TECHNISCHE UNIVERSITÄT CHEMNITZ

Redundancy Elimination for Data Aggregation in Wireless Sensor Networks

BACHELOR THESIS

By

Guillem Viñas Raventós

Bachelor of Elektrotechnik

Electrical Engineering and Information Technology

Department of Measurement and Sensor Technology

Tehcnische Universität Chemnitz

27 August 2015

Redundancy Elimination for Data Aggregation in Wireless Sensor Networks

Bachelor thesis approved:

Mrs Olfa KANOUN	*First auditor*
Mr Martin GÖTZ	*Second auditor*
Mr Christian VIEHWEGER	*Second auditor*
Ms Sabrine KHRIJI	*Supervisor*

Acknowledgements

The work presented in the next pages represents one of the most important goal reached in my life so far: the bachelor thesis. As important and meaningful this moment is for me, I would like to dedicate the next lines to the people who have contributed to this work and that have constantly motivated me during this period, hoping that my gratitude will fully reach them.

*Firstly, I would like to express my sincere thankfulness to **Pr. Olfa KANOUN**, a chair of Department of Measurement and Sensor Technology, for giving me the opportunity to accomplish my bachelor thesis in her department, and with her.*

*I would like to proceed with heartfelt thanks to the rest of my thesis committee: **Martin GÖTZ** and **Christian VIEHWEGER**. Their insightful comments and encouragement, but also for the hard question which have inspired me to widen my researches from various perspectives.*

*Especially, I would like to express my sincere gratitude to my supervisor **Ms Sabrine KHRIJI** for her continuous support, the patience, motivation, and awareness about the topic. Her guidance helped me constantly during the researches and the draft of this thesis. I could not have imagined having a better supervisor and mentor for my bachelor thesis.*

*Finally, I would also like to convey thanks to the **Technical University of Chemnitz** for having hosted me during these months and giving me the opportunity to complete my bachelor degree here as part of the Erasmus+ program.*

Dedication

This thesis is dedicated to my family,

For their endless support and encouragement

OUTLINE

CHAPTER 3: Proposed Solution - "Redundancy Elimination for Data Aggregation in WSN"........ 38

Conclusion and future works 69

REFERENCES 71

FIGURES LIST

TABLES LIST

ABBREVIATIONS LIST

WSN	Wireless Sensor Network
CH	Cluster Head
BS	Base Station
IP	Internet Protocol
QoS	Quality of Service
DD	Direct Diffusion
SPIN	Sensor Protocols for Information via Negotiation
LEACH	Low Energy Adaptive Clustering Hierarchy protocol
HEED	Hybrid Energy-Efficient Distributed clustering
STMA	Spatial and Temporal Multi-Aggregation for state-based sensor data
READA	Redundancy Elimination for Accurate Data Aggregation
EEBCDA	Energy efficient and Balanced Cluster-Based Data Aggregation
GA	Genetic Algorithm
PSO	Particle Swarm Optimization
ACO	Ant Colony Optimization
ABC	Artificial Bee Colony Optimization
ESPDA	Energy-Efficient and Secure Pattern based Data Aggregation
SRDA	Secure Reference based Data Aggregation
ERDA	Elimination Redundancy for Data Aggregation

General Introduction

1. Motivation

Wireless sensor network is a new generation of task-oriented distributed networks, which gathers micro-electromechanical systems, sensor technology, embedded computing technology, information processing technology, modern networks, wireless communication technology and digital electronics.

Wireless sensor networks have been a very important tool in many fields, including industrial control, environment, agriculture, disaster prevention and relief, health care, military defence, smart home network, space exploration, logistics, etc. However the main problem is, the sensor nodes are generally battery-powered devices, and exist the concern how to reduce the energy consumption of nodes, so the network lifetime can be extended to reasonable times.

Since battery life duration is one of the most important problems in the world of Wireless Sensor Networks, reducing energy consumption is the main motivation and objective to solve this problem.

Personally, perform a bachelor thesis about important and emerging technology as Wireless Sensor Network, make a proposed solution of data aggregation protocol, develop new ideas and algorithms for solving the main problem of this technology, has been the primary motivation of this thesis.

2. Objectives of the Bachelor Thesis

The project is based on the concept of intelligent agriculture, which stems from the need of improve, streamline and provide autonomy to agricultural crops.

This control is mainly based on intelligent techniques, such as WSN. This technique can provide periodically data reports via the wireless communication relating sensor nodes to each other. However, sensor nodes are energy constrained and limited to their remaining battery power. Therefore, prolonging their lifetime is an issue that should be highly considered. The communication between the nodes including transmissions and receptions is the major source of energy waste.

WSNs are mainly based on communication within any network the fact that makes the consumption of energy in question. The transmissions and receptions lead to power waste. To manage and reduce these wastes, the packets scheduling and their circulation within the network are to be highly considered in the wireless communication. To do so, a data aggregation algorithm should be well-designed.

The main goal of this bachelor thesis is to study different types of data aggregation techniques in Wireless Sensor Networks and develop a new data aggregation protocol, trying to improve existing protocols, providing new enhancements and new concepts trying to reduce as much as possible the quantity of data sent by sensor nodes. Therefore reduce the traffic inside cluster and among Cluster Heads, with main goal of achieving more efficient protocol in terms of amount of traffic transmitted and energy consumption.

When this method was proposed some goals were set, as following:

(1) Minimize the amount traffic of the network.

(2) Minimize the energy consumption of the network.

(3) Increase the network life time.

(4) Cluster Head selection must be better balanced.

(5) The Selection of which sensor node sends when redundant data is detected must be better balanced.

(6) Cases that pose a danger situation must be reported to the Base Station.

3. Contributions of the Bachelor Thesis

In this thesis, design a robust and efficient data aggregation scheme that considers the unique characteristics of WSNs is designed. All used processes and techniques have been focused in the field of agriculture.

In this thesis, a wireless sensor network with a data aggregation protocol in other to save energy is developed. MATLAB software is used to create the network and develop the protocol and its algorithms.

The main goal of the work presented in this thesis has been to provide possible improvements for aggregation data and energy consumption in WSN applications. To achieve that, different existing techniques and algorithms are studied to devise a protocol by using some techniques, improving another techniques and implementing new concepts.

4. Bachelor Thesis organization

This thesis is organized as follows:

Firstly, chapter 1 represents an introduction for the whole bachelor thesis. In Chapter 2, Wireless Sensor Network and its applications, different protocols of sensor network will be introduced. Also an analysis of different techniques in WSN which explain the different type of network, their differences and which protocols are used in each network will be presented. In Chapter 3, an explanation for the main objective of this thesis, related work about data aggregation algorithm in WSN will be given. In this section, an explanation for the objective of data aggregation and an analysis of different types of techniques, functions, strategies and intelligences methods of data aggregation will be presented. An analysis of some existing secure data aggregation protocols is posed in this same chapter. Two existing protocols and their main characteristics and processes are presented. These two protocols are the main key of new protocol development. Furthermore, in Chapter 4, a detailed description of the proposed data aggregation algorithm is performed. In this point is explained all the processes of the protocol, system model and assumptions, work contributions and the main characteristics of the protocol. To end the results are analyzed and compared with the protocols presented in the previous chapter. Finally, the conclusions of the thesis and future works are given in Chapter 5.

CHAPTER 1: Overview of Wireless Sensor Networks in Precision Agriculture

1. Wireless Sensor Networks: An overview

With the advancement of technology, it has managed to develop distributed sensing devices with small size, low cost and low consumption, able to process information locally and can communicate wirelessly, and work cooperatively.

A wireless sensor network [1] consists of a large number of devices, known sensors capable of collecting information from its environment, such as humidity, light, temperature, etc, by using sensors that embody these node devices. Besides these, there are also relay nodes to take charge of routing the data to the base station, is connected to a computer that can communicate to the outside via the Internet or a local area network (LAN).

The sensor nodes are scattered throughout the network in different types of topologies, explain that later will be explained later, depending on the application.

1.1. Wireless Communication

A WSN consists of a large number of sensor nodes. Each sensor node senses environmental conditions and sends the sensed data to a base station (BS), which is a long way off in general. Low energy consumption is very important for sensor nodes, since the sensor nodes are powered charged by limited power batteries. In order to reduce the energy consumption, a clustering and data aggregation approach has been extensively used. In this approach, sensor nodes are divided into clusters, and for each cluster, one representative node, which called cluster head (CH), aggregates all the data within the cluster and sends the data to BS. Since only CH nodes need long distance transmission via multi-hop, the other simple nodes only have to send data to CH via single-hop, whereby save the energy consumption. Efficient data collection in WSN plays a key role in power conservation.
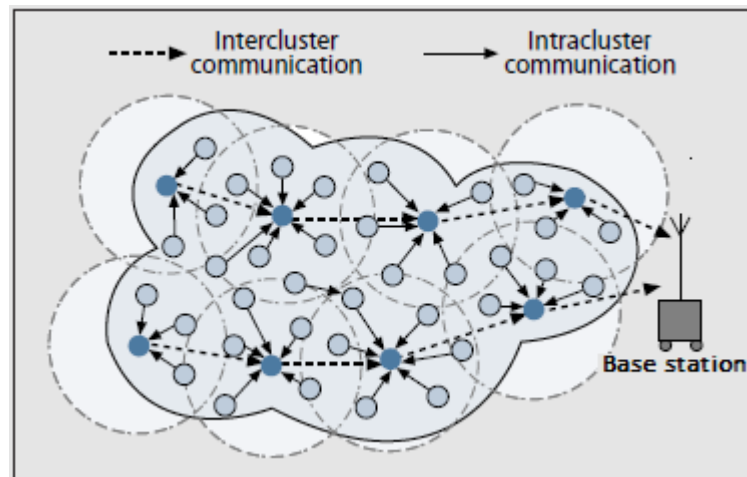
Figure 1. Wireless Sensor Network example

1.2. Characteristics of Sensor Network

When designing a network, it must taken into account various factors and / or restrictions depending on the application, design and aims of sensor networks. Factors that influence on design of routing protocol in a wireless sensor network are [2]:

Infrastructure less: A main characteristic of wireless sensor network is that every sensor node are communicate with each other without any fixed infrastructure so communication overhead will be less.

Mobility: In case of WSN, mobility of nodes in wireless sensor networks is low. Due to this kind of feature they are used in event tracking purposes like motion detection of video frames etc.

Multi-Hoping: Wireless sensor Networks composed of several nodes and they are communicating with each other and described several paths to several node to send data to Base Station. Thus the nodes sent the packet from one node to another node until reach the destination through several paths. Due to this Multi-hop features energy associated with each node can be conserved.

Openness: Wireless sensor network access information and services regardless of Geographic position.

Network size: In wireless sensor network the node are scattered in large geographical area. The number of nodes in Wireless sensor network may vary by application and geographic area.

Homogeneous Network: A Wireless sensor network composed of homogeneous devices that each node has the same features.

Location Awareness: Every node should know their position relative to other node by knowing their actual location. Is very important since the data are transmitted depending upon the location and de distance between nodes.

Reliable transmission of Data: To have a proper request-response communication model, the node should request and send the information to a proper node. To make this mechanism be reliable, this node should process the request efficiently and reply as soon as possible.

Fault tolerant: In wireless sensors, if one node fails then it does not affect the network operation because there are other adjacent nodes collecting similar data. But on the other hand the accuracy of data collected is reduced.

1.3. Challenges in Wireless Sensor Network

Continuous challenges in the wireless sensor networks affect in implementation of several services. There are a lot of controllable and uncontrollable parameter by which the implementation of wireless sensor network is affected, these parameters should be detected and controlled. Parameters such as [2]:

Node Deployment: Sensor nodes are densely deployed in the area of interest. These nodes can be deployed either manually or randomly. When nodes are manually, placed data is routed through pre-determined paths. In self organizing systems, sensor nodes are scattered randomly creating a topology in an ad hoc manner.

Network topology: Is the arrangement of the various sensor nodes of a network. Essentially, it is the topological structure of a network and may be depicted physically or logically. Physical topology is the placement of the various sensor nodes of a network, while logical topology illustrates how data flows within a network, regardless of its physical design.

Network dynamics: Normally, sensor nodes are assumed fixed, but in many applications the BS and/or sensor nodes can be mobile. This implies that routing messages are more challenging, in addition energy, bandwidth and so on are constantly changing. On the other hand, sensing fixed events allows the network to work in a reactive mode, while dynamic events in most applications require periodic reporting to the base station.

Energy conservation: As we know that the sensor node has small size. Due to this small size, the battery has low capacity and the available energy is less. And in that situation the refilling or replacing of battery is impossible. In order to reduce this problem some more energy efficient protocol are designed so that the sensor node communicate efficiently by increasing both throughput and network capacity.

Data aggregation: It is a combination of data from different sources .Similar packets from multiple nodes can be aggregated to reduce the transmission. This technique has been used to achieve energy efficiency and data transfer optimization in a number of routing protocols.

Communication quality: In WSN ,the communication quality depends on different situations, but normally have a very low quality communication. And also environment must be taken into account.

Transmission media: Generally, communication takes place through the wireless media, which is affected by fading which affects the operation of WSN.

Node Capability: A sensor node can have different roles or features depending on the application, such as relaying, sensing and aggregation. The problem appears if a sensor node have to be performed by all these functions, since this causes the energy of it would be drained more quickly. Different capabilities of sensor nodes raise multiple issues related to data routing and make routing more challenging.

Data processing: As in data aggregation is explained, data collected by many sensors can be redundant data. Whereby, data aggregation it required in network processing in order to avoid that the transmission of redundant data affects the efficiency of the network. It will help to conserve some amount of energy for further transmission.

Scalability: The deployment of sensor nodes is dependent on the nature of the application. It varies with respect to the demand of the application, this means that the number

of sensor nodes can be a variable value, hundreds, thousand or even more. To handle network scalability, routing algorithm should have the capability to cope with scalable network [3].

Application specific: Sensor network changes with change in application. For different type of application we have to design different sensor network.

Node Id: Every node in the sensor network doesn't have different Id. The overhead for Id maintenance is more important, it can be said that in sensor network data, is worth more than knowing the Id.

2. Applications in Wireless Sensor Networks

A sensor network is designed to perform a set of high-level information processing tasks such as detection, tracking, or classification. Measures of performance for these tasks are well defined, including detection of false alarms or misses, classification errors, and track quality [4].

Applications of sensor networks are wide ranging and can vary significantly in application requirements, mode of deployment, sensing modality, or means of power supply. Possible applications of sensor networks are of interest to the most diverse fields, as for example:

- Medicine and health care (e.g. Postoperative and intensive care, Long-term surveillance of patients)

- Logistics (e.g. Tracking of parcels during transportation, Inventory tracking in stores or warehouses)

- Precision Agriculture (e.g. Precise irrigation and fertilising of fields, – Temperature and brightness monitoring)

- Battlefield awareness (e.g. multi-target tracking)

- Context-aware computing (e.g. intelligent home, responsive environment)

- Industrial sensing and diagnostics (e.g. appliances, factory, supply chains)

- Infrastructure protection (e.g. power grid, water distribution)

- Intelligent Buildings/Bridges (e.g. Measurements about temperature, energy wastage, Monitoring of mechanical stress levels)

- Environmental monitoring (e.g. traffic, habitat, security)

• Disaster relief operations (e.g. Wildfire detection, Sensor nodes with thermometers are dropped from an air plane. Various temperature measurements are collected to produce a temperature map)

• Biodiversity Mapping (e.g. Gain an understanding about plants and animals)

3. Protocols for Sensor Networks

The protocol stack [5] integrates power and routing awareness, such as energy-aware routing, integrates data with networking protocols, as in the case of data aggregation, communicates power efficiently through the wireless medium, and it promotes cooperative efforts of sensor nodes, as for task management plane. The architecture of protocol stack used by the Base Station and sensor nodes is shown in Figure 2. The protocol stack for WSNs consists of five protocol layers: Application layer, Transport layer, Network layer, Data Link layer and Physical layer.
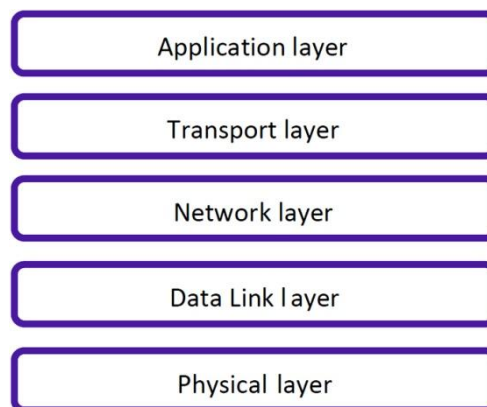
Figure 2. Protocol stack for sensor networks

(1) Application layer: It contains a variety of application, services and interface primitives available to generate various sensor network applications on every kind of platform. It means, that it is the layer that provides the interface that connects the platform in order to run deployed software. Moreover, all the services, tasks and other background applications used in the networking connectivity, data analyzing and etc. run at this layer.

(2) Transport layer: The transport layer is responsible for the delivery of reliable data required by the application layer. It helps to maintain the flow of data from the application layer to network layer if the sensor networks application requires it.

(3) Network layer: This layer is responsible for routing data and path selection from the transport layer. It directs the process of the selection paths along which to send data in the network. In addition, addressing, encapsulation and decapsulation are taken place in this layer. The data link layer is also primarily responsible for data stream multiplexing, data frame transmission and reception, medium access, and error control.

(4) Data Link layer: The Data Link layer is the layer that creates the link between software and medium. It is responsible for data stream multiplexing, data frame detection, medium access and error control in order to provide reliable transmissions.

(5) Physical layer: This layer is responsible for converting bit streams from the data link layer to signals that are suitable for transmission over the communication medium. For this purpose, it must deal with various related issues, for example, transmission medium and frequency selection, carrier frequency generation, signal modulation and detection, and data encryption. In addition, it must also deal with the design of the underlying hardware, and various electrical and mechanical interfaces.

4. Routing Techniques in WSN

Routing is one of the critical points in WSNs, since resources are limited in terms of power supply, processing capability and transmission bandwidth. A weakness is that is impossible to create an addressing scheme to route the traffic like Internet Protocol (IP). Due to the limited resources, it is hard to cope with unpredictable and frequent topology changes. It is normal to have a lot of redundant data, which must be considered by routing protocols. Nevertheless, keep it in mind that in WSNs the energy conservation is more important than quality of service (QoS) since in most applications all the sensor nodes are constrained with energy which is directly related to network lifetime [3].

To understand the differences between different types of network topology, routing protocols in WSN and can be divided into flat routing and hierarchical routing. The main difference between cases are: in flat based routing, all nodes are typically assigned equal roles or functionality whereas in hierarchical based routing, nodes will play different roles in the network.

```
┌─────────────────────────────────────────────────┐
│   Routing Protocols for Wireless Sensor Network   │
└─────────────────────────────────────────────────┘
              │                       │
              ▼                       ▼
    ┌──────────────────┐    ┌──────────────────┐
    │  Flat Protocols  │    │ Hierarchical     │
    │                  │    │ Protocols        │
    └──────────────────┘    └──────────────────┘
              │                       │
              ▼                       ▼
    ┌──────────────────┐    ┌──────────────────┐
    │ Sensor Protocols │    │ Low Energy       │
    │ for information  │    │ Adaptive         │
    │ via Negotiation  │    │ Clustering       │
    │ (SPIN)           │    │ Hierarchy        │
    │                  │    │ protocol (LEACH) │
    │ Direct Diffusion │    │                  │
    │ (DD)             │    │ Hybrid Energy-   │
    │                  │    │ Efficient        │
    │                  │    │ Distributed      │
    │                  │    │ clustering       │
    │                  │    │ (HEED)           │
    └──────────────────┘    └──────────────────┘
```
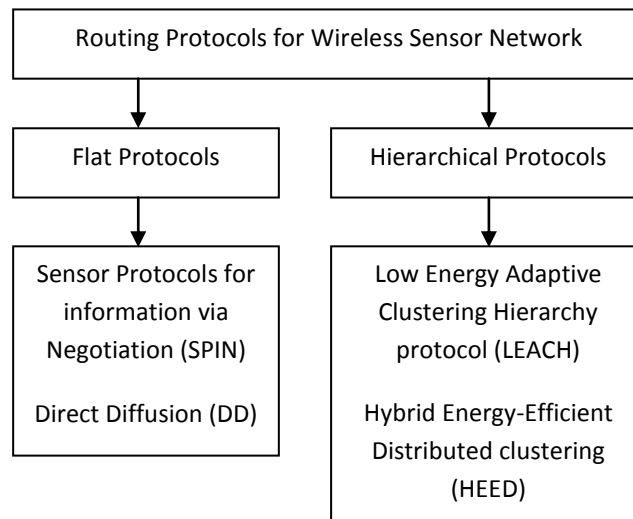
Figure 3. Routing Protocols classification

4.1. Flat Networks

A flat topology [3] consists in an equal and uniform network since all nodes have the same behaviour and perform the same tasks in the network. Data transmission among nodes is performed via hop by hop and by flooding system. In small-scale networks flat routing protocols are relatively effective. However, it is relatively undesirable in large-scale networks because resources are limited and all sensor nodes generate more data processing and bandwidth usage. In the next paragraphs Sensor Protocols are describe in detail for Information via Negotiation (SPIN) [6] and Directed Diffusion (DD) [3] protocols.

4.1.1. Senor Protocols for Information via Negotiation (SPIN)

SPIN consist in disseminating all the information from each node to every node in the network. The use of this technique, allows any user to consult any node and obtain the required information immediately. These protocols achieve that if a nearby sensor has similar data, it is only necessary to distribute the data to the other nodes that do not possess those data. Another strength is that SPIN protocol has access to the current energy level of the node and adapts the protocol.. that it is running , based on how much energy remains. The SPIN family is designed to address the deficiencies of classic flooding by negotiation and resource adaptation. SPIN is designed or based on two basic ideas:

(1) By sending data that describe the sensor data instead of sending all the data, this method achieves that sensor nodes operate more efficiently and they conserve energy.

(2) Using gossiping avoids the problem of implosion just by selecting a random node which the packet is sent to, rather than broadcasting the packet blindly. However, this causes delays in propagation of data through the nodes.

4.1.2. Direct Diffusion (DD)

The main idea of the DD protocol is to eliminate the redundancy by trying to combine the data that come from different sources, thereby minimizing the number of transmissions, thus saving network energy and prolonging its lifetime. Unlike traditional end-to-end routing, DD routing finds routes from multiple sources to a single destination that allows in-network consolidation of redundant data.

Directed diffusion differs from SPIN in two aspects. First, the BS sends queries to the sensor nodes by flooding some tasks. In SPIN, however, sensors advertise the availability of data, allowing interested nodes to query that data. Second, all communication in directed diffusion is neighbour to neighbour with each node having the capability to perform data aggregation and caching. Unlike SPIN, there is no need to maintain global network topology in directed diffusion.

4.2. Hierarchical Networks

In a hierarchical topology, sensor nodes perform different tasks in WSNs and typically are organized into lots of clusters according to specific requirements or metrics. Generally, each cluster comprises a leader referred to as cluster head (CH) and other simple nodes. The CHs can be organized into further hierarchical levels. In general, nodes with higher energy are defined as CH and the simple nodes perform the task of information sensing.

4.2.1. Low Energy Adaptive Clustering Hierarchy protocol (LEACH)

LEACH [6] [7] is a cluster-based protocol, it randomly selects a few sensor nodes as CHs and change this role in every session to evenly distribute the energy load among the sensors in the network. In order to reduce the amount of information that must be transmitted to the Base Station, in LEACH protocol the nodes working as a Custer Head have to compress data arriving from the nodes belonging to the same cluster, and send an aggregated packet to the BS.

The operation of LEACH is broken up into lots of rounds, where each round is separated into two phases:

(1) The set-up phase where the clusters are organized.

(2) The steady-state phase where data is delivered to the BS.

The clusters are organized during the set-up phase , each node decides whether or not to become a CH for the current session. This decision is based on the suggested percentage of CHs for the network and the number of times the node has been a CH so far.

When a node is elected CH, it broadcasts an advertisement message to the other nodes. The sensor nodes will receive different advertisement messages from different CHs, according to the received signal strength of the advertisement, nodes decide to which cluster it will join for this session and send a membership message to its CH. The sensor nodes sense and transmit data to the CHs during the steady-state phase. In each cluster, the CHs compress data arriving from nodes and send the aggregate data to the BS directly.

The advantages of LEACH include the following:

 (1) The nodes can equally share the load imposed upon CHs to some extent since any node that served as a CH in certain round cannot be selected as the CH again.

(2) Utilizing a TDMA schedule prevents CHs from unnecessary collisions;

(3) Cluster members can open or close communication interfaces to minimize energy consumption.
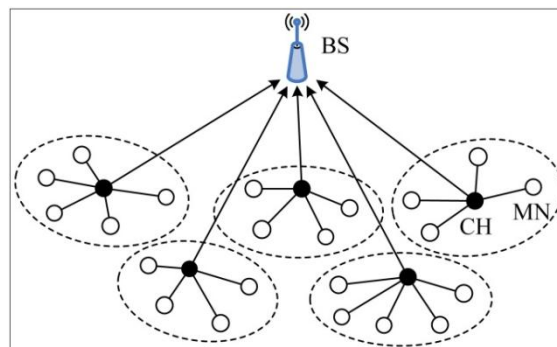
Figure 4. LEACH Network

However, LEACH has also some disadvantages as follows:

(1) It performs the single-hop inter-cluster, directly from CHs to the BS, and it is not applicable to large networks since the distance between a node and BS can be too long. Besides, a long-range communication can breed too much energy consumption;

(2) Cluster Heads are selected in terms of probabilities without energy considerations, whereby, LEACH protocol cannot ensure a balanced consumed of energy of the network.

4.2.2. Hybrid Energy-Efficient Distributed clustering (HEED)

HEED [7] achieves an energy-efficient clustering routing since is a multi-hop clustering algorithm in WSN. Different from LEACH, HEED protocol does not select nodes as CHs randomly. The manner of cluster construction is performed based on the hybrid combination of two parameters. One parameter depends on the node's residual energy, and the other parameter is the intra-cluster communication cost.

The advantages of the HEED protocol can be resumed in the following points:

(1) It is a fully distributed clustering method that benefits from the use of the two important parameters for CH election;

(2) Low power levels of clusters promote an increase in spatial reuse while high power levels of clusters are required only for inter-cluster communication.

(3) More energy conservation since communications are multi-hop between Cluster Heads and the Base Station.

However, there are some limitations with HEED as follows:

(1) Uncovered sensor nodes due the use to tentative CHs that do not become final CHs.

(2) The performing of clustering in each round imposes significant overhead in the network.

(3) It needs several iterations to form clusters, and this causes a overhead and it consequent energy consumption.

(4) Some CHs near the Base Station, may die earlier because these CHs have more work load, and the hot spot will come into being in the network.

4.3. Hierarchical Network vs. Flat Networks

In the following table the main characteristics of hierarchical and flat networks are summarized and compared:

Table 1. Comparison between Hierarchical and Flat Network

Hierarchical Network [7]	Flat Networks [3]
Data aggregation performed by cluster heads or leader node	Data aggregation is performed by different nodes along the multi-hop path
Overhead involved in cluster or chain formation throughout the network	Data aggregation routes are formed only in regions that have data for transmission
Even if one cluster head fails, the network may still be operation	The failure of sink node may result in the breakdown of entire network
Lower latency in involved since sensor nodes perform short rand transmission to the cluster heads	Higher latency is involved in the data transmission to the sink via multi-hop path.
Routing structure is simple but not necessary optimal	Optimal routing can guaranteed with additional overhead
Node heterogeneity can exploited by assigning high energy nodes as cluster heads	Does not utilize node heterogeneity for improving energy efficiency

CHAPTER 2: Background and Related work

1. Introduction

As wireless sensor networks become part of everyday life, one problem that still remains central to the operability and applicability of these networks is the limited energy and computational resources of the sensor nodes, which has a direct impact on the network lifetime. Thus there is a need for developing techniques and methods which increase the productivity and efficiency of these networks.

Given the present limitations of wireless sensor networks, data aggregation is an important paradigm that assists in decreasing the energy consumption, eliminating data redundancy and increasing the useful information flow from the source to the base station thereby prolonging the network lifetime. Data aggregation can be incorporated with other routing protocols or can be implemented as individual protocols/techniques that interact closely with routing protocols.

Data gathering is defined as the systematic collection of sensed data from multiple sensors to be eventually transmitted to the base station for processing. As discussed above, the problem is that the sensor nodes have a limited energy, for this reason it is energy inefficient that all the sensors transmitted the data. The point is that data generated from the same node in different time and data generated from neighbouring sensors are often redundant and highly correlated. On the other hand, if the traffic is enormous, the base station can't process all of the data generated. Hence, it is necessary to find methods for combining data or intermediate nodes able to reduce the number of packets transmitted to the base station, these methods allow a major conservation of energy and reduce bandwidth. This can be accomplished by data aggregation.

2. Data Aggregation

There are several proposed mechanisms with the main goal to reduce the power consumption of wireless sensor networks. Mechanisms such as radio scheduling, control packet elimination, topology control, and most importantly data aggregation [8].

Data aggregation is defined as the process of summarizing and combining sensor data in order to reduce the amount of data transmission in the network. With the aim of reducing power consumption, data aggregation is the global process of gathering and routing information through

a multi-hop network and processing data at intermediate nodes. It attempts to collect the most critical and important data from the sensors nodes and make it available to the Base Station in an energy efficient manner with minimum data latency and minimum possible bandwidth.

Key Points in data aggregation are as follows:

- Nodes sense attributes over the entire network and route to nearby nodes.
- Node can receive different versions of same message from several neighbouring nodes.
- Communication is usually performed in the aggregate.
- Neighbouring nodes report similar data.
- Combine data coming from different sources and routes to remove redundancy.

Advantages:

Data Aggregation uses the parameters of nodes joining the cluster so that the data attributes are selected and stored in an aggregated format for further evaluation and usage. For data collected from sensor nodes, data aggregation can reduce the existing redundant data by data fusion processing, whereby it reduces the traffic load and conserve energy of the sensors. Another advantage is that with data aggregation the robustness and accuracy of information obtained by the entire network are enhanced.

Disadvantages:

Data aggregation works with different clusters and sensor nodes send data to Cluster Heads and sends fuse data to the base station. A problem can be occurred that the Cluster Head may be affected by malicious attacks. If a cluster head is compromised, then the base station cannot be ensure the correctness of the fusion data that has been send to it. Several copies of the fusion result may be sent to the Base Station and this can be another problem, since it increases the traffic and the power consumed by the sensor nodes.

Nowadays there are different techniques, algorithms, etc. aimed to achieve an energy efficient data aggregation protocol explained in detail in the next paragraphs.

2.1. Data Aggregation Strategies

Centralized Approach, In-Network Aggregation, Tree-Based Approach and cluster-Based Approach are some of the existing strategies related used for data aggregation [9].

2.1.1. Centralized Approach

In this strategy the node sends data to a central node via the shortest possible route. These data are aggregated by the central node (header node) to reduce the redundancy.

2.1.2. In-Network Aggregation

There are two approaches in-network aggregation [10] :

• **With size reduction:** each node combines and compresses the data packets received from its neighbours in order to reduce the packet length which will be transmitted towards Base Station.

• **Without size reduction:** is defined as the process of merging data packets received from different neighbours into a single data packet. The process merging data packets received from different neighbours into a single data packet but unlike with size reduction process, it is without processing the value of data..

2.1.3. Tree-Based Approach

This strategy [11] is held by constructing an aggregation tree, in which Base Station is considered as roots and sensor nodes are the leaves. Each node has a parent node whose data are forwarded. Flow of data starts from sensor nodes (leaves) up to the Base Station (roots) and the aggregation is done by parent nodes.

2.1.4. Cluster-Based Approach

With this approach [12] the whole network is divided into different clusters. A Cluster Head is selected in each cluster among different sensor nodes or cluster members. The nodes selected as a Cluster Heads are responsible for the aggregation process of data received from cluster members and then transmit the result to the Base Station.

2.2. Aggregation Functions

With this approach the whole network is divided into different clusters. A Cluster Head is selected in each cluster among different sensor nodes or cluster members. The nodes selected as a Cluster Heads are responsible for the aggregation process of data received from cluster members and then transmit the result to the Base Station [8].

2.2.1. Lossless or lossless

Data aggregation operation can be lossy or lossless information at the BS [8]. Redundant data is compressed at each aggregation point. In lossy aggregation, several packets are combined into a single packet, and thus the amount of outgoing data is significantly smaller than the input. This approach is useful when communication load exceeds system capacity and energy saving is needed. Therefore, there is a trade-off between energy efficiency and data precision.

The lossless aggregation is effective when individual sensor readings are small in size since there are plenty room available for concatenation in each packet.

2.2.2. Duplicate sensitive or duplicate insensitive

If a sensor node receives multiple copies of identical information, the sensor node may do two different things: may process the data by taking into account the number of copies (duplicate sensitive) or considering one single copy of the data (duplicate insensitive) [10]. It must be taken into account that a duplicate sensitive aggregation functions will not provide the same result when the data set contains duplicate values.

2.2.3. Exemplary or summary

These two functions [8] differ as follows, exemplify aggregation only returns one representative value of the ones available in the dataset, for example MIN, MAX or MEDIAN.

2.2.4. Monotonic aggregates

With this method [8] only the value of the sensor node - whose value satisfies the query – is transmitted. For example, in a query that requests the MAX temperature reading in the network, all the source nodes report their values to the Cluster Head and only the nodes having their value greater than the current MAX will report. Using this method it is possible to reduce the traffic without affecting the result.

2.3. Transmission Energy Conservation Techniques

These techniques encompasses algorithms that process incoming sensed data with the aim to save transmission power. It try to combines data set so as to produce output that requires lesser energy for onward transmission. We divide this class in three sub-classes:
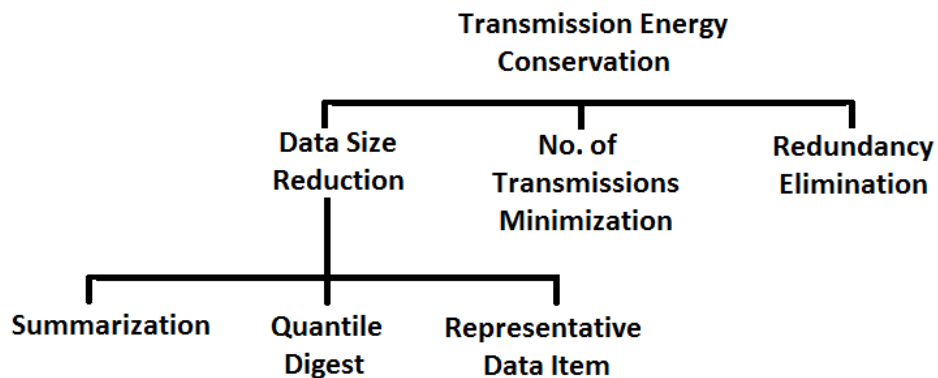
Figure 5. Transmission Energy Conservation techniques

2.3.1. Data Size Reduction

These techniques are useful when is convenient to have the size of aggregated data approximately equal to a single un-aggregated value.

- **Summarization Techniques:** Try different algorithms in order to find a single quantity extracted from the entire set of sensor readings. The common functions are average, min and max.

- **Quantile Digest:** Quantile is the division of ordered data into some equal sized data subsets. Digest dynamically partitions the data set aiming to have partitions of same size. The algorithm also provides a confidence factor along with the median value.

- **Representative data items:** This technique try to find relatively small parameters that represent an entire class of data items, to combine all the parameters in an unique packet.

2.3.2. Number of Transmissions Minimization

This technique repackages the received set of data into one packet to avoid multiple transmissions without information loss. The output packets differ for length that depends on number of input packets used for concatenation. This scheme attempts to reduce header overhead only. Using this technique an energy saving for a small payload size is achieved.

2.3.3. Redundancy Elimination

Data sensed by nodes is highly correlated, by this algorithm the transmission energy is saved by identifying and discarding spatially redundant data. One of the simplest forms of this technique is duplicate suppression, it tries to forward only one copy among the duplicates received from multiple sensors to higher level node or Base Station

2.4. Data Aggregation Techniques

The following points explain different data aggregation techniques valid for our case, that is to say, based in multi-hop and/or clustering [13].

2.4.1. Spatio-temporal correlation

The spatio-temporal correlation is a significant characteristic of the WSN which can be exploited to enhance the overall network performance.

• **Spatial correlation:** a spatially dense sensor deployment is required for WSN applications in order to achieve satisfactory coverage. This means that multiple sensors record information about a single event in the sensor field.

• **Temporal correlation**. Correlation between consecutive sensor measurements may vary according to the temporal variation characteristics of the phenomenon

2.4.2. An Ultra-Low Power And Distributed Access Protocol

This scheme takes the advantage of the redundancy built in the network. It allows the destination to recover lost data packets based on the data it receives from the neighbourhood of the original sender. Although this scheme works well to reduce the number of bits transmitted considerably, it does not check the flow of redundant packets in the network.

2.4.3. Interdependence of Routing and Data Compression

In this technique, the whole data generated by different sensor nodes are compressed and sent over multiple hops over the network. This helps in eliminating correlations in the representation of the sampled field.

2.4.4. Spatio-Temporal Sampling Rates and Energy Efficiency

The aim of this technique is to reduce the number of transmitting bits at the same time that ensures the number of transmitting packets remains unchanged, but which can be minimized by regulating the network access based on the spatial correlation.

2.4.5. STMA: Spatial and Temporal Multi-Aggregation for State-Based Sensor Data

STMA [14] technique tries to minimize the energy consumption and traffic load when a single node or multiple sensor nodes gather and transmit data from various sub-areas through multi-hop. To achieve spatial aggregation, an aggregation tree with an optimal intermediary between a target area and a base station is created. On the other hand to achieve temporal

aggregation this technique works with different intervals so that users can obtain the amount of data they need without suffering traffic excess.

2.4.6. READA: Redundancy Elimination for Accurate Data Aggregation

READA [15] applies a grouping and compression mechanism to remove duplicate data in the aggregated set before the data is sent to the Base Station, by this way is not largely loses the accuracy of the final aggregated data. This technique use a predictive model derived from cached values, by this way it can confirm if any outlier is a correct value or if it is an error.

2.4.7. An Energy Efficient Clustering Scheme for Data Aggregation

In this technique [16], all the energy consumptions are taken into account both in the phase of data transmissions like in the phase of cluster head rotations. By using this model they are able to obtain the solutions of optimal tier number and the optimal clustering scheme.

2.4.8. An Energy Efficient Spatial Correlation Based Data Gathering

The main objective of this algorithm [17] is to reduce the communication between source and base station. In it, the clusters are formed based on the spatial correlation in the network. The algorithm has two steps. Firstly, the base station collects the data from all the sensor nodes. Secondly, the sensor node transmits the data according to the schedule generated by the base station. If the predicted value is less than the predetermined threshold value, then the sensor node will not communicate the sensed value. However, if the predicted value is more than the threshold, then the sensor node communicates the current data to the base station.

2.4.9. EEBCDA: Energy Efficient and Balanced Cluster-Based Data Aggregation

EEBCDA [18] addresses the problem of unbalanced energy dissipation in cluster where cluster heads transmit data to base station by single-hop. Firstly, the network is divided into rectangular regions. Secondly , the operation is divided into rounds where in every round there are set-up phase and steady-state phase. In each region is selected as CH the node with the maximal residual energy. The region whose CH consumes more energy, means that it has more sensor nodes to take part in cluster head rotation and share energy load, by this way, it is able to balance energy dissipation.

2.4.10. Comparative study of data aggregation techniques

In the following table the characteristics of the techniques discussed above are summarized and compared:

Table 2. Comparative study of data aggregation techniques

[19]	Routing Protocol	Latency	Energy Efficiency	Accuracy
Spatio-Temporal Correlation	Cluster based	Yes	High	High
An Ultra-Low Power and Data Access Protocol	Cluster based	Yes	High	High
Interdependence of routing and Data Compression	Multi-hop based	No	High	Medium
Spatio-Temporal Sampling Rates	Cluster based	Yes	Medium	Medium
Spatio-Temporal Multi-Aggregation	Multi-hop based	No	Medium	Medium
Redundancy Elimination for Accurate Data Aggregation	Cluster based	No	Low	High
An Energy Efficient Clustering Scheme for Data Aggregation	Cluster based	No	High	Medium
An Energy Spatial Clustering Based Data Gathering	Cluster based	No	High	Medium
Energy efficient and Balanced Cluster-Based Data Aggregation	Cluster based	No	High	High

2.5. Data Aggregation using Computational Intelligence

There are many optimization algorithms which are developed based on nature-inspired concepts. For example, evolutionary algorithms (EA) [20] and swarm optimization algorithms are two categories of nature inspired algorithms. EA attempts to simulate the phenomenon of natural evolution, where each species search for beneficial adaptations in an ever changing environment. Genetic algorithms (GA) [21] and differential evolution (DE) [22] algorithms are the example of EA. Swarm optimization algorithms include Particle Swarm Optimization, Ant Colony optimization and Bee Colony Optimization [23].

2.5.1. Genetic Algorithm (GA)

GA is an optimization method which is based loosely on the Darwinian principles of biological evolution, reproduction and "the survival of the fittest". At first, the initial population is randomly generated. New populations are created in subsequent generations through the use of four fundamental mechanisms: selection, crossover, repair and mutation operations [24]. Selection mechanism selects individuals (parents) for crossover and mutation. Crossover is another important technique that causes the exchange of genetic materials between parents to produce offspring, whereas mutation incorporates new genetic traits in the offspring. GA maintains this population and repeatedly modifies it to produce a new generation of

chromosomes. This procedure is repeated until convergence is reaches or until a maximum number of generations are achieved [25].

Advantages:

- It can solve every optimization problem which can be described with the chromosome encoding.
- Genetic algorithms can be easily transferred to existing simulations and models.
- GA is not dependent on the error surface, so we can solve many problems of multi-dimensional, non-differential, non-continuous, and even non-parametrical nature.

2.5.2. Particle Swarm Optimization (PSO)

Swarm Intelligence is a branch of Artificial Intelligence. These systems consist of simple interacting agents organized in small societies, called swarms, which exhibit traits of intelligence, such as the ability to react to environmental threats and decision making capacities. In PSO the individuals called particles change their position with time which fly around in a multidimensional search space. Each particle uses its own experience, and the experience of a neighbouring particle, to adjust its position.

Advantages:

- It is easy to implement.
- Only few parameters need to adjust.
- It is efficient in global search.
- Good quality solutions are possible because of its ability to escape from local optima.
- It has quick convergence [26] [27].

2.5.3. Ant Colony Optimization (ACO)

The ant colony algorithm is used for finding optimal paths from a node to Base Station. This method is based on the behaviour of ants, while searching for food [28]. At first, the ants move randomly. When food source is found, ants walk back to the colony and leave some "markers" (pheromones) that show the path of the food. When other ants come across these markers, they are likely to follow the same path with a certain probability. If they do, then they populate the path with their own markers as they bring the food back. The path gets stronger as many ants follow the same path. Because the ants drop pheromones every time they bring food,

shorter paths are more likely to be stronger. In the meantime, some ants are still randomly searching for closer food sources.

Advantages:

- Inherent parallelism
- It can be used in various dynamic applications.
- Positive Feedback leads to rapid discovery of good solutions.
- They react quickly to the changes in the environment [29] [30].

2.5.4. Artificial Bee Colony Optimization (ABC)

Is a swarm-based artificial intelligence algorithm which is inspired by the intelligent foraging behaviour of honey bees. The position of a food source denotes a possible solution to the optimization problem and the nectar amount of a food source represents the quality of the associated solution. The ABC algorithm has three types of bees: onlookers, scouts, and employed bees [31]. The bee which carries out random search is known as a scout. The bee which is going to the food source and visited by it previously is employed bee. The bee which is waiting on the dance area is an onlooker bee.

Advantages:

- It has few control parameters, i.e. population size, limit and maximum cycle number.
- It is simple, flexible and robust.
- It has fast convergence speed.
- It can be easily used with other optimization algorithms .

2.5.5. Comparison [23]:

In the following table the characteristics of the algorithms discussed above are summarized and compared:

Table 3. Comparison of Data Aggregation using Computational Intelligence

PROBLEM DOMAINS	PSO	ACO	GA
Optimal Node Deployment	Centralized nature of PSO minimizes the area of coverage holes of stationary node positioning.	Distributed nature of ACO is better in solving mobile node deployment issues.	Good for random as well as for deterministic node deployments.
Data Aggregation	It is quite suitable for PSO.	In case of large scale and dynamic WSN it can	Suitable in finding the minimum number of

		perform better.	aggregation points while routing data to the BS.
Energy Efficient Clustering and Routing	PSO shows better performance in selecting the high energy node as CHs in each round and can find an optimal route effectively.	Performs better in maximizing both network lifetime and data delivery to the base station.	GA is used in the formation of a number of pr-defined clusters, which helped in reducing the overall minimum communication distance.

Comparing ACO between GA, we must emphasize that ACO has better solutions for the same computational power are identified, especially when the computational time can be afforded to search for the best possible solution, and that ACO may be a good alternative for solving very complicated networks. ACO approaches demonstrates the potential for implementation in a parallel computing configuration, which could significantly reduce the processing time to find solutions when compared to the present processing time and to solve, discrete optimization problems. The PSO technique is a robust and efficient technique for solving difficult robust and population optimization problems. Both the ACO and PSO algorithms are the data clustering algorithms by implementing swarm behaviour. Whereas the ACO is more applicable for problems where source and destination are predefined and specific. At the same time PSO is a clustering algorithm in the areas of multiple objective, dynamic optimization and constraint handling. In PSO and ABC algorithms Cluster-Heads are optimally distributed. These two algorithms give minimum distance between CH and non-CH nodes. Artificial Bee Colony gives more alive nodes as compare PSO algorithm.

ACO algorithms have an advantage over simulated annealing and genetic algorithm approaches of similar problems when the graph may change dynamically; the ant colony algorithm can be run continuously and adapt to changes in real time.

3. Security Analysis of Secure Data Aggregation Protocols in Wireless Sensor Network

In order to increase wireless sensor network lifetime, data aggregation is applied, is important to consider the security and propose secure data aggregation protocols. The essential of those secure approaches are to make sure that the aggregators aggregate the data in appropriate and secure way. Below two secure data aggregation protocols are explained.

3.1. ESPDA: ENERGY-EFFICIENT AND SECURE PATTERN BASED DATA AGGREGATION

3.1.1. Introduction

In conventional data aggregation methods, Cluster Heads receive all the data from sensor nodes and then eliminate the redundancy by checking the contents of the data as shown in Figure 6(a).

In ESPDA [32] is an energy and bandwidth efficient technique. In it, Cluster Heads prevent the transmission of redundant data from sensor nodes. In ESPDA, instead of transmitting the entire data with redundancy, the sensor nodes send the corresponding pattern codes to CH. If more than one sensor node send the same pattern code to the CH, then only one of them is permitted to send the data to the CH. Thus, ESPDA achieve that data aggregation is performed before the actual data is transmitted from the sensor nodes as illustrated in Figure 6(b).

Figure 6(a). Data Transmission using convetional data aggregation
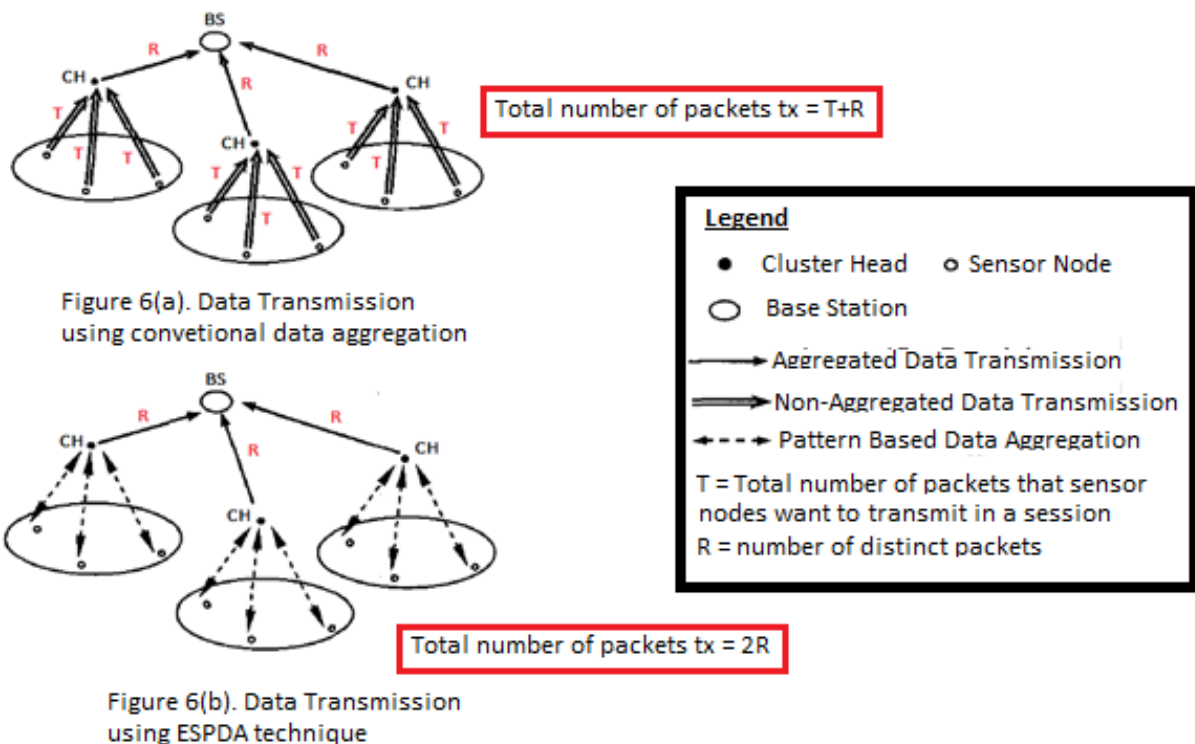
Figure 6(b). Data Transmission using ESPDA technique

Figure 6. Data Transmission using ESPDA

In ESPDA the number of transmitted packets is much less than the conventional data aggregation technique. Therefore ESPDA has more energy efficient.

If let us consider T as the total number of packets that sensor nodes want to transmit in a session, and R as the number of distinct packets, where R less than or equal to (T).

In conventional data aggregation algorithms, the total number of packets transmitted from sensor nodes to CH would be (T) since Cluster Head receives all data packets prior to eliminating redundant data. CH eliminates redundancy and sends R packets to Base Station. Therefore, the total number of packets transmitted in the network is (T+R).

In ESPDA, firstly, CH receives (T) pattern codes from all sensor nodes. Secondly, CH eliminates redundancy based on pattern codes and selects the sensor nodes to transmit the data. Since selected nodes are the nodes that have distinct packets, the total number of packets transmitted from sensor nodes to CH would be R, which are transmitted to base station later. Therefore, the total number of packets transmitted in the network is (2R).
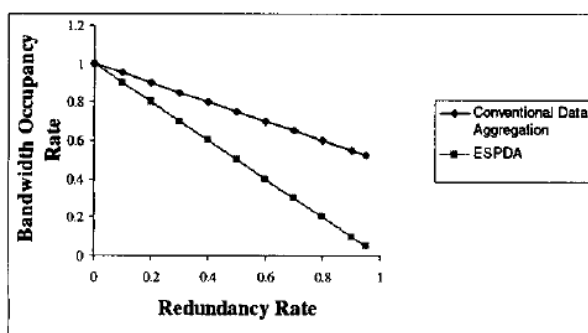

Figure 7. Occupied bandwidth versus redundancy rate

Figure 7 shows that as the redundancy increases the bandwidth efficiency of ESPDA also increases. ESPDA eliminates redundancy before sensor nodes transmit the actual data packets and for this reason, at 100% redundancy, the bandwidth occupancy is close to zero. On the other hand, in conventional data aggregation, bandwidth occupancy is more than 50% of the total bandwidth since all sensor nodes transmit the actual data to be aggregated at CH.

3.1.2. ESPDA Protocol

The main idea of ESPDA is that instead of directly sending the actual sensed data to the CH, the sensor nodes send their pattern codes at first. The pattern code is a representative data which is periodically distributed by the CH. The CH then compares those pattern codes, selects only the unique ones and requests the actual data from these selected nodes. The CH does not need to decrypt the received encrypted data because the data aggregation process is done prior

to the actual data transmission. This reduces the overhead of the CH and thus contributes to the energy efficiency. After that CH can forward the message to BS [33].

In ESPDA, each sensor node is assigned a unique ID (id_i), a node specific secret key (k_i) and a secret key common to all nodes (k). In addition, BS periodically broadcasts a random session number (r_b) in encrypted format using key k ($enc_k(r_b)$). Upon receiving r_b any node i computes the node specific session key ($k_{i,b}$) for data communication by XOR-ing its built-in secret key k_i with r_b, $k_{i,b}=k_i\otimes r_b$. Later on node i uses $k_{i,b}$ to encrypt its actual data. Accompanying this encrypted data, node i also sends its timestamp and its ID number. Those two data will help BS to choose the right rb and compute the right $k_{i,b}$ to decrypt the message. In order to provide data integrity, message authentication code (MAC) of the message using $k_{i,b}$ is also included.

Figure 8 shows the summary of ESPDA protocol. This figure depicts the protocol run in one session. In every new session BS broadcasts a new r_b.

1. i	: compute $k_{i,b} = k_i \otimes r_b$
2. i $\to$ H	: $pattern\_code_i$, $timestamp_i$, id_i
3. H	: compare pattern codes and select unique pattern codes based on timestamps
4. H $\to i_{selected}$	: actual data request message
5. H $\to i_{de-selected}$	: ack message to discard the data (optional)
6. $i_{selected} \to H$	: id_i , $timestamp_i$, $enc_{k_{i,b}}(d_i)$, $mac_{k_{i,b}}(d_i)$
7. H $\to BS$	: id_H , id_i , $timestamp_i$, $enc_{k_{i,b}}(d_i)$, $mac_{k_{i,b}}(d_i)$

Figure 8. Process of ESPDA protocol

3.2. SRDA: SECURE REFERENCE-BASED DATA AGGREGATION

3.2.1. Introduction

The data aggregation paradigm is essential for the lifetime of the network due to the reduced number of broadcasts and collisions.

Secure Reference-Based Data Aggregation (SRDA) [34] protocol that incorporates both data aggregation and security concept together in cluster-based wireless sensor networks. The main idea of SRDA protocol is, nodes transmit the differential data rather than the raw sensed data. It means, the actual data of each node is compared with a reference data, and only the difference between data are transmitted to the Cluster Head in encrypted form.

The basic motivation behind differential data aggregation is that significant changes in sensor measurements occur only when an important event (e.g., a fire event for a temperature network) happens in the environment. In general, these so-called important events occur much less frequently than ordinary events in sensor networks.

SRDA provides a key distribution scheme with low memory overhead to establish secure communication links in the network and then save energy reducing network traffic, since differential aggregation has great potential to reduce the amount of data to be transmitted from sensor nodes to CH.

3.2.2. SRDA Protocol

In SRDA protocol after the key distribution process is implemented. In every new session, any node i computes its reference value M_t by taking the average of last N sensed data value, where $N \geq 1$. The node i then sends encrypted M_t to CH. It uses a secret key shared with CH, kr. CH creates a reference entry for the node i with value M_t. For the subsequent transmission, node i then transmits only the encrypted differential value of the next raw data to the reference value, $M(t+j)-Mt$ where $j>1$. If a new session is started, CH removes the correspondent reference entry. The same concept is applied to CH when it sends the data to higher CH or BS [33].

1. i	: Computes reference value for the new session, Mt
2. $i{\rightarrow}H$	: $enc_k(M_t)$
3. H	: Create reference entry for node i with value $M1$
4. $i{\rightarrow}H$	: $enc_k(M(\text{t+j})-M_t)$, where $j>1$
5. H	: if new session starts, remove correspondent reference entry

The main steps of Reference-Based Aggregation can be summarized as follows

1) Cluster Head reports to a higher level Cluster Head or Base Station and transmits the data packet containing the raw measurement value, Mt, at time instant t, for the first packet in a session.

2) Cluster-head creates a reference entry for that node with value Mt.

3) For subsequent readings $M(t+j)$; $j > 1$, sensor node sends differential data ($M(t+j)$ i Mt) instead of raw data $M(t+j)$.

4) When the session ends for the sensor node, the cluster head removes the reference entry for the sensor node.

The efficiency of this technique is greater when the reference value is larger as compared to the differential values due to the fact that these raw data are not transmitted in packets. Another important factor for the performance of this technique is the variance of the value of successive packet contents, as the variance gets smaller the gains achieved by differential data aggregation increase since a smaller amount number of bits are needed to represent the differential data values. Reference values are kept only during a session of data transfer and the storage lookup overhead on CHs are deleted since reference values are kept only during a session of data transfer.

CHAPTER 3: Proposed Solution - "Redundancy Elimination for Data Aggregation in WSN"

1. Introduction

The proposed solution, called ERDA (Elimination Redundancy for Data Aggregation), tries to combine the strengths of ESPDA [32] and SRDA [34] protocols and implement its own improvements in order to create a more efficient protocol and that able to provide better results. The main and most important improvement in this proposed solution is based on the concept of selection of Cluster Head and which node sends the information when redundant data are detected.

On the other hand, by combining features of SRDA and ESPDA protocols is allowed not to sent the redundant data within a cluster and among different iterations, i.e. redundancy is eliminated first among nodes in the same cluster, and later from the same nodes among consecutive iterations, thus we eliminate 100% redundancy.

In other words, within the same cluster, we achieve that if more than one node has the same or similar data (data in the same range), only one node sends the information, and also that if the same node yields the same or similar information among consecutive iterations, it does not send data. In this way the traffic on the network will be 100% useful, thereby reducing the bandwidth used and what is most important, energy consumption.

2. Work Contribution

(1) **Data Aggregation:** Data aggregation [35] is defined as the process of aggregating the data from multiple sensors to eliminate redundant transmissions. The main objective of ERDA is to create a data aggregation protocol as efficient as possible, using techniques of existing protocols, improving these techniques and improve different parts of the protocol with ideas and own studies always focused on the application of ERDA to the field of agriculture.

(2) **Energy Consumption:** The main goal of data aggregation algorithms is to gather and aggregate data in an energy efficient manner so that network lifetime is enhanced. Whereupon the clear objective of creating a new data aggregation protocol is trying to reduce as much as possible the traffic of redundant information, so reduce as much as possible energy consumption, try to reduce the traffic load and conserve energy of the sensors.

(3) **Improve ESPDA and SRDA protocol:** As mentioned before, ESPDA and SRDA are two important security and data aggregation protocols, but as far as data aggregation is concerned,

they are incomplete or at least could be improved. ESPDA eliminate redundancy between different nodes in the same cluster, but not inside the same node. In contrast, SRDA removes redundancy of the same node, but not among neighbouring nodes. ERDA tries to combine these two characteristics and implements several improvements and new concepts in order to obtain better data aggregation protocol and therefore reduce network energy consumption and extend the life of the nodes. These improves are:

- Minimize the amount traffic of the network.
- Minimize the energy consumption of the network.
- Increase the network life time.
- Cluster Head selection must be better balanced.
- The Selection of which sensor node sends when redundant data is detected must be better balanced.
- Cases that pose a danger situation must be reported to the Base Station.

These improvements are explained in more detail further on.

3. System Model and Assumptions

(1) The network is divided into n levels (multi-hop communication)

(2) The initial number of nodes of each cluster are n nodes and are dispersed randomly.

(3) All the nodes start with the same energy level, 1 (100%)

(4) All the nodes generate randomly a temperature between the value of 40 and 70 centigrade.

(5) The base station (BS) is a node with no energy constraint and enhanced computation capabilities and placed in the center next to the level 1.

(6) Energy consumption of each node who transmit data can be estimated as model which proposed by Heinzelmann et.al [36] :

4. Energy Considerations

Clustering method is used in our wireless communication network. In this case, the data is not delivered directly from CH to BS. Nodes send their data to Cluster Head and it forwards the data to another Cluster Head of upper level up to the base station. Cluster Heads, closer to BS, receive data from the others Cluster Heads and have to send the data from their its own cluster and the data from other clusters. It means that CHs closer to Base Station have more computational load, that means, more energy consumption [37]. On the other hand, clustering enables bandwidth reuse and can, thus, increase system capacity. Using clustering enables better resource allocation and helps to improve power control.

In WSN analysis, energy consumption formulations widely used by researchers are the three part model which proposed by Heinzelmann [36]:
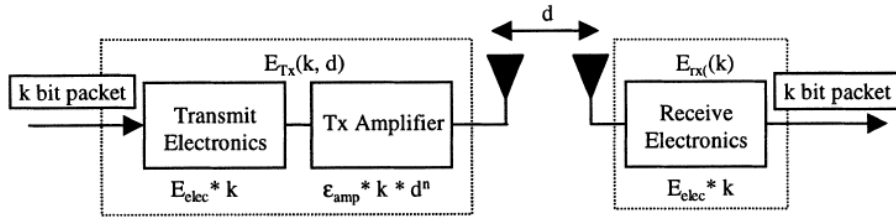

Figure 9. Radio energy dissipation model

In the proposed solution is used the following process to calculate the energy consumed for each node who sends. There are four types of energy: energy for receiving data from the node, energy for transmitting data, energy for receiving data from another cluster, and energy to transmit data to another cluster or to the base station.

1. Energy used to transmit a *l*-bit message a distance d, the radio expends:

$$E_{TX} \begin{cases} l \cdot E_{elec} + l \cdot \varepsilon_{\text{fs}} \cdot \text{d}^2 & \text{for} \quad \text{d} < \text{d}_\text{o} \\ l \cdot E_{elec} + l \cdot \varepsilon_{\text{fs}} \cdot \text{d}^4 & \text{for} \quad \text{d} \geq \text{d}_\text{o} \end{cases} \tag{1}$$

The right formula for our case is the first one, since in our case the data transmission among sensor nodes and Cluster Head, and data transmission between Cluster Heads or Clusters Head and base Station are considered as a small distance transmission.

2. Energy for receiving this message, the radio expends:

$$E_{RX} = l \cdot E_{elec} \tag{2}$$

where "l" is number of bits for each transmission, in our case in each transmission of data is sending 1 bit. E_{elec} it refers to the constant of electrical energy spending in each transmission, its value is $E_{elec} = 50 \cdot 10^{-9}$ J.

Energy consumption of a Cluster Head is illustrated as a simple model as shown in Figure 10. There are two different cases: The CH A (Cluster Head of cluster A) receives and aggregate data from cluster members and sends them to BS via CH B (Cluster Head of cluster B). Thus, CH B receives relaying data from CH A and also receives aggregate data from its members.
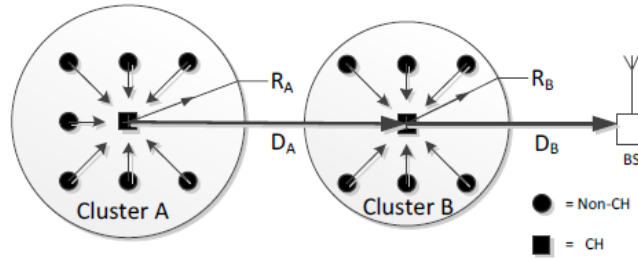

Figure 10. Two levels network model

The energy consumed for each node that is transmitted inside each cluster is:

$$E_{TX}(Node) = l \cdot E_{elec} + l \cdot \varepsilon_{fs} \cdot d^2 = l(E_{elec} + \varepsilon_{fs} \cdot d^2) \tag{3}$$

Where, d is the distance between nodes and ε_{fs} is a constant used in transmission in a short distance. Its value is $\varepsilon_{fs} = 10 \cdot 10^{-6}$J

In the energy consumed by a CH, E_{tot}^{CH}, two components can be identified: 1) energy needed for handling the data of internal cluster E_{own} and 2) energy for relaying data from outer CHs, E_{relay}, so:

$$E_{tot}^{CH} = E_{own} + E_{relay} \tag{4}$$

In the first case, as we can see in Figure 10, there are no data that must be relayed by CH A, therefore the energy component of CH A is only E_{own} . For one sensing cycle, the energy consumed by CH A for sending 1 bit data is:

$$E_{tot}^{CH\_A} = E_{own} = \left(N_{mbr}^A - 1\right)E_{rx}^{CH\_A} + \Upsilon \cdot N_{mbr}^A \cdot E_{tx}^{CH\_A} \tag{5}$$

where N_{mbr}^A is the number of total sensor nodes in cluster A, $(N_{mbr}^A - 1)$ is the number of member nodes which transmit data in cluster A, and $E_{rx}^{CH\_A}$ and $E_{tx}^{CH\_A}$ express the energy for

receiving and transmitting 1 bit data, respectively. The parameter Υ is compression factor whose value range is from 0 to 1. The first part of the equation corresponds to the energy for receiving data from node from the same cluster, and the second part corresponds to the energy used for transmitting to Cluster Head B.

$$E_{Consumed}^{CH\_A} = E_{own}^{CH\_A} = \left(N_{mbr}^A - 1\right)E_{elec} + \Upsilon \cdot N_{mbr}^A(E_{elec} + \varepsilon_{fs} \cdot d_A^2) \tag{6}$$

On the other hand, for CH B case, the total energy is the sum of internal data processing energy and external data relaying energy:

$$E_{tot}^{CH\_B} = E_{own}^B + E_{relay}^B \tag{7}$$

By using the energy model, the own energy consumption of Cluster Head B is:

$$E_{own}^B = \left(N_{mbr}^B - 1\right)E_{elec} + \Upsilon \cdot N_{mbr}^B\left(E_{elec} + \varepsilon_{fs} \cdot d_B^2\right)$$

$$\tag{8}$$

As in eq. (20), N_{mbr}^B is the number of total sensor nodes in cluster B, $(N_{mbr}^B - 1)$ is the number of member nodes which transmit data in cluster B. The first part of the equation corresponds to the energy for receiving data from node from the same cluster, and the second part corresponds to the energy used for transmitting to Base Station.

and the relay energy consumption of Cluster Head B is:

$$E_{relay}^B = \Upsilon \cdot N_{mbr}^A \cdot E_{elec} + \Upsilon \cdot N_{mbr}^A(E_{elec} + \varepsilon_{fs} \cdot d_B^2) \tag{9}$$

In this case, the first part of the equation corresponds to the energy for receiving data from cluster A and the second part refers to the energy for transmitting relay data to the Base Station.

5. Software

MATLAB [38] is the software chosen to develop the proposed solution. MATLAB (abbreviation of MATrix LABoratory) is a mathematical software tool that offers integrated development environment (IDE) with a programming language itself (language M).

MATLAB is a computing environment and fully integrated application development aimed to perform projects where there are involved high mathematical calculations and graphical display of the same. MATLAB integrates numerical analysis, matrix computation, signal processing and graphical display in a complete environment where problems and solutions are expressed in the same way, without using traditional programming.

Consistently and without any cracks, it integrates the key requirements of a system of technical computing: numerical computation, graphics, tools for specific applications and ability to execute on multiple platforms.

6. Process

In the following flowchart, you can see step by step the different processes in order to explain in a simple way the different steps and concepts that have been carried out in the development of the solution proposed by Matlab. These steps will be explained in more detail below flowchart.
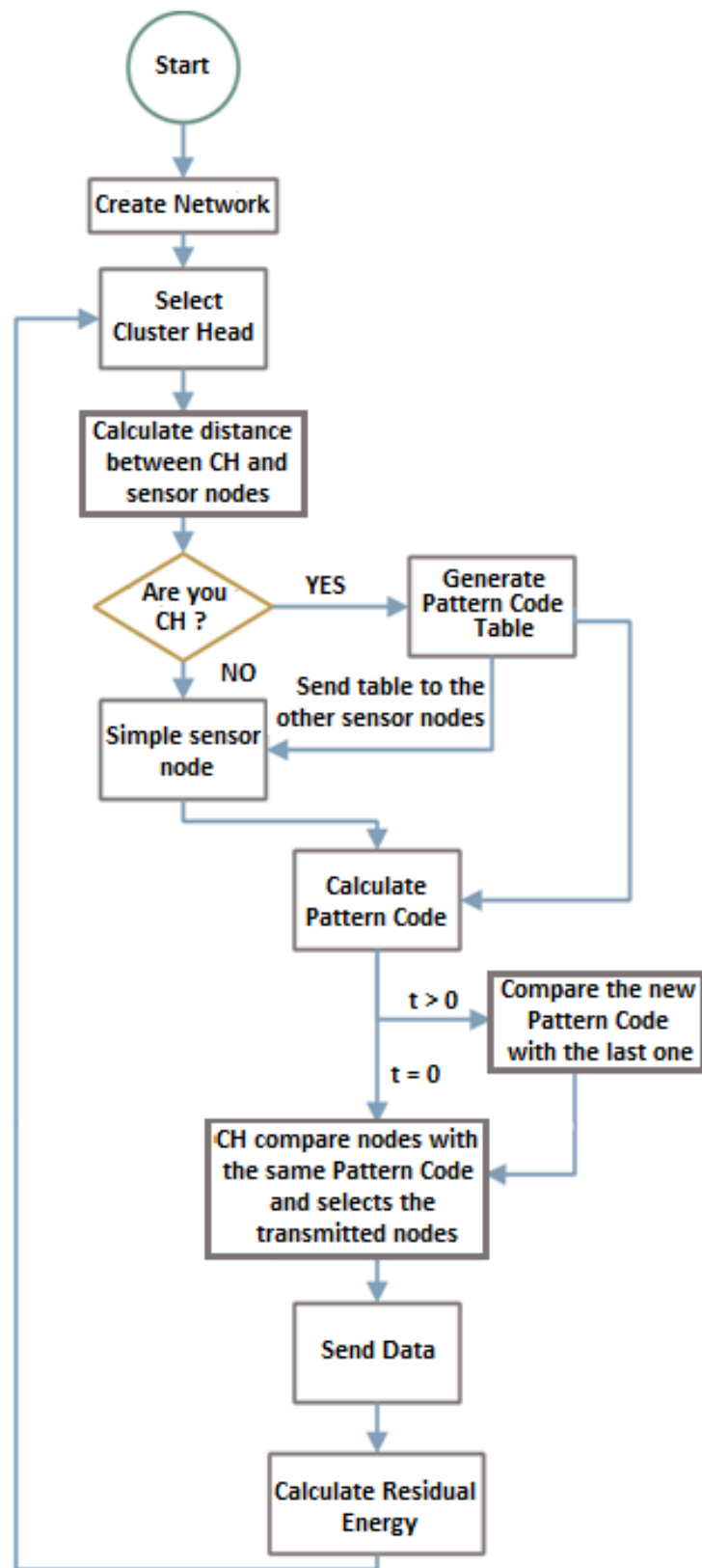
Figure 11. Process of proposed solution

Each of the steps are explained in more detail below:

(1) Network generation: 2 levels, 2 clusters in the first level, 3 clusters in the second level and a Base Station.

(2) Generate number of sensor nodes inside each cluster. Number of nodes randomly between 4 and 10 in each cluster.

(3) All the nodes starts with the same energy level, 1 (100%)

(4) All the nodes generate randomly a temperature between the values 40 and 70 centigrade.

(5) Cluster Head selection: Inside each cluster is selected as a Cluster Head the node who has the most among of energy. If there are more the one node with the same energy, Cluster Head is selected randomly.

(6) Each Cluster Head creates its own Pattern Code Table. This table will be unique for each cluster. Each Cluster Head creates this table by generating 4 random numbers and without repetition 40 and 70. Whereupon 5 temperature ranges are obtained, in which each one of these ranges will be a code pattern, from 1 to 5.

For t = 0;

(7) Each node calculates its own Pattern Code, depending on which range has its own temperature and send the pattern code to Cluster Head.

For t > 0;

(7) Each node calculates the new Pattern Code and compares this new Pattern Code with the last one. If is the same, the node will not send the Pattern code to the Cluster Head. On the other hand, if the new pattern code is different the node will send it to Cluster Head.

(8) Cluster Head compares and finds equal pattern codes. If there are more than one node with the same pattern code, Cluster Head select the node with more energy. If the nodes have the same energy level this selection is randomly.

(9) Selected nodes sent to Cluster Head the information.

(10) Nodes that have transmitted the information to the Cluster Head, calculate the energy consumed in the transmission. (nodes that transmit information to the Cluster Head also calculate the energy consumed in the transmission)

(11) Cluster Head sends the aggregation data to the next level. For the Clusters Head of the level 2, the data will be sent to the closer Cluster Head of the level 2. In the case of level 1, all the aggregation data will be send to the Base Station

(12) Emergency case. If there are temperatures inside the last range (5) of pattern code table, the Cluster Head of this cluster will send an emergency message informing that node and from which cluster comes the danger.

(13) Each Cluster Head calculates the energy consumed in the transmission. For the CHs of level 2, the energy consumed in the transmission between CHs of different levels, and for level 1 case, between CHs and Base Station.

(14) End of itineration.

(15) Each node generate randomly again a temperature between the values 40 and 70 centigrade.

(16) Cluster Head selection. In each iteration is carried out the selection process of the cluster head. This is very important because cluster head node is the node with the highest level of energy consumption and it is desirable that this would be the node with the highest amount of energy.

(17) Continue from point (6).

7. Proposed Solution steps

To develop ERDA protocol, the process has been divided into three distinct phases.

The process development starts with a single cluster in which the parts of the protocol work like ESPDA protocol , as to say the first iteration when t = 0. The steps taken are shown in the following Flow Chart (Figure 12). This step ends with the calculation of the energy consumed for each node after data have been sent to Cluster Head.
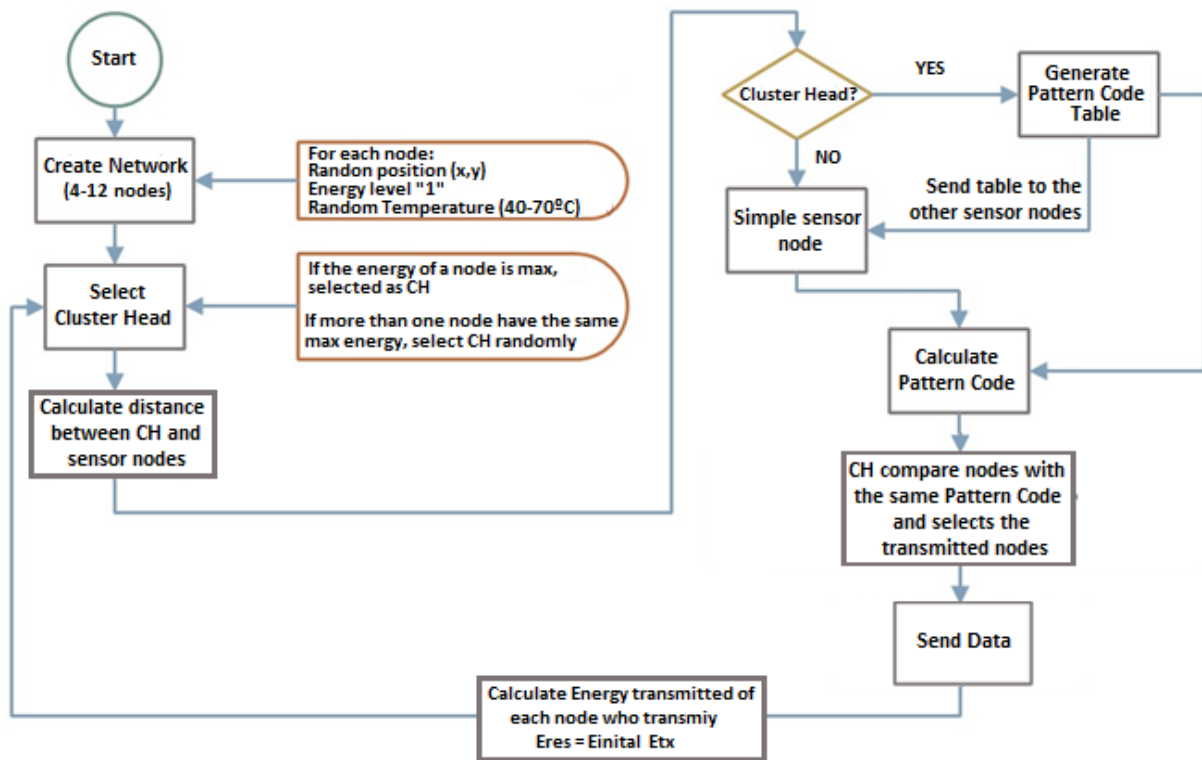
Figure 12. Flow Chart, Step 1

The second phase to develop algorithms and processes that take place when the first iteration is completed. This process is repeated for iterations left. This is the process in which each sensor node compares the current Pattern Code with last one (Figure 13). This phase ends with the calculation of the energy consumed for each node after data are sent to Cluster Head.
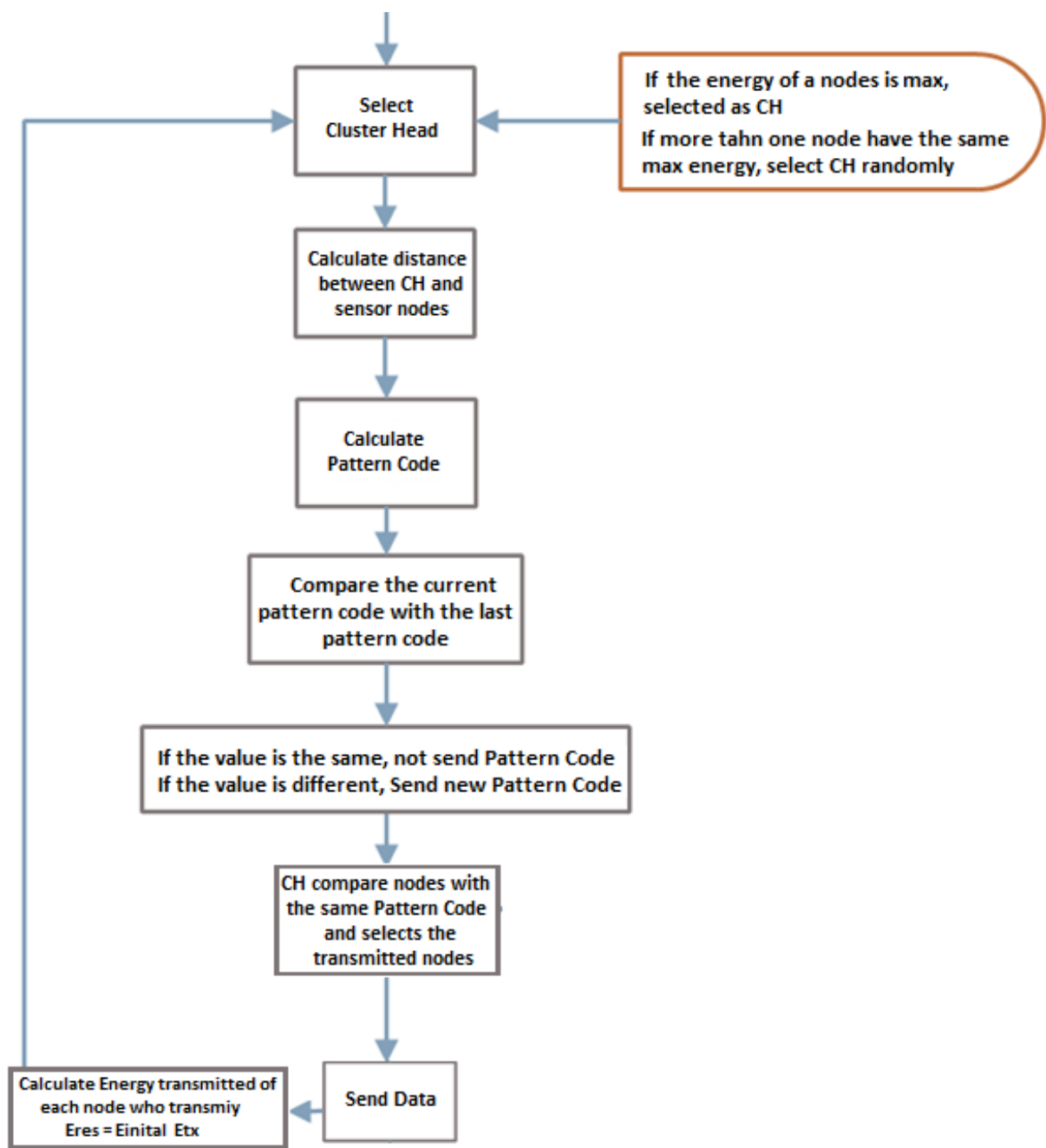
Figure 13. Flow Chart, Step 2

In the third and the last phase of the protocol the previous two steps are held together in one cluster, in order to build a network with five clusters whose two clusters are at the first level, three clusters on the second level and the Base Station. After that the communication between Cluster Heads and also between Clusters Heads and the Base Station is allowed, so it is possible to calculate the energy consumption of the Cluster Head after sending aggregation data to the next level cluster or base station.
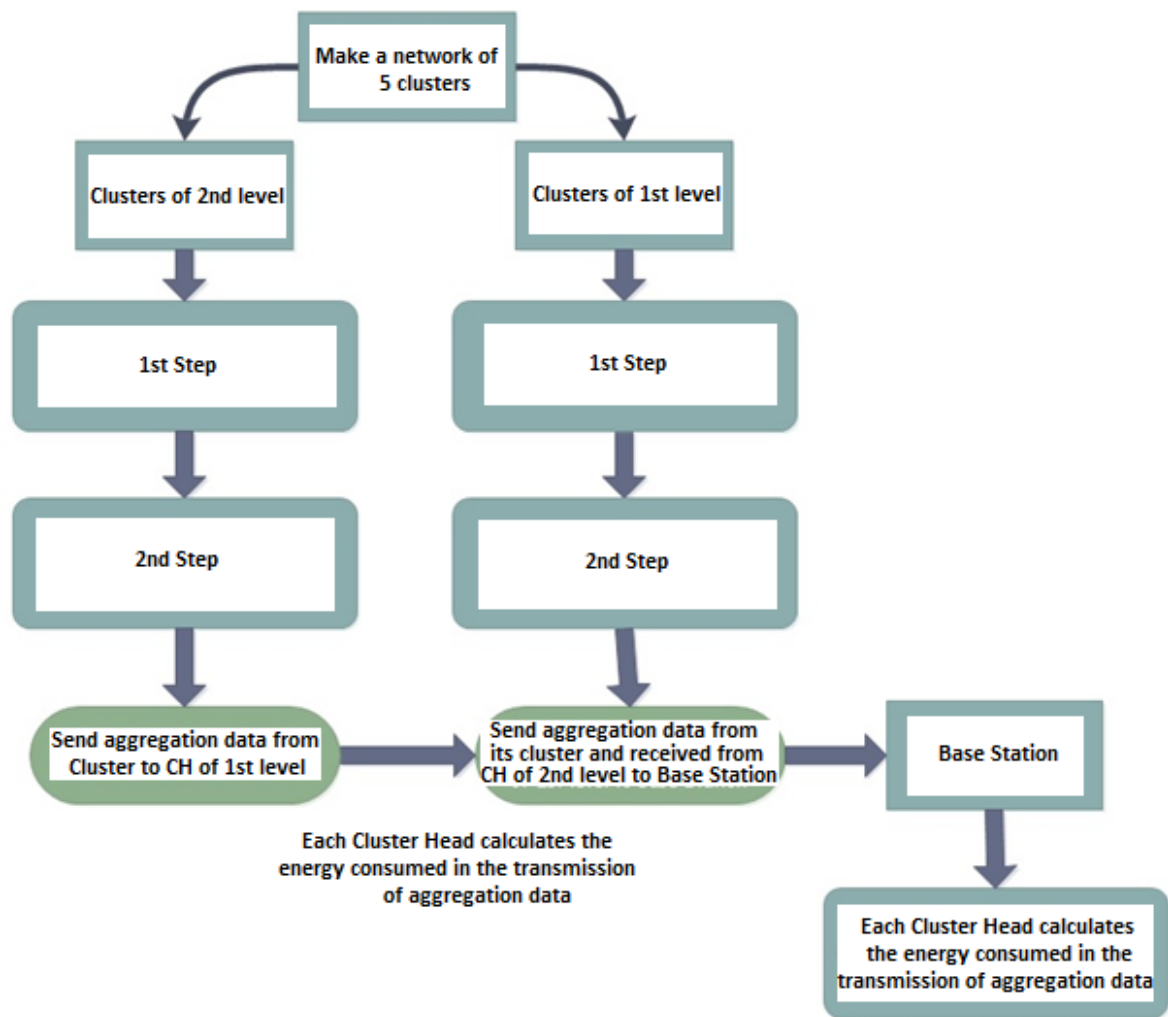
Figure 14. Flow Chart, Step 3

The networking shows in Figure 15, where one of the iterations is captured.
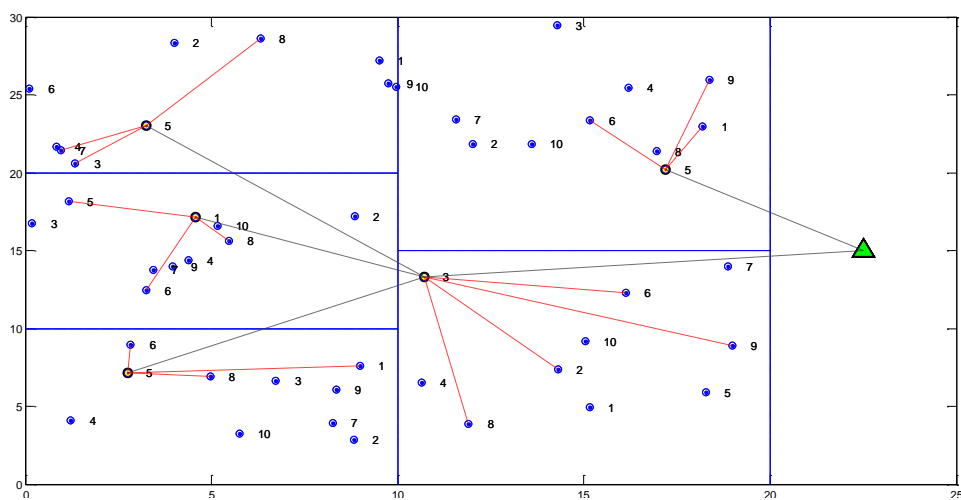

Figure 15. Network example of "Proposed Solution" by MATLAB

Figure 15, shows a network generated by proposed solution, where the sensor nodes are represented by blue points, the sensor nodes selected as a Cluster Head are represented by yellow points and the base station by a green triangle. The network is divided in two levels, two clusters are generated in the first level - next to de base station - and the second level is formed by three clusters.

8. Operating example " Proposed Solution "

In this example are considered six sensor nodes inside the same cluster that are sensing temperature (d1). This sensed parameter is assumed to have threshold values between the ranges 40 to 70 centigrade.

On the figure 16, the nodes sensors are represented from s1 to s5 from left to right and Cluster Head is sensor 6.

In a new session:

(1) Network generation

(2) Generate number of sensor nodes inside each cluster (6 nodes)

(3) All the nodes start with the same energy level, 1 (100%). In this example all nodes have different levels of energy.

Figure 16. Network generated

(4) All the nodes generate randomly a temperature between the values of 40 and 70 centigrade.

(5) Cluster Head selection. Inside the cluster is selected as Cluster Head the node who has the most amount of energy. If there are more the one node with the same energy, Cluster Head is selected randomly. (Sensor 6)
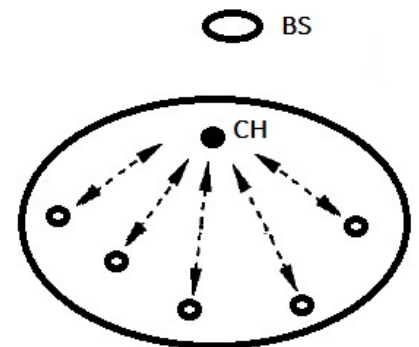
(6) Cluster Head generates Pattern Code Table.

Table 4. Pattern Code Table

Interval values	40 - 47	47 - 53	53 - 60	60 - 65	65 - 70
Pattern Code	1	2	3	4	5

At instant t=0.

(7) Each sensor has different values and with this values any node calculate and send the pattern code using Table 4 as show in Table 5.

Table 5. Pattern Code calculation

	Sensor values	Critical values		Send data
	Temperature	**Pattern code**	**id**	**Energy level**
Senor 1	56	3	1	0,9
Sensor 2	70	5	2	0,81
Sensor 3	58	3	3	0,72
Sensor 4	68	5	4	0,86
Sensor 5	69	5	5	0,84
Sensor 6	51	2	6	0,77

(8) CH compares pattern codes and selects one node among different nodes with the same pattern code. In this case, data sensed by sensor 1 and sensor 3 are redundant, similarly data sensed by sensor 2, sensor 4 and sensor 5 are redundant and data from sensor node 6 is unique. Hence the CH compares the energy levels of different sensor nodes and select only sensor 1, sensor 4 and sensor 6 to transmit the data from each redundant set.

(9) CH reports the selected codes and If sensor node is selected it requires to send actual data, if it doesn't require to drop actual data and not the pattern code.

Table 6. Pattern Code Comparison

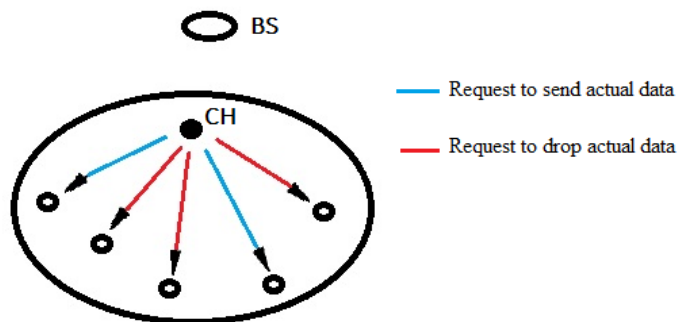| Sensor	Pattern Code	Energy level
s1	3	0,9
s2	5	0,81
s3	3	0,72
s4	5	0,86
s5	5	0,84
s6	2	0,77

Figure 17. Who sends selection

(10) Nodes selected (s1, s4 and s6) send the data.

(11) Cluster Head sends data to Base Station and informs that sensor 2, 4 and 5 are in danger (emergency case).
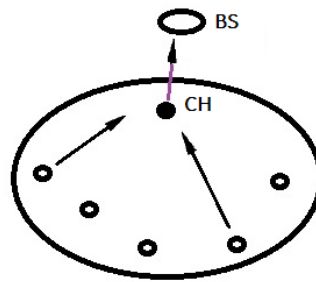
Figure 18. Sending data 1

(12) Recalculate energy level of each node.

Table 7. Nodes energy level

Node	1	2	3	4	5	6
Energy level	0,87	0,81	0,72	0,83	0,84	0,73

At instant t>0.

(13) Any sensor nodes generate a new pattern code.

Table 8. Pattern Code Generation

	Sensor values	Critical values		Send data
	Temperature	**Pattern code**	**id**	**Energy level**
Senor 1	58	3	1	0,87
Sensor 2	66	5	2	0,81
Sensor 3	52	2	3	0,72
Sensor 4	63	4	4	0,83
Sensor 5	64	4	5	0,84
Sensor 6	50	2	6	0,73

(14) Each node compares this new pattern code value with its own previous value.

- If the value is the same (previous value - new value = 0), it is nothing to do.

- If the value is different (previous value - new value ≠ 0), it sends new pattern code.

Table 9. Pattern Codes Comparison

Sensor	New Pattern Code	Last Pattern Code	Difference	Energy level
s1	3	3	0	0,87
s2	5	5	0	0,81
s3	2	3	-1	0,72
s4	4	5	1	0,83
s5	4	5	1	0,84
s6	2	2	0	0,73

53

After the new pattern codes are generated, it can be seen that the values of s1, s2 and s6 haven't changed while the values of s3, s4 and s5 are different. Only these sensors send their pattern code to BS.

(15) Cluster Head receives and compares pattern codes and selects one node among different nodes with the same pattern code.

In this case, pattern code sensed by sensor 3 is unique, but pattern code sensed by sensor 4 and sensor 5 is redundant and taking in to account the energy level, CH selects sensor 5 to transmit the data from each redundant set since it has more energy than sensor 4 and sensor 3.

(16) If sensor node is selected-set requires to send actual data, if it does not it drops actual data.
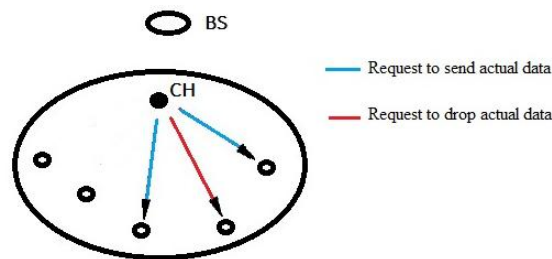

Figure 19. Sending data 2

(17) Nodes selected (s3 and s5) send the data.

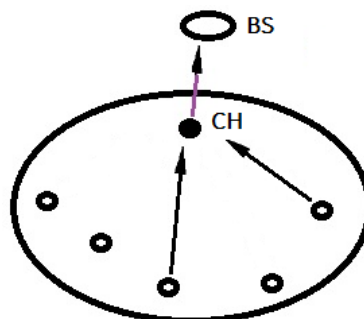(18) Cluster Head sends data to the Base Station and informs that sensor 2 is in danger (emergency case).

Figure 20. Who sends selection

54

(19) Recalculate energy level of each node

Table 10. Nodes energy level

Node	1	2	3	4	5	6
Energy level	0,87	0,81	0,69	0,83	0,81	0,70

As we can see, this protocol at instant t=0, works like ESPDA in which after redundancy is eliminated -based on pattern codes - cluster requests selected sensor nodes to transmit their data. Since selected nodes are the nodes that have distinct packets, the total number of packets transmitted from sensor nodes to cluster-head would be R which are later transmitted to the base station. Therefore, the total number of packets transmitted from sensor nodes to base station is (2R).

On one hand, from instant t>0 (in the same session) the protocol yields that only the pattern codes are sent if they are different from the previous pattern code of the same node. That is to say that they will only be sent if the sensor values have had a significant change. This idea is based on the operation of SRDA protocol.

However, the protocol continue to work like at instant t=0, as to say that if one of this pattern codes is the same, CH will remove the data.

In this way we reduce further data aggregation, so if the data of a node are the same, this node doesn't send its pattern code, and if the pattern code of a node is the same as another node, CH eliminates the redundancy. The main feature of this new protocol, is that the node selected as a CH is always the node with more energy, and also Cluster Head selects which node will send the data depending on the level of energy of each node, and always the node with more quantity of energy sends the data. This is the major improvement of ESPDA and SRDA protocols

9. Algorithm simulation

The following paragraphs will explain the main characteristics of processes and algorithms used in proposed solution.

(1) Cluster Head selection: As explained above, in ERDA, the selection of the Cluster Head is performed by comparing the energy of all nodes that form a cluster. In the event that more than one node has the same energy and this is the maximum, Cluster Head selection will be made

randomly among the candidates. This concept is very important because the nodes that are working as Cluster Head are the nodes that consume more energy, which is the reason why it should be the node with more energy.

Input: Nodes id, Energy node

Output: Cluster Head

```
for i = 1 to n
      if energy equal energy max
            node is selected as a Cluster
      end if
end for

if There are only once candidate for Cluster Head
      node = Cluster Head
end if



if There are more than one Cluster Head candidate
      Select Cluster Head randomly
end if
```

(2) Pattern Code Table Generation: The Pattern Code Table consists in 5 denominated Pattern Code values, each one refers to a range of values between 40 and 70 Celsius. These 5 ranges are randomly generated at t = 0 and in each new session. 4 random numbers are generated in 41 and 69 and are ordered from the smallest to the biggest, so that five temperature ranges can be achieved, each one corresponding to a value from 1 to 5. It achieve that data aggregation is performed before the actual data is transmitted from the sensor nodes.

Input: Cluster Head id

Output: Pattern Code Table

```
if t == 0
      for i = 1 to n
            while n<4
                  Generate 4 numbers randomly between 41 and 69
            end while
      end for

      Order values from small to large
      Pattern Code Table = Interval from 40, Ordered Numbers and 70

end if
```

(3) Calculate Pattern Code and Compare Pattern Codes: In the first iteration (t = 0) each sensor node calculates its own pattern code using the pattern code table generated by the Cluster Head. From the following iterations (t> 0) each sensor node calculates the new pattern code and also makes a comparison of the current pattern code with the last pattern code, if the difference between the two codes is different from zero, this sends the code to Cluster Head, but if the difference is zero, it means that the two adjacent codes are the same and the sensor node doesn't send anything. In this way we eliminate all the redundant data ,meaning both the redundancy between nodes and the redundancy of the values in a sensor node in consecutive iterations. Thus saving a lot of energy consumption.

Input: Number of nodes, Pattern Code Table, Sensor node id

Output: Pattern Code of each node

```
for i = 1 to n
      if temperature is between 40 and the first value of Pattern
      Code Table
            Pattern Code= 1;
      end if
      if temperature is between first value and second value of  Pattern
Code Table
            Pattern Code = 2;
      end if
      if temperature is between second value and third value of  Pattern
Code Table
            Pattern Code = 3;
      end if
      if temperature is between third value and fourth value of  Pattern
Code Table
            Pattern Code = 4;
      end if
      if temperature is between fourth value and 70 value of Pattern
      Code Table
            Pattern Code = 5;
      end if
 end for
       Pattern Code
if t == 0
      Pattern Code
end if

if t > 0
      Calculate difference of the current pattern code with the last
            pattern code
      for i = 1 to
            if difference is equal to cero
                  Sensor node doesn't send
            else
                  Sensor node send pattern code to Cluster Head
            end if
      end for
 end if
```

(4) Who sends selection: The Cluster Head also has equal responsibility as the sensor nodes in data aggregation. It sends the pattern code periodically to all the active sensor nodes to maintain the confidentiality of the pattern codes. After receiving pattern codes from the sensor nodes at the period T, the entire set of codes is classified based on redundancy. Unique patterns are then moved to the 'selected-set' of codes [32]. If more than one node has the same code, the Cluster Head node chooses the node with the greater amount of energy to send the data, if the nodes have the same energy this selection is done randomly. In this way it is obtained that the nodes who send are always the sensor nodes with more energy, whereby we achieve to keep the energy of the sensor nodes with less energy and thus a longer life to whole of the network.

Input: Pattern Code, Sensor node id, Energy Node

Output: Nodes who transmit

```
for i = 1 to number of pattern code values
      Pattern Code Repetition = sensor nodes with the same pattern
      code

      if node with a pattern code value is unique and its energy is
      more than cero
           Node sensor transmit data
      end if

      if more than one node with the same pattern code
           Compare energy among the nodes with the same pattern code
           if energy node is maximum
                 Node sensor transmit data
           end if
           if more than one node have the same energy max
                Select who transmit randomly
           end if

      end if
end for
```

(5) Emergency case: To keep the network safe and to prevent all hardware components from a possible dangerous situation, as in the case of agriculture, a too high temperature could be a case to consider as it could cause a fire. In ERDA" is created a state of emergency in which the Cluster Head informs the Base Station about all the sensors on which the temperature is in the last threshold of the Pattern Code Table, that means at the value of 5.

Input: Pattern Code, Sensor node id

Output: Emergency nodes

```
Cluster Head
for Pattern Code Value equal 5
      Find sensor nodes with Pattern Code value equal 5
end for
```

10.   Proposed Solution enhancements compared with SRDA and ESPDA

(1) **Cluster Head selection**: Inside of each cluster is selected as a Cluster Head ,the node that has the highest amount of energy. Only if there are more the one node with the same energy, Cluster Head is selected randomly. In SRDA and ESPDA protocols this selection is always randomly regardless of the amount of energy of the node, which might involve nodes ran out of power more quickly because the Cluster Heads are responsible of sending the data set of all the cluster to the CH of the next level or to the Base Station, which are the nodes that consume more energy.

This process is very important to do in each itineration because cluster head node is the node with the main level energy consumption and it is desirable that this be the node with the highest amount of energy.

(2) **Who sends selection**: As in the point (1), in ESPDA protocol, if more than one sensor node has the same pattern code, Cluster Head selects which node sends randomly. This concept can be improved because as the energy level is known, it would be desirable for the sensor that has the biggest amount of energy, so that avoid the energy consumption of those sensors with little energy and have a network more energetically balanced. In the ERDA this concept is improving and in this case is always the sensor node with more energy level who send.

(3) **No encryption of data transmission**: SRDA and ESPDA protocols are based on data encryption, and this consumes a lot of energy. Agriculture field is not a kind of critical application in terms of protection of information, thus we consider it's not necessary to encrypt data to transmit. In this is further reduced the energy consumption while maintaining the secure network.

(4) **Emergency case**: If there are temperatures inside the last range (5) of pattern code table, the Cluster Head of this cluster will send an emergency message informing that node and from which cluster comes the danger. SRDA and ESPDA protocols do not have any detection

system emergencies. In this case a too high temperature could pose fire hazard. These emergency cases are solved by ERDA.

11. Comparison and analysis of results

To asses and compare the energy performance of proposed solution, we have carried out various simulations, where they have compared different processes of ERDA with the same process using a protocol without data aggregation, and also there have been different comparisons with SRDA and ESPDA protocols.

(1) Energy Consumption sensors nodes depending of redundancy: Simulation results show that ERDA protocol improves energy efficiency significantly by reducing the number of packets transmitted in data communication, as shown in Figure 21. In this simulation we can see that as bigger is the amount of redundant data, the less data are transmitted and less energy is consumed.

Furthermore, in the protocol that is not using data aggregation, the energy consumption of the network in each iteration is always the same regardless of the amount of redundant data because all the sensor nodes send data, and will be the same number of transmissions as sensor nodes inside the network and never depending if the data are redundant or not.

The pattern code generation requires negligible amount of energy as the algorithm is not complex. The energy required for transmission of pattern codes in ERDA is also negligible since pattern codes consist of few bits.

All the simulations below have been realized with a network of 5 clusters, 10 nodes per cluster - in which one of these nodes is the Cluster Head - and 6 iterations. In the follow figure, the energy consumption is calculated in different cases of redundancy rate, all of these cases are independent, and are not the sum of the above cases, for this reason, as we can see, the case of energy consumption without data aggregation, the amount of energy consumption for each case of redundancy rate is the same regardless of redundancy rate, on the other hand, in the case of energy consumption using the proposed solution, the energy consumption is reduced with increasing de redundancy rate.
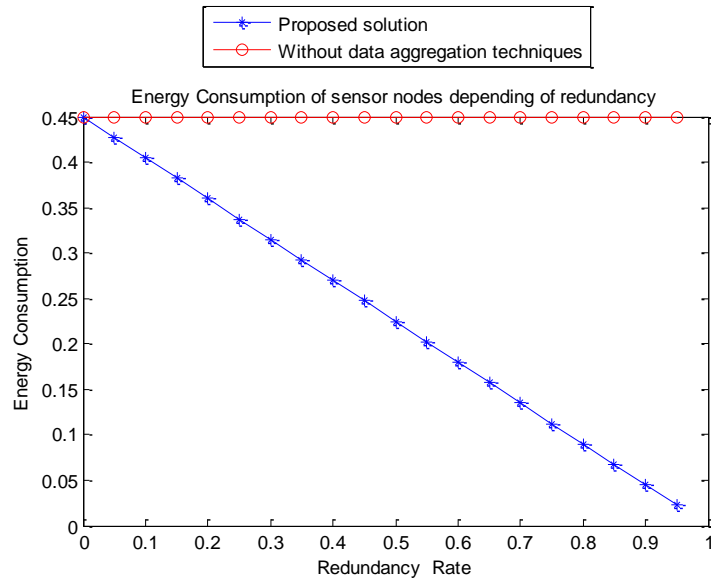
Figure 21. Energy Consumption if sensors nodes depending of redundancy

(2) Energy Consumption of sensor nodes depending of iterations with 40% of redundancy: This difference in energy consumption can be seen even more clearly if we compare the amount of energy consumed in successive iterations. Figure 22 shows that with a level of data redundancy of 40%, in the case of the protocol without use of aggregation of data, power consumption increases linearly. In the case of ERDA protocol this consumption increases more slowly , consuming almost half of the energy in only 6 iterations. Figure 22 shows the power consumption of the nodes, regardless of consumption of Cluster Heads, which have a higher consumption, which will be studied and analyzed later.
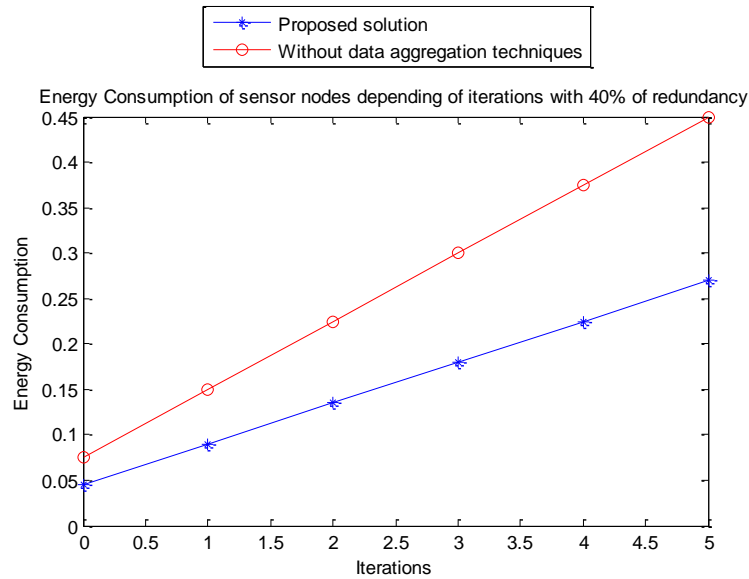
Figure 22. Energy Consumption of sensor nodes depending of iterations with 40% of redundancy

(3) Energy consumption of Cluster Heads depending of redundancy: In the case of energy consumption of the Clusters Heads, we have to keep in mind that the nodes chosen as Cluster Heads has an higher energy consumption, since as explained above, these have their own energy consumption and the energy consumption caused to receive the data transmitted by the other nodes. In Figure 23 we can see the consumption of Cluster Heads in our network, consisting of two levels, the first level with two clusters and the second with 3 clusters. In case of second-level, Cluster Head energy consumption is composed from data received from different nodes and data sending to the next level Cluster Head. On the other hand, the Cluster Heads of the first level, also have energetic consumption of receiving data of CHs from the second level and send the data of your cluster and data received by other CHs to the base station.

As we can see, after 6 iterations, the overall consumption of 5 CHs using a protocol without data aggregation is always the same, regardless of the amount of redundant data, since all nodes are sending always data, and this suppose a huge quantity of energy consumed for Cluster Heads. In the case of ERDA, the energy consumption of CHs is reduced as the amount of redundant data increase, this lead to achieve a significant reduction of energy consumption in the Cluster Heads, a very important thing, because the nodes selected as Cluster Head are those who suffer the most in terms of energy consumption.
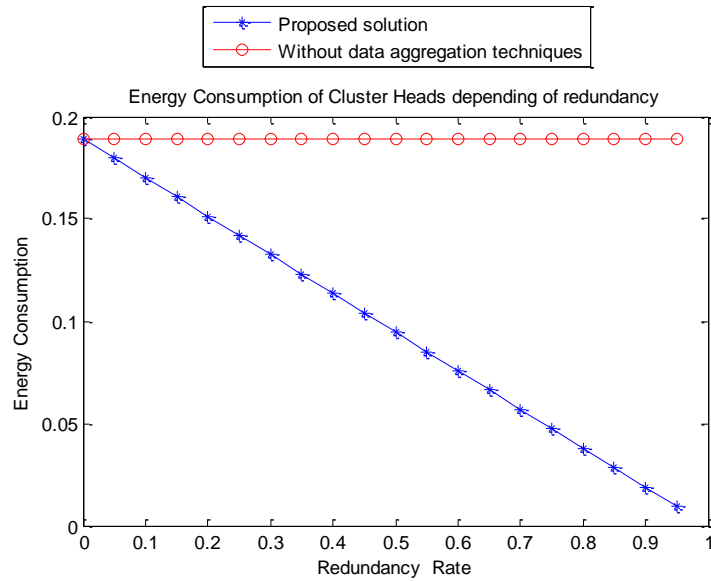
Figure 23. Energy consumption of Cluster Heads depending of redundancy

(4) Energy consumption of cluster Heads depending of iterations with 40% of redundancy: This difference of energy consumption in the CHs can be clearer by giving a view at the power consumption of the Cluster Heads as the iterations are increasing. Figure 24 shows the energy consumption of the nodes selected as Cluster Heads with 40% of data redundancy. As it can be seen there is a big difference between the energy consumption of CHs using ERDA and without using any type of data aggregation protocol. Remember that the Cluster Heads are the nodes with the highest quantity of data received and transmitted, this means that are the sensor nodes with more energy consumption. Using ERDA is possible to reduce CHs energy consumption to about a third.
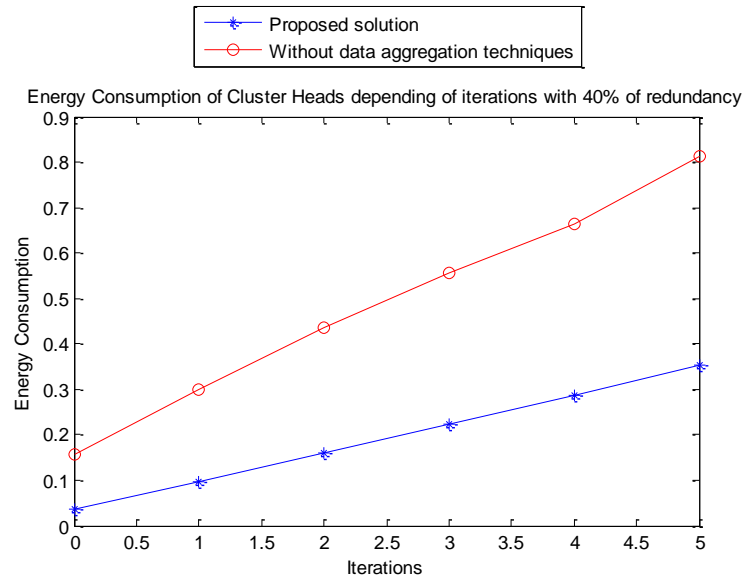
Figure 24. Energy consumption of cluster Heads depending of iterations with 40% of redundancy

(5) Occupied bandwidth rate versus redundancy rate of each sensor node: In our simulations we have considered the communication channel bandwidth between the sensor nodes and the base station. The occupied bandwidth rate is the ratio of bandwidth occupancy and the total available bandwidth. When compared to conventional data aggregation algorithms, as the redundancy increases the bandwidth efficiency of ERDA protocol also increases (Figure 25). In this case of proposed solution and ESPDA yield the same results, since this simulation is realized only at t=0, since ERDA is working like a ESPDA protocol [32].

Later we will see how these two protocols differ as they will increase the number of iterations. At 100% redundancy, the bandwidth occupancy of proposed solution and ESPDA is close to zero, since both eliminates redundancy before sensor nodes transmit the actual data packets. However, in conventional data aggregation bandwidth occupancy is more than 50% of the total bandwidth since all sensor nodes transmit the actual data to be aggregated at the cluster-head.
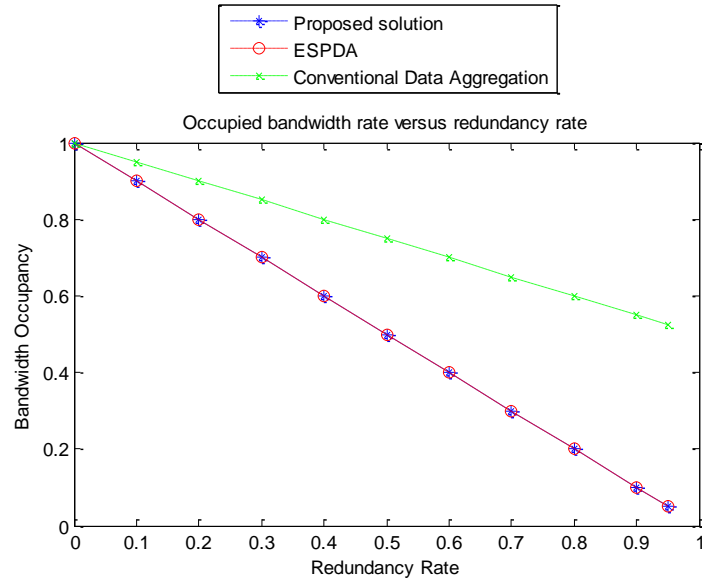
Figure 25. Occupied bandwidth rate versus redundancy rate of each sensor node

(6) Bandwidth occupancy for each node if the data are the same that the last one in the own node: In the following chart (Figure 26) we can see the bandwith used by each node when data from one node are redundant or not between consecutive iterations. In case of SPDA protocol, it ignores the consecutive data from the same node, thus, the node will always send pattern code, in the event that no other node has a temperature within the same range, it will always send to CH. On the other hand, using ERDA protocol if a node has the same code in consecutive iterations, this, will not send the pattern code to the Cluster Head therefore will not send the data. Thereby achieving further reduce traffic and in the end energy consumption
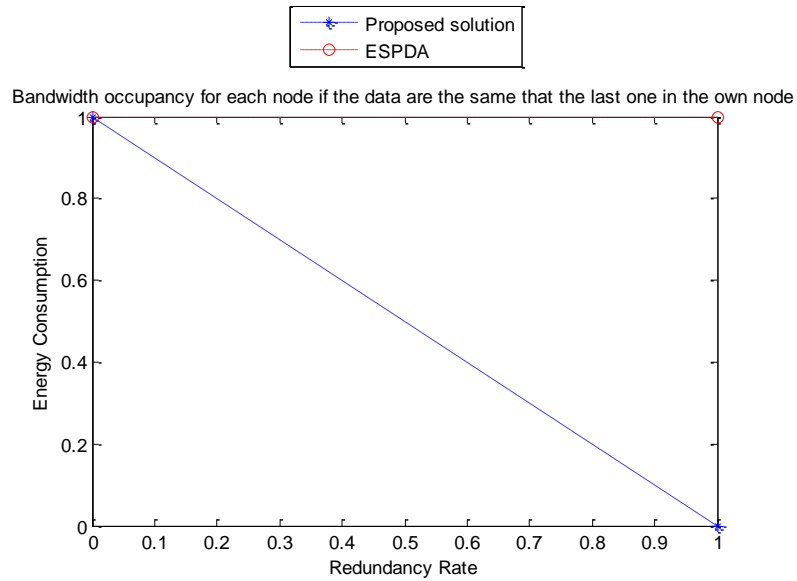
Figure 26. Bandwidth occupancy for each node if the data are the same that the last one in the own node

(7) Occupied bandwidth rate versus redundancy rate: To clarify the concept explained previously, Figure 27 shows the results of energy consumption as increase the amount of redundant data between nodes and with 40% redundancy of data in the same node and between iterations consecutively (intra-node). As we can see if redundancy between nodes is zero, but the redundancy between consecutive data from all nodes is 40, ERDA and SRDA eliminate this redundancy, but not ESPADA protocol. As the redundancy between nodes increases, the SRDA protocol cannot delete this redundancy so it cannot further reduce the bandwidth used. Moreover the ESPADA protocol eliminates all redundancy between nodes, so that redundancy between nodes increases, it reduces the bandwidth used for the transmission, but it never eliminates this 40% redundancy between consecutive data within a same node. Instead, this redundancy intra-node can be eliminated by ERDA protocol is why it always uses less bandwidth to transmit than the ESPDA protocol.
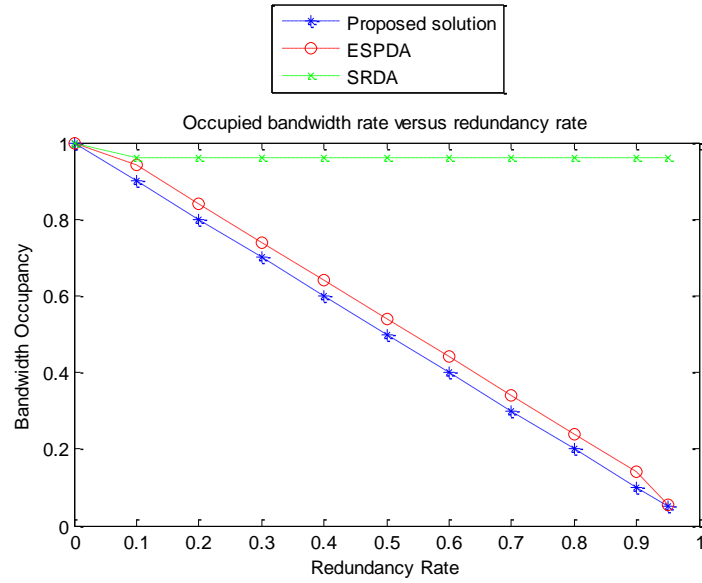
Figure 27. Occupied bandwidth rate versus redundancy rate

(8) Total of occupied bandwidth rate on different iterations: As in the past outcomes, these results can be better visualized by observing the occupation of the total bandwidth in different iterations.

In Figure 28, the values are set as the following: 40% redundancy data between different nodes and 60% redundancy between consecutive data from the same node. We can observe, as SRDA and ESPDA protocols, in order to consider only the redundancy between nodes in the same iteration, and redundancy between the data from the same node in consecutive iterations, respectively. The bandwidth used to transmit is progressively increasing. In this case, redundancy between nodes (40%) is smaller than the redundancy between consecutive data from the same node (60%), for this reason, at the beginning, the ESPDA protocol occupies less bandwidth than SRDA, but after few iterations, ESPDA protocol occupies less bandwidth than the SRDA protocol. On the other hand, proposed solution uses both techniques, whereby the bandwidth used is kept low, with a straight with a slope equal to the average of the two slopes of the other two protocols.
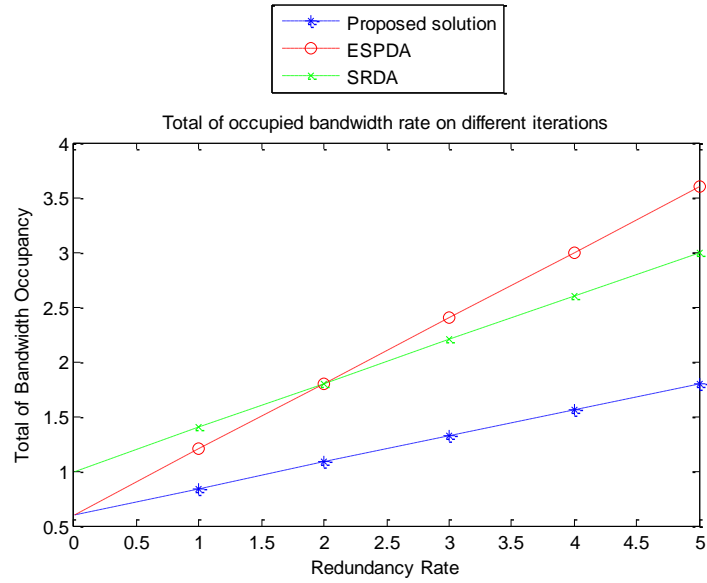
Figure 28. Total of occupied bandwidth rate on different iterations

12.   Conclusion

Data aggregation is one of the critical concepts related to sensor networks in order to reduce the large amount of raw data being transmitted inside the clusters. Designing secure and efficient data aggregation protocols remains a key challenge to be solved.

This point has introduced an energy-efficient and secured data aggregation protocol. In order to provide an efficiency data aggregation protocol, it is proposed a new protocol using the main characteristics of SRDA and ESPDA protocols and new techniques to achieve an efficient and secure protocol. A new data aggregation technique employing the differential data among sensor nodes and inside the same node in consecutive iterations is proposed. Whereupon, ERDA protocol avoids the transmission of redundant data from the sensor nodes to the Cluster Head. In agriculture field is not required to encrypt data to make the data transmission and aggregation more secured and efficient.

Simulations and results show that the proposed solution improves the energy and bandwidth efficiency, the protocol reduces the number of packets transmitted. In this way better results are achieved if compared with ESPDA and SRDA protocols and it helps to achieve the primary goal of energy efficiency.

Conclusion and future works

1. Conclusion

In this project different techniques and data aggregation methods have been studied. In particular, these techniques have been analyzed and suitable selected for wireless sensor network in the field of agriculture. In addition, starting from two of the main concepts of SRDA and ESPDA protocols it has been provided a combination of these concepts with others studied techniques and new ideas. The result has been to achieve a protocol that allows the elimination of redundancy for data aggregation in WSN, which has been very successful in eliminating redundant data, by reducing network traffic and therefore the reduction of energy consumption.

The ideas enlightened in the project, such as the selection of the Cluster Head and the selection of the transmitting nodes, lead to have a more energetically balanced network, and furthermore a secure network which ensure a notification of danger in case of emergency.

In this project, a new data aggregation protocol has been successfully developed and implemented. This protocol contributes to decrease the main problems in wireless sensor systems, as energy consumption, especially in Cluster Head nodes, that nowadays is considered as one of the biggest problems in WSN. In addition, the proposed protocol not only reduces power consumption of each sensor, but also achieves a more energetically balanced network, by choosing the sensor nodes with more energy consumption that are always sensor nodes with more level of energy.

Although with the new concepts provided and implemented and the results obtained, WSN is a technology that is growing and gaining more applications every day. Therefore, there are many concepts that have to be investigated in order to have a more reliable, safer, energetically efficient and with more data integrity.

2. Future work

The main objective of this project is to develop a new protocol of data aggregation in order to reduce as much as possible the network traffic, thus the energy consumption of sensor nodes. As future work to continue in this proposed solution, it is proposed to perform data compression between transmissions of aggregation data between Cluster Heads and between Cluster Head and the Base Station. By compressing data the traffic among CHs and BS will be reduced, it means network traffic, and thus further reduce the total energy consumption.

REFERENCES

[1] Aqeel-ur-Rehman, A. Z. Abbasi, N. Islam y Z. A. Shaikh, «A review of wireless sensors and networks' applications in agriculture,» Computer Standards & Interfaces, 2014.

[2] D. R. sethi, «IMPLEMENTATION OF ENERGY EFFICIENT PROTOCOL FOR WIRELESS SENSOR NETWORK.,» Department of Computer Science and Engineering, 13 May 2010.

[3] J. N. AL-KARAKI y A. E. KAMAL, «ROUTING TECHNIQUES IN WIRELESS SENSOR NETWORKS: A SURVEY,» de *IEEE Wireless Communications*, December 2004.

[4] D. D. Hogrefe y D. O. Alfandi, «Wireless Ad Hoc & Sensor Networks,» TELEMATICS GROUP, WS 2010/2011.

[5] P. D. MARCO, «Protocol Design and Implementation for Wireless Sensor Networks,» April 2008.

[6] Geetu y S. Juneja, «Performance Analysis of SPIN and LEACH Routing Protocol in WSN,» *International Journal Of Computational Engineering Research,* vol. 2, nº 5, Setember 2012.

[7] X. Liu, «A Survey on Clustering Routing Protocols in Wireless Sensor Networks,» 9 August 2012.

[8] K. Maraiya, K. Kant y N. Gupta, «Wireless Sensor Network: A Review on Data Aggregation,» *International Journal of Scientific & Engineering Research,* vol. 2, nº 4, April - 2011.

[9] S. Anavatti y S. Sirsikar, «Data Aggregation Techniques to Remove Redundancy in Wireless Sensor Networks: Brief Overview.,» *International Journal of Advance Foundation and Research in Computer (IJAFRC),* vol. 1, nº 12, December 2014.

[10] E. FASOLO, M. ROSSI, J. WIDMER y M. ZORZI, «IN-NETWORK AGGREGATION TECHNIQUES FOR WIRELESS SENSOR NETWORKS: A SURVEY,» *IEEE Wireless Communications,* April 2007.

[11] O. D. Incel y B. Krishnamachari, «Enhancing the Data Collection Rate of Tree-Based Aggregation in Wireless Sensor Networks», 2008.

[12] S. K. Singh, M. P. Singh y D. K. Singh, «A Survey of Energy-Efficient Hierarchical Cluster-Based Routing in Wireless Sensor Networks,» *Int. J. of Advanced Networking and Applications ,* vol. 2, nº 2, pp. 570-580, August 11, 2010 .

[13] A. Tripathi, S. Gupta y B. Chourasiya, «Survey on Data Aggregation Techniques for Wireless Sensor Networks,» *International Journal of Advanced Research in Computer and Communication Engineering,* vol. 3, nº 7, July 2014.

[14] H. S. KIM y W. H. KWON, «Spatial and Temporal Multi-Aggregation for State-Based Sensor Data in Wireless Sensor Networks,» *Telecommunication Systems,* p. 161–179, 2004.

[15] K. Khedo, R. Doomun y S. Aucharuz, «READA: Redundancy Elimination for Accurate Data Aggregation in Wireless Sensor Networks,» *Wireless Sensor Network,* nº 2, pp. 300-308, 2010.

[16] J.-T. Meng y S.-Z. Feng, «An Energy Efficient Clustering Scheme for Data Aggregation in Wireless Sensor Networks,» *JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY 28,* p. 564–573, May 2013.

[17] C.Tharini y P. V. Ranjan, «AN ENERGY EFFICIENT SPATIAL CORRELATION BASED DATA GATHERING ALGORITHM FOR WIRELESS SENSOR NETWORKS,» *International Journal of Distributed and Parallel Systems,* vol. 2, nº 3, May 2011.

[18] J. Yue, W. Zhang, W. Xiao, D. Tang y J. Tang, «Energy Efficient and Balanced Cluster-Based Data Aggregation Algorithm for Wireless Sensor Networks,» *Procedia Engineering 29 ,* 2012.

[19] A. Tripathi, S. Gupta y B. Chourasiya, «Survey on Data Aggregation Techniques for Wireless Sensor Networks,» *International Journal of Advanced Research in Computer and Communication Engineering,* vol. 3, nº 7, July 2014.

[20] Khalil, E. A, Attea y B. A., «Energy-aware evolutionary routing protocol for dynamic clustering of wireless,» *Swarm and Evolutionary Computation,* pp. 195 - 203, 2011.

[21] S. Hussain, A. W. Matin y O. Islam, «Genetic Algorithm for Hierarchical Wireless Sensor Networks,» *JOURNAL OF NETWORKS,* vol. 2, nº 5, SEPTEMBER 2007.

[22] X. Li, L. Xu, H. Wang, J. Song y S. X. Yang, «A Differential Evolution-Based Routing Algorithm for Environmental Monitoring Wireless Sensor Networks,» *Sensors,* nº 10, pp. 5425-5442, 2010.

[23] L. Kaur y D. Kumar, «Optimization techniques for Routing in Wireless Sensor Network,» *International Journal of Computer Science and Information Technologies,* vol. 5, pp. 4719-4721, 2014.

[24] D. S. Dief y Y. Gadallah, Classification of Wireless Sensor Networks Deployment Techniques, Communications Surveys & Tutorials, IEEE, (vol. 16,issue 2), 2014.

[25] O. Islam y S. H. , «Genetic Algorithm for data aggregation trees in Wireless sensor networks,» *3rd IET International Conference on Intelligent Environments,* pp. 312-316, Jan. 2007.

[26] R. V. Kuknarni y G. Kumar, «Particle Swarm Optimization in Wireless Sensor Networks: A Brief Survey,» *Man and Cybernetics,* vol. IEEE Transactions on Systems, March 2011.

[27] K. Veeramachaneni y L. A. Osadciw, «Dynamic sensor management using multi objective particle swarm optimizer,» *Swarm Optimizer, SPIE Defence and Security Symposium,* April 16-20, 2004.

[28] R. Geetha y G. U. Srikanth, «Ant Colony optimization based Routing in various Networking Domains – A Survey,» *International Research Journal of Mobile and Wireless Communications,* vol. 3, nº 1, 2012.

[29] R. Misra y C. Mandal, «Ant aggregation: ACO for optimal Data aggregation in Wireless Sensor Networks,» de *Proceedings of the Third IEEE and IFIP International Conference on Wireless and Optical Communications Networks(WCON 2006)*, Le meridian, Bangalore, India, Apr 11-13, 2006.

[30] Y. Kao, W.-H. Liao y Chien-Ming, «Data Aggregation in wireless sensor networks using Ant Colony algorithm,» *Journal of network and computer applications,* pp. 387-401, 31(2008).

[31] F. S., Abu-Mouti, M. E. y El-Hawary, «Overview of Artificial Bee Colony (ABC) Algorithm and Its Applications,» de *International Systems Conference, IEEE 2012*.

[32] H. cum, S. Ozdemir, P. Nuir y D. Muthuavinashiappun, «ESPDA: ENERGY-EFFICIENT AND SECURE PATTERN BASED DATA AGGREGATION FOR WIRELESS SENSOR NETWORKS», Oct 2003.

[33] T. M. Rahayu, S.-G. Lee y H.-J. Lee, «Security Analysis of Secure Data Aggregation Protocols in Wireless Sensor Networks», Feb 2014.

[34] H. O. Sanli, S. Ozdemir y H. C. am, «SRDA: Secure Reference-Based Data Aggregation Protocol for Wireless Sensor Networks», Sept 2004.

[35] K. Maraiya, K. Kant y N. Gupta, «Wireless Sensor Network: A Review on Data Aggregation,» *International Journal of Scientific & Engineering Research,* vol. 2, nº 4, April -2011.

[36] A. Suharjono, Wirawan y G. Hendrantoro, «A NEW UNEQUAL CLUSTERING ALGORITHM USING ENERGY-BALANCED AREA PARTITIONING FOR WIRELESS SENSOR NETWORKS,» *INTERNATIONAL JOURNAL ON SMART SENSING AND INTELLIGENT SYSTEMS,* vol. 6, nº 5, pp. 1808 - 1829, DECEMBER 2013.

[37] W. B. Heinzelman, A. P. Chandrakasan y H. Balakrishnan, «An Application-Specific Protocol Architecture for Wireless Microsensor Networks,» *IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS,* vol. 1, nº 4, pp. 660 - 670, OCTOBER 2002.

[38] MathWorks, «MATLAB R2015a», March 2015.