Escuela Politécnica Superior de Linares
Telecommunication Engineering Master

UNIVERSIDAD DE JAÉN
Escuela Politécnica Superior de Linares

Master Thesis

______

CLUSTERING IN MOBILE WIRELESS SENSOR NETWORKS

Student: **Antonio José León Sánchez**

Supervisors: Juan Carlos Cuevas Martínez – Antonio Jesús Yuste Delgado
Dept.: Ingeniería de Telecomunicación

March, 2020

Escuela Politécnica Superior de Linares
Máster en Ingeniería de Telecomunicación

UNIVERSIDAD DE JAÉN
Escuela Politécnica Superior de Linares

Trabajo Fin de Máster

CLUSTERING EN REDES DE SENSORES INALÁMBRICOS MÓVILES

Alumno: **Antonio José León Sánchez**

Tutor: Juan Carlos Cuevas Martínez – Antonio Jesús Yuste Delgado
Depto.: Ingeniería de Telecomunicación

Marzo, 2020

“A Antonio José Sáez Castillo, por haberme enseñado mucho más que estadística, y por haber compartido conmigo mucho más que una investigación. Siempre estarás presente en este y cada uno de mis trabajos y decisiones en la vida. Tu serenidad y buena voluntad siempre serán un faro guía en mi vida. Gracias por todo, amigo.”

ABSTRACT

Nowadays, wireless sensor networks are a technology with a wide use. They allow humans to monitor parameters without building a specific infrastructure. One of its advantage is that they use little low-cost micro-sensors that are programmed to obtain information from the environment and to send it later to a base station, which has constant energy supply unlike sensors. The fact that the sensors have limited energy is the main challenge that this technology has to face. Most of the research tries to minimize the energy use so the network lifetime is longer. In order to solve this, the mobility concept is introduced by many authors in order to avoid long range communications. In this work, some related work is going to be presented and classified in a taxonomy, and later a method that supports mobility in the base station and based on Type-2 Interval Fuzzy logic will be developed and compared to the results obtained by the methods studied in the bibliography.

CONTENT TABLE

FIGURE INDEX

TABLE INDEX

REPORT

1 INTRODUCTION

In the last years, the evolution of the technology has allowed the development of new microprocessor based devices called sensors. The main features of this sensors are their small size, their energy source, usually batteries, their wireless communication capabilities and also their low manufacturing cost. Thanks to these factors, companies have built tons of them with the aim of measuring any kind of parameter, from a biomedical radioactivity indicator to a temperature probe in an isolated forest, and so on, without interrupting the normal working of the environment in which it is deployed.

Then, some researchers proposed the possibility of grouping nodes in what they know as Wireless Sensor Networks (WSNs). This technology aims to collect data from a specific place, whose dimensions could vary, to communicate the results to a Base Station (BS), where the data is processed and usually allows connections by remote clients, so it can be easily accessed. In general, for most common applications, nodes are placed across the field, usually randomly, and have limited battery, which means that it is important to optimize the lifetime of the sensor, and thus the network lifetime. This is an important point for applications related to the natural environment, like for example those that try to measure volcanic events [1], because it is hard place to deploy sensors manually by people and the sensors are randomly placed, so it is fundamental to ensure a good optimization of the energy. Another important application that needs an efficient energy resource management is the forest fire detection [2]. Additionally, this last application also looks for methods that ensures the real time, which means that the method need to collect data from all the nodes each period of time, not being able to store the data and send it when it could be optimal. Additional applications [3] that are covered with this technology are air pollution monitoring [4] where the data measured could be ozone, fine particle matter, nitrogen dioxide, carbon monoxide etc. Ocean pollution prevention [5] is aimed to cover underwater places monitoring temperature and density in those places. Intrusion detection and tracking for military applications [6] is also an interesting application that tracks movement using magnetic and motion sensors. One last example could be tsunami, earthquake or flood detection [7] with seismic sensor and tide gauges placed underground, similarly to the underwater sensors.

This fact led many researchers to create ways of improving the use of the battery, in such a way it was proved that by reducing the number of messages sent by nodes, the sensors would live longer. In this point, a mechanism called **clustering** was introduced. Although it will be explained later in detail, the aim of clustering is to divide the network, and concentrate the communication tasks in a reduced set of nodes, that will send only a message per cluster to the BS. Thus, the optimal election of those set of nodes, known as Cluster Heads (CHs), is one of the main topics in clustering when trying to expand the lifetime of the network. That is the reason why stabilising a taxonomy is an interesting way to have a better point of view of this research field. After that, a new system based on a Type-2 Fuzzy technology will be developed in which mobility is introduced as a novelty. The results will show whether is it worth to make mobile the BS or not, depending of the results obtained and the energy needed to support that mobility.

1.1 Objectives

Through the development of this work, it is intended to propose a method based on Type-2 Fuzzy logic and also that supports mobility, trying to see if it is worth or not mobility. Specifically, these are the **objectives** defined to this work:

1- Related work about topology control and clustering in mobile sensor networks.
2- Taxonomy of the most relevant methods: advantages and disadvantages.
3- Design and implementation of several methods that are representative of the taxonomy developed.
4- Create a testbed environment.
5- Design and implementation of a new solution.
6- Performance of tests and analysis of results.

In order to achieve them, the following **methodology** will be followed:

1. Determination of the bibliography. Study of a classification of:
   - Scenarios (dimensions, number of nodes, status of the base stations, initial energy, energy model, variables used, etc.)
   - Mobility models.
   - Description of state of the art.
   - Fuzzy system proposal.
2. Determination of scenarios for evaluation of the proposal.
   - Selection of the scenario (s).
   - Selection of algorithms with which we will compare.
   - Development of a new fuzzy system.
3. Simulations and results. Statistical analysis.
4. Elaboration of final document and material for thesis defense.

1.2 Background

This thesis is framed inside the research line of the supervisors, PhD. Juan Carlos Cuevas Martínez and PhD. Antonio Jesús Yuste Delgado. It is aimed to follow the work started in the Bachelor Thesis, trying to improve the results obtained there and, in the paper, published later.

2 STATE OF THE ART

In this section will be shown some of the proposed methods by different authors that aim to solve the clustering problem and also justify the proposed taxonomy.

In WSNs is a common issue in research to find the optimal method in order to collect the data from the group of sensors using efficiently the available resources, like battery energy, which would mean a larger life time. In most of the cases, clustering and data aggregation is the preferred option by the authors. Then, we find different options in the establishment of the clusters, the path selection, etc.

In the following sections some of the most representative methods are classified attending to different parameters of their methodology.

2.1 Clustering

In the last years, most of the wireless sensor networks research have focused on how to expand the life time of the network, among them a high percentage of them are based on clustering. Clustering is based on the division of the nodes of the network into groups called clusters, from which a special node is selected called Cluster Head (CH). Therefore, the methods for getting clusters and CHs are those under research.

The main reason of using clustering is to expand the lifetime of the network. If the nodes communicate with nodes that are in their range like a CH, they do not need to use a lot of energy, so in this way they life longer. Moreover, it is easier to manage the working periods, so with a low number of nodes, they know better when to enter in power saving status, when to communicate or if they have to measure.

This CH is the node in charge of receiving the information from the nodes in the cluster, aggregating that data[1], which is explained in section 2.1.1 and sending a package to the BS with all the information gathered in the cluster. By doing this, it is avoided the energy waste of every node sending a message directly to the BS, because in each round, only a set of nodes, those who are CH, need to send a message to the sink, that usually has a larger size than a control packet. Figure 1 shows an example of clustering, where the nodes that belongs to each cluster are represented with a different colour, and its corresponding CH as a cross.

[1] Data aggregation is a mathematical operation that is used to join data and reduce their size without losing information.

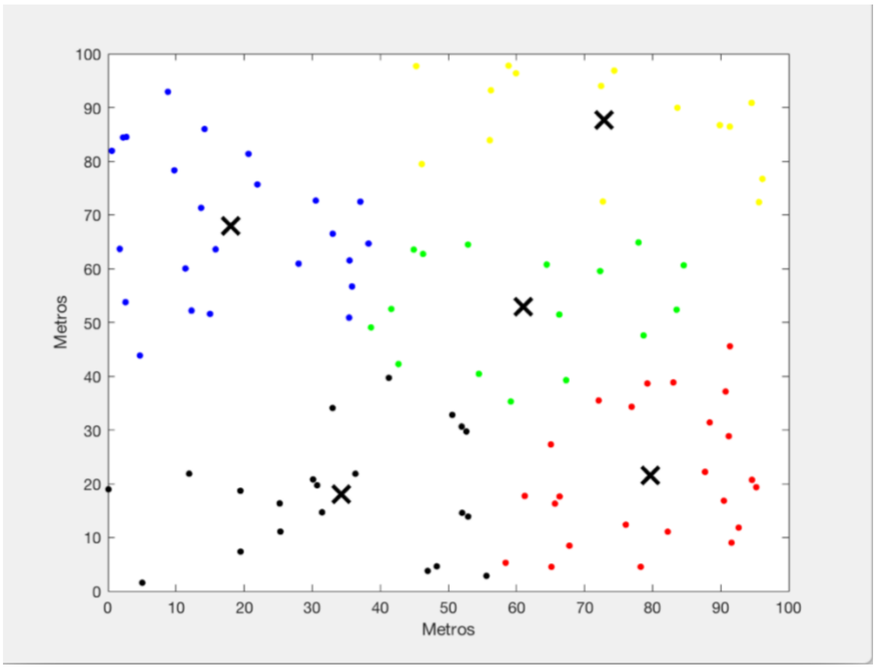

Figure 1. Clustering example.

The process of creating a cluster and dividing the network into them is based on different parameters depending on the application, like for example the current energy of the node, the number of neighbors, distance to the BS and so on. In the next section is presented a classification attending different features that each method could have, and also a description of an example that represents each different category.

2.1.1 Data Aggregation

One of the most used techniques together with the clustering is the data aggregation, used to reduce the energy waste during the communication. In a WSN that can be from ten to hundreds or thousands of nodes, the amount of generated traffic is so huge, so sending that amount of information would take all the energy in a short period of time. To face this issue, researchers have develop mathematical techniques that offer the user the possibility to aggregate the data from different sources but have a high correlation, dude to proximity or the studied magnitude, so it reduces the generated information [38]. Thus, when a CH receives the data from its cluster's members, it applies these algorithms and obtains a representative value or a small set of values. A good example could be a WSN aimed to measure the temperature of a forest, those nodes that are geographically near, will measure similar values, so in this case, instead of sending four different values, a single value could represent the information of that cluster, so the size would be reduced four times. If this example is extended into hundreds of nodes, the energy save increases a lot.

2.2 Clustering method classification according mobility

In this section it is going to be presented a taxonomy that is made in base of a set of features that the studied methods in the bibliography have. There are three characteristics that are able to define most of those methods, which are related to mobility, the mathematical base of the mechanism to select the CH and finally the distribution of the responsibilities. These categorization is usually found in the bibliography, because it is able to define most of the methods. Figure 2 collects the proposed taxonomy and the name of one representative example of each type.

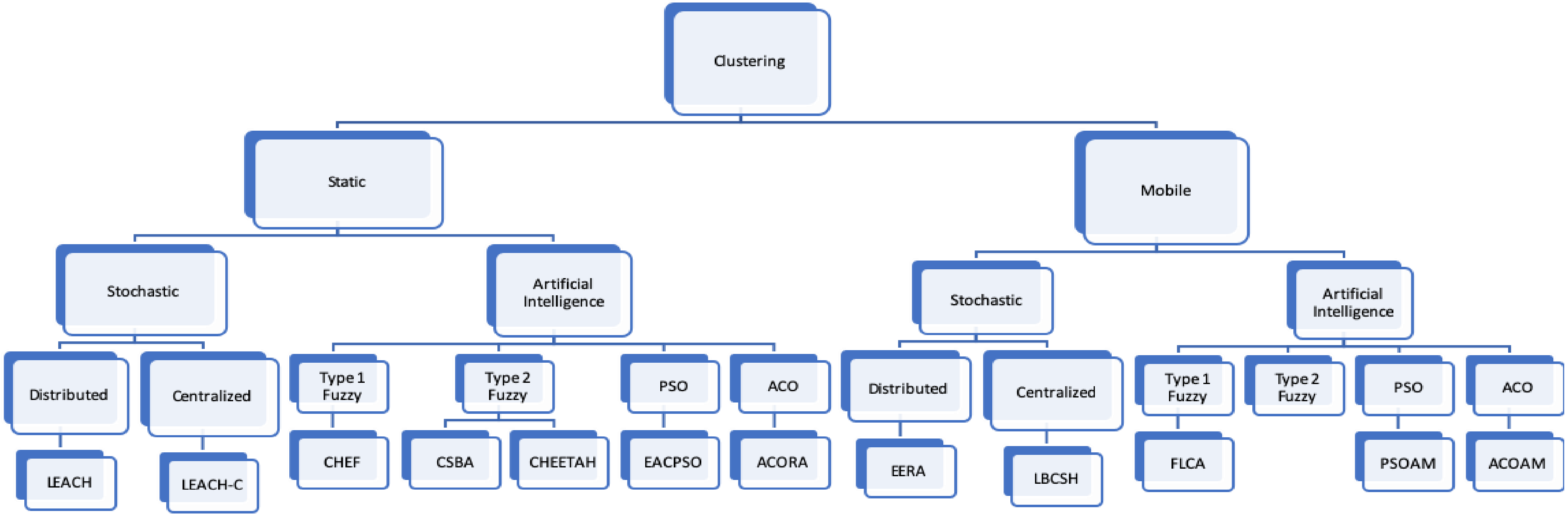

Figure 2. Taxonomy

When talking about mobility, it is observable that both sensors and sink are capable of movement in order to optimize the data collection. In the consulted bibliography authors propose many different methods, most of them focus on how the sink will move, because sensors usually move randomly due to the application of the WSNs, so in those cases it is desirable to use additional infrastructure to track the movement and the data collection, like we can see in [1].

Coming back to sink mobility, the BS path election is not the only thing that can be taken into account. Some authors use mathematical models to define the path used [9][10]. Others use algorithms to know the optimal rendezvous points, which will be visited by the BS in order to make the CHs use the lowest energy when communicating. However, there are also some that use artificial intelligence techniques in order to find the path, using Unmanned Aerial Vehicles (UAVs) that have more energy and processing skills, like for example in [11][12].

The second division focuses in the criteria employed to select the CHs. Stochastic methods usually rely on some random parameter that the base station or each node computes and compare with the rest of the network, like for example in LEACH, so in that way the one with the highest value of the parameter, in general, is the CH of their zone. In the other hand, we can find a wide range of methods based on artificial intelligence, such as Fuzzy Systems, Ant Colony Optimization (ACO) or Particle Swarm Optimization (PSO). Each author selects in which type will be based their proposal, attending to their necessities and what parameter do they want to optimize.

In the following sections some of those clustering methods referred in the taxonomy are further detailed.

2.2.1 Low-energy adaptive clustering hierarchy (LEACH)

The first proposal that is going to be analyzed, *Low-energy adaptive clustering hierarchy*, known as LEACH [13], is considered the most important method and the referent when talking about WSNs and clustering. LEACH is a complete **distributed stochastic clustering method for static WSN** that works based on rounds, in a way that for each round, a different set of CH is selected, and changed in the next round. By doing this, the energy cost of being a CH is balanced across the network, because each CH has to receive and aggregate data from all the nodes of the cluster, and then send it to the Base Station or sink, which consumes the most energy, especially if the BS is far from the CH.

It is considered a distributed method because each node calculates a parameter known as T, defined in Equation 1. After the computation, the node selects a random number between 0 and 1. If the number is less than T, the node becomes a cluster head, and will communicate that fact to their neighbors, allowing them to join him by proximity.

$$T = \frac{P}{1 - P * (r \bmod \frac{1}{P})} \quad (1)$$

Where:

- T is the probability of being a CH.
- P is the desired CH to nodes rate.
- r is the current round.

Once the nodes have selected their CH, they send their measurement to them. After receiving all the data, the CH aggregates it and send a packet to the sink.

The order in the communication is stablished with Time Division Media Access, that means that each node has a time slot in a round to send their data, as we can observe in Figure 3.:

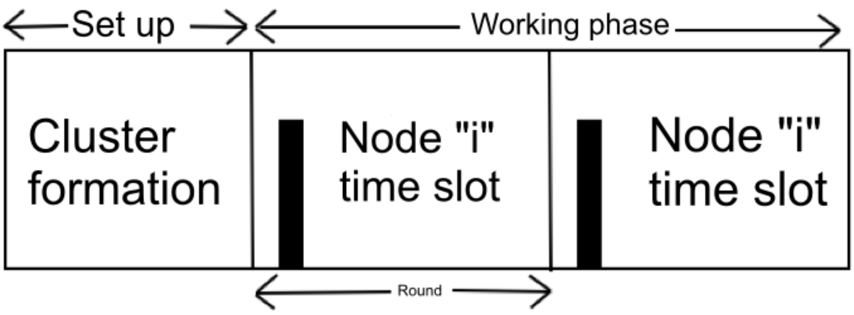

Figure 3. TDMA example.

2.2.2 LEACH-C

The method reviewed before, has a famous centralized variant known as LEACH-C [14]. In this case, the base station is in charge of selecting the different CH, so LEACH-C is a **static stochastic and centralized clustering method**. To do that, the BS is able to compute all the *T* parameter for the different nodes, and knowing the position of the network, it defines which cluster each node must join. Then, the Base Station broadcasts a message in which appears the configuration of the network, telling each node their CH and their TDMA position.

With this change, the authors provided a way to save energy, because the node do not need to broadcast a message with their parameter, and also do not need to listen all the adjacent nodes, which also consumes energy, it only needs to listen the configuration message and then, send or receive data from the nodes.

2.2.3 Static based on Type-1 Fuzzy system

In this category, Cluster Head Election using Fuzzy logic [15], also known as CHEF is going to be presented. This method is implemented in a **centralized** way. In this case, the Base Station needs to know for each node three parameters that will be used as input for the fuzzy system:

- Energy for each node.
- Concentration: amount of nodes that are near of the desired node.
- Centrality: known as the total amount of distances from the node to the rest of the network.

With all these parameters ready, the system provides a list of scores for each node, that will be useful to sort them. In this case, the node with the highest score is the one chosen to be the CH, which means that all the nodes in the network will send their data to it, and it will aggregate them into a single message that will send to the sink.

2.2.4 Static based on Type-2 Fuzzy system

The next method is known as Cluster Head Enhanced Election Type-2 Fuzzy Algorithm [16] or just CHEETAH. In this case the authors choose a **Type-2 Fuzzy** because it is easier to design rules with the knowledge base and also fits better to the application in any case. The difference is that each node runs the Fuzzy system by their own, with the following inputs:

- Energy.
- Distance to the BS (obtained with the signal level).
- Number of nodes subscribed when the node has been CH.
- Efficiency, that is the energy used being CH compared to the energy wasted if there were no CH.

The output is used as a parameter, so the node chooses a random number, and compares to the output, and if it is lower, it will be a CH, as LEACH does, following the same methodology in the communication.

2.2.5 Static based on Type-2 Fuzzy system and centralized

This method, called Centralized Skip Base Algorithm (CSBA) [17] proposes a centralized Fuzzy system that is run by the BS. This means that it has knowledge of the current status of the network and each node. In this case, authors include a new parameter that they call "*Skip*", explained later. The inputs of the Type-2 fuzzy system are explained next:

- Energy represents the current energy of the node, estimated by the BS.
- Normalized distance to the BS.
- Energy ratio compare to the rest of the network.
- Number of rounds without being a CH.

After knowing all the scores for each node, the BS computes to which CH each node must join, done by proximity. With this information, the BS computes the *Skip* parameter. It represents the number of rounds that the worse node will remain without die. In other words, it computes how many rounds can stay as CH. It is compared to a manual threshold stablished before running the network.

After that, the BS communicates the topology and the value of Skip, and works until that value is reached in terms of rounds. With this, the network saves a lot of energy that would be made with control packets, and also uses more time good nodes, so the energy is distributed better, showing a really good performance in their results.

2.2.6 Static based on Particle Swarm Optimization

In this method [18] the **PSO algorithm is used at the Base Station**. It uses the energy of the node as a threshold to decide if it goes in the computation or not. An initial particle is randomly placed and used as solution to choose the CH. The authors defined a cost function taking into account the sum of distances of the nodes to its CH, the inter-cluster distance and also the energy use of the current set up.

After obtaining a solution, like in most of the centralized methods, the BS broadcasts the role of each node, defining the topology of the network, using also a TDMA algorithm to enable the nodes to communicate without disturbing each other's.

2.2.7 Static based on Ant Colony Optimization

This algorithm has a peculiarity, because it is based on the LEACH clustering, so the CH are elected randomly, but the **ACO** is applied to solve the problems that this method presents [19]. They use this optimization to find an optimal path for each CH to arrive to the sink using the minimum energy. Here is introduced the multi-hop communication. With this method the authors are able to avoid the increase of energy waste due to long distance communications, but they do not solve the energy hole problem, which means that nodes near to the BS will die fast compare to far nodes, because they have an important amount of traffic during their work as CH.

2.2.8 Mobile stochastic and distributed method

This method, known as EERA is proposed in [20] , introduces the concept of **mobility,** in the base station specifically. In this case, the clustering is made dividing the field of simulation in sectors, depending on how many clusters do we want. After that, every sensor calculates a parameter called Weight, based on their energy, and broadcasts it. This organization is used until one of the nodes of the cluster has a weight value higher than the cluster head, saving in that way much energy that would be wasted otherwise with control messages.

After that, the nearest CH to the sink, will receive all the information got from the other clusters, creating a path from those who are far from the sink. It is important to know that the sink is moving with constant speed in circles. Doing this, it can balance the energy use from every cluster, so each round the path changes in function of the position of the sink.

2.2.9 Mobile stochastic and centralized method

In this case [10], the authors propose a model where the BS decides which nodes are those that collect the data, in a phase called pre clustering, meaning that is a **centralized method**. This pre-clustering is made in base of the energy of the nodes, in order to obtain a parameter and then, with a random number obtain the pre-clusters. After this, the BS will communicate the results, as usual, and will start the movement, following the Hilbert Curve. It divides the network virtually based on the path taken, and in each vertex of the curve, will stop, and collect the data from the CH that are close to that stop.

2.2.10 Mobile based on Type-1 Fuzzy

This method [11], as well as before, proposes a system using a **Type-1 Fuzzy** model. It is important to mention also that the authors introduce a new element called Super Cluster Head, explained later. The cluster formation is done by using LEACH approach, which means using a random number to determine whether is CH or not. To select the SCH commented before, a Fuzzy system is used with thee inputs:

- Energy
- Distance to BS
- Centrality

After selecting a SCH, all the CH send the data to that SCH that is now able to send the data message to the sink. In order to avoid holes of energy, the BS will be moving following two paths proposed by the authors, one in star and one in spiral.

2.2.11 Static based on Particle Swarm Optimization

At the start of this method [21], all the nodes in the network send their information to the BS, including position and energy. With these parameters, the BS starts running the **PSO algorithm,** in such a way that it obtains virtual lines to divide the network, that will be the clusters.

To select the CH, in each cluster, the nodes send a parameter taking into account their distance to the centre of the cluster and their energy. The best node is chosen to receive the information and aggregate it. Now is the turn of the BS to move. It requests with a Hello packet the average energy of the near clusters. Then, It will move to the one with the highest energy, until all the CH are visited.

2.2.12 Static based on Ant Colony Optimization

Here [22], the model is based on a fixed division of the network by doing squares. After that, the node with the less distance to the center and with the maximum energy, is chosen as CH. After the node is a CH, it will be CH until its energy remains above a threshold. When that threshold is reached, the process starts again inside the cluster.

Then the authors use the **ACO algorithm** to solve a Travelling Salesman Problem, to determine which of the CH is the best option to visit in each moment. It takes into account the number of nodes in the cluster, the scores of the near clusters and also the remaining energy of the cluster.

2.3 Summary

After reviewing some of the studied methods, it is observable that many of them solve an specific problem, but creates new ones. One problem that is observable, but not much of the authors take into account is the amount of energy that the BS need to use in the case of the mobile methods, or the time that it needs in for example ACO or PSO systems.

Static methods show good results in scenarios where there are no large distances and the nodes are distributed near the BS. That is the reason why the mobility was introduced as a concept, in a try of solving the distance problem, when a system that offers good results, like Fuzzy in terms of time, energy and metrics, is used in a larger map.

In the Table 1 we can find a summary with the advantages and drawbacks of the different features of the methods.

Table 1. Taxonomy summary.

Features	Strength	weakness
Static	Small maps	Big maps
Mobile	Big maps	Energy
Stochastic	Simple	Bad results
PSO/ACO	Good results	Time/Energy
fuzzy	Adaptability	Need expert system
Distributed	Simple	High overload
centralized	Low overload	Distances

3 METHOLODOGY

After giving a vision about the problem and the different proposed solutions, it is going to be presented the steps followed in order to design and implement a new solution as well as the procedure to simulate and compare the results with the previous one obtained. In this case, it is going to be presented the procedure that the network is going to follow each round until the end, focusing in the communications and also in the development of the Fuzzy system. After having selected the CH, the BS will move each round to cover different sectors in different rounds.

3.1 Method description

In this section the proposal of this work is going to be described. In this case it is selected a distributed method, based on an Interval Type-2 Fuzzy logic system , in which, as a new feature, the base station is able to support mobility, with different paths, so it can offer a comparison between them and take the best one for each situation.

In Figure 4 it is presented the flow diagram of the method that represents how the network must work. After deploying the nodes, randomly most of the times, the BS will exchange control messages with the nodes to help them to know their position with respect to it. Then, the nodes will run the Fuzzy system, explained later, in order to choose which nodes are going to be CH for the current round.

When a node is selected as CH, it broadcast a control message that fact with a message in a range that the nodes can listen to it. Nodes that are not CH, receive those messages and choose the one with the higher RSSI as their CH, joining them for that round. This means, that while running the TDMA mechanism (described before in LEACH section), this node will have a time slot to send its measures to the CH selected. If a node does not receive any CH advertisement, it sends the message directly to the BS, which usually implies more energy than a transmission to a CH, so it is important to avoid that situation.

After receiving the information, the CHs need some extra energy to aggregate the data (this mechanism will be explained later, in section 4.1) in order to save energy when sending the information to the base station. Then, each CH sends its aggregated message to the BS, that will start a new round if there are alive nodes in the network.

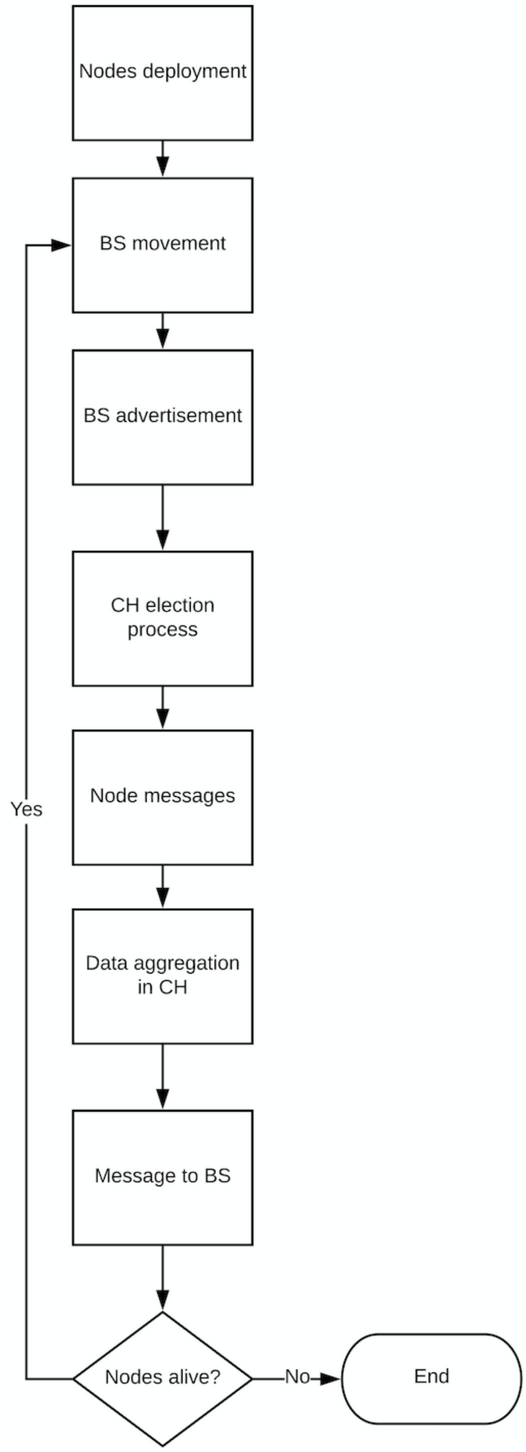

Figure 4. Method flow diagram.

The reason of using artificial intelligence is because of the high degree of uncertainty that appears in telecommunication systems, which make it suitable for a wide range of different applications, such as wireless networks without infrastructure [23][24], grid computing [25], smart cities [26], vehicular networks [27], wireless security [28] and so on.

Specifically, in WSNs it can be used in different fields not only on clustering [32], it is used also in node location [30],best gateway election [29], security [31], etc.

Fuzzy logic is mostly used in situations related with controlling a parameter of the system: waste reduce, flux control and so on. This type of intelligence should be used when traditional mathematic functions do not fit with the specifications, so it cannot be model the behavior of the parameter with a function. That is other reason of using this type of systems when the relation between input and output is not clear, are difficult to stablish or extremely complex to calculate. Fuzzy logic is a good option in WSNs due to the high uncertainty in the parameters. Inside this technology, Type-2 [33] works better compared to its predecessor, Type-1, when the information is imprecise, unreliable and incomplete [34]. In the presented algorithm, the information is taken by each node in their own, like for example the distance between nodes is a parameter that cannot be computed precisely, even if using GPS a little mistake would be produced. In this case, it is measured through a wireless communication, that has a degree of uncertainty [35]. Another parameter is the estimated energy, that is an approximation that the node does in his own, that is not exact [36]. The last reason of using the Type-2 fuzzy logic over the Type-1 is that it is easier to design the rules of the system from the expert knowledge, which increases robustness, and also it has better adaptability due to the complex relationships established between inputs and outputs [37].

Compared to ACO and PSO algorithm, Fuzzy logic has a really important advantage, and it is the time that it takes to find a solution. In problems like this one that it aims to find a good path, ACO could be a good option if there were no time restrictions, but it takes a long time to find the optimal solution, so the data cannot be collected on time, that is the aim of clustering in this type of networks, collect data from the entire network constantly. Also, the implementation in this case would take a little portion of storage, as previously said, Type-2 logic is able to be implemented with a matrix, so the node only has to map the position in the four dimension matrix, and obtain the output, while the other mechanism need to run them and compute solutions until obtain a good one, spending energy and processing resources.

To select the CH, each node runs its own Interval Type-2 Fuzzy system. Those systems have the architecture presented in Figure 5, which consists of four blocks:

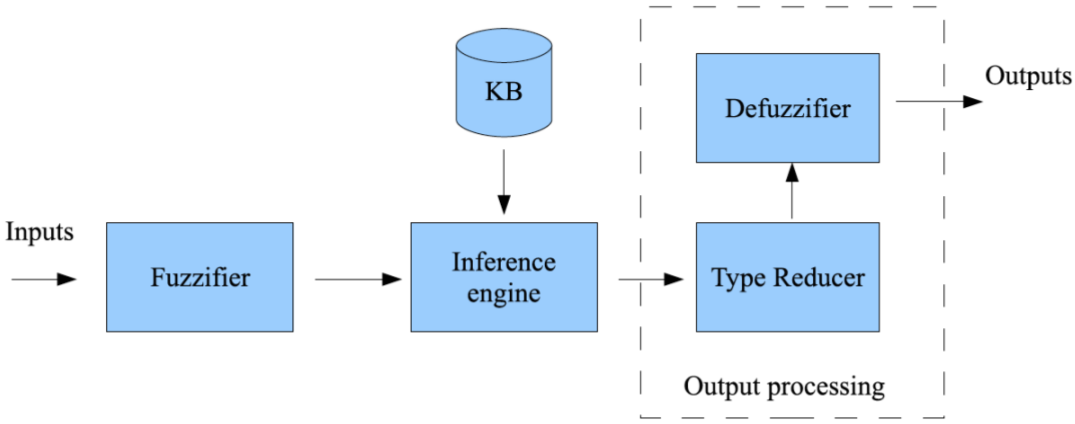

Figure 5. Fuzzy system.

- Fuzzifier: This block receives the inputs of the system and is in charge of normalizing those values, so in that way they are in an interval that is appropriated for the inference engine.
- Inference engine: In this case, the engine is able to obtain a solution from the normalized inputs, taking into account the knowledge base.
- Knowledge base: This block contains the brain of the system. Here is where the previous knowledge in the field has to appear in order to create a set of IF-THEN rules, that join inputs and outputs.
- Defuzzifier block, produces an output interval in the case of Type-2 fuzzy. These values are de-normalized in order to obtain a crisp interval.
- Type reducer, this block uses the Karnik-Mendel Algorithm in order to solve the problem [39]. It takes a long time.

Next, it is necessary to define the set of variables that will act as inputs of the system. It is important to select those variables in such a way that they fit and give to the best nodes, the best score so the network is able to live longer. Taking into account the studied bibliography and the previous knowledge [16], the selected variables are the following:

- ***R*** is defined as the Residual Energy of the current node, measured with respect to the initial energy, in other words, the percentage of energy.
- ***D,*** defined as the distance to the BS at the start of the network lifetime. It is measured through the received signal strength indication (RSSI) obtained from a message sent by the BS.
- ***H*** is the historical contribution when being a CH. This variable measures the number of nodes that contribute to it when it acts as CH. By doing this, the nodes that show a good behavior in terms of number of nodes collecting data, will be preferred in future rounds. To compute this value, it is presented Equation 1:

$$H = \frac{\sum_{i=0}^{r} h_i}{r * k * n} \quad (1)$$

Where:

- r is the number of rounds where the node has been CH.
- h_i represents the number of nodes that chose the node as CH in round i.
- k is defined as the optimal factor for the number of CH with respect to the total number of nodes. It was defined and computed in LEACH, with a value of 0.05, which means that for each 100 nodes, there should be 5 CH.
- And n is the total number of nodes of the network.

This value has a maximum defined as $k * n * 1.5$, in order to normalize in the fuzzifier.

- ***E*** is defined as the efficiency. In this case it is defined as the energy that the network would spend if the nodes should send their message to the BS directly instead of to the CH. This value is updated when a node is selected as CH, following Equation 2:

$$E = 1 - \min\left(\frac{\sum_{i=1}^{h} e_i}{\sum_{i=1}^{h} e'_i}, 1\right) \quad (2)$$

Where:

- h is defined as before, the number of nodes that chose the current node as CH.
- e_i is the energy spent by node i when sending a message to the CH.
- e'_i is the estimated energy that the node would spend if it sends the message directly to the BS instead of to the CH.

It is important to keep in mind, that due to the restrictions of the system, it is impossible to know the real value for the parameter E, because it is computed in base of the RSSI received from the BS. This parameter, shows that a node that saves low energy, is a bad node to act as CH, and therefore will obtain a low efficiency.

All the variables have a similar layout, presented in Figure 6 in order to map the input value into three groups: low, medium and high. Due to the nature of Interval Type-2 Fuzzy systems, it accepts some uncertainty defined in previously mentioned layouts, represented by the width of the area of the function, in such a way that the wider the function is, more uncertainty the variable has, showing robustness to some possible noise that could appear in the system, like the commented before due to the base of the computation of the parameter E.

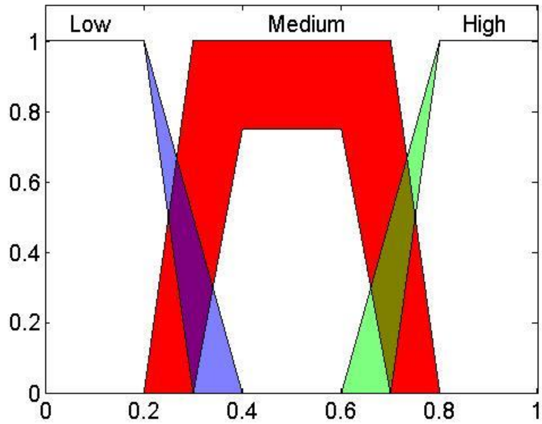

Figure 6. Fuzzy input and output layout.

When talking about the output, there are defined two more groups, very low and very high, in order to have a better performance and location of the results. There are 81 rules defined in the knowledge base, trying to define the expert knowledge in a “IF – THEN” format. These rules are modeled with two matrixes, that map the four input variables into the output interval, helping in this way to the software a much faster simulation, compared to the time of a Fuzzy implementation on MATLAB, and being similar to the real implementation in a node, reducing drastically the battery needed to compute the solution.

A few rule examples could be the following:

- **IF** *R* is low *D* is high *H* is low *E* is low **THEN** the output is very low.
  This rule represents a non-desired node to be CH, because it is far from the BS, has low energy, and it is not efficiency in terms of nodes attached and energy saving.
- **IF** *R* is high *D* is low *H* is high *E* is high **THEN** the output is very high.
  In the other hand, it is defined this rule, representing the most desired node, because it is almost full energy, is near the BS (saving energy in the message broadcast), and has a lot of neighbors, increasing in that way the energy saved if it is selected as CH.
  In Table 2 it is shown the entire description of the rules of the system, where L means low, M medium, H high, VL is very low and VH is very high:

Table 2. Fuzzy rules.

Rule	*R*	*D*	*H*	*E*	*score*
1	L	L	L	L	L
2	L	L	L	M	L
3	L	L	L	H	M
4	L	L	M	L	M

5	L	L	M	M	M
6	L	L	M	H	L
7	L	L	H	L	L
8	L	L	H	M	M
9	L	L	H	H	M
10	L	M	L	L	VL
11	L	M	L	M	VL
12	L	M	L	H	L
13	L	M	M	L	VL
14	L	M	M	M	L
15	L	M	M	H	L
16	L	M	H	L	VL
17	L	M	H	M	L
18	L	M	H	H	L
19	L	H	L	L	VL
20	L	H	L	M	VL
21	L	H	L	H	L
22	L	H	M	L	L
23	L	H	M	M	L
24	L	H	M	H	L
25	L	H	H	L	L
26	L	H	H	M	L
27	L	H	H	H	M
28	M	L	L	L	L
29	M	L	L	M	M
30	M	L	L	H	H
31	M	L	M	L	M
32	M	L	M	M	M
33	M	L	M	H	H
34	M	L	H	L	M
35	M	L	H	M	H
36	M	L	H	H	VH
37	M	M	L	L	L
38	M	M	L	M	M
39	M	M	L	H	M
40	M	M	M	L	M
41	M	M	M	M	M
42	M	M	M	H	H
43	M	M	H	L	M
44	M	M	H	M	H
45	M	M	H	H	H
46	M	H	L	L	L
47	M	H	L	M	L
48	M	H	L	H	L
49	M	H	M	L	VL

Rule	*R*	*D*	*H*	*E*	*score*
50	M	H	M	M	L
51	M	H	M	H	M
52	M	H	H	L	M
53	M	H	H	M	H
54	M	H	H	H	VH
55	H	L	L	L	VL
56	H	L	L	M	VL
57	H	L	L	H	L
58	H	L	M	L	VL
59	H	L	M	M	VL
60	H	L	M	H	L
61	H	L	H	L	VL
62	H	L	H	M	VL
63	H	L	H	H	L
64	H	M	L	L	M
65	H	M	L	M	M
66	H	M	L	H	H
67	H	M	M	L	M
68	H	M	M	M	M
69	H	M	M	H	H
70	H	M	H	L	M
71	H	M	H	M	H
72	H	M	H	H	VH
73	H	H	L	L	VL
74	H	H	L	M	L
75	H	H	L	H	L
76	H	H	M	L	L
77	H	H	M	M	L
78	H	H	M	H	M
79	H	H	H	L	M
80	H	H	H	M	M
81	H	H	H	H	H

When every node has run the fuzzy system, it is the moment where in basis to the obtained score, the CHs should be chosen. The first step is done already, from which it obtained two values, the high and the low probability, that are the limits of the interval obtained from the de-fuzzifier. After that, it runs a random function that provides the node with a random number between 0 and 1. If the node has been CH in the round before, it will compare its random number to the lower output, meanwhile if it has not been a CH, will check if the number is lower than the higher output, giving more chances of being CH to a node that has not been CH before, and balancing the energy use. In Figure 7 it is presented the flow diagram that a node follows in order to be or not CH:

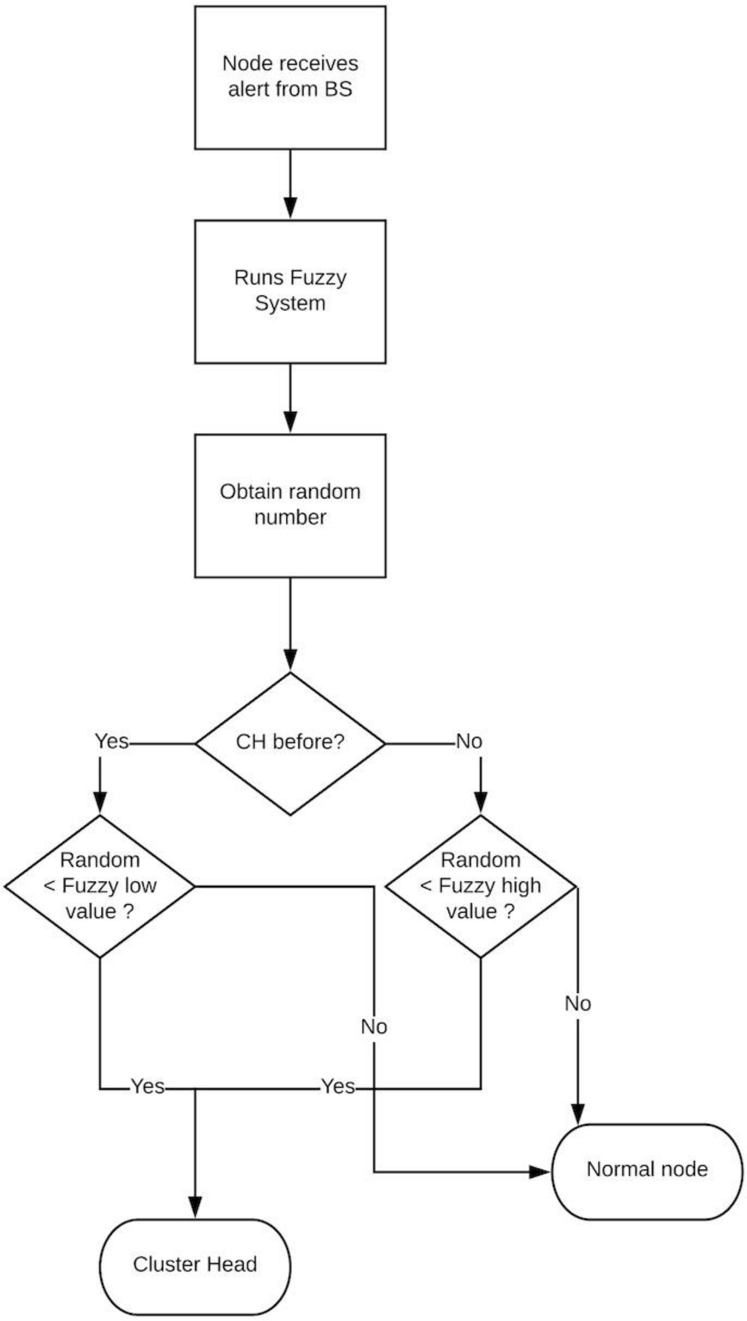

Figure 7. CH election flow diagram.

3.1.1 Mobility support

As previously said, in this work, it is introduced the possibility of supporting mobility in the side of the BS. After exploring the different options that the bibliography offers, three types of paths have been chosen:

- **Random movement:** For each round, the BS computes a random point in the map and moves to it, like in Figure 8. Random movement.:

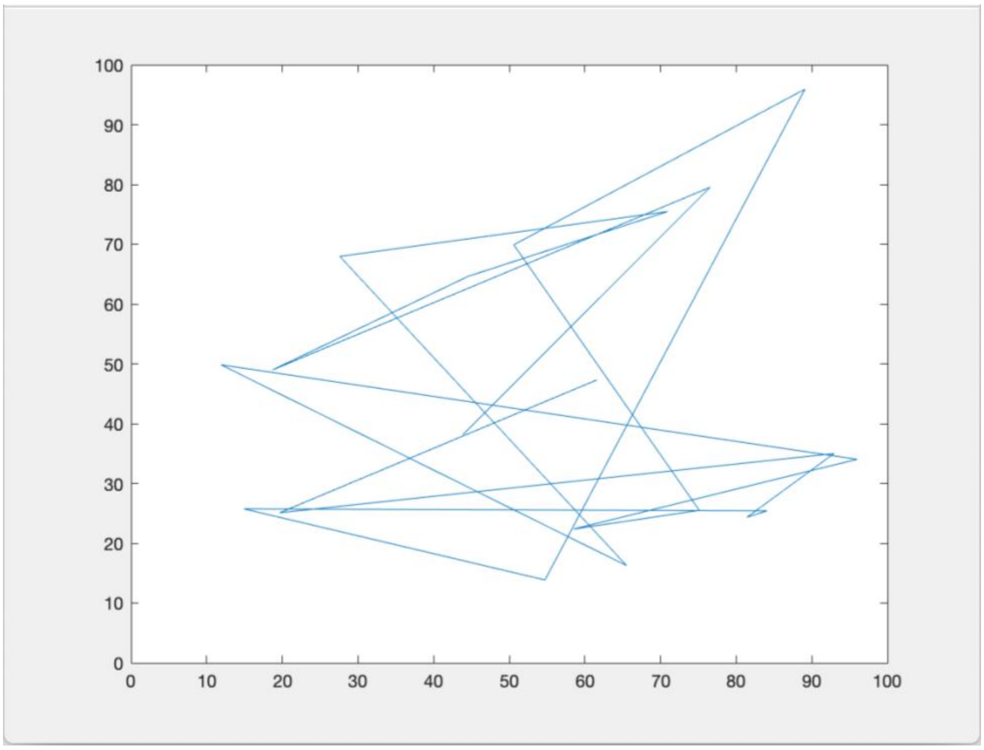

Figure 8. Random movement.

- **Circular movement:** In this case, the BS follows a movement with a determined speed, taken from the bibliography in order to cover the most possible area during its movement, as in Figure 9:

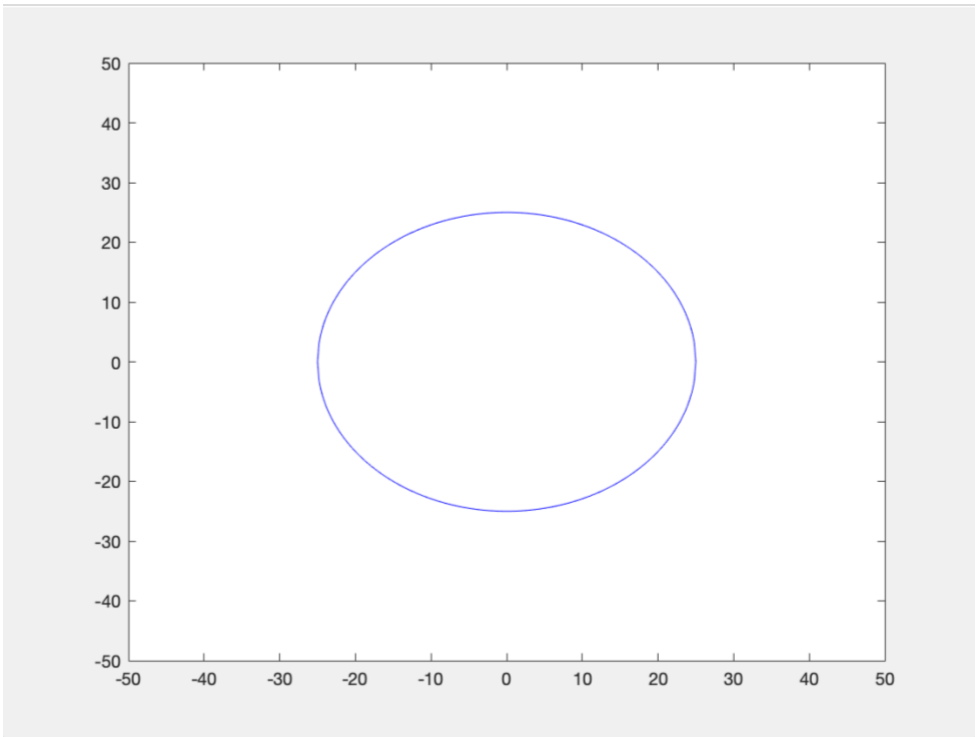

Figure 9. Circular movement.

- **Squared movement:** This one is made by defining a virtual grid, that divides the map into sectors, in this case, 4 sector were taken. During the movement, the BS visits the centre of each sector, as shown in Figure 10 :

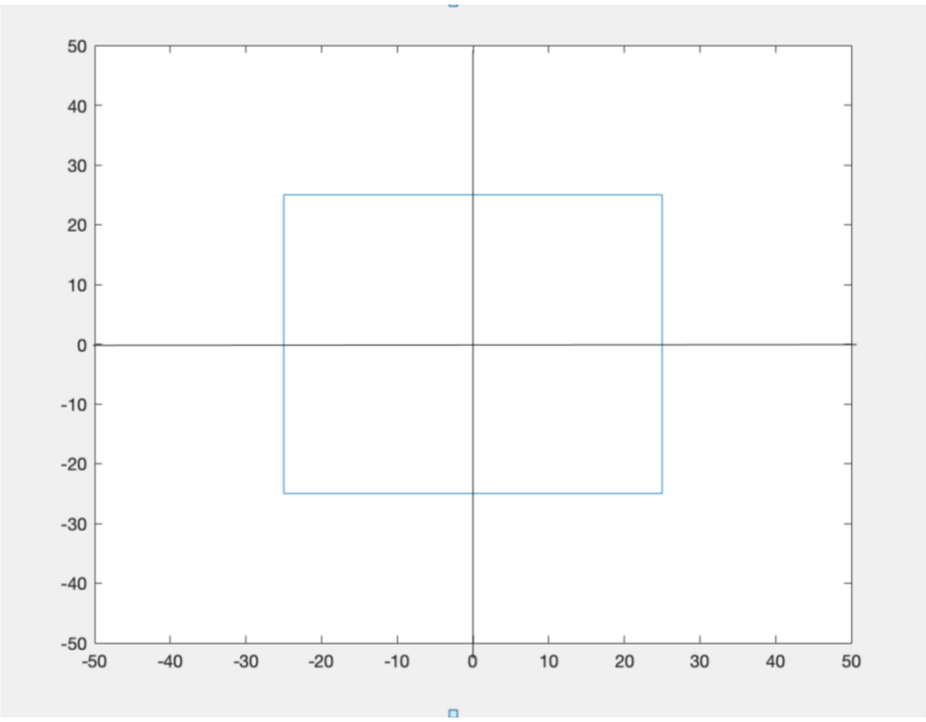

Figure 10. Squared movement.

After each movement, the BS must indicate its new position, to realize this task, it must send a new packet, in such a way that the nodes are able to compute the new RSSI level, updating their inputs for the Fuzzy system, and also knowing the strength of the signal that they need in order to send the data packet if they are CH or they are not attached to any CH.

3.2 Justification

In this section, the main elections done during the design of the method, are justified attending different reasons. The main features of the system, are the Type-2 Fuzzy CH election and the mobility of the BS. These two features are going to be justified in order to understand why are they selected in base of their advantages or their use in the bibliography.

First, a good election of the CH is a key moment during the process as seen before. It means that the design of the system and the variables chosen are important to ensure a good performance of the network. In this case, the inputs of the system are selected because of their direct impact in the battery use of the nodes. In the case of the variable that represent the energy is clear, because the system prefers a node with high energy, so nodes with low battery do not need to receive and send data to the BS or aggregate data. The second variable also has a huge impact on the battery, the distance to the BS. Nodes that need to send a message to the BS that is far from them, the energy that consumes to send it is huge. The variable that collects the history of the node as CH is used to award nodes that ensure a good behavior. The last variable, the efficiency, one more time is directly related with the energy, because a node that saves a lot of energy compare to the direct send, should be good for the network.

When talking about mobility, after studying the bibliography and methods proposed by different authors, this kind of mobility, in which the path is predefined offers the networks some features. The first is the simplicity, this means that the BS does not need to implement an additional module to compute the mobility and support it, in the case of the square or the circle it can be added with an additional infrastructure, that makes easier the movement, and provides an energy saving, because if the BS computes the path with another parameter, such as using an ACO or PSO, the path will be computed for each round based on the current situation of the network, so it is needed extra energy to move to the next point. Also, the use of algorithms like the commented before, shows the same problem as in clustering, the time spent. If the BS needs to know the situation of each node, first it needs to receive the energy and the parameters that model each node in each round, increasing the energy wasted in control message and the time of sending them. Moreover, the BS needs a period of time to find the solution of the optimal point to visit it, then communicate with those nodes, and so on. It loses the point of the WSNs that is collecting data in “real time”. That is the reason of using a simple mobility model, the resource saving, because that is one important advantage of this type of networks, the use of a small amount of resources and obtain a lot of information from the environment.

3.3 Simulation software

Simulation software is a key part in the process of comparing two models. It has to allow the researcher to define their proposal perfectly, in terms of communications, node behaviour, energy and radio model. Taking into account the studied bibliography, two software appear in a high percent of the papers or articles, MATLAB and Network Simulator version 2 (NS2).

NS2 is an open source software design to implement network topologies and simulate the behaviour along the time. In this field, the researcher is able to define the radio model and the energy consumptions by scripting. It is recommended when the CH election process is stochastic, which means a simple behaviour and processing. Due to the nature of this work, a software that allows the implementation of fuzzy systems and process them without problems is needed.

MATLAB is the preferred option by most of the authors of the field, due to its flexibility. It is easy to code and work with large amount of data, in a matrix format. Also it provides some toolboxes that make easier the implementation of artificial intelligence, that is the case of Fuzzy, ACO and PSO toolbox. Due to its flexibility and native support for the features commented before, MATLAB will be the selected software for this work.

4 RESULTS

In this section the results obtained from the simulation are going to be described. The simulations were made taking into account the description of the methods in order to be implemented according to their specifications. In this case, the proposed method is compared to the methods LEACH, CSBA, EERA and CHEETAH. Additionally, it has been develop a modified version of the proposal, in which the nodes that are near the BS are able to send directly the message to it.

In the following sections there are presented the conditions taken to obtain the results, being important to describe the communication model, the scenarios used and finally the parameters selected to simulate the different parameters. After this, the results for each type of map are presented.

4.1 Communication model

After explaining the behaviour that the network is going to have, in this section it is going to be presented the elements that appear during the communication process of the network, in this case the radio model.

So the mechanism would be, receiving the information from all the nodes that belong to the cluster, aggregate the data reducing a lot the total size of the information, and sending that packet to the BS.

The next step is to define the model that is used to simulate the communications between the nodes. In this case, the most used model is the one known as **first order radio model [13]**, due to its simplicity and it represents very good the reality. This model has the following parts shown in Figure 11:

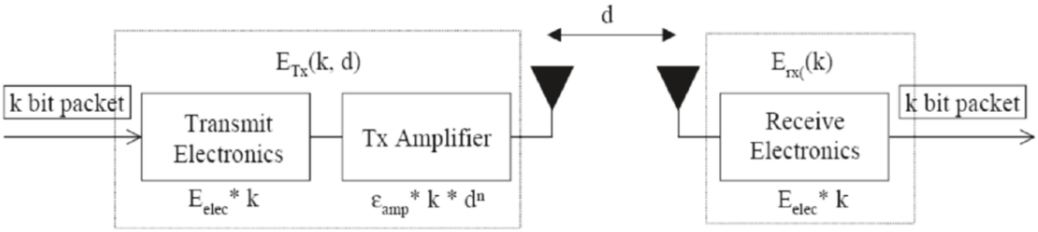

Figure 11 . First order radio model.

- Transmitter
- Amplifier
- Receiver

In this model, transmitter and receiver are considered symmetric, this means that they spend the same energy during their processes. Their consumption only depends in their architecture, parameter E_{elec} and the number of bits that pass through it, known as *k*.

In the case of the amplifier, its consumption depends one more time on the architecture, represented with the parameter ε_{amp} but also and more important, on the distance that the packet must cover, represented by d_a. This means that depending on a parameter, defined as d_0 defined in Equation 3, a could take two values, defined in Equation 4 and 5:

$$d_0 = \sqrt{\frac{Eelec}{Eamp}} \tag{3}$$

$$E_{TX} = E_{elec} * n + E_{fs} * n * d^2 \text{ when } d_0 > d \tag{4}$$

$$E_{TX} = E_{elec} * n + E_{amp} * n * d^4 \text{ when } d_0 < d \tag{5}$$

$$E_{RX} = E_{elec} * n \tag{6}$$

$$E_{RX-cluster\ head} = (E_{elec} + E_{DA}) * n \tag{7}$$

At this point, the clustering takes a big importance due to Equation 5, so if the network avoids a communication using that distance, the energy used will be really smaller. In Equation 6 it is represented the energy wasted by a node when receiving a packet. If that node is CH, the consumption of the data aggregation must be added, so in Equation 7 it is represented.

Typical values for these parameters, and the one used in this work are presented in Table 3:

Table 3. First order radio model typical values.

Parameter	Typical value
$\boldsymbol{E_{elec}}$	$50\ nJ$
$\boldsymbol{E_{fs}}$	$10 \frac{pJ}{bit * m^2}$
$\boldsymbol{E_{amp}}$	$0.0013 \frac{pJ}{bit * m^4}$
$\boldsymbol{E_{DA}}$	$5\ nJ$

4.2 Simulation scenarios

One feature of the WSNs, is that the nodes, in general, are not placed following any deterministic function. This means that the nodes are randomly placed across the map. Due to this fact, nodes do not know in advance their location, and that is why they must communicate with the BS to know it. For this work, it was agreed to use three types of network layouts, or maps, according to the bibliography and previous research:

- The first scenario includes a set of 30 square area maps of 100 meters wide with 100, 150, 250 and 350 nodes respectively, shown in additional results.
- The second also includes the same amount of maps, but with 200 nodes and 200 meters wide.
- The last, has the same features, but larger, 300 nodes and 300 meters wide.

As a novelty compared to previous work, the maps only differ in their size and number of nodes, not in the position of the base station. This is due to the mobility, so in every map the BS starts in the centre of the map and follows the commented path for each case. The increase in the size of the map, is in an attempt to check if the mobility is a good solution for applications that need to cover a larger area and expecting that with mobility the distance that the message has to cover will be reduced.

4.3 Simulation parameters

As previously said, the different proposals have been tested in 180 maps, using the same set of maps for every method, because a map could be better for a determined method, so in this way it is removed the map effect that could appear during the simulation.

To compare the performance of the methods, five parameters are taken, that are representative of the performance of the network:

- Last Node Death (LND), represent the round in which the first node depletes it battery (also referred as death). This moment is known as the stability period for many authors and is one significant metric, because until this point the network takes precise measures.
- Half Node Death (HND), collects the round in which the network only has the half of the nodes active (or alive). At this point the collected data is still useful but it is not as accurate as before.
- Last Node Death (LND) is the moment where the proportion of deaths is high and the network is considered to be useless. In this work, this moment is when the 90% of the nodes are death.
- Distance covered by BS. This parameter is included to compared if the BS moves a lot or not, because some methods show good performance but making

the BS cover long distances each round, so it has to use a lot of energy, that not so many authors take into account but in reality it would be a problem.

- No CH rounds (no-CH). This parameter is taken in order to represent the number of rounds in which there is no CH selected in the network, which means that all the nodes need to send directly the message to the BS, so the clustering is not working in that round.

As commented before, there were selected four methods, which have been implemented with the following parameters:

- LEACH, using ratio of CH to nodes of 5%.
- CHEETAH uses the same threshold as in the paper, as well as the probability commented before.
- CSBA is used with a skip = 19, that is the optimal value given in the paper.
- EERA is used with 5 equal sectors.
- The modified version of the proposal does send directly to the BS if the distance to it is smaller than d_0, defined in (3).

4.4 Small size maps

As previously said, this set of 30 maps in which 100 nodes are randomly placed across a square area of 100 meters wide. For the methods that do not include mobility, the BS will be placed in a corner of the map, in this specific case in the point (100,0) m. In .

Table 4 the results are presented with the described metrics.

Table 4. Simulation results for the first set.

METHOD	FND	HND	LND	noCH	Distance
LEACH	752,7	902,7	1003,4	34	0
CHEETAH	936,5	1140,7	1298,5	25,2	0
Proposal (random movement)	951,2	1144,4	1260,5	14,2	6434,9
Proposal (grid movement)	944,4	1117,3	1236,6	14,3	6185,9
EERA	1509,5	1630,2	1751,6	0	104811
CSBA	1701,8	2148,4	2409,6	0	0
Modified proposal (random)	2165,1	2680,1	3174,7	2014,7	16644
Modified proposal (grid)	2418,7	2942,9	3460,6	2594,8	17306

We can observe that, in a first instance that the method outperforms LEACH and CHEETAH, both distributed and static methods, in terms of lifetime and also in terms of the rounds that there are no CH, so there are less rounds in which the nodes use a high amount of energy to send the message directly to the BS. Comparing to the method based on mobility, EERA, it is clear that the method has a worse performance, but observing the distance covered by the BS in this method, it is higher than the performance obtained by the method. This implies that the BS will need much more energy per round, as it is proved if it is calculated the ratio between the distance covered and the rounds, as in Equation 8.

$$Proposal_{ratio} = \frac{6434}{1260} = 5{,}1 \, , EERA_{ratio} = \frac{104811}{1751} = \; 59{,}85 \qquad (8)$$

So it is more than 10 times higher, but the number of rounds it is not 10 times higher, this means that the efficiency obtained is worse than the proposed method. CSBA is the only centralized method of the proposed, so it has a better know of the network topology and current state for each round. It shows a better performance without moving the BS even though it is on a corner.

Looking at the results of the modified version, it is observable that has the best results in every metric regarding lifetime. The reason is that due to the movement of the BS, if it is in the range of the node, it can be better to send to it directly, saving energy in the CH and also in the node, because it does not need to send to the CH that is far from it. Inside the proposal, the grid path has better results, because it is in the half of the map most of the times, so there are higher possibilities of being near a node compared to its CH.

4.5 Medium size maps

In this case, the second set is the one with 200 nodes and an square area of 200 meters wide. The results are shown in Table 5:

Table 5. Second set results.

METHOD	FND	HND	LND	noCH	Distance
LEACH	235,1	577,2	730,6	10,83	0
CHEETAH	136,4	238,13	292,76	2,5	0
Proposal (random movement)	138	231,9	285	1,63	30044
Proposal (grid movement)	141,3	232,1	285,03	1,63	14378
EERA	210	1239,3	1307,7	0	250822,4
CSBA	541,5	1591,1	1980,8	0	0
Modified proposal (random)	196,2	359,6	489,8	27,6	51429
Modified proposal (grid)	206,1	363,6	505,1	33	25382

It is observable, that due to the increase in the distances, the nodes will need a larger amount of energy to communicate with the rest of the network. The results here are not the expected, because the concept of mobility was introduced in an attempt to cover longer distances, but it is not the case. This is because, being larger the map, the BS will have more terrain to cover, so in a round in where the BS is in a corner, a node in the contrary corner will need a huge amount of energy, due to Equation 5.

One more time, EERA show better results but in this case, the distance covered by the BS increases a lot, being even higher than the previous set, and obtaining worse results. In the case of the proposal with different movements, the grid movement show a better performance, because they have the same metrics, in terms of number of rounds, but the distance covered is lower, so it means that in the same time, it will spend less energy to do the same work. In this case, the long distances also affect the results in the modified version, because the threshold is lower than the distance to the BS in most of the cases, that is the reason why the results are better but not comparable with the rest.

4.6 Large maps

This is the case with will show the performance with higher distances, 300 m of side in this case, and more nodes, 300. In this case, the physical medium would have a high amount of traffic so it would be necessary to define an additional mechanism to do it, or extend the TDMA time slots. In Table 6 are shown the results obtained in this case:

Table 6. Results for large maps.

METHOD	FND	HND	LND	noCH	Distance
LEACH	55	253,7	440,4	4,4	0
CHEETAH	15,5	58,2	94,96	0,3	0
Proposal (random movement)	13,9	55,5	89,5	0,1	14170
Proposal (grid movement)	15,4	55,5	89,86	0,033	4719,5
EERA	45,6	708,9	799	0	22489,2
CSBA	10,1	902,8	1595,9	0	0
Modified proposal (random)	17,9	69,4	114,3	1,5	18094
Modified proposal (grid)	16,6	69,1	114,6	1266,7	5959,5

The reduction of the performance of the methods is generalized, due to the increase of the distances. In this case, LEACH shows the better performance in the FND metric. This is due to the simplicity that this method has. But increasing in time, the CSBA method is the one that shows better performance with long distances. Looking to mobile methods, one more time the proposal does not fit the expected results, not being able to give a better performance. The reason was exposed before, and this set of maps gives more consistency to it. As it happened before, the modified version do not show better results, due to the long distances.

Additionally, the results confirm the fact that EERA obtain a decent performance using a lot of energy when comparing with the rest of the mobile methods. Comparing the proposed methods, in this case the random has a lower distance covered, due to the nature of the protocol, which computes a random position and visits it, but the performance is near the same to the grid one.

4.7 Additional results

Additionally, three extra simulations have been made in a small map, with 100 meters of side, but with three different number of nodes, specifically 150,250 and 350 nodes. This has been made to check the behaviour of the method if the node density increases. The results are shown in .

It is important to see that in a small map, the modified version has the best results for the different number of nodes. This means that the only problem that this method has is the management of long distances, not the number of nodes in the network. This fact can be used in order to obtain better results.

Table 7, Table 8 and Table 9.

It is observable that every method has a lower performance with the increase of the number of nodes, but it is not a huge decrease. This lead to think that the node density is not a problem in this kind of networks if the method is correctly defined. This is done by using a different number of CH, so the load balance is better.

It is important to see that in a small map, the modified version has the best results for the different number of nodes. This means that the only problem that this method has is the management of long distances, not the number of nodes in the network. This fact can be used in order to obtain better results.

Table 7. 150 nodes simulation results.

METHOD	FND	HND	LND	noCH	Distance
LEACH	719,2	890,6	994,6	19,16	0
CHEETAH	849,5	1037,6	1170,1	9,8	0
Proposal (random movement)	852,3	1008,8	1118,4	4,4	58532
Proposal (grid movement)	853,8	1010,8	1120,8	5,6	56071
EERA	1458,4	1654,8	1692,2	0	59376
CSBA	1694,9	2195,1	2444	0	0
Modified proposal (random)	2071,1	2613,8	3097,6	1778,4	16240
Modified proposal (grid)	2337	2912,2	3416,4	2276,2	17085

Table 8. 250 nodes simulation results.

METHOD	FND	HND	LND	noCH	Distance
LEACH	700,2	839,7	936,5	5,4	0
CHEETAH	731,1	883,5	982,5	2,6	0
Proposal (random movement)	721,7	856,1	938,2	1,4	49167
Proposal (grid movement)	726,6	857,7	941,5	0,9	47104
EERA	1466,4	1501,6	1536,6	0	42956
CSBA	1573,2	2253	2498,5	0	0
Modified proposal (random)	1985,3	2531,6	2997,4	1546,8	15713
Modified proposal (grid)	2201	2875,1	3379	1904	16898

Table 9. 350 nodes simulation results.

METHOD	FND	HND	LND	noCH	Distance
LEACH	659,5	784,4	871,8	0	0
CHEETAH	631,8	742,5	845,7	1,1	0
Proposal (random movement)	624,0	734,5	813,4	0,23	42674
Proposal (grid movement)	625,1	738	817,5	0,46	40906
EERA	1206,8	1367,4	1418,9	0	11327
CSBA	1462,5	2275,6	2514,8	0	0
Modified proposal (random)	1877,8	2428,2	2876,8	1358,6	15084
Modified proposal (grid)	2068,8	2810,1	3311,9	1560,8	16562

5 CONCLUSIONS AND FUTURE WORK

Mobility is an interesting feature for some applications because it allows an increase of the efficiency. In this case, it has been observed that compared to the mobile protocol from the bibliography, it shows a better performance if it is taken into account the energy needed to cover the distance from the protocol EERA.

The main advantage from the method, appeared when the modified version was simulated. It offered better results when the nodes send directly the message to the BS if they are close to it, so the mobile feature is used to take advantage. But the problem appeared when the map was larger. As explained before, the long distances are the main problem of this type of network and this type of maps. The number of nodes do not affect the results obtained with this method, so this mean that is has a high scalability in terms of number of nodes. A good way of solve this problem is to deploy additional BSs which cover a sector each one. By this way, each sector is a reproduction of the small map architecture, in which the method has proven to have good results.

It was observable also that, for this type of maps, the path selected did not have a direct influence in the results, except for the modified version, that the grid path was way better. This lead to think that the selected path only has a direct influence if all the nodes can take advantage of it, not only the CH. This means, that if the BS is moving around locations near to the center, it can be easily accessed each round by a high number of nodes which can send their data to it directly.

From the conclusion obtained above, some ways of thinking are opened in order to be tested in future works, as for example the addition of new infrastructure that act as collectors or secondary BS. By dividing the entire network in sectors, each BS would cover a portion of the network with paths like shown before. This solution would reduce drastically the distance each node would need to cover with their message, and would solve the problem shown in large maps by using the modified version of the proposal method. Also, this increase in performance would be attached to a higher cost of the network, because of additional infrastructure which presumably is more expensive than adding energy to the nodes, for example.

Another solution without adding cost could be using multi-hop schemes. This solution added to the current proposal could solve the problem of the long distance communications. It would consist on using near CH to send the data message, and this CH could aggregate or resend the message. This is a good option, but would add a lot of complexity in terms of communication, and also in energy wasted, because a lot of control message would be needed.

In the case of not needing real time, in different applications a solution with the same implementation would be an accumulative scheme. This mean that each CH stores the data until the BS arrives to a near point, advising with a message and the CH send the data, so the communication range would be reduced.

These solutions are thought in base of methods that were studied in the bibliography. Thinking in the current proposal, some new features could be added.

The first one is to redefine the Fuzzy system taking into account the mobility, this includes a new research to know which parameter could be a good option and also to get the necessary knowledge to define it well. This option also could lead to a redesign of the behavior, having a hybrid method. This option is made by redefining the features of the fuzzy system with the information gotten of a message received from the BS. This could be made by changing the normalization distance in base of the communication or using in a different way the interval gotten by the Fuzzy system.

These options are possible lines of research for future works that could be explored, because this mobility is a relatively new feature in this field, and there are not big advances in this field that could be taken as reference, each author offer a solution that solves a problem but appear different problems, so exploring the different possible options is the way.

BIBLIOGRAPHY

1. WERNER-ALLEN, Geoffrey, et al. Deploying a wireless sensor network on an active volcano. *IEEE internet computing*, 2006, vol. 10, no 2, p. 18-25.
2. AL-FARES, Mohammad S.; SUN, Zhili. Self-organizing routing protocol to achieve QoS in wireless sensor network for forest fire monitoring. In *2009 IEEE 9th Malaysia International Conference on Communications (MICC)*. IEEE, 2009. p. 211-216.
3. MOHAMED, Reem E., et al. Survey on wireless sensor network applications and energy efficient routing protocols. *Wireless Personal Communications*, 2018, vol. 101, no 2, p. 1019-1055.
4. KHEDO, Kavi K., et al. A wireless sensor network air pollution monitoring system. *arXiv preprint arXiv:1005.1737*, 2010.
5. KHAN, Arijit; JENKINS, Lawrence. Undersea wireless sensor network for ocean pollution prevention. En *2008 3rd International Conference on Communication Systems Software and Middleware and Workshops (COMSWARE'08)*. IEEE, 2008. p. 2-8.
6. BOKAREVA, Tatiana, et al. Wireless sensor networks for battlefield surveillance. En *Proceedings of the land warfare conference*. 2006. p. 1-8.
7. KUMAR, Prashant, et al. Underwater acoustic sensor network for early warning generation. In *2012 Oceans*. IEEE, 2012. p. 1-6.
8. SILVA, Ricardo; SILVA, Jorge Sá; BOAVIDA, Fernando. Mobility in wireless sensor networks–survey and proposal. *Computer Communications*, 2014, vol. 52, p. 1-20.
9. ABUARQOUB, Abdelrahman, et al. Dynamic clustering and management of mobile wireless sensor networks. *Computer Networks*, 2017, vol. 117, p. 62-75.
10. PANDEY, Sudhakar; ANAND, Veena. Load-Balanced Clustering Scheme with Sink Mobility for Heterogeneous Wireless Sensor Networks. *National Academy Science Letters*, 2017, vol. 40, no 5, p. 335-341.
11. ALOMARI, Abdullah, et al. Dynamic fuzzy-logic based path planning for mobility-assisted localization in wireless sensor networks. *Sensors*, 2017, vol. 17, no 8, p. 1904.
12. KRISHNAN, Muralitharan; YUN, Sangwoon; JUNG, Yoon Mo. Enhanced clustering and ACO-based multiple mobile sinks for efficiency improvement of wireless sensor networks. *Computer Networks*, 2019, vol. 160, p. 33-40.
13. HEINZELMAN, Wendi Rabiner; CHANDRAKASAN, Anantha; BALAKRISHNAN, Hari. Energy-efficient communication protocol for wireless microsensor networks. In *Proceedings of the 33rd annual Hawaii international conference on system sciences*. IEEE, 2000. p. 10 pp. vol. 2.

14. Heinzelman, W. B., Chandrakasan, A. P., & Balakrishnan, H. (2002). An application-specific protocol architecture for wireless microsensor networks. *IEEE Transactions on wireless communications*, *1*(4), 660-670.
15. Kaswan, A., Nitesh, K., & Jana, P. K. (2017). Energy efficient path selection for mobile sink and data gathering in wireless sensor networks. *AEU-International Journal of Electronics and Communications*, *73*, 110-118.
16. Cuevas-Martínez, J. C., Yuste-Delgado, A. J., & Triviño-Cabrera, A. (2017). Cluster head enhanced election type-2 fuzzy algorithm for wireless sensor networks. *IEEE Communications Letters*, *21*(9), 2069-2072.
17. Cuevas-Martinez, J. C., Yuste-Delgado, A. J., Leon-Sanchez, A. J., Saez-Castillo, A. J., & Triviño-Cabrera, A. (2019). A New Centralized Clustering Algorithm for Wireless Sensor Networks. *Sensors*, *19*(20), 4391.
18. Latiff, N. A., Tsimenidis, C. C., & Sharif, B. S. (2007, September). Energy-aware clustering for wireless sensor networks using particle swarm optimization. In *2007 IEEE 18th international symposium on personal, indoor and mobile radio communications* (pp. 1-5). IEEE.
19. Wang, G., Wang, Y., & Tao, X. (2009, October). An ant colony clustering routing algorithm for wireless sensor networks. In *2009 Third International Conference on Genetic and Evolutionary Computing* (pp. 670-673). IEEE.
20. Wang, J., Gao, Y., Liu, W., Sangaiah, A. K., & Kim, H. J. (2019). Energy efficient routing algorithm with mobile sink support for wireless sensor networks. *Sensors*, *19*(7), 1494.
21. Wang, J., Cao, Y., Li, B., Kim, H. J., & Lee, S. (2017). Particle swarm optimization based clustering algorithm with mobile sink for WSNs. *Future Generation Computer Systems*, *76*, 452-457.
22. Wang, J., Cao, J., Sherratt, R. S., & Park, J. H. (2018). An improved ant colony optimization-based approach with mobile sink for wireless sensor networks. *The Journal of Supercomputing*, *74*(12), 6633-6645.
23. YUSTE, Antonio J., et al. An adaptive genetic fuzzy control gateway discovery to interconnect hybrid MANETs. In *2009 IEEE Wireless Communications and Networking Conference*. IEEE, 2009. p. 1-6.
24. YUSTE-DELGADO, Antonio Jesús, et al. Improving hybrid ad hoc networks: The election of gateways. *Applied Soft Computing*, 2016, vol. 41, p. 1-14.
25. SANTIAGO, AJ Sánchez, et al. A dynamic-balanced scheduler for Genetic Algorithms for Grid Computing. *WSEAS Transactions on Computers*, 2009, vol. 8, no 1, p. 11-20.
26. AL RIDHAWI, Ismaeel, et al. Providing Secure and Reliable Communication for Next Generation Networks in Smart Cities. *Sustainable Cities and Society*, 2020, p. 102080.

27. VALIVETI, Hima Bindu; KUMAR, Budati Anil. Handoff strategies between wireless fidelity to light fidelity systems for improving video streaming in high-speed vehicular networks. *International Journal of Communication Systems*, p. e4285.
28. REDDY, S. Sai Satyanarayana; CHATTERJEE, Priyadarshini; MAMATHA, Ch. Intrusion detection in wireless network using fuzzy logic implemented with genetic algorithm. In *Computing and network sustainability*. Springer, Singapore, 2019. p. 425-432.
29. CABRERA, A. Triviño; YUSTE-DELGADO, Antonio J.; MACÍAS, D. Cintrano. A fuzzy logic-based and distributed gateway selection for wireless sensor networks. In *Highlights in Practical Applications of Agents and Multiagent Systems*. Springer, Berlin, Heidelberg, 2011. p. 243-248.
30. YUN, Sukhyun, et al. A soft computing approach to localization in wireless sensor networks. *Expert Systems with Applications*, 2009, vol. 36, no 4, p. 7552-7561.
31. KAPITANOVA, Krasimira; SON, Sang H.; KANG, Kyoung-Don. Using fuzzy logic for robust event detection in wireless sensor networks. *Ad Hoc Networks*, 2012, vol. 10, no 4, p. 709-722
32. KULKARNI, Raghavendra V.; FORSTER, Anna; VENAYAGAMOORTHY, Ganesh Kumar. Computational intelligence in wireless sensor networks: A survey. *IEEE communications surveys & tutorials*, 2010, vol. 13, no 1, p. 68-96.
33. ZADEH, Lotfi A. Fuzzy sets. *Information and control*, 1965, vol. 8, no 3, p. 338-353.
34. MENDEL, Jerry M. Uncertain rule-based fuzzy systems. In *Introduction and new directions*. Springer International Publishing, 2017. p. 684.
35. LEE, Wang-Chien. Uncertainty in wireless sensor networks. In Workshop on AFRL. 2010.
36. MOSCHITTA, Antonio; NERI, Igor. Power consumption assessment in wireless sensor networks. In ICT-energy-concepts towards zero-power information and communication technology. IntechOpen, 2014.
37. WU, Dongrui. Two differences between interval type-2 and type-1 fuzzy logic controllers: Adaptiveness and novelty. In *Advances in type-2 fuzzy sets and systems*. Springer, New York, NY, 2013. p. 33-48.
38. Hall, D. L., & McMullen, S. A. (2004). *Mathematical techniques in multisensor data fusion*. Artech House.
39. WU, Dongrui; MENDEL, Jerry M. Enhanced karnik--mendel algorithms. *IEEE transactions on fuzzy systems*, 2008, vol. 17, no 4, p. 923-934.

FIRST APPENDIX. SPANISH DOCUMENTATION

In this section it is presented the Spanish compulsory sections that are requested.

Resumen

En la actualidad, las redes de sensores inalámbricos es una tecnología de uso extendido. Permiten a los humanos monitorizar cualquier parámetro sin la necesidad de usar una infraestructura específica grande. Una ventaja es el reducido coste de los micro sensores que las componen, programados para medir su entorno y enviar posteriormente a una estación base dicha información, la cual a diferencia de los sensores, tiene un suministro constante de energía. Esta restricción energética supone un reto que debe afronta esta tecnología. La mayoría de investigadores intentan reducir el gasto energético en las tareas comunicativas y de gestión para así alargar la vida de la red. Para ello, la movilidad se ha propuesto por diferentes autores como una solución para reducir las distancias en las comunicaciones, y así, reducir el gasto energético. En el presente trabajo se va a exponer una taxonomía con la idea de clasificar los métodos estudiados en la bibliografía para posteriormente implementar un método que se basa en la movilidad de la estación base y con lógica difusa de Intervalo Tipo 2, y será comparado a un grupo de métodos seleccionados de la comentada bibliografía.

Introducción

En los últimos años, la evolución de la tecnología ha permitido el desarrollo de unos nuevos dispositivos llamados sensores. Sus principales características son su reducido tamaño y el bajo coste de fabricación. Gracias a ellos, muchas empresas los fabrican en grandes cantidades para distribuirlos y medir cualquier parámetro, ya sea radioactividad o algún parámetro biomédico, sin la necesidad de alterar el normal funcionamiento del sistema.

Es por ello que un grupo de investigadores propuso la opción de agrupar los sensores en lo que es conocido como Redes de Sensores Inalámbricos (WSN de sus siglas en inglés). Esta tecnología esta enfocada a la recolección de datos en un determinado lugar, por lo general a gran escala, y comunicar dichos resultados a un dispositivo conocido como Estación Base (BS de sus siglas anglosajonas), donde se pueden acceder a los datos más fácilmente. La disposición geográfica de los nodos suele ser aleatoria y tienen una batería limitada, por lo que la optimización de la misma es fundamental para obtener una buena eficiencia temporal. Esto es muy importante para aplicaciones relacionadas con el medio ambiente, como puede ser la monitorización de la actividad volcánica. Suelen ser lugares de difícil acceso por lo que el despliegue de los sensores no sería sencillo. Esto hace necesario que los nodos que se lancen duren el máximo posible. Así mismo, en otras aplicaciones como la detección de incendios en bosques, necesitan que estos datos recogidos sean procesados en tiempo real, por lo que los métodos que se desarrollen deben tener en cuenta que el nodo ha de enviar su dato en cada ronda de funcionamiento.

Para ello, se han investigado diferentes aspectos en el comportamiento de este tipo de redes, como por ejemplo la reducción del numero de paquetes de datos y control. En este punto, aparece un mecanismo propuesto por la inmensa mayoría de métodos y es el ***Clustering***. Aunque es explicado en detalle a lo largo del trabajo, su objetivo es dividir la red en grupos de nodos que se comunicarán con un nodo de cabecera conocido como ***Cluster Head***, que será el encargado de comunicarse con la estación base.

La elección optima de dicho *cluster head* es el foco principal de la investigación en este campo. Es por ello, que realizar una taxonomía para clasificar los aspectos fundamentales en que se basa cada método a la hora de tomar esta decisión es un paso necesario antes de proponer algún método, para intentar obtener las ventajas de cada tipo de método. En este trabajo se desarrolla un método en el que los *cluster heads* son elegidos mediante lógica difusa de intervalo tipo 2, y novedad se propone un esquema en que la estación base es capaz de seguir distintos patrones de movimiento para recortar las distancias en la comunicación.

Conclusión

La movilidad es una característica interesante para algunas aplicaciones porque permite un aumento de la eficiencia. En este caso, se ha observado que, en comparación con el protocolo móvil de la bibliografía, muestra un mejor rendimiento si se tiene en cuenta la energía necesaria para cubrir la distancia recorrida por el protocolo EERA.

La principal ventaja del método apareció cuando se simuló la versión modificada. Ofreció mejores resultados cuando los nodos enviaron directamente el mensaje a la BS si están cerca, por lo que la movilidad es aprovechada por los nodos, aunque aparecieron problemas con mapas grandes. Como se explicó anteriormente, las largas distancias son el principal problema de este tipo de red y este tipo de mapas. El número de nodos no afecta a los resultados obtenidos con este método modificado, por lo que esto significa que tiene una alta escalabilidad en términos de número de nodos. Una buena forma de resolver este problema es desplegar BS adicionales que cubran un sector cada una. De esta manera, cada sector es una reproducción de la arquitectura de mapas pequeños, en la cual el método ha demostrado tener buenos resultados.

También fue observable que, para este tipo de mapas, la ruta seleccionada no tuvo una influencia directa en los resultados, a excepción de la versión modificada, en el que la ruta de la cuadrícula era mucho mejor. Esto lleva a pensar que el camino tomado solo tiene una influencia directa si todos los nodos pueden aprovecharlo, no solo el CH. Esto significa que si el BS se mueve en posiciones cercanas al centro, se puede acceder a él mediante una gran cantidad de nodos en promedio cada ronda por envío directo.

A partir de la conclusión obtenida anteriormente, se abren algunas formas de pensar para ser probadas en trabajos futuros, como por ejemplo la incorporación de nueva infraestructura que actúe como colectores o BS secundaria.. Al dividir toda la red en sectores, cada BS o recolector cubriría una parte de la red con rutas como se muestra anteriormente. Esta solución reduciría drásticamente la distancia que cada nodo necesitaría cubrir con su mensaje y resolvería el problema que se muestra en mapas grandes. Además, este aumento en el rendimiento estaría asociado a un aumento en el precio de la red, porque agregar infraestructura adicional es más costoso que agregar energía a los nodos, por ejemplo.

Otra solución sin agregar costes, podría ser utilizar esquemas de múltiples saltos. Esta solución agregada a la propuesta actual podría resolver el problema de las comunicaciones de larga distancia. Consistiría en usar CH cercano para enviar el mensaje de datos, y este CH podría agregar o reenviar el mensaje. Esta es una buena opción, pero agregaría mucha complejidad en términos de comunicación y también en el desperdicio de energía, porque se necesitarían muchos mensajes de control.

En el caso de no necesitar tiempo real, en diferentes aplicaciones una solución con la misma implementación sería un esquema acumulativo. Esto significa que cada CH almacena los datos hasta que la BS llega a un punto cercano, avisando con un mensaje y el CH envía los datos, por lo que el rango de comunicación se reduciría.

Estas soluciones están pensadas en base a métodos que fueron estudiados en la bibliografía. Pensando en la propuesta actual, se podrían agregar algunas características nuevas.

El primero es redefinir el sistema Fuzzy teniendo en cuenta la movilidad, esto incluye una nueva investigación para saber qué parámetro podría ser una buena opción y también para obtener los conocimientos necesarios para definirlo bien. Esta opción también podría conducir a un rediseño del funcionamiento, proponiendo un método híbrido. Esta opción se realiza redefiniendo las características del sistema difuso con la información obtenida de un mensaje recibido de la BS. Esto podría hacerse cambiando la distancia de normalización en la base de la comunicación o usando de otra manera el intervalo obtenido por el sistema Fuzzy.

Estas opciones son posibles líneas de investigación para futuros trabajos que podrían explorarse, ya que la movilidad es una característica relativamente nueva en este campo, y no hay grandes avances en este campo que puedan tomarse como referencia, cada autor ofrece una solución que resuelve un problema pero aparecen diferentes problemas, por lo que se han de explorar todavía diferentes opciones al respecto.