Research Progress Report

Botao Zhu

May 13, 2019

1   Reading and Research Activities

1.1   Reading Summary

1.1.1   Energy-efficient modified LEACH protocol for IoT application

- **Algorithm description**

  [1] extends the application of LEACH to IoT for the first time, which modifies the original LEACH protocol to maintain uniform energy throughout the network.

  Firstly, it proposes a new LEACH algorithm with hard and soft threshold strategies to select cluster heard (CH) effectively. According to conventional LEACH protocol, a CH is not eligible for selection of CH for the next $1/P$ rounds if it had been the CH. However, in some cases, a CH maybe does not consume much energy in this round and it still can remain CH in the next rounds. This paper solves the above problem by setting a threshold value, which means any CH with more energy than the threshold value can continue to be CH with the same cluster for the next round. In this way, the energy cost in the phase of set-up for a new CH selection can be saved in each round, because there is no need to send information to re-select the CH and notify members to join the cluster. Assuming there are n nodes in a WSN, the energy of CH replacement process can be denoted by P_{HR} during the set-up phase in each round

  $$P_{\mathrm{HR}} = (P_{k\mathrm{Tx}}P_{\mathrm{Tx}} + P_{k\mathrm{Rx}}P_{\mathrm{Rx}}\,(nC - 1))\,RnC \tag{1}$$

  where C is the percentage of clusters, R is the count of CH replacement, $P_{k\mathrm{Tx}}$ and $P_{k\mathrm{Rx}}$ are the transmitted and received packet sizes respectively, P_{Tx} and P_{Rx} are the energy cost of transmission and reception of 1 byte data. The total energy of each cluster

  $$P_{\mathrm{WEC}} = E_{\mathrm{Init}} \times nC \tag{2}$$

  where P_{WEC} is the initial energy of each node. The energy consumption of each cluster 'i' during each round, denoting $N = nC$

  $$P_{\mathrm{HR}}\,(i) = ((N_i - 1)\,P_{k\mathrm{Tx}}P_{k\mathrm{Rx}} \times P_{k\mathrm{Rx}}P_{\mathrm{Rx}}) + ((N_i - 1)\,P_{k\mathrm{Rx}}P_{Rx} + (N_i - 1)\,P_{k\mathrm{Tx}}P_{\mathrm{Tx}}) \tag{3}$$

  The number of rounds can be calculated by

  $$\mathrm{Count}_{Rnd} = \frac{P_{HR}}{P_{WEC}} \times 100 \tag{4}$$

  So, the minimum level of energy can be estimated by the following equation, which can be considered as the threshold value

  $$P_{\mathrm{Th}} = \mathrm{Count}_{Rnd}\,(P_{k\mathrm{Tx}} + P_{k\mathrm{Rx}})\,P_{\mathrm{Tx}} \tag{5}$$

  Secondly, because the energy requirement of intra-cluster transmission and inter-cluster transmission is different, a CH uses a high energy amplification level. When it becomes a common node in the following rounds, it will choose a low power level. Fig. 1 shows the whole algorithm flowchart.

- **Simulation analysis**

  Fig. 2(a) shows that the average residual energy of LEACH comes to zero after 500 rounds, however, the average residual energy of IoT LEACH comes to zero after around 1250 rounds. In Fig. 2(b), the throughput of IoT LEACH is much higher than that of LEACH. After completion of approximately only 750 rounds, the number of active nodes goes completely to 0 for LEACH, whereas with IoT LEACH, some nodes are still alive till 1500 rounds, shown in Fig. 3.
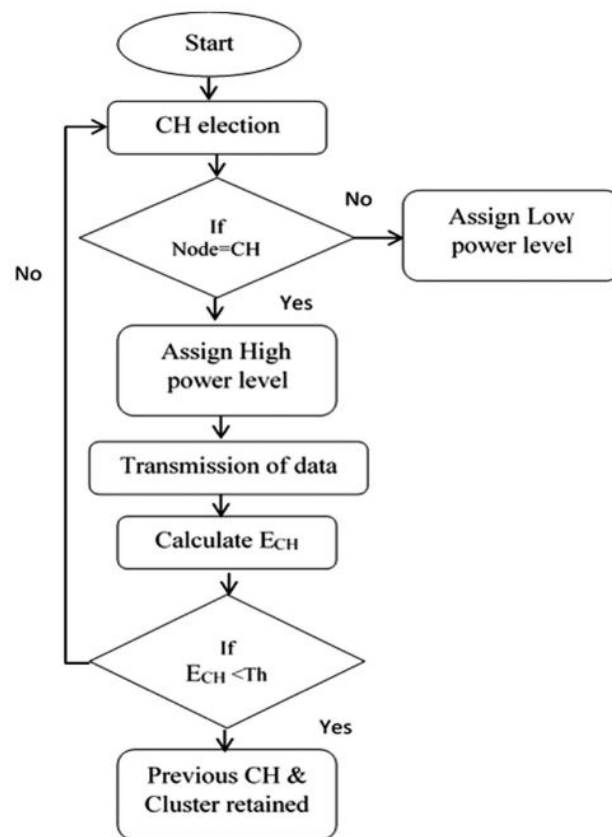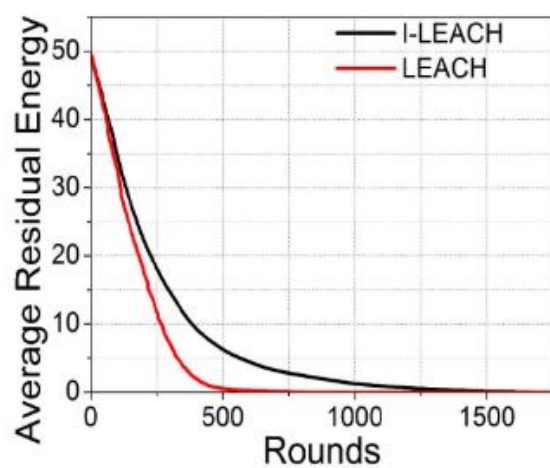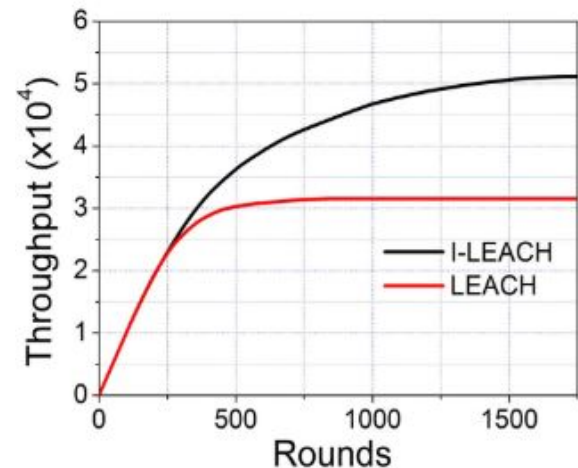
Figure 1: IoT LEACH algorithm

(a) Average residual energy

(b) Throughput

Figure 2: Network performance

1.1.2 Energy efficient clustering protocol based on kmeans-midpoint algorithm for enhanced network lifetime in wireless sensor network

- **Algorithm description**
  [2] considers residual energy as the parameter in addition to euclidean distance used in basic K-means algorithm for approximate CH selection. It mainly has three contributions:
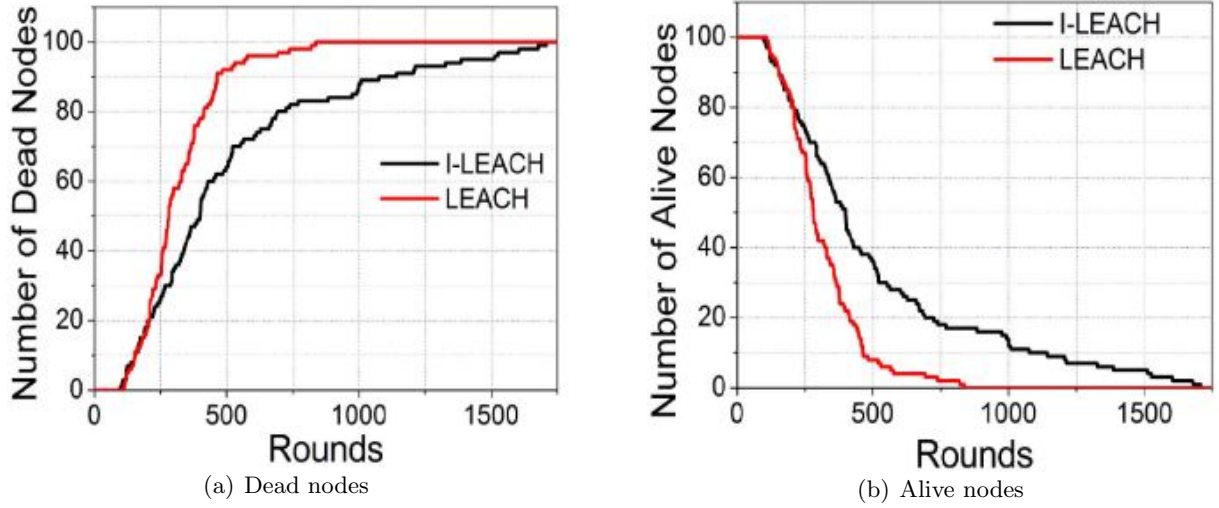
(a) Dead nodes

(b) Alive nodes

Figure 3: Lifetime metrics

1. It gives the optimum number of desired clusters based on the size of region and the number of nodes.

$$k_{opt} = \frac{\sqrt{N}}{\sqrt{2\pi}} \sqrt{\frac{\varepsilon_{fs}}{\varepsilon_{mp}}} \frac{M}{d_{BS}^2} \quad (6)$$

where M is the length and width of region, N is the total number of nodes, d_{BS} is the distance from CH to BS, ε_{fs} is the parameter for free model and ε_{mp} is the parameter for multipath model.

2. For CHs selection, midpoint method is used in order to balance the number of cluster and the load of CHs, which prolongs the network lifetime.

Algorithm 1: Midpoint algorithm for initial CH selection

Input:

D = set of n data points

k_{opt} = optimum number of desired clusters

Output:

k_{opt} number of initial centroids

  1: Compare the distance from origin for each data point

  2: Sort the distance obtained in step 1, sort the data points in accordance with the distance

  3: Partition the sorted data points into k_{opt} equal sets

  4: In each set, take the middle point as the initial centroid

---

3. The residual energy of node is considered as the parameter of CH selection in addition to the euclidean distance used in basic kmeans. To maintain the connectivity of the network, residual energy of the CH is checked every round. If the node's residual energy is less than the threshold value, it cannot be selected as CH. The threshold energy is given by

$$E_{threshold} = lE_{elec}\left(\frac{N}{k_{opt}} - 1\right) + lE_{DA}\frac{N}{k_{opt}} + lE_{elec} + l\varepsilon_{fs}d_{toBS}^2 \quad (7)$$

The proposed algorithm has three phases:

Phase 1: initial CH selection, shown in algorithm 1.

Phase 2: balanced cluster formation.

Algorithm 2: Balanced cluster formation

Input:

D = set of n data points

k_{opt} = optimum number of desired clusters

$E_{threshold}$ = threshold energy

Output:

k_{opt} A set of k_{opt} clusters

1: Apply midpoint method in algorithm 1 to choose k_{opt} out of D sensor nodes as initial cluster heads;

2: Repeat

3: Each of the remanining nodes decides to join its nearest CH according to the euclidean distance;

4: Centroid of each cluster is calculated as

$$Centroid\,(X,Y) = \left(\frac{1}{s}\sum_{i=1}^{s} x_i, \frac{1}{s}\sum_{i=1}^{s} y_i\right)$$

5: After cluster formation, based on the distance from the centroid, and ID number is allotted to each node of a cluster, assigning smaller number to the closer one

6: **for** *all selected cluster heads* **do**

   if *Residual energy of cluster head¿$E_{threshold}$* **then**

   | The node will remain as cluster head

   else

   | Check ID number of all nodes in that cluster.

   | The node in the next order of ID number is selected as a new CH

   end

   end

7: The newly selected CHs inform other nodes about the CHs change.

8: Until the cluster heads are not changed any more.

Phase 3: Data communication.

Algorithm 3: Data communication

Input:

D = set of n data points

$\left(CH_1, CH_2, ..., CH_{k_{opt}}\right)$ = a set of k_{opt} clusters

$d_{threshold}$ = threshold distance

1: Nodes send data packet to their CHs.

2: Calculate the distance between each elected CH and BS(d_{BS}).

3: **if** $d_{BS} < d_{threshold}$ **then**

   | Cluster head directly commuication to the BS

   else

   | It selects the nearest neighbour cluster head whose d_{BS} is less than $d_{threshold}$ to communicate to the BS.

   end

- **Simulation analysis**

  Fig.4(a) shows uneven number of nodes by using basic kmeans, CH of the heavily loaded cluster with 39 nodes will be exhausted much earlier than the other clusters. After applying the proposed approach, a particular cluster contains maximum 29 nodes and minimum 21 nodes, which is much closer to the average number of nodes.

  The number of nodes which are alive using LEACH-B, BPK-means, basic Kmeans, Mk-means

---

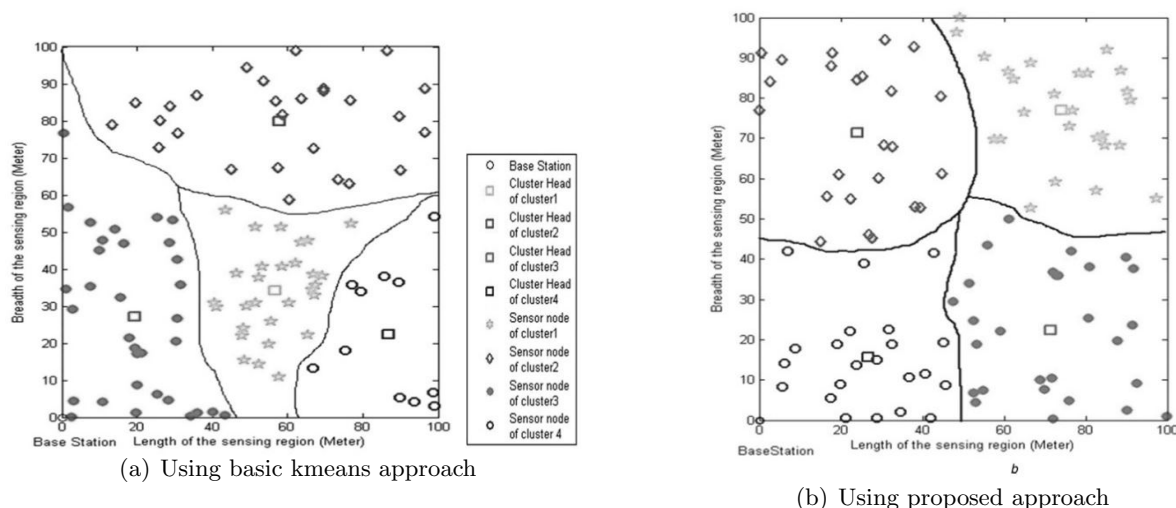(a) Using basic kmeans approach

(b) Using proposed approach

Figure 4: Lifetime metrics

and the proposed protocol are compared with respect to the number of rounds as shown in Fig. 5. It clearly shows that proposed protocol provides better network lifetime compared to the above mentioned algorithms and helps to provide enhanced network lifetime to a great extent. In Fig. 6, it shows the comparison of the energy consumption of the proposed protocol with
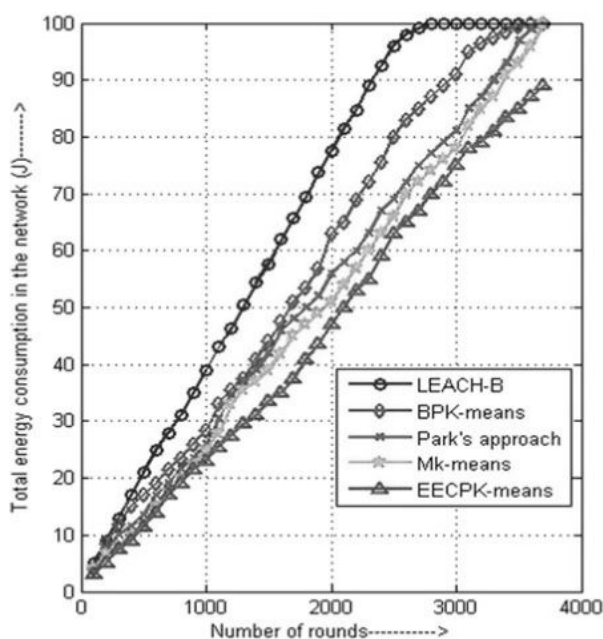

Figure 5: Number of alive nodes with respect to number of rounds

LEACH-B, BPK-means, basic Kmeans and Mk-means algorithm with respect to the number of rounds. It shows that the proposed protocol can reduce energy consumption significantly compared to above mentioned algorithms.
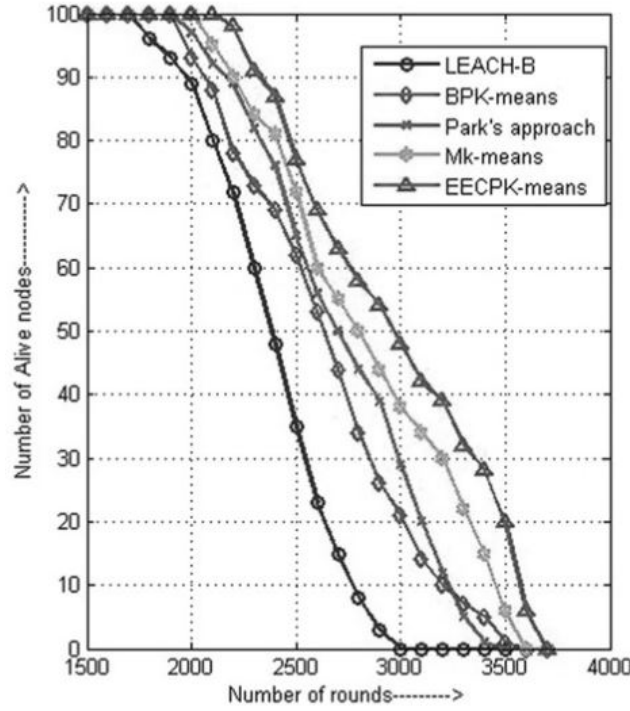
Figure 6: Energy consumption with respect number of rounds

1.1.3 EEM-LEACH: Energy efficient multi-hop LEACH routing protocol for clustered WSNs

[3] presents a cluster based multi-hop routing protocol EEM-LEACH characterized by
1. Cluster head selection based on residual energy and average energy consumption of nodes.
2. Multi-hop inter cluster communication that choose a choose a multi-hop path with minimum communication cost from each node to the base station.
3. Direct communication by nodes near to the base station.

- **Algorithm description**
  At the beginning of set-up phase, each node chooses a random number of 0 and 1, every node computes the threshold $T(n)$ as follows:

$$T(n) = \begin{cases} \frac{p}{1-p \times (r \bmod 1/p)} \times P\left(RE\right), n \in G \\ 0\,, otherwise \end{cases} \tag{8}$$

$$P(RE) = \begin{cases} \frac{E_{res} - E_{avg}}{E_{res}}, if\ E_{res} > E_{avg} \\ 1 - E_{avg}, otherwise \end{cases} \tag{9}$$

where p is the desired percentage of cluster heads, r is the current round, G is the set of nodes that have not become the cluster head in last $1/p$ rounds, E_{res} is the residual energy. If the chosen random number $< T(n)$, then n becomes the cluster head for the current round r.
Once cluster heads are selected, base station at the center sends CHADV message to its neighbor. CHADV message contains a communication cost field that consists of cost metric of the base station/cluster head that has sent the message. Initially, cost metric is zero at the base station and large value at each cluster head. When a cluster head v receives CHADV messages

from base station/cluster head u, communication cost metric c is computed as follows:

$$c = \begin{cases} C(u) + d_{uv}^2, d_{uv} > d_0 \\ C(u) + d_{uv}^2, d_{uv} < d_0 \end{cases} \qquad (10)$$

where $C(u)$ is the cost metric at u and d_{uv} is the distance between u and v. The cost metric at v is

$$C(v) = min(C(v), c) \qquad (11)$$

Each cluster head v receives CHADV message from multiple cluster heads/base station, computes cost metric for each and choose the minimum cost metric c' after t' time from the receipt of first CHADV message where t' is the propagation delay for the first CHADV message received. If cost metric c' is computed when cluster head v received CHADV message from u, then u is selected as the next hop of v in the multi-hop path to the base station. Also, CHADV message with communication cost field updated as $C(v) = c'$ is sent to $v's$ neighbors. If a cluster head receives at least one CHADV message after sending the first CHADV message, cost metric is computed. If the cluster head's cost metric is updated, then a CHADV message is sent with updated communication cost field. Thus, by propagating cost via CHADV message, multi-hop path with minimum communication cost from each cluster head to base station is determined. The set-up phase is shown in Fig. 7. During the steady state phase, once cluster
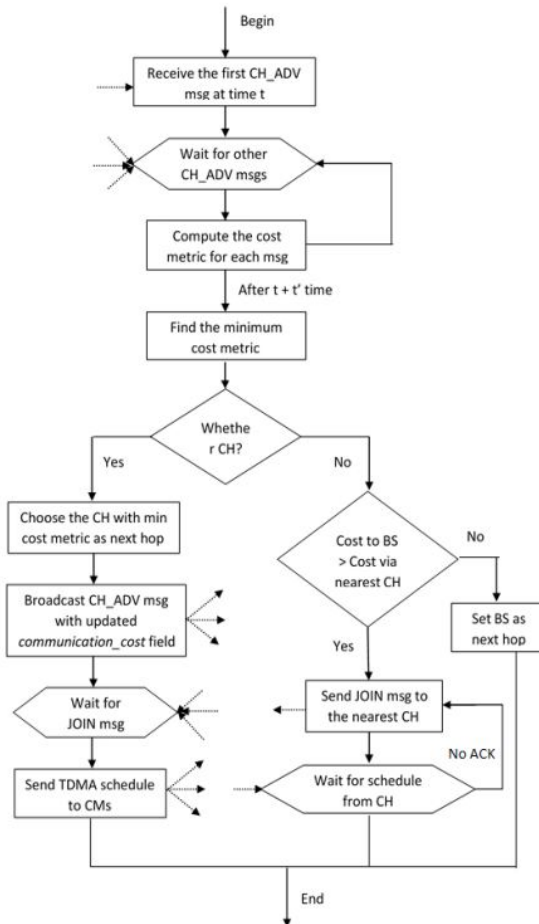

Figure 7: Set-up phase of EEM-LEACH

members receive BEGIN message from the cluster head, each member node sends data in the
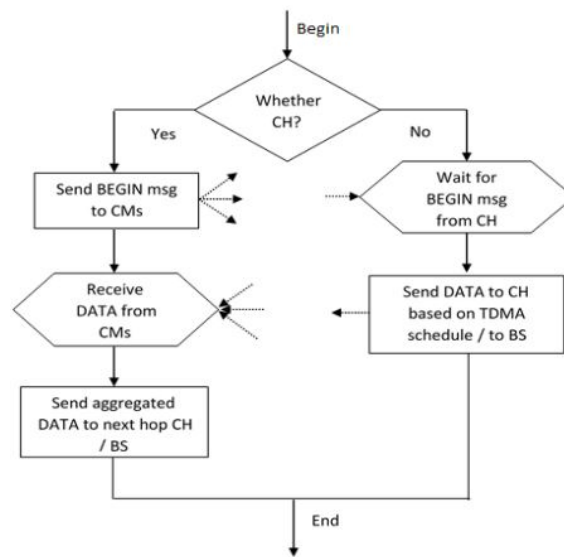
allotted time slot, shown in Fig. 8.

Figure 8: Steady state phase of EEM-LEACH

- **Simulation analysis**
  Fig.9 shows that the proposed EEM-LEACH protocol has improved lifetime than existing protocols. For a smaller network with area 100m × 100m, until the first node's death, EEM-LEACH has 42% and 41% better lifetime than LEACH and M-LEACH respectively.
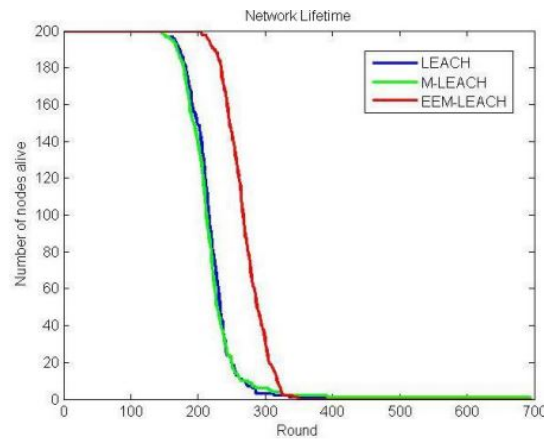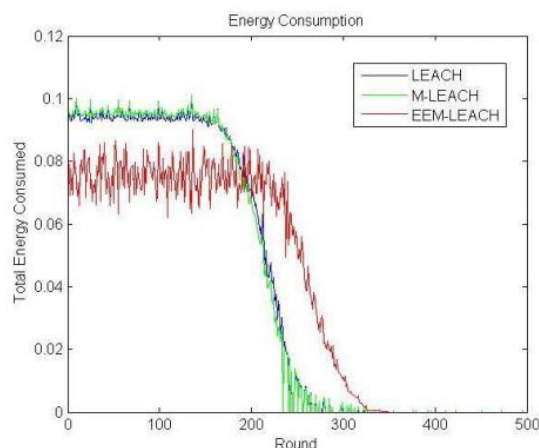

Figure 9: Network lifetime for 100m × 100m size network

The proposed EEM-LEACH protocol has less energy consumption than LEACH and M-LEACH because of choosing a multi-hop path with minimum communication cost, shown in Fig.10.

Fig. 11 shows data packet delivery is better for proposed protocol than LEACH and M-LEACH.

Figure 10: Energy consumption for 100m × 100m size network
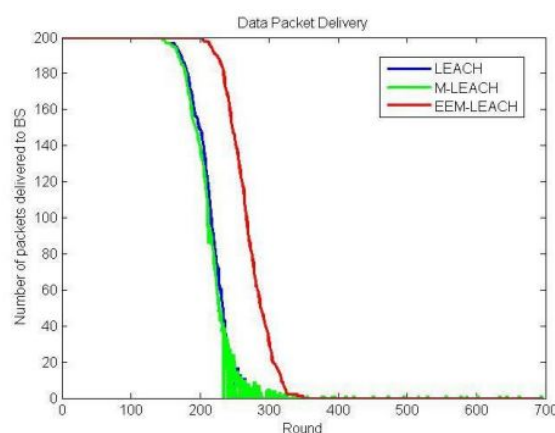
Figure 11: Data packet delivery for 100m × 100m size network

1.1.4   Other papers

[4] has presented the LEACH-SAGA routing protocol that is based on optimization techniques of simulated annealing and genetic algorithms for better energy distribution among the sensor nodes in the WSN. The LEACH-SAGA routing protocol is a completely centralized control algorithm which is executed at the BS. Initially it forms optimal numbers of clusters accomplished by simulated annealing and genetic algorithms. After the formation of clusters, the BS calculates the centroid of each cluster. In the selection of the CH, the BS considers the residual energy of the sensor nodes and the distance of the sensor node from the centroid. The sensor nodes with energy greater than the average energy of the cluster become the set of possible CHs in each cluster. The BS selects the final CHs from the possible CHs set based on the minimum distance from the centroid. In the steady state phase, member nodes of the clusters communicate with their respective CHs and CHs then transmit the aggregated data to the BS via other CHs. The LEACH-SAGA routing protocol outperforms LEACH in terms of network lifetime. It also claims even distribution of the CH in the network with less energy consumption. The main issues with this protocol are scalability, delay and complexity.

[5] proposes a genetic algorithm (GA) based adaptive clustering protocol with an optimal probability for cluster formation and selection. Initially, all nodes participate in the candidate CH (CCH)

selection process by generating a random number r and comparing this r with threshold $T(s)$. If the value of r is less than $T(s)$, based on a probability value p_{sat}, then the node is selected as CCH. After the selection of initial CCH, all nodes send their status messages containing their id, location information and CCD information. Based on this information, the BS finds the optimal probability P_{opt} for formation of optimal clusters K_{opt} with the help of GA. The GA searches the solution space to determine the P_{opt} using an evolutionary optimization process including probabilistic transitions and non-deterministic rules with crossover and mutation operators. After selecting P_{opt}, the BS broadcasts the value of P_{opt} to all sensor nodes n. The set up and steady state phase are the same as in LEACH. The performance of this protocol is compared in two scenarios based on the BS position. In the first case, the BS is located in the centre of the network and in the second case, it is situated outside of the network. In both cases, LEACH-GA performs better than LEACH in terms of energy efficiency but it suffers from message overhead and scalability.

1.2 Course

Finished the chapter 1 of Structuring Machine Learning Projects, Coursera, https://www.deeplearning.ai/deep-learning-specialization/

2 Objectives for the Next 2 Weeks

2.1 Reading

By reading the papers, it is found that the most important parts of LEACH and improved LEACH protocols are the selection of cluster heads and the formation of the clusters. Some papers have studied to solve above problems by Kmeans. For the next two weeks, I will continue reading papers that use Kmeans in LEACH and try to propose an improved Kmeans LEACH protocol.

3 Advisor's Comments

Bibliography

[1] T. M. Behera, U. C. Samal, and S. K. Mohapatra, "Energy-efficient modified leach protocol for iot application," *IET Wireless Sensor Systems*, vol. 8, no. 5, pp. 223–228, 2018.

[2] A. Ray and D. De, "Energy efficient clustering protocol based on k-means (eecpk-means)-midpoint algorithm for enhanced network lifetime in wireless sensor network," *IET Wireless Sensor Systems*, vol. 6, no. 6, pp. 181–191, 2016.

[3] A. Antoo and A. R. Mohammed, "Eem-leach: Energy efficient multi-hop leach routing protocol for clustered wsns," in *2014 International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT)*, July 2014, pp. 812–818.

[4] H. Zhang, S. Zhang, and W. Bu, "A clustering routing protocol for energy balance of wireless sensor network based on simulated annealing and genetic algorithm," *International Journal of Hybrid Information Technology*, vol. 7, pp. 71–82, 03 2014.

[5] J.-L. Liu and C. V. Ravishankar, "Leach-ga: Genetic algorithm-based energy-efficient adaptive clustering protocol for wireless sensor networks," in *Int. J. Mach. Learn. Comput.*, 2011, pp. 79–85.