algorithm and Its Application," *Indonesian Journal of Electrical Engineering and computer science*, 10(5), 1081-1086, 2012.
[28] H Nasution, "Development of Fuzzy Logic Control for Vehicle Air Conditioning System," *Telkomnika Indonesian Journal of Electrical Engineering*, 6(2); 73-82, 2008.
[29] Rana K M, Zaveri M A. "ASEER: A Routing Method To Extend Life of Two-Tiered Wireless Sensor Network," *International Journal Smart Sensor Network System*, 2011; 11(2), 1-16.

BIOGRAPHIES OF AUTHORS

Pankaj Kumar Kashyap recevied the M.tech degree in Computer science and technology from Jawaharlal Nehru University, new delhi , in 2014. He is currently pursuing Ph.D from the same university. His research area interest is wirless sensor networks using machince learning approach, Internet of things or vehicles.

Sushil Kumar received the Ph.D. degree from Jawaharlal Nehru University, New Delhi, in 2014. He is currently Assistant Professor in JNU, new delhi. He is member of IEEE and published numerous research papers and conduct technical program all over the country. His research area is wireless sensor networks, internet of things, remote monitoring and control and hybrid system.