Final Project Report
EECE 5698 - Wireless Sensor Networks

Mauro Belgiovine

April 24, 2018

1 Introduction

In this work I've chosen to analyze and simulate LEACH, a seminal distributed clustering-based routing protocol specifically proposed for Wireless Sensor Networks (WSN) by MIT researchers in early 2000's. LEACH is very popular in WSN research community and has inspired several variants and new algorithms, because of its simplicity and proven effectiveness in improving the network energy efficiency. Although the very first embodiment of the protocol relied on some peculiar assumptions that prevent to fully exploit the true potential of this communication scheme. In this work I wanted to propose a few variations in the original Cluster Heads selection scheme and compare the performances of the new and original algorithm by implementing and simulating it in Omnet++.

2 LEACH

LEACH (Low-energy Adaptive Clustering Hierarchy) [] is a clustering-based protocol that aims to minimize energy dissipation in WSNs. The key features of LEACH are the following:

- Localized coordination and control for cluster set-up and operation;
- Randomized rotation of Cluster Heads (CH) and corresponding clusters;
- Local compression of data collected to reduce communications with a Base Station (BS).

The idea behind clustering is to leverage small transmit distances for most of the nodes in the network and limit long distances transmissions, the ones toward BS, in order to reduce the energy dissipated by sensors for communicating data over time and extend the network lifetime. LEACH outperforms classical clustering algorithms by using an adaptive cluster selection and rotating CHs over time to evenly distribute the energy spent for high power transmission to all nodes in the network.

2.1 Algorithm

LEACH works in discretized time steps called *rounds*, where each round begins with a clusters *set-up* phase and a *steady-state* phase, when data transfers to CHs and subsequent BS occurs. Here I briefly describe what operations are included in each step.

2.1.1 Cluster set-up

Within the cluster set-up phase, several steps are performed by each node:

1. **Advertisement Phase**
   Each nodes decides to become or not CH based on a threshold function defined as

$$T(n) = \begin{cases} \frac{P}{1-P(r \bmod \frac{1}{P})}, & \text{if } n \in G \\ 0, & \text{if } n \notin G \end{cases}$$

   where P is the desired ratio of CH per node in the network (e.g. $P = 0.02$), r is the current round and G is the set of nodes that have not been CH in the last $\frac{1}{P}$ rounds. This threshold defines the probability of each node at each round to become CH and ensures that each node will be a cluster-head at some point within $\frac{1}{P}$ rounds. This probability increases over time since less nodes will be eligible to become CH at each round and $T(n) = 1$ when $r = \frac{1}{P} - 1$. It becomes again $T(n) = P$ when $r = \frac{1}{P}$. Fig. 1 shows the plot of $T(n)$ for a given value of P. After CHs have been self-elected, they broadcast and advertisement (`ADV`) message to all

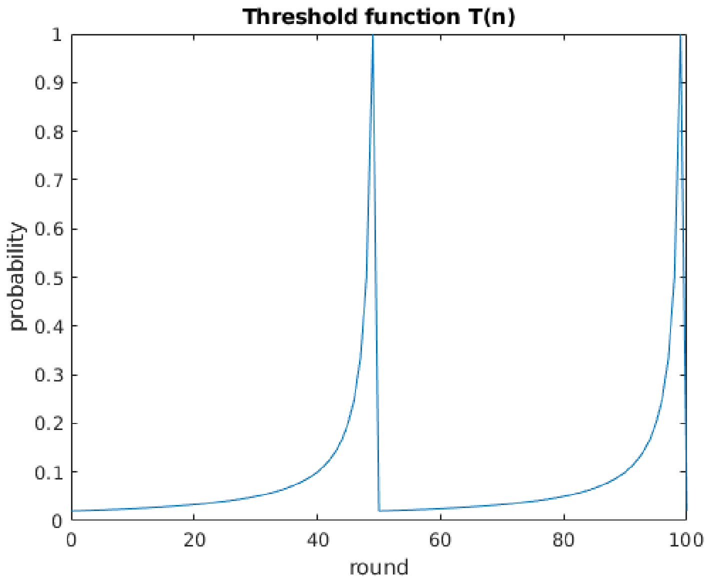

Figure 1: Threshold function with $P = 0.02$ from round 0 to 100.

other nodes. CHs use CSMA/CA MAC protocol to transmit their advertisement.

2. **Cluster set-up**
   After `ADV` messages are received by each node, they will select their CH based on the received signal strength, in order to choose the one that requires less energy to transmit data to (i.e. the closest one). Each node then will send a `JOIN` message to the selected CH in order to notify that is joining its cluster. Again, CSMA/CA MAC is used to access the channel for transmission.

3. **Schedule creation**
   Each CH, upon reception of `JOIN` messages, creates a TDMA schedule based on the number of nodes in the cluster and tells each node when they are allowed to transmit their data. The schedule is then broadcast back to the nodes in a `SCHED` message.

2.1.2 Steady-state

Once the clusters are created and the TDMA schedule is fixed, data transmission can begin. Assuming nodes always have data to send, they send it during their allocated transmission time. This transmission uses a minimal amount of energy, chosen based on CH's received signal strength, and the radio is ON only when it's their turn to transmit data, thus minimizing the energy required to transmit `DATA` messages. The CH keeps it's radio ON during data transmission in order to receive all data from sensors. When data transfer is complete, the SH performs data compression into a single signal and send it to the BS.

After a predefined amount of time, a new round begins with the selection of new CHs and starting this whole procedure again.

2.2 Simulation of LEACH

In order to evaluate the performance of LEACH, I used Omnet++ to implement and simulate the protocol based on the algorithm description of the original paper. LEACH clustering performance is then compared to a *Direct Transmission* scheme, where essentially all nodes at each round join a single cluster held by the BS itself that directly collects `DATA` messages from sensors using TDMA transmission scheme, in order to reduce the energy consumption and avoid collisions for long messages[1].

Sensor nodes are equipped with a *First order radio model* (Fig. 2) and the energy dissipation accounted in the simulation regards the three most expensive operations carried out by sensors:

- **Transmit**
$$E_{TX}(k, d) = (E_{elec} * k) + (e_{amp} * k * d^2)$$
- **Receive**
$$E_{RX}(k) = E_{elec} * k$$
- **Data Aggregation**
$$E_{DA}(k, N) = E_{comp} * k * N$$

where E_{elec} is the energy needed to run the transmitter or receiver circuitry, e_{amp} is the energy needed for the transmitter amplifier, k is the number of bits in the message, d is the distance in meters and N is the number of messages of k bits to be processed.

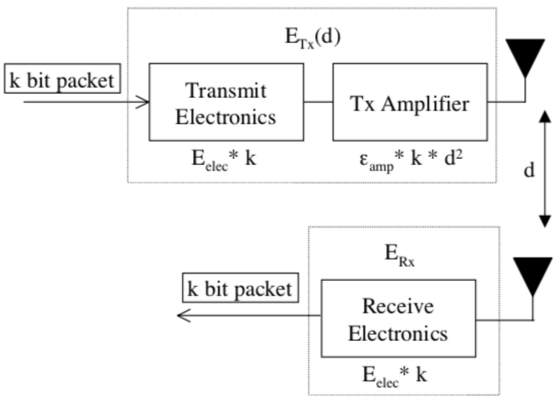

Figure 2: First order radio model.

For sake of simplicity and to better compare the two transmission schemes, the simulation is based on the following assumptions, based on original paper:

- Sensor nodes start all at the same power level at the beginning of the experiment;
- Sensor nodes are fixed and uniformly distributed in a square area with no obstacles;
- The *maximum range* of transmission is defined as the diagonal of the square area and depends on its edge size;
- The BS position is assumed to be unknown by nodes. Therefore, transmissions to the BS always happens at the highest transmission power (i.e. based on max. range of transmission)[2].
- The cost for cluster formation phase is not taken into account, since control messages are assumed to be very small and don't impact much the energy consumption compared to `DATA` messages;
- Channel access during clusters formation is assumed free of collisions[3]. Collision are avoided by TDMA scheme during `DATA` transmissions.
- Only one `DATA` transmission to CH/BS is performed at each round by all nodes;

[1]Note that is possible to obtain Direct Transmission scheme with LEACH having $P = 0$.
[2]This is because the position of BS is assumed to be "far away from sensors" in the original paper.
[3]CSMA/CA MAC takes care of channel access.

- Nodes are assumed to be synchronized and start a new round after a predefined amount of time. In this case, the round duration is selected in order to guarantee the cluster formation and transmission of all nodes as if they were all transmitting to the BS using TDMA, since this is the case when the highest amount of time is needed in a round to let all nodes perform their transmission;
- If no `ADV/JOIN` messages are received by sensors (i.e. no CHs have been self-elected in this round or no sensor joined the cluster of a CH), the sensor falls back to a Direct transmission scheme, joining the BS cluster as described earlier.

2.3 LEACH vs. Direct TX comparison

For these simulations, the following parameters have been used, except where explicitly stated:

Table 1: Simulation parameters

Parameter	Value
E_{elec}	50 nJ/bit
e_{amp}	100 pJ/bit/m^2
E_{comp}	5 nJ/bit/message
`DATA` size	2000 bit
Num. of nodes (N)	100
Init. Battery charge	0.5 J
Sensed area	200 m^2
P (CH ratio)	0.05

In these simulations it will be shown the improvement in the energy efficiency of the network in terms of network average lifetime extension and the relative average energy dissipation. Each experiment consist of 100 different configurations of sensors in network.

2.3.1 Network lifetime vs. Ratio of CH

Fig. 3 and Fig. 4 shows respectively the network lifetime extension and the relative total energy dissipation achieved by LEACH with respect to the Direct Transmission method, varying the desired ratio of CH in the sensor network. Note that 0% and 100% of CH is the same as Direct Transmission. From these two plots it is possible to see that exists an optimal ratio of nodes P^* that guarantees the maximum longevity of the network, which in this case is $P = 0.05$. For lower values of P, there are fewer CHs (or none) and sensors are most likely to send their data with radio signals covering long distances, whereas for higher values of P there are more CHs that performs energy intensive tasks at each round, draining very quickly the overall energy available in the network. Both these results are in line with the performance showed in the original paper.

2.3.2 Network lifetime vs. Size of the network

Fig. 5 and Fig. 6 shows respectively the network lifetime extension and the relative total energy dissipation achieved by LEACH with respect to the Direct Transmission method, varying the size in m^2 of the sensed area for the optimal CH ratio $P = 0.05$. By increasing the network diameter, communications to CHs and BS require more transmitting power and therefore reduce drastically the energy autonomy of the sensor network. Especially important is to note how the energy dissipation efficiency is kept at a good level with respect to Direct Transmission scheme even for very large sensed areas. Again, this these results are in line with the ones described in the original paper.

3 Variations on CH election scheme

By looking at the threshold function $T(n)$ described in the previous section, the CH election policy depends only on the actual round number and whether the nodes have already been selected or not as CH in the last $1/P$ rounds. This scheme is good to randomly choose an adequate and dynamic

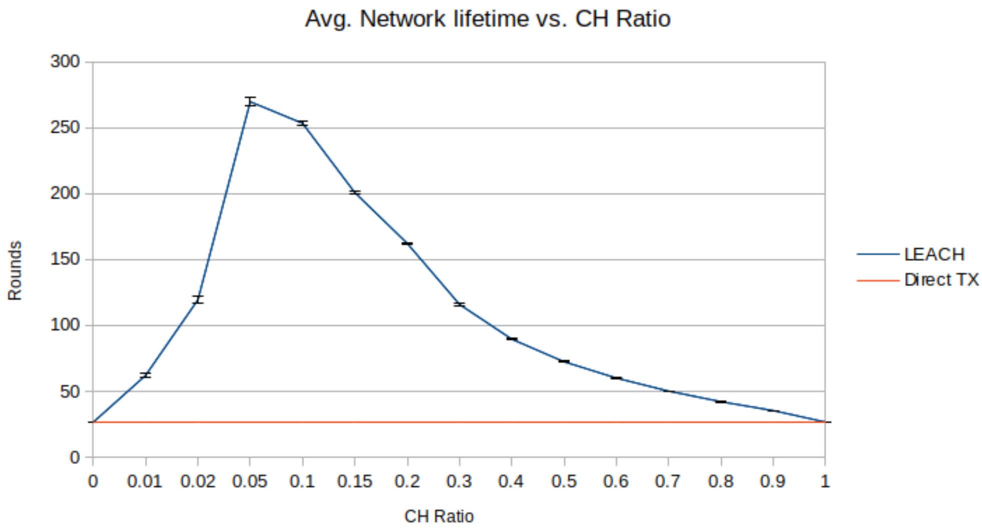

Figure 3: Average network lifetime varying the ratio of CH in the network (P).

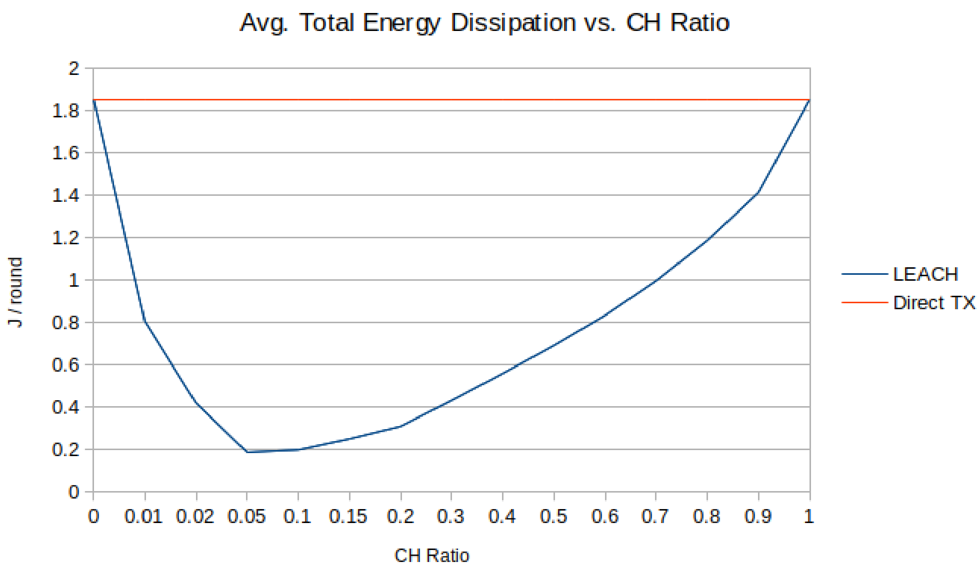

Figure 4: Average total energy dissipation for each round varying the ratio of CH in the network (P).

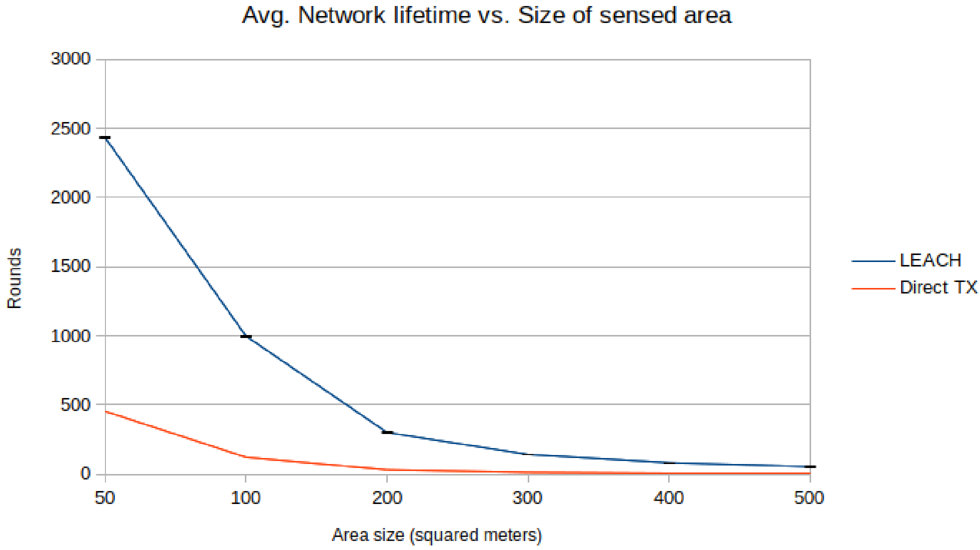

Figure 5: Average network lifetime varying size of the sensed area.

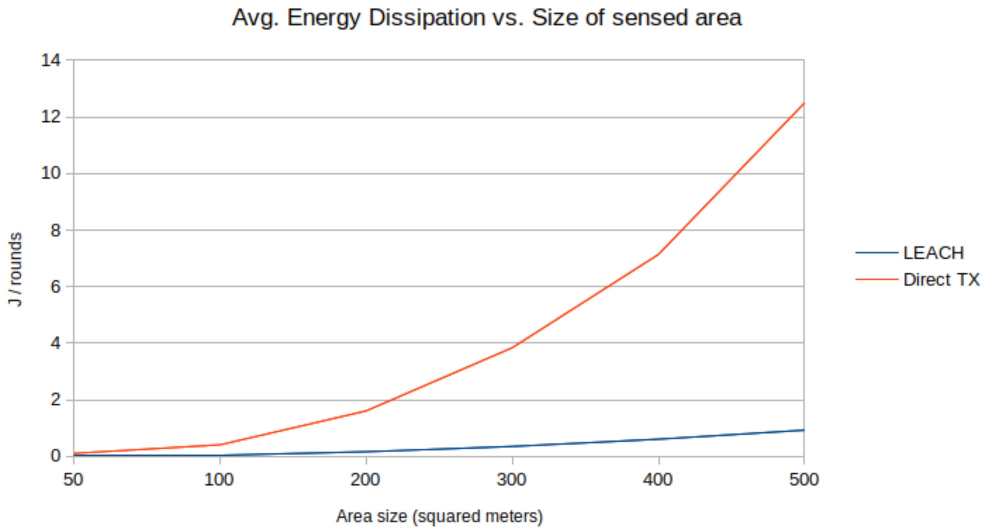

Figure 6: Average total energy dissipation for each round varying size of the sensed area.

amount of clusters and CHs among the network to reduce overall energy dissipation, but has several weaknesses that prevent LEACH to improve even more energy saving or other particular network metrics by fully exploiting informations about the network topology.

3.1 Weaknesses of threshold function

In particular, the following weak points in the cluster head selection scheme have been identified in this work.

3.1.1 Energy awareness

The cluster selection doesn't depend on the energy left at each node at a given moment in time. This is because all nodes are assumed to start with the same amount of energy and the CH rotation should guarantee a fair election of CH within each set of $1/P$ rounds, evenly distributing the energy intensive reception, compression and transmission of data to BS operations involved in the CH role.

Although due to the dynamic nature of cluster formations and randomized network topology, the energy spent by each node is not evenly distributed as time goes by: CH energy consumption varies with the number of sensors joining the cluster, so it can happen that some CHs spend less energy than others during their activity; the same dynamic nature of clusters impacts differently in the range of communications and consequently on the energy spent at each round when nodes are operating as normal sensor nodes. With the current CH election scheme it can happen that a node that has been involved in particularly intensive sensor/CH roles can be elected as CH sooner than others, leading to a quick drain of its energy and reducing the overall network lifetime.

For example, a node that is close the border of the sensed area is in general more involved in longer distance transmission to CHs and is more likely to die sooner than others if selected as CH repeatedly over time. Intuitively, the selection of these nodes for energy intensive CH roles should be pushed further in time to increase the longevity of the network.

3.1.2 Spatial awareness

The second issue identified in this selection scheme is that it doesn't take in consideration the position of CHs in the network and in each cluster. For instance, it can happen that CHs elected in a round are too close to each other, leading to a cluster formation that penalizes sensors far away from their CH. This kind of situation leads to highly unbalanced transmission costs within nodes in each cluster, where a big portion of nodes transmit at long distances and few to close ones. Fig. 7 shows the cost of communication for sending 2000-bit `DATA` messages at different distances. Moreover, this also leads to the creation of *isolated* CHs that would not be joined by any sensor node, ending up in transmitting directly to BS during such rounds, which causes more energy inefficiency.

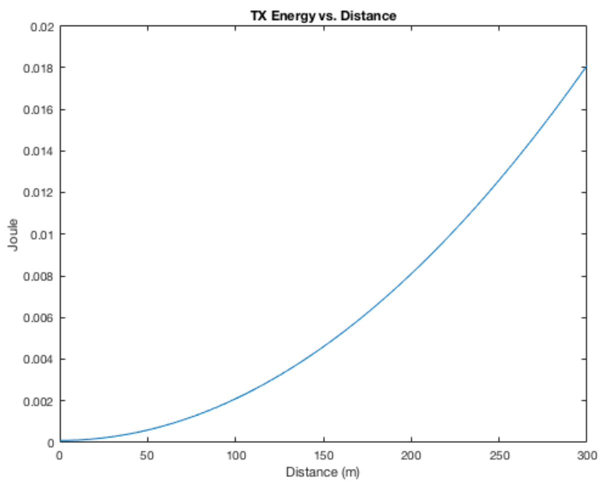

Figure 7: Energy consumption for transmission of 2000-bit `DATA` messages ($E_{elec} = 50$ nJ/bit, $e_{amp} = 100$ pJ/bit/m^2).

Ideally, regardless of how clusters are formed, in order to balance the cost of transmission for sensors within the same cluster, one would like to select CH that are "in the center" of the cluster: the reason is that, if CH is positioned on the border of the cluster (which happens especially when two or more CH are selected very close to each other), there's a major imbalance in distance of transmissions and, since the cost of transmission power grows exponentially with the distance, this leads to reduced efficiency in the overall energy dissipation.

3.1.3 Guarantee of CH selection

The current scheme relies on certain probability of selection of CH, which increases over time as explained previously but for instance it can also never happen. In this simulation, when no CHs are elected in a round, the nodes will behave as for Direct Transmission scheme, by sending their data to the BS that will coordinate the TDMA schedule of transmissions. In the original paper it is not clear how they manage this kind of situation, but in this case this is of course a source of inefficiency and reduces the longevity of the network substantially.[4]

3.2 Proposed solution

The main contribution of this work, beside the LEACH and Direct Transmission schemes simulation effort, consists in a modified CH selection procedure that involves a few but effective changes in the original algorithm in order to better control which node is better eligible to become CH based on some user-defined performance metric. In this case, changes proposed aim to increase the network longevity by addressing the *energy awareness* and *spatial awareness* issues explained above. The proposed solution is designed on top of the current protocol, it doesn't add any additional message overhead and only relies on a small addition in the `JOIN` and `SCHED` messages respective payloads.

3.2.1 *Controlled CH selection*

The key idea behind the proposed solution is to choose a better suited CH after a cluster has been formed within a LEACH round, *deferring* to a better suited node, if any, based on some strategy defined by the network administration. The original *Cluster set-up* phase is modified like this:

1. After a CH has been advertised, sensors will select the CH they want to join by sending them a `JOIN` message following the usual received signal strength metric. This message, beside containing the ID of the joining sensor node, will also carry *some other informations* about the node that will be used by the CH to decide whether there is a better suited node for CH role within the cluster or not. Such additional informations will depend on the selection strategy that are discussed later in this section.

2. The CH receives the `JOIN` messages and read the additional informations carried out.

3. The CH will *sort* for a given criterion (the one that reflects a better suited CH node within the cluster) the joining nodes (including itself), using the additional informations retrieved from their requests. Then it will pick the best element from the sorted list of nodes (either the first or last, according to the sorting criteria) and advertise it as the new CH in the `SCHED` message that it will broadcast. To do this, it will simply append its ID or the ID of the newly selected CH to the `SCHED` message.

4. When the nodes in the cluster receive the `SCHED` message, they will read the CH ID additional field and compare it to their ID: if a node notices that it has been elected as the new CH, it immediately starts listening for `DATA` messages from other nodes; otherwise, they simply follow the TDMA schedule informations and transmit to the ultimately designed CH.

The threshold function is therefore used only for the purpose of dynamically create a certain number clusters at each round of the simulation, based on the desired ratio of CH, and leaves to this procedure the task of identifying, if any, the right CH.

[4]This issue is not addressed in this work, but it is worth to mention in order to explain the simulation results obtained.

3.2.2 Selection strategies

As anticipated, the better CH selection strategies are based on the weak points identified in the threshold function presented in the last section, but can be further extended to meet other network metrics dependent on the application. Precisely, the sorting criteria proposed are the following:

- *Node with highest energy left* (requires sensor nodes to append the energy left information to the `JOIN` message);
- *Node closer to the centroid of the cluster* (requires sensor nodes to append their XY coordinates to the `JOIN` message);
- *Node with highest energy left* ***and*** *closer to the centroid of the cluster* (requires sensor nodes to append both energy left and XY coordinates informations to the `JOIN` message);

For all of these strategies a simple sorting criteria has been created, which is used to sort the nodes in ascending order so that the first element is the most desirable CH node, based on the selection strategy.

3.3 Different CH selections simulation

In order to evaluate how the behavior of the network differs from one CH selection strategy to another, the same analysis performed on network lifetime for different network diameters has been performed using the optimal $P = 0.05$ ratio of CHs.

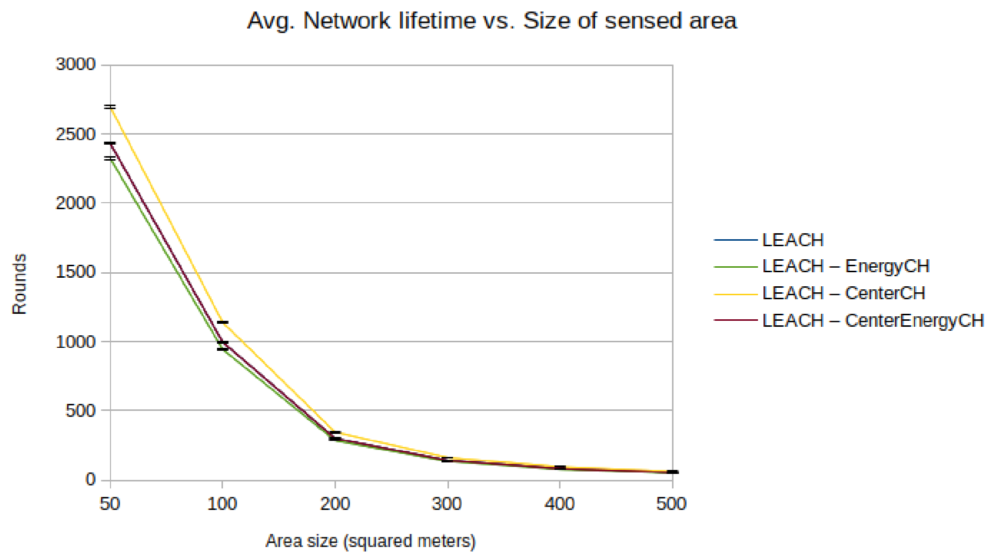

Figure 8: Average network lifetime varying size of the sensed area and the CH selection strategy.

Fig. 8 and Fig. 9 show that, in terms of network longevity, both the CH selection strategies that prefer the node closer to the centroid of the cluster are the one that performs better in this simulation scenario, followed by the basic LEACH approach and then by the energy aware strategy. This shows that, balancing the distance of transmissions within clusters impacts more on network longevity with respect to other strategies. The pure energy aware strategy, on the contrary, puts too much pressure over nodes that shows higher energy levels, which actually causes the network to fall back to a suboptimal energy efficiency state if the performance metric is the round when the last node dies. Nevertheless, such strategies may be still suitable for other applications and network metrics different by lifetime, as described in the next section.

3.4 Impact of different selection strategies

In order to understand why one strategy is more effective than others, let's compare the energy levels of sensor nodes over time for each of the proposed CH selection strategies for a same simulation configuration. Fig. 10 shows how the energy level of cluster is affected over time by traditional

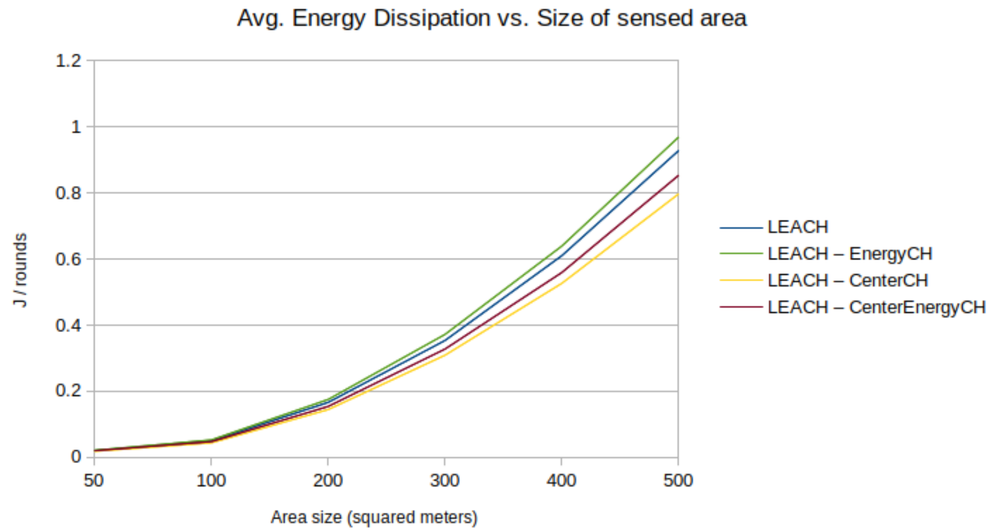

Figure 9: Average total energy dissipation for each round varying size of the sensed area and the CH selection strategy.

CH selection based on threshold alone. Fig. 11 show how the network lifetime is increased by the centroid strategy, which tend to select always nodes that are in the central area of the network and is able to extend the network lifetime by distributing the energy consumption more evenly in the other nodes, that perform transmissions at more moderate power levels with respect to the first scenario. This is the most desirable strategy from the network longevity point of view. Energy aware CH selection, depicted in Fig. 12, shows a very interesting behavior, even if not suitable to extend further the network lifetime: it shows a tendency to keep all the sensor in the network to a very similar energy level over time, increasing by a great factor the time when the first node dies. If we have an application for WSN where it's important to keep all the nodes alive as much as possible, this would be the best strategy. In the end, Fig. 13 shows features of both *energy aware* and *centroid* strategy and would be more suitable for WSN applications in which a tradeoff between both strategies is desirable.

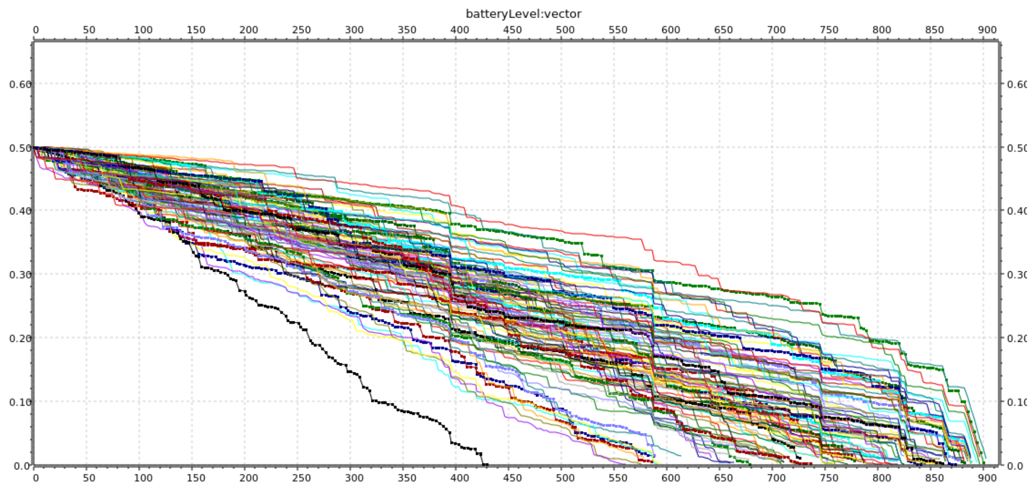

Figure 10: Usual LEACH CH selection strategy.

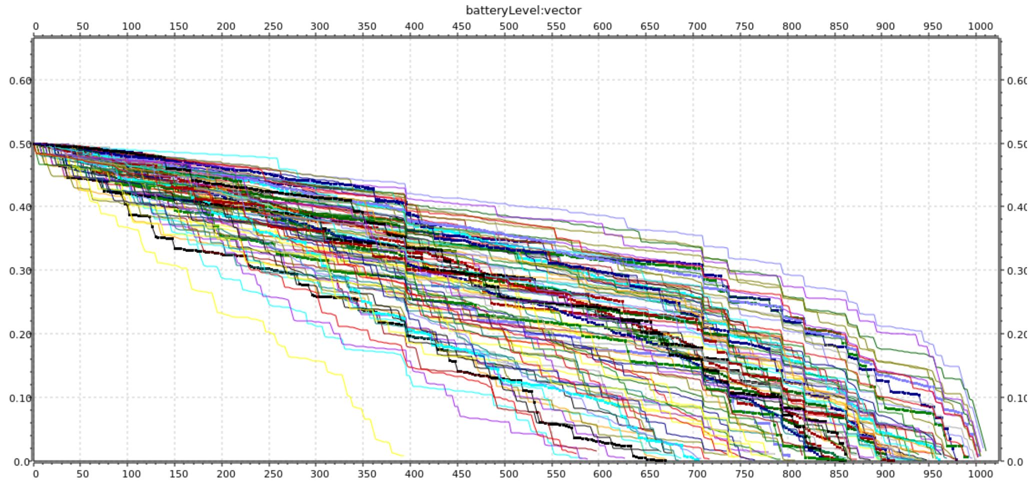

Figure 11: *Node closer to the centroid of the cluster* strategy.

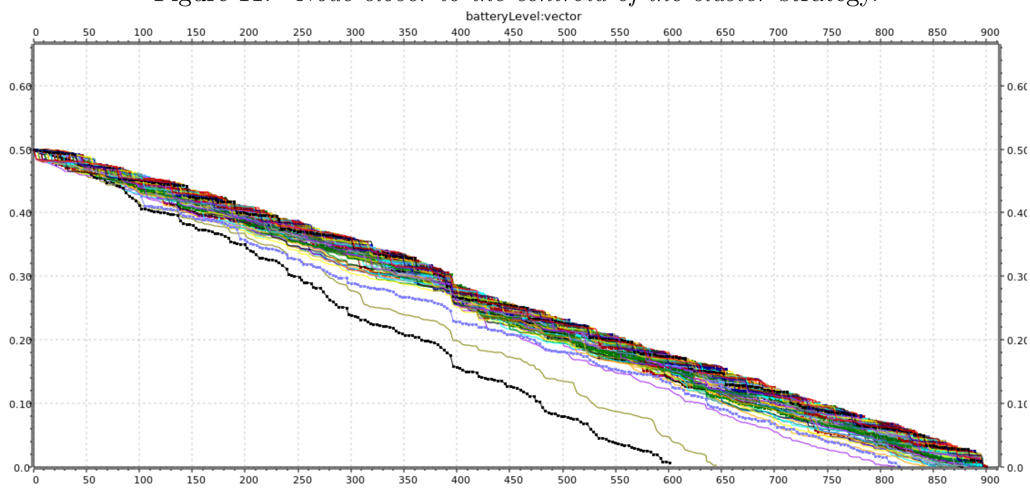

Figure 12: *Node with highest energy left* strategy.

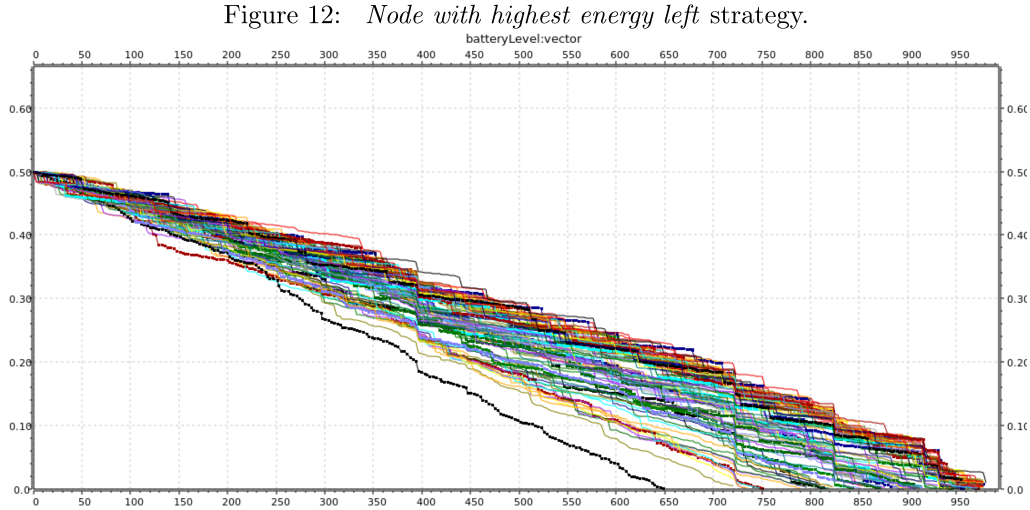

Figure 13: *Node with highest energy left **and** closer to the centroid of the cluster* strategy.