International Journal of Scientific Engineering and Research (IJSER)
www.ijser.in
ISSN (Online): 2347-3878, Impact Factor (2014): 3.05

GMMC: Gaussian Mixture Model Based Clustering Hierarchy Protocol in Wireless Sensor Network

Shaveta Gupta[1], Vinay Bhatia[2]

Baddi University of Emerging Sciences and Technology Baddi (HP), India

Abstract: ***Nowadays energy efficiency is a major issue for distributed wireless sensor networks (WSNs) deployed in varied environments. The energy efficient cluster-based protocols play a vital role for energy saving in hierarchical wireless sensor networks. In modelling, probability plays its important role. The probability definition is different for different approaches. The present paper is an approach to incorporate a model in the form of Gaussian Mixture Model (GMM) instead of an algorithm or a formula. The present approach seems to improve the performance when compared to existing algorithms with their results. Computer simulations are carried out to demonstrate and compare the performance of the proposed algorithms.***

Keywords: Clustering, Energy efficiency, GMM, Probability, WSN

1. Introduction

With the advent of WSN and its utilitarian practical applications, nowadays it is speedily introduced in researches to investigate the environmental conditions of remote and esoteric regions. WSN is the collection of wireless sensor nodes consists of small powered batteries with limited resources. These sensor nodes used to capture the data from the environment, process it and sends the data packets to the destination. While doing the process of reception and transmission large amount of the energy has been consumed. Due to limited resources of the power, energy consumption becomes a major issue for WSN design. At the same time number of energy efficient routing algorithms have been developed to minimize the energy consumption and enhance the lifetime of sensor network [1].

Clustering is one of the important hierarchical techniques used for enhancing the energy efficiency where the load has been distributed between the sensor nodes to increase the lifetime of the network. In clustering, the region of the network is divided into groups consisting of sensor nodes. The sensor nodes form the clusters on the basis of their position from the cluster head (CH) where CH are formed on the basis of distributed algorithm. After clustering and CH selection, the sensed data transmitted by the sensor nodes to the CH where processing of the data takes place and finally data packets transmitted to the sink/base station (BS). The transmission of data from CH to the BS is either by single-hop or multi-hop communication. In single-hop communication the CH sends data packets directly to the BS but in this case energy consumption is high if the distance between the two is large. On the other hand in multi-hop communication the CH far from the BS uses the intermediate CHs for data transmission but energy consumption the CH close to the BS become very high.

Different optimisation techniques have been used to improve the energy efficiency and enhance the network lifetime. Low Energy Adaptive Clustering Hierarchy (LEACH) is the first optimized protocol used to distribute the energy uniformly by selecting CH on the basis of round robin [2]. In this protocol few nodes in a cluster are randomly selected with a certain/fixed probability to become CHs per round. It is also assumed that all the nodes have same amount of energy in every round and all the nodes are electing themselves to become CH with a constant predesigned probability.

LEACH-C is a centralized approach where BS has the energy level and distance information of all the nodes. The probability of CH selected is fixed by the BS [2]. In a Stable Election Protocol (SEP) weighted election probabilities based on residual energy of each node decides the node to become CH [3]. In Distributed Energy Efficient Clustering (DEEC), CHs are formed on the basis of the ratio between residual energy of each node and the average energy of the network [4]. There are so many protocols used to save the energy of the nodes to enhance the network lifetime.

The existing protocols initially at the first most round assume that all the nodes in a cluster have same probability to become CH and as the rounds proceeds, probability varies. But practically the nodes could not have same probability because of various reasons. First it is not necessary that all the nodes in a cluster are in range with the BS which may cause more energy consumption due to longer distance between CH and the BS. Secondly some nodes might be at the boundary of the cluster which may cause high intra-cluster energy consumption. Thirdly the nodes might have different energy levels depending upon different power sources that may cause different probability for each node to become CH.

Above mentioned reasons imply that the probability of all the nodes initially to become CH should not be same. In this paper we proposed a novel method in which the Gaussian mixture models (GMM) are often used for data clustering and clusters are formed by selecting the component that maximizes the posterior probability which indicate that each data point has some probability of belonging to each cluster. That is why clustering done by GMM is also called soft clustering.

This paper is organised as follows: Section 2 describes energy model for the network, Section 3 describes GMM approach, Section 4 evaluates the simulation results and Section 5 offers some conclusion along with some future research.

2. Energy Model

The energy model used is a simple first order model where the energy consumption is calculated for the transmission of data in the form of bits from transmitter to the receiver. This model used to calculate the amount of energy consumed for transferring the data from simple node to CH, from CH to intermediate CH, from CH to BS and also from node to BS. The radio dissipation energy model consists of transmitter having transmit electronics (E_{elec}) which depends upon factors like coding, modulation, filtering and transmit the signal and amplifier depends on the distance to the receiver and the tolerable bit-error rate.

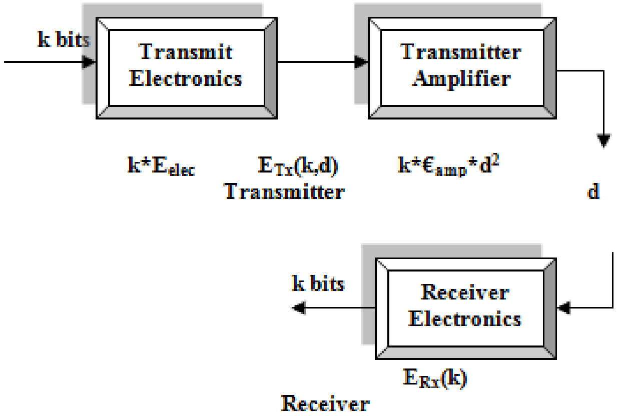

Figure 1: Energy Model

If the distance between transmitter and receiver is less than threshold distance (say d_o) then free space (d^2 power loss) channel model used and if distance between transmitter and receiver is greater than threshold distance (say d_o) then multi path fading (d^4 power loss) channel model used. If the distance between transmitter and receiver is less than threshold distance (say d_o) then free space (d^2 power loss) channel model used and if distance between transmitter and receiver is greater than threshold distance (say d_o) then multi path fading (d^4 power loss) channel model used [5].

The energy consumed by the specific nodes/CH for transmitting k bits of data is:

Energy consumed by transmitter (for $d<d_o$)

$$do = \sqrt{\left(E_{fs} \setminus E_{mp} \right)} \quad (1)$$

$$E_{tx}(k,\ d) \ = E_{elec} * k \ + k * (E_{fs} * d^2) \quad (2)$$

Transmission energy for intermediate node

$$E_{tx}(k,\ d) \ = ((E_{elec} + E_{DA}) * k) \ + \ (E_{fs} * k * d^2)$$
(3)

Energy consumed by transmitter (for $d \geq d_o$)

$$E_{tx}(k,\ d) \ = E_{elec} * k + \ k * (E_{mp} * d^4) \qquad (4)$$

Transmission energy for intermediate node

$$E_{tx}(k,\ d) \ = ((E_{elec} + E_{DA}) * k) \ + \ (E_{mp} * k * d^4)$$
(5)

Energy consumed by Receiver

$$E_{rx}(k) \ = E_{elec} * k \quad (6)$$

Table 1 contains first order radio model parameter used to calculate the energy consumed by each node in a cluster at various distances.

Table 1: Radio Parameters

Parameters	*Operation*	*Values*
Transmitter / Receiver Electronics	E_{elec}	50 nJ/bit
Transmit amplifier (if d to BS<do)	E_{fs}	10 pJ/bit/4m^2
Transmit amplifier (if d to BS>do)	E_{mp}	0.0013 pJ/bit/m^4
Data aggregation energy	E_{DA}	5 nJ/bit/signal

3. GMM Approach

GMM is a probability model of mixture of Gaussians. It seems to be one which is closer to the natural distribution and an easy model to do mathematical manipulation for having Gaussian function. If the distribution is not Gaussian in nature then there are different methods that can be used to form the Gaussian like clusters having Gaussian distribution and this analysis is based on GMM. The univariate Gaussian distribution (or "normal distribution," or "bell curve") is the distribution in which the result is the average of the events occurs again and again.

$$G(x|\mu,\sigma) = \frac{1}{\sqrt{(2\pi\sigma^2)}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (7)$$

where G is the Gaussian function, μ is the mean and σ^2 is the variance. The mean (μ) indicates the maximum likelihood and variance (σ^2) is the deviation from the maximum likelihood with in a univariate Gaussian distribution field. While multivariate Gaussian distribution is a generalization of the univariate normal with two or more variables [6], [7]. It is parameterized with a mean vector μ and a covariance matrix Σ.

$$N(x|\mu,\Sigma) = \frac{1}{\sqrt{(2\pi|\Sigma|)}} e^{[-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)]} \quad (8)$$

The mean (μ) indicates the maximum likelihood and Σ is the covariance between different Gaussian distribution fields [8]. Taking the log of equation 8, the analysis can be simple.

$$\ln p(x|\mu,\Sigma) = -\frac{1}{2}\ln(2\Pi) - \frac{1}{2}\ln|\Sigma| - \frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)$$
(9)

It is simply a single Gaussian distribution where then we can set the derivative of ln p(x|μ,∑) to zero

$$\frac{\delta \ln p(x|\mu,\Sigma)}{\delta\mu} = 0 \quad (10)$$

$$\frac{\delta \ln p(x|\mu,\Sigma)}{\delta\Sigma} = 0 \quad (11)$$

Solve directly for μ and ∑.

$$\mu = \frac{1}{N}\left(\sum_{n=1}^{N} x_n\right) \quad (12)$$

$$\Sigma = \frac{1}{N}\sum_{n=1}^{N}(x_n - \mu)(x_n - \mu)^T \quad (13)$$

However, it is impossible for various practical problems to find such analytical expressions and there is requirement of more elaborate techniques.

GMM with EM Algorithm

The GMM is the collection of mixture of the multivariate Gaussian distribution. It is an appropriate model for the clusters having different size and they have correlation with each other. The clustering in GMM evaluates the corresponding posterior probabilities for each node and describes that how each node relates to each cluster i.e. the mean of the corresponding nodes [9]. The clustering done by GMM also called soft clustering where nodes are not restricted to one cluster only.

It uses an iterative optimization technique which is operated locally. For maximum likelihood an Expectation–Maximization (EM) optimization technique is used. The mixing coefficient is considered as the prior probabilities of the data and for a given number of the clusters the corresponding posterior probabilities can be calculated by

$$p(x) = \sum_{k=1}^{K}\Pi_k N(x_k|\mu_k,\Sigma_k) \quad (14)$$

where Π_k is the mixing coefficient i.e. the weightage of each Gaussian distribution and K is the number of clusters.
A mixture model with maximum likelihood has the following advantages:

1. The data in a cluster are firm i.e. node distributions have high peaks or mean.
2. Choice of selecting the cluster for the nodes.
3. The node density can be calculated.
4. Soft clustering is used.
5. The unobserved data can be iterated.

The EM used in GMM has two important steps:

Estimation (E) step: for given parameter values we can compute the expected values of the latent variables.

Maximization (M) step: update/maximize the joint distribution of the data and the hidden variables.

The EM algorithm of GMM comprises of following steps:

Given a GMM, the goal is to maximize the likelihood function with respect to the parameters comprising the mean and covariance of the components and the mixing coefficients.

1. Initialize the mean μ_i, covariance Σ_i and mixing coefficient Π_i and evaluate the initial value of the log likelihood.
2. E-Step: Evaluate the posterior probability/responsibilities using the current parameter values

$$\gamma_k(x) = \frac{\pi_k N(x|\mu_k \Sigma_k)}{\sum_{j=1}^{K}\pi_j N(x|\mu_j \Sigma_j)} \quad (15)$$

3. M-Step: Re-estimate the parameters using the current responsibilities

$$\mu_j = \frac{\sum_{n=1}^{N}\gamma_j(x_n)x_n}{\sum_{n=1}^{N}\gamma_j(x_n)} \quad (16)$$

$$\Sigma_j = \frac{\sum_{n=1}^{N}\gamma_j(x_n)(x_n - \mu_j)(x_n - \mu_j)^T}{\sum_{n=1}^{N}\gamma_j(x_n)} \quad (17)$$

$$\pi_j = \frac{1}{N}\sum_{n=1}^{N}\gamma_j(x_n) \quad (18)$$

4. Evaluate log-likelihood

$$\ln p(x|\mu,\Sigma,\pi) = \sum_{n=1}^{N}\ln\{\sum_{k=1}^{K}\pi_k N(x_n|\mu_k,\Sigma_k)\} \quad (19)$$

if there is no convergence, repeat step 2.

This process evaluates the varying probabilities for the clusters [6].

Cluster Head Formation

After the evaluation of the probabilities, the nodes select themselves to the CH on the basis of threshold value , $T_{s(i)}$. . The sensor node chooses a random number r between 0 and 1. The node becomes CH for that current round, if the value of r is less than the threshold value, $T_{s(i)}$. The threshold value is evaluated by:

ISSN (Online): 2347-3878, Impact Factor (2014): 3.05

$$T_{s(i)} = \begin{cases} \dfrac{p(i)}{1 - p(i)\left(r \bmod \dfrac{1}{p(i)}\right)} & \text{if } s(i) \in G \\ 0 & \text{otherwise} \end{cases} \quad (20)$$

where is the set of nodes that have not been CH in the last $1/p_i$ rounds.

Cluster Formation

After the selection of CHs, the CH nodes broadcast an advertisement (ADV) message to the whole network using non-persistent carrier sense multiple access (CSMA) MAC protocol. This small message contains the CH's ID and a header which specifies the type of the message. After receiving the message, the non-CH nodes choose the CH on the basis of minimum communication distance by calculating the signal strength of the ADV message from different CHs.

After selecting the CH by the non-CH nodes they send a join-request (JOIN-REQ) to their corresponding CH for which they form the cluster. Again the join message is a short message consisting of the CH node's ID, non-CH node's ID and the header. The clustering done in this algorithm is a soft clustering where the nodes are not restricted to a cluster but in every round the clusters changes dynamically.

Data Transmission

The data then sensed by the environment and transmits to the CH where the data aggregated and processed so that only useful data get transmitted to the BS and small data transmission must preserves the energy. The CHs close to the BS uses single hop or direct transmission where as the CHs far away from the BS uses multi-hop transmission/communication i.e. they transmit there data to the next CH close to the BS and soon.

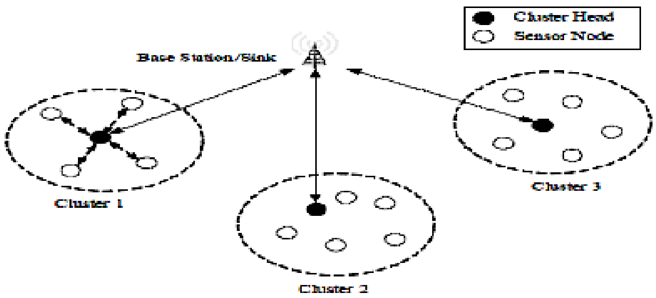

Figure 2: Single-hop Communication

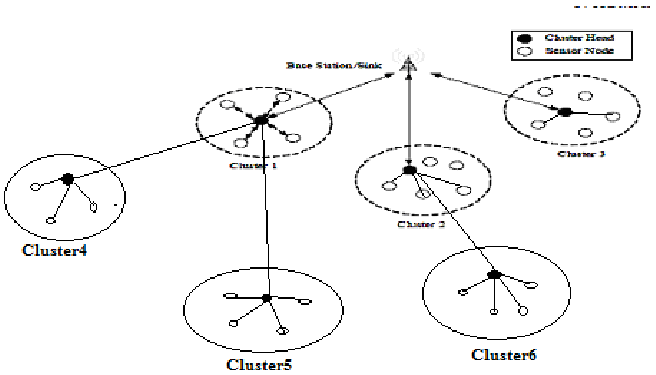

Figure 3: Multi-hop Communication

4. Simulation Results

The simulation is performed on MATLAB version 7.9.0.529 with intel (R) Core (™) 2 Duo CPU with 2GB RAM. The results are as follows:

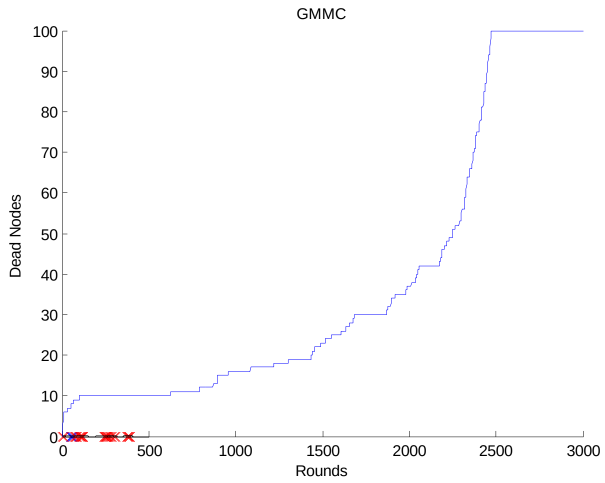

Figure 4: Stability

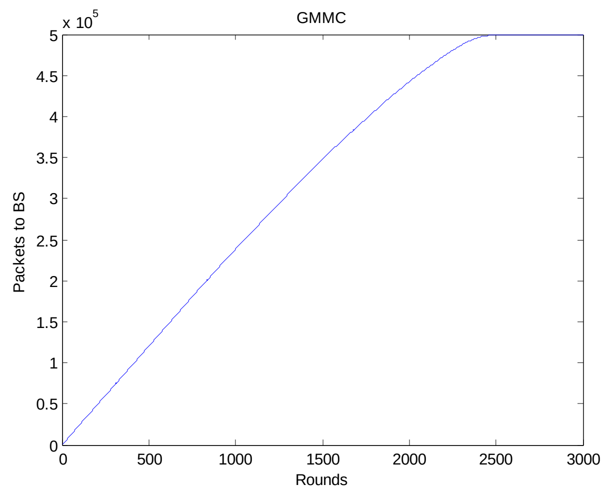

Figure 5: Packets to BS

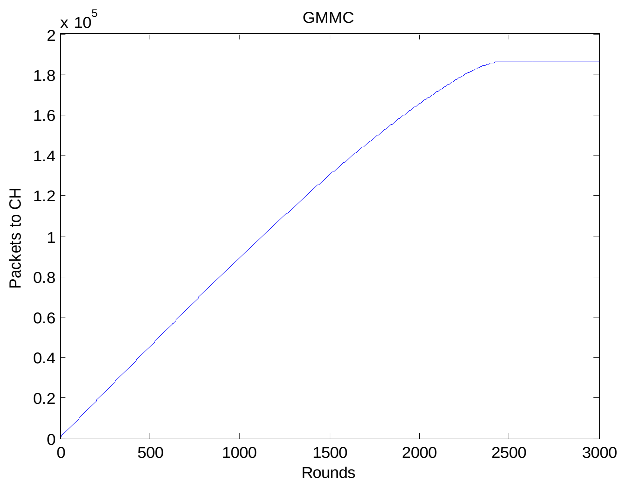

Figure 6: Packets transmitted to CHs

Table 2: Results

S. No	*Parameters*	*GMMC*
1	All Node Dead	2450th round
2	Packets send to BS	$5x10^5$
3	Packets send to CH	$18x10^4$

5. Comparison with Established Model

In this method, our effort was to enhance the energy efficiency of the network. We focused on the lifetime of the network by enhancing the overall lifetime of the nodes. We also took into account the number of CHs formed, number of packets transmitted from node to CH and from CH to the BS. The evaluation of GMM is performed on MATLAB. In our opinion the proposed model enhances the lifetime and energy efficiency of the network. The performance of the GMM is compared with the LEACH protocol.

We consider a wireless network with 100 nodes distributed randomly within a region of 100m x100m and assume that the BS is at the centre of the region. The number of rounds considered to be 3000. Various factors used to be compared between GMM and LEACH protocols.

Stability

It is not necessary that initially the probability of the nodes to become CH is same. In GMM, our approach was to evaluate the probabilities of the nodes on the basis of certain criteria whereas in LEACH all the nodes assumed to have equal probability.

Figure7 shows that by varying the probability in the initial stage of the network cause an overall enhancement of the network. In case of LEACH, the last node dead at 1100th round while in cadse of GMMC the last node dead at 2450th round. This approach shows that the last node dead in GMMC is 1350 rounds more than the LEACH protocol.

Considering the varying probability into account the overall stability of the network has been enhanced.

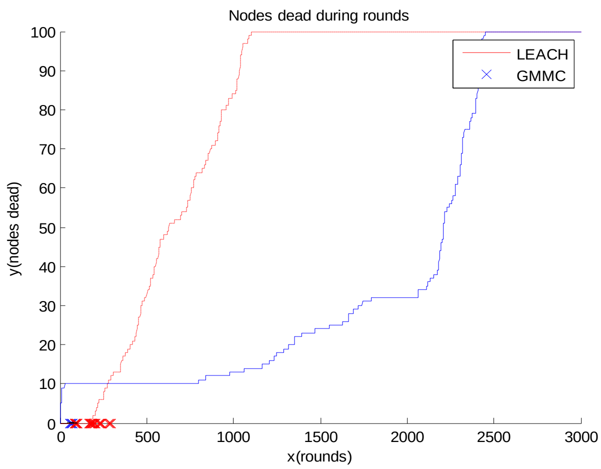

Figure 7: Comparison of the Stability

Packets to BS

Figure 8 shows that the number of data packets transmitted from the CHs to BS is more in case of GMMC as compared to the LEACH protocol. The number of packets transmitted by the GMMC is approx. $5x10^5$ while in case of LEACH it is approx. 7300. This shows that GMMC transmits $49x10^4$ data packets more from CHs to BS then the LEACH.

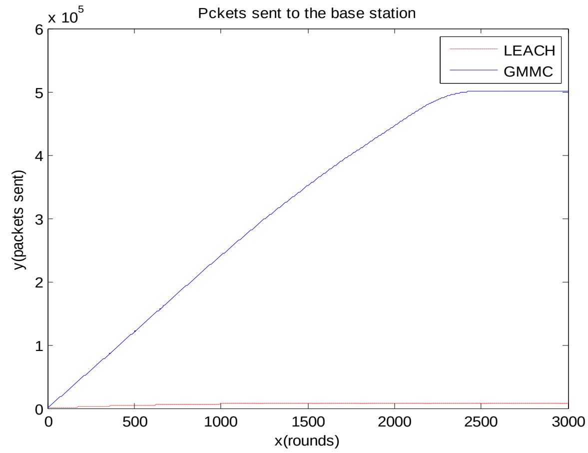

Figure 8: Comparison of the packets sent to the BS

Packets to CHs

Figure 9 shows that the number of data packets transmitted from the non-CH nodes to the CHs is more in case of GMMC as compared to the LEACH protocol. The number of packets transmitted by the GMMC is approx. $18x10^4$ while in case of LEACH it is approx. $58x10^3$. This shows that GMMC transmits $12x10^4$ data packets more from non-CH nodes to the CHs then the LEACH.

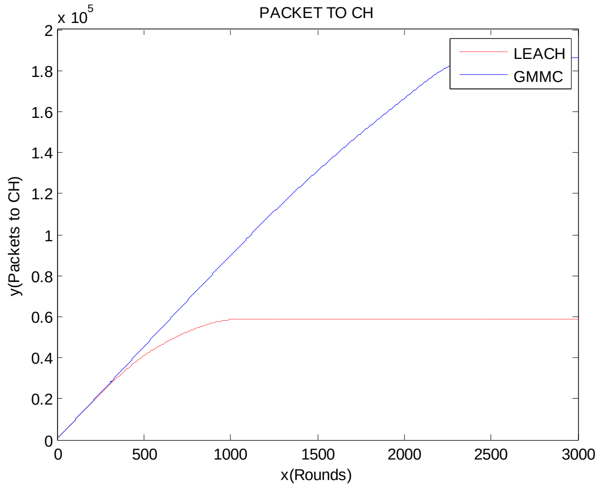

Figure 9: Comparison of the packets sent to the CHs

CH Formation

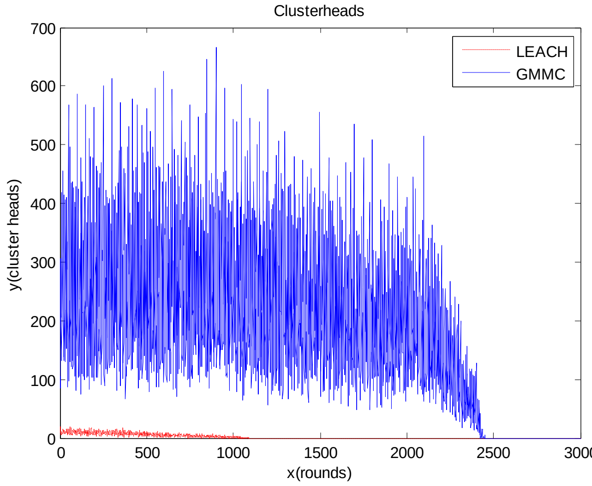

Figure 10: Comparison of the CHs formation

Figure 10 shows the comparison of CH formation between the GMMC and the LEACH protocol. The CHs formed in the GMMC is much more than the formation in LEACH protocol.

Number of Nodes Alive

Fig.11 shows that the last node alive in case of GMMC is approx. at 2400^{th} round while in case of LEACH protocol last node remains alive at 1100^{th} round.

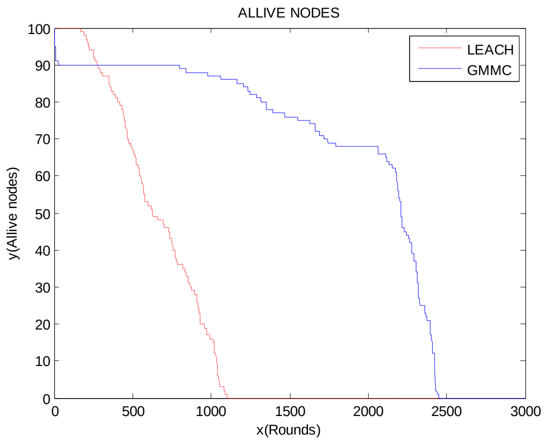

Figure 11: Comparison of the last node alive

6. Conclusion

The work is an attempt to improve Energy efficiency of wireless system by incorporating a model into a wireless sensor system. Discussing design issues and taking practical constraints into account, the attempt is made to develop application oriented system to increase Energy Efficiency and improve the performance of Wireless Sensor network.

With the application of Gaussian Mixture Model into the wireless sensor network protocol the system becomes more practical. The results establish that there is improvement in Energy performance of system.

The work is compared with established protocols in terms of results for various parameters in order to establish model based approach to the protocol architecture practically .With better results, practical wireless based applications will be benefited by incorporation of this model and will provide improvement in the system performance facing acute Energy constraints.

References

[1] A. Nikolaos Pantazis, A. Stefanos Nikolidakis and D. Dimitrios Vergados, “Energy-Efficient Routing Protocols in Wireless Sensor Networks: A Survey” IEEE communication survey & tutorials, VOL. 15, No.2, secondquarter2013.

[2] W. 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.

[3] Georgios Smaragdakis and Ibrahim Matta, “SEP:Stable Election Protocol for Clustered Heterogeneous Wireless Sensor Networks” 2004.

[4] L. Qing, Q. Zhu, M. Wang, “Design of a distributed energy-efficient clustering algorithm for heterogeneous wireless sensor networks”,In ELSEVIER, Computer Communications 2006.

[5] T. Rappaport, Wireless Communications: “Principles & Practice. Englewood Cliffs”, NJ: Prentice-Hall, 1996.

[6] Reynolds, D.A.: “A Gaussian Mixture Modeling Approach to Text-Independent Speaker Identification”. PhD thesis, Georgia Institute of Technology (1992).

[7] Jeff A. Bilmes,” A Gentle Tutorial of the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models” International Computer Science Institute

[8] Gaussian mixture models and the EM algorithm Ramesh Sridharan.

[9] Reynolds, D.A., Rose, R.C.: Robust Text-Independent Speaker Identification using Gaussian Mixture Speaker Models. IEEE Transactions on Acoustics, Speech, and Signal Processing **3**(1) (1995) 72–83

[10] Reynolds, D.A.: A Gaussian Mixture Modeling Approach toText-Independent Speaker Identification. PhD thesis, Georgia Institute of Technology (1992)

Author Profile

Shaveta Gupta received the B.E. degrees in Electronics and Communication Engineering from Government College of Engineering and Technology, Jammu in the year 2003. During 2005-2010, she stayed in IIIM Jammu. From 2010 onwards she is working in BGSB University and at present pursuing M.Tech in Electronics and Communication from Baddi University of Emerging Sciences and Technology Baddi (HP) University.

Dr. Vinay Bhatia received his PhD from MM University in the year 2014 and B.Tech. and M.Tech. degrees from Kurukshetra University and Punjab Technical University in the year 2002 and 2006, respectively. At present he is working as Associate Professor and Head of Department, Electronics and Communication, Baddi University of Emerging Sciences and Technology Baddi (HP) University.