An Energy-Efficient and Load-Balancing Cluster-Based Routing Protocol for Wireless Sensor Networks

Rwan Ibrahim

Department of Electrical & Computer Engineering
McGill University
Montreal, Quebec, Canada

May 2013

A thesis submitted to McGill University in partial fulfillment of the requirements for the degree of Master of Engineering.

Abstract

Clustering is a popular routing technique in configuring Wireless Sensor Networks (WSNs). It can determine the communications between all nodes to collect data in an efficient manner. It handles the main challenge of energy-efficiency in WSNs, and can be used to re-configure the network according to changes in the nodes' conditions. This thesis contributes to the routing in WSNs by proposing an Energy-efficient and Load-balancing Cluster-based (ELC) routing algorithm for Carrier Sense Multiple Access (CSMA)-based WSNs. In particular, both distance and residual energy are taken into consideration in developing the cluster-head selection procedure while ensuring that the network has a desired number of cluster heads. In addition to distance, cluster size is also used in formulating the cost function for cluster forming in order to balance load and energy consumption among the nodes, and hence, enhancing the network lifetime. Besides, ELC employs multi-hop inter-cluster routing based on a lowest-cost path approach that considers both energy efficiency and load balancing. Illustrative simulation results show that ELC consumes less energy and offers longer network lifetime as compared to other existing cluster-based routing algorithms such as Low-Energy Adaptive Clustering Hierarchy-Centralized (LEACH-C) protocol and Central Base Station Controlled Density Aware Clustering Protocol (CBCDACP).

Abrégé

Le clustering est une technique de routage populaire utilisée dans la configuration d'un réseau de capteurs sans fil. Cette technique peut établir les paramètres de communication entre tous les nœuds du réseau pour une collecte de données plus efficace. Elle traite l'obstacle principal à la performance des réseaux de capteurs sans fil, l'efficacité énergétique, et peut être utilisée dans la reconfiguration du réseau selon le changement de conditions des nœuds. La contribution de cette thèse au domaine de routage dans les réseaux de capteurs sans fil consiste dans la présentation d'un nouvel algorithme de routage à base de clustering écoénergétique et d'équilibrage de charge (en anglais, Energy-efficient and Load-balancing Cluster-based routing algorithm ou ELC) pour les réseaux de capteurs sans fil à base de accès multiple avec écoute de porteuse. Particulièrement, les critères de distance et énergie résiduelle sont pris en considération dans la formulation de la procédure de sélection des Cluster Heads (CHs) tout en garantissant que le réseau est formé en tout temps par un nombre désirable de CHs. Outre que la distance, la taille du cluster est de même utilisée dans la formulation de la fonction du coût de la formation des clusters. Ceci vise à équilibrer la répartition de charges et l'énergie des nœuds du réseau, et par conséquence, à aboutir à une plus longue durée de vie du réseau. En outre, ELC emploie une technique de routage inter-cluster avec sauts multiples qui se base sur une approche au moindre coût qui prend en considération l'efficacité énergétique et l'équilibrage de charge dans le réseau. Les simulations démontrent que ELC consomme moins d'énergie et aboutit à une plus longue durée de vie du réseau par rapport à d'autres algorithmes de routage à base de clustering comme LEACH-C et CBCDACP.

Acknowledgments

Foremost, I would like to express my gratitude to my supervisor Prof. Tho Le-Ngoc for his continuous support for my M.Eng. study and research, for his motivation and enthusiasm. His guidance helped me in all the time of research and writing of this thesis.

Besides my supervisor, I would like to thank Dr. Quang Dung-Ho for his detailed and constructive comments, and for his important support throughout this work. Furthermore, I would like to thank my colleagues Leonardo Jimenez Rodriguez and Duy T. Ngo for their patience and helpful suggestions throughout writing this thesis.

Last but not the least, I would like to express my deepest appreciation to my parents for their unconditional support throughout my degree. Their love provided my inspiration and was my driving force. I would also like to thank my brother and sister for standing by me through all the hard times.

Contents

List of Figures

List of Tables

List of Acronyms

ADEEC	Adaptive Energy Efficient Clustering
ADRP	Adaptive Decentralized Re-clustering Protocol
BCDCP	Base-station Controlled Dynamic Clustering Protocol
BIDRP	Base station Initiated Dynamic Routing Protocol
BS	Base Station
CBCDACP	Central Base station Controlled Density Aware Clustering Protocol
CDMA	Code Division Multiple Access
CH	Cluster Head
CM	Cluster Member
CPMS	Clustering Protocol with Mode Selection
CSMA	Carrier Sense Multiple Access
CSMA/CA	Carrier Sense Multiple Access/Collision Avoidance
CSMA/CD	Carrier Sense Multiple Access/Collision Detection
CTPEDCA	Cluster-based and Tree-based Power Efficient Data Collection and Aggregation
dB	Decibel
DF	Density Factor
ELC	Energy-efficient and Load-balancing Cluster-based
FCS	Frame Check Sequence
GHz	Gigahertz
IEEE	Institute of Electrical and Electronics Engineers
kbps	kilobits per second
LEACH	Low-Energy Adaptive Clustering Hierarchy
LEACH-C	Low-Energy Adaptive Clustering Hierarchy-Centralized
MAC	Media Access Control

MHR	MAC Header
MHz	Megahertz
MPDU	MAC Protocol Data Unit
MSDU	MAC Service Data Unit
NAN	Neighborhood Area Network
NS-2	Network Simulator version 2.35
PHY	Physical
PPDU	Physical Protocol Data Unit
PSDU	Physical Service Data Unit
TDMA	Time Division Multiple Access
WSN	Wireless Sensor Network

Chapter 1

Introduction

Recent advances in technology encourage the wide deployment of Wireless Sensor Networks (WSNs). WSNs have been first considered for military applications where battlefields information is gathered at sensor nodes, collected via wireless links, and interpreted at a Base Station (BS). With the development of low-cost and smaller sensor nodes, WSNs have been recently considered for various civilian applications. Thus, more interest is directed towards improving different aspects of WSNs to provide better services for the public [1]. One of the most important aspects in WSNs is the routing techniques that are used to relay data among the nodes in a WSN. Routing has a major effect on the performance and efficiency of WSNs. Energy efficiency is one of the main challenges in developing routing techniques since sensor nodes have limited amount of energy. A popular technique in saving energy and extending network lifetime is clustering, which has the advantage of being able to configure the network based on the nodes energy requirements.

The rest of this chapter is structured as follows. Section 1 defines and describes WSNs. Section 2 presents the motivation behind the research in this thesis. Section 3 outlines the contributions and structure of this thesis.

1.1 Wireless Sensor Networks

A typical WSN consists of a number of sensing nodes that collect information from the surrounding environment and forward it to a collector (referred to as the BS) for further processing. Sensing nodes are normally small, inexpensive, and have limited processing, computing and energy resources. Each of the sensing nodes has a data processing unit, a communication unit and a power unit. Depending on the type of application, these nodes are equipped with different kinds of sensors, such as temperature, humidity and motion detectors. The sensors gather information from the environment which will be processed and transferred to the BS or another node via a wireless link. An example of a WSN is shown in Fig. 1.1 in which the system components of a sensor node are also illustrated [2]. WSNs have been widely used in recent years due to the low-cost and ease of deployment. Moreover, a WSN is self-organizing and can be left unattended once deployed [3]. WSNs have great contributions in many applications such as home applications which will be discussed next.

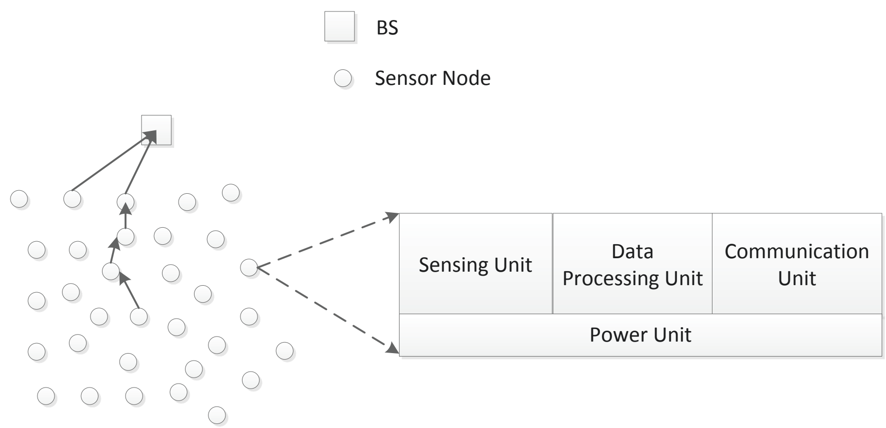


Fig. 1.1: A WSN and system components of a sensor node.

1.1.1 Home Applications

The demand for smart homes has increased with the need of enhancing life quality at homes and their surrounding area. A smart home is equipped with smart objects that interact with each other as part of a home network. The home network is connected to the outside world through a gateway that processes the collected data from the smart objects [4]. Since the scale of such networks is not large, there is less concern about scalability issues in these networks.

The smart objects deployed at a home and its surrounding area are associated with smart sensors. The sensors, for instance, can be used to measure temperature, CO_2 level, water level, humidity or light as well as motion detectors or smoke detectors [4]. In order to establish a home network, a WSN, which consists of a number of sensor nodes and a BS as a gateway, can be used.

The most important types of smart home applications can be classified into safety and security, monitoring and healthcare.

1.1.1.1 Safety and Security

A WSN is capable of providing the technology to improve home safety. This can be done through home monitoring and alarm systems which are deployed at homes and their surrounding area. It can provide information about different dangerous situations, such as gas leakage, thief intrusion, water leakage and the presence of fire [5] [6].

Traditional fire detection system has many problems. It uses cables and wires that can be easily damaged when a fire occurs, and hence leads to signal failure. Moreover, wired systems have maintenance problems and require long construction period. Therefore, wireless fire alarm systems can be employed using a WSN, since it provides reliability and stability [6] [7]. An example of a WSN used to provide a fire detection system at homes is shown in Fig. 1.2. Here, sensors are used to detect smoke and CO gas concentration as well as to measure temperature. The information from the sensors is forwarded towards the BS for data processing [7]. If a fire is suspected, either an alarm is turned on or a message is sent to the owner on his mobile phone. Similar process can be applied to gas leakage, thief intrusion and water leakage.

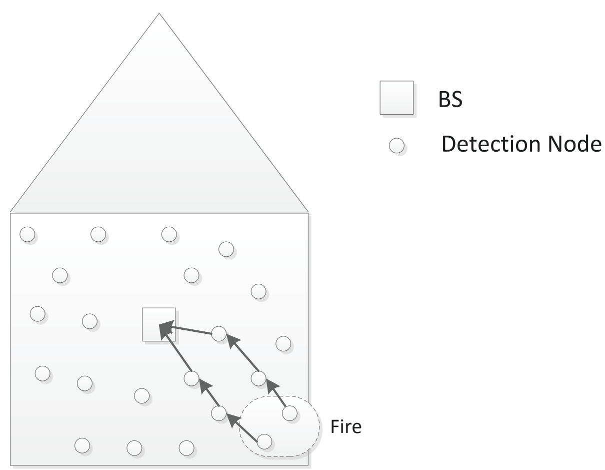


Fig. 1.2: An example of fire detection using WSN.

1.1.1.2 Monitoring

The high amount of energy consumed by home appliances contributes to the negative impact on the environment. Therefore, more effort is focused on energy management at homes in order to reduce the consumption especially during peak hours. By using WSN, the amount of energy consumed by each appliance can be monitored over time and suitable actions can thus be determined. From the collected data, an energy management system can be set according to the owner's habits. Moreover, the monitored data of energy consumption can be delivered to the energy utility to improve the efficiency of energy generation [4] [8].

In addition, the conditions of a small-scale farm associated with a home can be monitored to improve production efficiency and quality of products [9]. By monitoring crops and their requirements, farmers can identify the right amount of water, fertilizer and pesticide to enhance yield standard. Climate and soil sensors can be used to prevent damages to crops and enhance production [9] [10]. An example of a WSN used to monitor a farm associated with a home is shown in Fig. 1.3. Data collected from the different sensors is delivered to the BS (associated with a personal computer) located inside the home of the farmer [9]. By monitoring the farm conditions, the farmer can adopt the appropriate measures that suit the requirements of the crops.

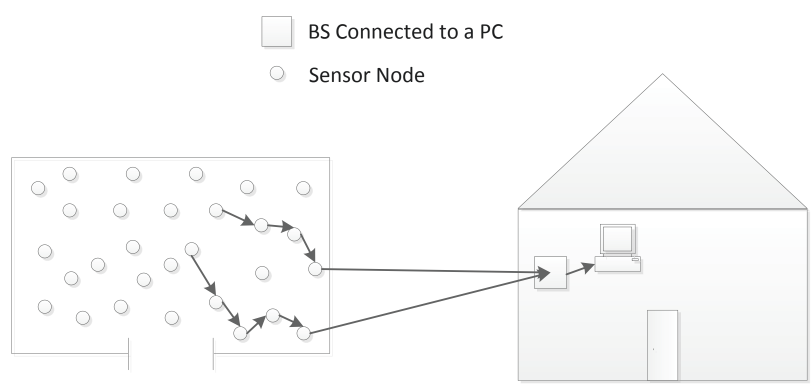


Fig. 1.3: An example of a WSN monitoring a farm associated with a home.

1.1.1.3 Healthcare

WSNs have many applications related to health care. In some circumstances, deploying a home care system would be essential for living at home, especially for elderly people. Monitoring their daily activities provide information about their well-being. This can help to determine if an elderly person is at risk or if an accident has occurred. For example, appropriate sensors can indicate if the elderly person has fallen or if he is having trouble in his sleeping. An example of a home care system is shown in Fig. 1.4. These systems are preferred by customers since there is no privacy invasion caused by the usage of cameras [11].

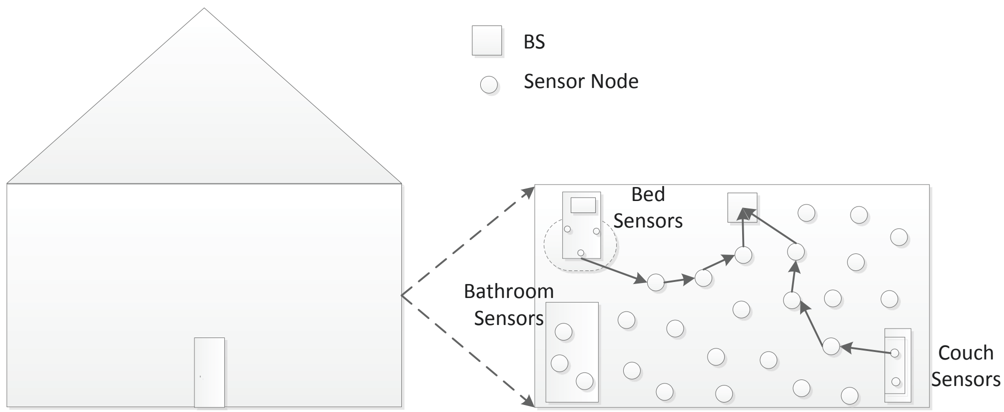


Fig. 1.4: An example of a home care system.

1.1.2 Routing in WSNs

In a WSN, the distances between sensor nodes and the BS can be largely different. This implies that data from a sensor node may need to traverse multiple hops to reach the BS and therefore, routing protocols are required. A routing protocol defines and determines the methods for relaying data from sensor nodes to the BS based on different requirements. Below we briefly present some of the routing challenges:

- **Energy Consumption:** Nodes have limited amount of energy and it is nearly impossible to replace their batteries in a short period of time especially in areas that cannot be easily accessed. Hence, energy efficiency is of great importance for WSNs and research has been carried out to achieve energy-efficient routing [2] [3].
- **Scalability:** The number of sensor nodes employed in a network can be very large. As a result, the routing algorithm should be able to accommodate such large number of nodes without degrading the network performance [3].

- **Transmission Media:** Since the nodes are linked wirelessly, there are wireless-associated issues such as fading and high error rate. These issues might affect the performance of the network [3].

1.2 Research Motivation

Clustering is one of the important routing techniques that are able to re-configure the network based on the desired requirements. By configuring the network, all the communications and routing paths are specified among all the nodes and the BS. In clustering, the network is divided into clusters, each of which has one Cluster Head (CH) that is responsible for its Cluster Members (CMs). A CH collects information from its members, aggregates and forwards such information to the BS. Clustering is an energy efficient routing technique that can extend network lifetime [12] [13] and it will be discussed in details in Chapter 2. The beneficial characteristics of clustering motivated us to focus the research in this thesis on clustering and cluster-based routing protocols.

In order to design a cluster-based algorithm, there are three main aspects that need to be considered, namely, CH selection, cluster forming and cluster communications. These different processes should produce a network configuration that delivers data to the BS in an efficient manner. As mentioned above, reducing the network's energy dissipation is the primary concern. Besides, fast energy depletions in some nodes might cause network partitions and shorten the lifespan of the network. Through our review of the existing cluster-based algorithms, we observed a number of issues that motivated us to do the research and address them in this thesis. These issues are mainly concerned with balancing load and energy consumption among nodes in a WSN. Selecting CHs is one of the critical decisions in configuring clustered-networks since CHs have more responsibilities and functions than CMs. To form clusters, one of the main concerns is how to balance the load generated by CMs among the different CHs. Delivering data from CHs to the BS need to take into consideration energy efficiency and load balancing. Configuring networks requires signaling messages which may increase overhead and energy consumption, which is another concern that needs to be considered.

1.3 Thesis Contributions and Structure

To overcome the issues observed in the existing cluster-based protocols, this thesis contributes to to the area of routing in WSNs by designing and implementing an Energy-efficient and Load-balancing Cluster-based (ELC) routing algorithm. The key contributions of this algorithm are summarized as follows.

- It introduces a new method to determine the neighbors of each node in the network to be used in the CH selection.
- It balances the load among CHs by introducing the number of members that belong to a CH as one of the parameters in cluster forming.
- It incorporates multi-hop routing paths (which are based on distance, energy and number of members) into the process of forwarding data from CHs to the BS to balance load and energy consumption among CHs.
- It introduces a method where only the nodes with significant change in their status information send updates to the BS to be used in network configuration, as opposed to all the nodes updating the BS.
- It employs only Carrier Sense Multiple Access (CSMA) in data transmission, with no Time Division Multiple Access (TDMA) or Code Division Multiple Access (CDMA).

Simulation results show that ELC consumes less energy and extends network lifetime, while sustaining the amount of data received at the BS.

The rest of this thesis is structured as follows. Chapter 2 presents a literature review of clustering and describes the first cluster-based routing protocol as well as the recently proposed protocols. It discusses the issues in these protocols that this thesis is exploring. The proposed ELC algorithm is introduced in Chapter 3 as well as the details of its operation and functions. The performance of ELC algorithm is evaluated and discussed in Chapter 4 using simulation. Finally, Chapter 5 concludes this thesis.

Chapter 2

Clustering Routing Techniques

This chapter presents a literature review of clustering and cluster-based routing protocols. Section 2.1 introduces the concept behind clustering. Section 2.2 discusses the advantages of clustering. Section 2.3 presents the main aspects to be considered in designing a cluster-based routing algorithm. Section 2.4 provides a brief review of the existing distributed and centralized cluster-based protocols that deal with energy efficiency. Section 2.5 summarizes our view on the existing cluster-based algorithms and their issues.

2.1 The Concept of Clustering

To relay information from many sensor nodes towards one BS in an efficient manner, a hierarchical routing technique is usually employed. Cluster-based routing aims to form a network routing hierarchy based on a number of clusters as shown in Fig. 2.1. Grouping sensor nodes into clusters can simplify the method used in determining their communication, which can help to deal with scalability issues. Each cluster has a special node called Cluster Head (CH) and a number of Cluster Members (CMs). A CH is responsible for collecting the data of the CMs in the corresponding cluster. This will enable the data of each node to travel locally in its corresponding cluster rather than through the whole network [12] [14]. Each CH has the knowledge about CMs in its corresponding cluster and about other CHs that it might use to deliver data to the BS. On the other hand, each CM only needs to know about its own CH (but not about other nodes in the entire network). In this way, the network appears smaller in the view of each node in the network, whether it is a CH or a CM [15].

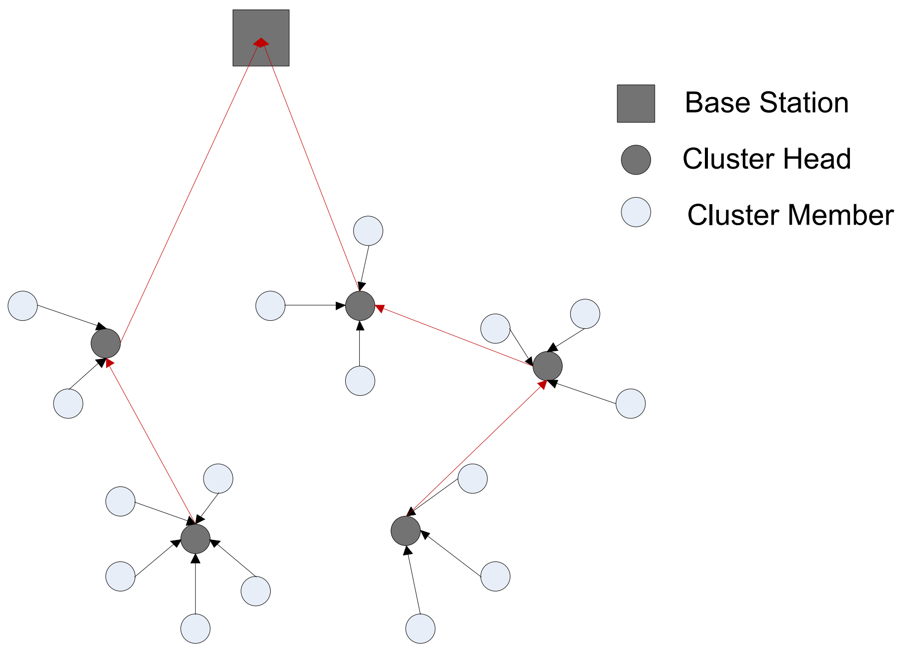


Fig. 2.1: A clustered network

2.2 Advantages of Clustering

Since CMs are usually located close to each other, the sensed data may have correlation due to similarities. Hence, a CH can perform data aggregation to eliminate redundant information by using functions such as suppression, maximum, minimum and average. This allows to differentiate between raw sensor data and useful data. For example, if CMs are sensing and sending the air temperature value, the corresponding CH can determine the average value and forward it in a single message. Data aggregation reduces the number of data transmissions in the network as well as the amount of data forwarded towards the BS. This contributes to reducing bandwidth demands, resulting in a better utilization of bandwidth [12] [14]. By recognizing that the energy consumed in computation is usually less than that used for communication, data aggregation can also reduce the total energy consumption in the network and extend network lifetime [14].

Since only CHs perform routing, there is less exchange of routing information among the nodes in the network. The computations to determine the routing paths to deliver data to the BS are reduced. Also, the sizes of the routing tables as well as the routing overhead are reduced. This can enhance the network's energy consumption profile and thus extends network lifetime

[12] [14] [15].

In flat routing techniques where all the nodes have the same role in the network, data may end up going through many hops and nodes before reaching the BS. This can increase the travel time. On the other hand, in hierarchical routing techniques, data covers larger distances as it travels from one level into the other. This may indicate that data travels faster and latency can be reduced [12].

The above discussion highlights some important characteristics of the clustering routing techniques. Note that, clustering also has some drawbacks. Configuring clusters and their maintenance require the exchange of messages among the nodes in the network, which may introduce overhead [13] [15]. Also, creating clusters requires resources that are not being used for data transmission. Furthermore, CHs may act as bottlenecks due to the extra functions they have to carry out. Overall, the benefits of clustering outweigh its drawbacks in terms of energy efficiency.

2.3 Aspects Considered in Designing a Cluster-Based Algorithm

A cluster-based algorithm can be executed either in a distributed manner or in a centralized manner. A distributed algorithm does not require a central point in order to be executed. The task of configuring the network is shared among all the nodes in the network. The nodes use only local information to make network configuring decisions. This might be advantageous since there is no need to keep global information about the network. However, nodes require computation resources to make decisions about network configuration. As for a centralized algorithm, the BS is usually employed as the central point that configures the network. This is because the BS does not have the issue of limited energy and computation capability, and hence, it is usually used to execute the operations that require a large amount of energy. Indeed, utilizing a centralized algorithm can provide a better control of the network routing.

Whether the clustering algorithm is distributed or centralized, there are three important aspects to be considered: cluster head selection, cluster forming, and cluster communications.

2.3.1 Cluster Head Selection

CHs can be nodes with more resources or regular nodes. In case of CHs being regular nodes, it is important to perform re-clustering where new CHs are selected. This is to avoid the fast depletion of CHs' energy due to the different functions that CHs perform. Re-selecting CHs can also be a means of fault tolerance. For instance, if a CH dies, re-selecting CHs helps in establishing new communication links for the CMs that have lost their CH [14]. There are different considerations in selecting CHs such as nodes' location and nodes' residual energy.

The number of selected CHs has an effect on the total energy dissipation in the network. A network with few CHs can result in some CMs being very far away from the corresponding CHs. This increases the energy consumed by CMs to reach their CHs. On the other hand, a network with many CHs can cause collisions and redundancy in their transmitted data. This is because CHs can be located very closely to each other, and data is assumed to be correlated. Also, as CHs tend to consume more energy, having a large number of CHs increases the total energy dissipation [16]. From this discussion, the number of CHs to be selected should be taken into consideration.

2.3.2 Cluster Forming

In cluster forming, the CMs of each of the selected CHs are identified to form different clusters. Clusters should be formed while ensuring that CMs do not consume a large amount of energy in transmitting data to their CHs [14].

CMs contribute to the load of their CHs since each CH has to collect, aggregate and forward the data of its CMs. Thus, more CMs introduce more load and more energy consumption on the corresponding CH. To make sure that CHs do not deplete their energy too fast, clusters should be formed such that they have similar sizes and balanced load [14].

2.3.3 Cluster Communications

There are two types of cluster communications, namely, intra-cluster communication and inter-cluster communication. Intra-cluster communication covers the interactions between each of the CMs and its corresponding CH. Typically, this type of communication includes a one-hop transmission between each of the CMs and its CH, which simplifies the communication. In some special cases where CMs have a very short transmission range and they cannot reach their CHs, multi-hop transmissions would be beneficial [14].

Inter-cluster communication determines the path between each of the CHs and the BS to deliver data. Single-hop transmission can be used to deliver data to the BS, which can be beneficial in small networks where the BS is close to the CHs. When CHs are far away from the BS, multi-hop transmission can be more advantageous in forwarding data to the BS [14].

2.4 Existing Cluster-based Routing Protocols

As previously discussed, the nodes in WSNs have limited amount of energy and thus, developing energy-efficient routing protocols for WSNs is necessary. This section focuses on the existing cluster-based routing protocols that target at achieving energy efficiency in WSNs. We will first describe LEACH, one of the first distributed cluster-based algorithms. Then, we will present algorithms that improve the CH selection, cluster forming and inter-cluster communication in LEACH. After that, the centralized LEACH-C will be described, followed by algorithms that work on improving LEACH-C.

2.4.1 LEACH

Low-Energy Adaptive Clustering Hierarchy (LEACH) [17] [18] is one of the first and most popular distributed cluster-based routing algorithms. In LEACH, the network activity is organized into rounds with a fixed duration of time, where at the beginning of each round new CHs are selected and clusters are formed. The objective of LEACH is to have a network with a desired number of CHs and to ensure that each node act as a CH approximately the same number of times as the other nodes in the network. Based on these constraints, the threshold value in Equation (2.1) was developed for use in the CH selection.

$$T(n) = \begin{cases} \frac{p}{1-p\times[r \ mod \ \frac{1}{p}]} & , \quad \text{if } n \in G \\ 0 & , \quad \text{otherwise} \end{cases} \tag{2.1}$$

where p is the desired percentage of CHs, r is the current round, and G is the set of nodes that has not been a CH in the last $1/p$ rounds. If a node has been a CH in the last $1/p$ rounds, the threshold is set to 0 and the node cannot be a CH in the current round. Otherwise, the threshold is calculated according to the expression $\frac{p}{1-p\times[r \ mod \ \frac{1}{p}]}$. This ensures that a node is a CH only once in the last $1/p$ rounds. Using the calculated threshold value, a node decides whether it is a CH or not by generating a random number uniformly distributed between 0 and 1. If this number is less than the threshold value, this node decides to be a CH; otherwise, it is not a CH.

After all the CHs are selected, CHs broadcast advertisement messages indicating that they are CHs. Each of the other nodes determines the advertisement message with the best signal strength and joins the corresponding CH by sending a join-request message. After forming the clusters, CMs send their data to the corresponding CHs that aggregate the data messages and forward it to the BS in a single-hop transmission.

In LEACH as well as in many subsequent cluster-based algorithms, Time Division Multiple Access (TDMA[1]) and Code Division Multiple Access (CDMA[2]) codes are used to support data exchanges between CMs, CHs and the BS. Each cluster has its own TDMA schedule, where each CM in this cluster has its own time slot to transmit. In this way, each CM can transmit its own data while avoiding collisions with other CMs in the corresponding cluster. Moreover, the TDMA schedule allows CMs to sleep when it is not their turn to transmit. As for the CDMA codes, each cluster has its own CDMA code that is used for transmitting data from CMs to CHs. Another CDMA code is associated with the BS to transmit data from CHs to the BS. The use of CDMA codes helps reduce the interference between different clusters [19]. TDMA schedules and CDMA codes require synchronization between the nodes in the network in order to determine the exact transmission time and the code used in transmission.

The Desired Number of Clusters: It is the number of clusters that results in a network with the lowest energy dissipation. In order to determine the desired number of clusters, experiments for LEACH in [17] were conducted to find out the percentage of CHs that would produce a network with the lowest energy dissipation. In these experiments, the number of CHs was varied between 1% and 11% of the total number of nodes. For each case, the network energy dissipation was calculated. From the results of the experiments, the percentage of CHs that minimizes the network energy dissipation is in the range of 1% to 6% of the total number of nodes. Percentages lower than this range would cause the CMs to transmit longer distances to reach their CHs, which increases the network energy dissipation. On the other hand, percentages of CHs that are higher than the range of $1 - 6\%$ means that the network has many CHs that consume more energy than CMs. This results in increasing the network's energy dissipation. Also, having many CHs implies that some CHs might end up with a few CMs. This indicates that not as much local data aggregation is being performed in these CHs. As a result, redundant data is being delivered to the BS since data is assumed to be correlated.

While LEACH aims to have each node act as a CH approximately the same number of times as the other nodes, the probabilistic selection of CHs does not result in a good distribution of CHs. Nodes that are adjacent may end up to be CHs. This is a waste of energy since both has to aggregate data and send it to the BS. Also, having adjacent CHs indicates that some areas in the network may not have close-by CHs. The nodes in these areas will be forced to join the far

[1]See Appendix A for more details on TDMA
[2]See Appendix A for more details on CDMA

away CHs, consuming a large amount of energy to communicate with their CHs. Moreover, CHs selection and cluster forming are done without considering the nodes' residual energy. This may result in fast energy depletions, which shortens the network lifetime. In addition, as LEACH employs one-hop communication to deliver data from CHs to the BS, a large amount of energy may be required especially in the case where CHs are far away from the BS. This can also shorten the lifespan of the network.

2.4.2 Modified-LEACH

The authors in [20] propose an Energy-efficient Clustering Scheme for Self-organizing Distributed WSNs that focuses on enhancing the CH selection and cluster forming of LEACH. We will denote this algorithm as Modified-LEACH. The objective of this algorithm is to balance the energy consumption among the nodes in the network by considering the nodes' residual energy and the number of times the nodes have been elected as CHs. The CH selection is done probabilistically as in LEACH using the following modified threshold value

$$P_n(r) = \left(\frac{k}{N - k \times [r \; mod \; \frac{N}{k}]} \left(\frac{E_{current}}{E_{init}} \right) \right) \left(\frac{\frac{k}{N}\sqrt{r}}{C_{ch} \; mod \; \frac{N}{k}} + 1 \right) \tag{2.2}$$

where k is the expected number of clusters, N is the total number of nodes, r is the number of the current round, $E_{current}$ is the current energy of node n, E_{init} is the initial energy of node n, and C_{ch} is the number of times that node n was selected as a CH until the current round. The value of k/N is the same as the value of p in LEACH, i.e., the desired percentage of CHs.

Three parameters are considered in the CH selection, namely, the residual energy, the current round number, and the number of times a node has been selected as a CH. Considering the residual energy helps in selecting nodes with higher energy as CHs. However, after a number of rounds, all the nodes would have small values of residual energy causing to have small threshold values according to Equation (2.2). If all the nodes have small threshold values, only a few nodes would decide to be CHs. In this case, the number of selected CHs can be less than the desired number of CHs. To compensate for the decrease in the threshold values as a result of having small energy values, the current round number r is included in the numerator of the threshold expression. To have a gradual increase in the threshold values as the round number r increases, the square root of r is taken and multiplied by k/N. As for the third parameter (the number of times a node has been elected as a CH in the last N/k rounds), it is used to be able to distribute the burden of being a CH among the nodes in the network. As this parameter increases, the threshold value decreases and there is less chance for the corresponding node to be a CH. After the CH selection, the rest of the nodes join the CHs according to the signal strength of the advertisement messages as in LEACH.

While this algorithm helps in balancing energy consumption and extending network lifetime by considering the nodes' residual energy and their count as a CH, the probabilistic selection of CHs does not guarantee a balanced distribution of CHs. CHs can be concentrated in a certain area of the network, forcing some nodes to consume a large amount of energy to reach their CHs. Moreover, as in LEACH, the size of clusters is not considered in the process of cluster forming, where a cluster with many CMs produces a high load on the corresponding CH. Furthermore, CHs forward data to the BS in a one-hop manner, which may consume a large amount of energy.

2.4.3 ADEEC

Another protocol that considers the nodes' residual energy in the CH selection is proposed in [21], which is a novel Adaptive Energy Efficient Clustering protocol (ADEEC). Unlike the protocol discussed previously, ADEEC ensures a fair distribution of CHs by selecting CHs deterministically instead of using a probabilistic selection. More specifically, each node decides to be a CH if it has higher residual energy than its neighbors. In ADEEC, it is assumed that there are two values of transmission power, one used inside clusters and another used between CHs and the BS. Using the transmission range inside a cluster, the neighbors of each node can be identified to be used in the CH selection. Then, each elected CH broadcasts an advertisement message declaring it is a CH, where each of the other nodes joins the CH that has the highest residual energy by sending a join-request message.

An example with 6 sensor nodes is shown in Fig. 2.2, where nodes n_1 and n_6 have been selected as CHs. The details of the CH selection process that lead to this decision are described in Table 2.1. It is shown that a node is selected as a CH if its residual energy is higher than that of its neighbors.

The results in [21] state that ADEEC can provide a number of CHs that is smaller than other algorithms. This can reduce the total energy dissipation since CHs tend to consume more energy than CMs. However, ADEEC does not guarantee that the produced network has the desired number of CHs (described in Subsection 2.4.1). Furthermore, in ADEEC, only the residual energy is used in cluster forming. This may result in CMs being far away from their CHs, which

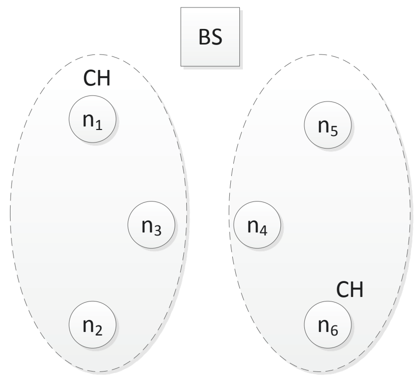


Fig. 2.2: An example of a clustered-network using ADEEC.

Table 2.1: CH selection in ADEEC.

Node ID	Energy Values of the Node and its Neighbors (J)	The Node with Maximum Energy	Is the Node a CH
n_1	$(n_1, 10), (n_3, 8)$	n_1	Yes
n_2	$(n_2, 7), (n_3, 8)$	n_3	No
n_3	$(n_3, 8), (n_1, 10), (n_2, 7), (n_4, 4)$	n_1	No
n_4	$(n_4, 4), (n_3, 8), (n_5, 3), (n_6, 6)$	n_3	No
n_5	$(n_5, 3), (n_4, 4)$	n_4	No
n_6	$(n_6, 6), (n_4, 4)$	n_6	Yes

requires large transmission power. Also, it may produce clusters with unbalanced size affecting the load of CHs. In addition, ADEEC employs one-hop transmissions to deliver data from CHs to the BS. This may consume a large amount of energy.

2.4.4 CTPEDCA

Another routing protocol that is also based on LEACH is proposed in [22]: Cluster-based and Tree-based Power Efficient Data Collection and Aggregation protocol (CTPEDCA). It focuses on improving the inter-cluster communication in LEACH in order to reduce nodes' energy consumption and extend network lifetime. CTPEDCA utilizes a minimum spanning tree to forward data among CHs in a multi-hop manner instead of depending only on the single-hop communication used in LEACH. Specifically, each of the selected CHs broadcasts an advertisement message to the other CHs stating that it is a CH. The message also includes the residual energy value of the corresponding CH. The CH with the highest residual energy would act as the root of a tree formed based on the distance between CHs. In this way, all CHs send their data towards one CH (the root of the tree) that will forward the data directly to the BS. An example of a multi-hop tree between CHs is shown in Fig. 2.3. The reason for choosing one CH to forward data of all other CHs is to limit the communication with the far away BS. This is observed as compared to the case in LEACH where all CHs have to send their data directly to the far away BS. As for the CH selection and cluster forming, CTPEDCA employs the same principle as LEACH.

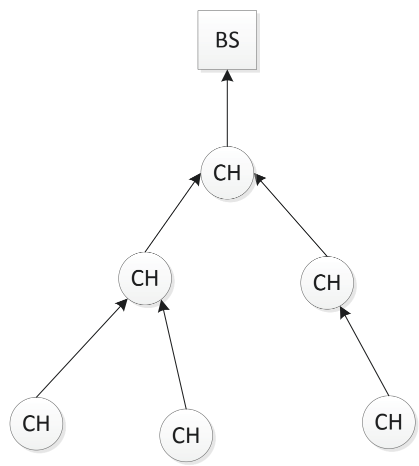


Fig. 2.3: An example of a multi-hop tree between CHs and the BS.

By employing multi-hop inter-cluster communication between CHs, CTPEDCA can reduce the network's energy consumption and prolong network lifetime compared to LEACH. The multi-hop paths are constructed based only on the distance between CHs with the aim of reducing the transmission power as well as the energy consumed. However, depending only on the nearby CHs to relay data can increase their load and deplete their energy faster than the other CHs. Moreover, depending only on one CH to forward the data of all other CHs directly to the BS can quickly deplete its energy. This is because such a CH has to handle a very high load and it represents a bottleneck of the network. Furthermore, the load and energy of CHs are not considered in relaying data between CHs, which affects network lifetime.

2.4.5 LEACH-C

LEACH-C [17] [18], the Centralized version of LEACH, is the most popular centralized cluster-based routing algorithm. In LEACH-C, the network activity is organized into rounds where at the beginning of each round, the nodes in the network send their status information to the BS to be used in configuring the network. The CH selection and cluster forming are performed at the BS using the simulated annealing algorithm. Here, the number of selected CHs is predetermined and is in the range of the desired number of CHs (that was described in Subsection 2.4.1). The objective is to determine the clusters that minimize the energy used by the CMs to send any data to their corresponding CHs. This is done by minimizing the total sum of squared distances between each of the CMs and the closest CH. Thereby, the cost function of a certain set (S), which includes the selected CHs, is calculated using the following equation

$$f(s) = \sum_{i=1}^{N} \min(dist^2(i,s)) \tag{2.3}$$

where $s \in S$ and N is the total number of nodes. After forming the clusters, the BS broadcasts a message to the nodes in the network indicating the ID of each node's CH. Then, the CMs send their data to the corresponding CH that aggregates the data messages into one message and forwards it to the BS in a single-hop transmission.

The simulated annealing algorithm is a slow and approximative algorithm. It does not always provide the optimal solution [23]. Moreover, it considers distance only as a parameter in cluster forming. This can produce clusters with unbalanced sizes and thus, unbalanced loads among CHs. This is illustrated in an example in Fig. 2.4, where the network has 60 sensor nodes and 3 of them are selected as CHs (5% of CHs). As the three clusters have different sizes, their load is unbalanced. In LEACH-C, data is forwarded to the BS in a one-hop manner which may consume a large amount of energy. In addition, the clustering setup is done at the beginning of every round, leading to increased energy consumption and communication latency [24].

2.4.6 CBCDACP

A Central Base station Controlled Density Aware Clustering Protocol (CBCDACP) is proposed in [23]. It aims to improve LEACH-C by considering another parameter (besides distance) in CH selection and cluster forming. Besides minimizing the sum of squared distances between CMs and CHs, CBCDACP employs the density factor in CH selection and cluster forming. The density factor (DF) of a CH is represented by the number of neighbor nodes at closest distance to that CH. Thereby, nodes with high density factor have a higher probability to be selected as CHs. By minimizing the sum of squared distances and considering the density factor, CBCDACP

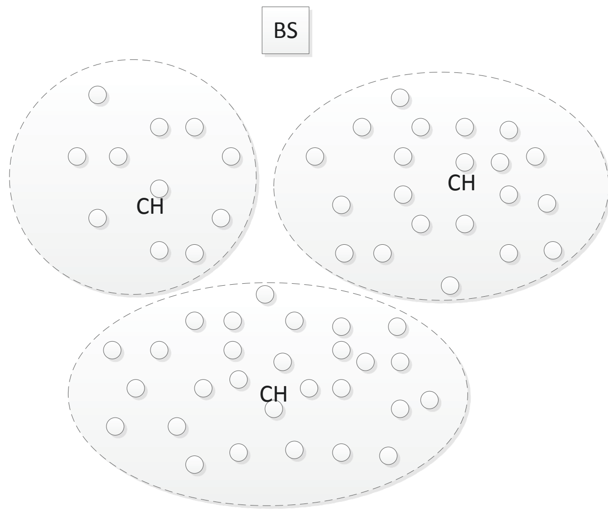


Fig. 2.4: Unbalanced clusters sizes in LEACH-C.

reduces the amount of energy required for the CMs to transmit data to the corresponding CHs. Therefore, a better set of CHs is produced, as compared to LEACH-C, in terms of the total energy consumption in the network. CBCDACP modifies the cost function of LEACH-C as follows

$$f(s) = \frac{\sum_{i=1}^{N} \min(dist^2(i, s))}{\sum_s DF} \tag{2.4}$$

where $s \in S$, and N is the total number of nodes.

Even though CBCDACP is able to reduce the energy consumption as compared to LEACH-C, the number of members in different clusters is still variable. Thus, the load is not balanced among CHs which will have an effect on the network lifetime. Also, since CHs use one-hop communication to the BS, a large amount of energy is consumed, especially in the case where CHs are located far away from the BS.

2.4.7 ADRP

One way to reduce the number of times the clustering set up has to be done, which is one of the issues in LEACH-C, is to select CHs for the current round as well as the subsequent rounds. This is accomplished with the ADRP [24], an Adaptive Decentralized Re-clustering Protocol. Each round in ADRP has two phases: initial phase and cycle phase. In the initial phase, the BS utilizes the location and energy information of the nodes to select CHs and clusters that minimize the total sum of squared distances between the CHs and their members, similar to LEACH-C. Also, only nodes with residual energy higher than the average residual energy of the nodes in the network participate in the CH selection. Moreover, in this phase, ADRP selects the $next - heads$, i.e., the nodes that will become CHs when the re-clustering occurs. The next head in each cluster is a node that has higher residual energy than the average residual energy of the nodes in the cluster. In the cycle phase, data is sent, aggregated at CHs and forwarded to the BS. At the end of the cycle phase, the re-clustering occurs where the next heads take the role of CHs. In this case, there is no need to communicate with the BS to acquire the new CHs, which can save energy. If there is no $next - head$ available, the BS executes the initial phase again.

ADRP introduces a method to reduce the number of times required to communicate with the BS. However, as all the nodes send their status information to the BS to configure the network in the initial phase, the amount of required communication increases. In ADRP, because the CHs send their data directly to the BS, a significant amount of energy is required. Furthermore, the size of the produced clusters is unbalanced and so is the load among the different CHs.

2.4.8 BCDCP

Another protocol that utilizes the BS to configure the network is proposed in [25], Base-station Controlled Dynamic Clustering Protocol (BCDCP). Here, the CH selection and cluster forming are based on an iterative cluster splitting algorithm, where the splitting process is repeated until the predetermined number of clusters is attained. At first, the two nodes in the network that have the maximum separating distance are identified. Then, the rest of the nodes are associated with the closer of these two nodes. In this way, two groups are formed. After that, a similar process is followed for each of the two groups to split them until the predetermined number of clusters is acquired. Hence, as in LEACH-C, the distance between the CMs and their CHs is considered while forming the clusters. For the inter-cluster communication, a multi-hop routing technique is used to forward data from CHs to the BS, as opposed to the one-hop communication in LEACH-C. The multi-hop routing paths are constructed based on the residual energy of CHs, where each of the CHs forwards its data to a CH with high residual energy in the direction of the BS.

In the splitting process of BCDCP, some of the resulted clusters can have nodes that are much farther to the corresponding CH than the other nodes in the same cluster. This is because the network has a fixed number of clusters. Instead, the far-away nodes can form their own cluster to have short transmissions between CMs and their CH. Moreover, while BCDCP utilizes a multi-hop communication to forward data to the BS in order to reduce the energy consumption, the routing paths are constructed based on the energy of CHs. Forwarding data based only on the residual energy can result in long range transmissions which may result in more energy consumption. Also, the load of CHs is not considered in forwarding data between CHs, which is important for balancing the energy consumption among the CHs.

2.4.9 BIDRP

The centralized protocols discussed previously perform clustering at the BS and send the results to the nodes. There are other protocols that process clustering decisions in both the BS and the nodes. Here, we discuss one of these protocols: Base station Initiated Dynamic Routing Protocol (BIDRP) [26]. It considers heterogeneous networks, as opposed to the homogeneous networks considered in the previously described protocols. In BIDRP, the stronger nodes in terms of energy and computational capability are designated as CHs while the rest of the nodes are used as CMs. After the CHs are determined, the network is divided into levels. The BS is at level 0 and the CHs close to the BS have a lower level than the CHs that are farther away. To determine the level of each CH, the BS broadcasts a packet to all of the CHs, and according to the signal strength of that message, the CHs identify their level. Then, each CH broadcasts an advertisement message indicating that it is a CH as well as its level. Based on this, the other CHs choose their next hop to forward data to the BS. CHs at higher levels will choose their next hop from the lower levels based on the signal strength of the advertisement messages from the CHs in the lower levels. This indicates that the distance between CHs is used to construct the routing paths between CHs and the BS.

In BIDRP, only the stronger nodes are considered as CHs. However, some of them might be close to each other, and there would be no point in having both of them as CHs since it would create redundancy. To avoid this situation, one can employ a mechanism that select strong CHs while considering their distribution in the network. Moreover, each CH selects the closest CH to be its next-hop without considering the residual energy of the next-hop CH. This can quickly deplete the energy of the next-hop CH. Furthermore, executing cluster forming and determining the routing paths between CHs at the nodes, instead of having them at the BS, can require a large number of signaling messages between the nodes. This may increase the consumed energy.

2.4.10 CPMS

From the discussions above, BCDCP forwards data to the BS according to the energy of CHs, while in BIDRP, the closest CH to the BS will always be the last CH that forwards data to the BS (which acts as a leader CH). A protocol in [27] considers both the energy level and the distance in selecting the leader CH, to improve the performance of network lifetime. This protocol is called the Clustering Protocol with Mode Selection (CPMS). Here, the selection of CHs and cluster forming are the same as in BCDCP. To be able to forward data between CHs, CPMS uses the idea of creating levels, similar to BIDRP. Then, the leader CH is determined while considering its residual energy as well as its distance to the BS. The leader CH is selected using the following energy-distance ratio

$$R_{ED} = \frac{E_{Rm}}{Level_n} \tag{2.5}$$

where E_{Rm} is the remaining energy of the CH and $Level_n$ is the level of that CH. The CH with the highest ratio is selected as the leader CH. After that, the previously determined levels are transformed into another notion, i.e., layers. Layers are created in order to determine the direction of traveling data. The level at which the leader CH reside is identified as layer 0 and the adjacent level becomes layer 1 and so on. Data travels from higher layers to lower layers until it reaches the leader CH, which will forward the data to the BS.

Depending on one node (i.e., the leader node) to deliver data from the entire network to the BS can quickly deplete its energy. Fast depletion of nodes' energy may cause disconnections between parts of the network, thus shortening the lifespan of the network. Moreover, in CPMS, the load of CHs is not considered when relaying data between CHs. Considering the load of CHs in forwarding data can contribute to balancing the CHs' load and energy consumption which can then extend network lifetime.

The cluster-based algorithms discussed above are summarized in Table 2.2 where the key features of each of these algorithms are highlighted. Other energy-efficient cluster-based routing protocols, which have similar considerations to the algorithms discussed in this chapter, can be found in [12], [13] and [28].

Table 2.2: Summary of the algorithms discussed in Chapter 2.

Algorithm & Type	CH Selection		Parameters in Cluster Forming	Inter-cluster Routing		Notes
	Type	Parameters		Transmission Type	Parameters	
LEACH [17] [18], Distributed	Probabilistic	Number of times a node is a CH	Distance	Single-hop	NA	Most popular distributed cluster-based algorithm
Modified-LEACH [20], Distributed	Probabilistic	Number of times a node is a CH, residual energy, current round	Distance	Single-hop	NA	Improve CH selection in LEACH
ADEEC [21], Distributed	Deterministic	Residual energy	Residual energy	Single-hop	NA	Improve CH selection and cluster forming in LEACH
CTPEDCA [22], Distributed	Probabilistic	Number of times a node is a CH	Distance	Multi-hop	Distance	Improve inter-cluster routing in LEACH
LEACH-C [17] [18], Centralized	Deterministic	NA	Distance (to select CHs and form clusters)	Single-hop	NA	Most popular centralized cluster-based algorithm
CBCDACP [23], Centralized	Deterministic	NA	Distance, density factor of CHs (to select CHs and form clusters)	Single-hop	NA	Improve CH selection and cluster forming in LEACH-C
ADRP [24], Centralized	Deterministic	NA	Distance (to select CHs and form clusters)	Single-hop	NA	Reduce number of times to do clustering as compared to LEACH-C by selecting $next\_head$
BCDCP [25], Centralized	Deterministic	NA	Distance (to select CHs and form clusters)	Multi-hop	Residual energy	Improve Inter-cluster routing in LEACH-C
BIDRP [26], Centralized	Deterministic	Strong nodes in energy and computation	Distance	Multi-hop	Distance	For heterogeneous networks. BS and nodes participate in configuring the network
CPMS [27], Centralized	Deterministic	NA	Distance (to select CHs and form clusters)	Multi-hop	Distance, residual energy to select a leader CH	Based on BCDCP and BIDRP

2.5 Our View on the Existing Cluster-based Protocols

In centralized protocols, the BS exploits its global knowledge of the network to produce better decisions, as opposed to distributed protocols where decisions are made locally and may not be the most efficient. This can help in avoiding uncertainty at the nodes since the BS is making decisions on their behalf while taking into consideration the conditions in the entire network. Utilizing a centralized algorithm can provide robustness where the BS reacts to any changes in the nodes' conditions by re-configuring the network and providing the new routing paths. Moreover, centralizing the algorithm can offer flexibility in computing entire routes as opposed to determining only the next-hop at the nodes themselves which is the case in distributed algorithms. In centralized algorithms, the required computations are shifted from the nodes, which have limited computation capability and energy supply, to the BS, which has a much higher computation capability and no energy limitation. Having a centralized algorithm can avoid the need for exchanging messages between the nodes in the association process, e.g., nodes announcing they are CHs, and nodes choosing and joining their CHs, which are used in distributed algorithms. This can reduce the amount of signaling messages, helping to reduce collisions in the network. However, centralized algorithms still require exchanging messages between each of the nodes and the BS, which needs to be considered in designing a centralized cluster-based algorithm.

Selecting CHs in a deterministic manner can better control the decision and can produce better distributed CHs. This is different from employing a probability function where nodes make the decision independently like in LEACH [17] and CTPEDCA [22]. Employing the residual energy as a parameter in CH selection can ensure that only nodes with high remaining energy are elected as CHs. This can balance the energy consumption among the nodes, as in ADEEC [21]. On the other hand, ADEEC does not guarantee that the produced network has the desired number of CHs, which affects the network's energy dissipation. For cluster forming, LEACH-C [17] and ADRP [24] consider only the distance parameter in determining the CMs of each CH, whereas ADEEC [21] considers only the residual energy of the CHs. This may result in unbalanced cluster sizes, and thus unbalanced energy consumption among CHs. Furthermore, besides the distance parameter, CBCDACP [23] includes the density factor to identify the CMs of each CH. Nevertheless, this still cannot balance the cluster sizes and the energy consumption among CHs, and thus cannot effectively extend network lifetime.

In many of the existing cluster-based protocols, CHs deliver their data to the BS in a single-hop manner, like in LEACH-C [17], CBCDACP [23] and ADRP [24]. This requires high transmission power when CHs are far away from the BS. To address this issue, other protocols constructs multi-hop routing paths between CHs and the BS based on residual energy alone (BCDCP [25]) or distance alone (CTPEDCA [22]). However, forwarding data based on the residual energy of the next-hop CH alone can result in long range transmissions, which will require high transmission power. On the other hand, considering only the distance to the next-hop CH can deplete its energy fast. CPMS [27] employs both energy and distance parameters in selecting one CH to forward the data of all other CHs to the BS. However, depending on one CH to relay data to the BS depletes its energy faster than the other CHs, which results in unbalanced energy consumption among CHs. Moreover, CPMS do not consider the issue of balancing the load of CHs when forwarding data between CHs, where the load of a CH includes both data of its cluster and relayed data of other CHs. Unbalanced load results in unbalanced energy consumption, which shortens the network lifetime. This is one of the issues that need to be considered in the case of a multi-hop inter-cluster communication.

Furthermore, in the existing centralized cluster-based algorithms [17] [23] [24], all the nodes in the network send their status information in every round to the BS. This information is used in configuring the network. However, in practice, the network might suffer from a large amount of traffic because of these transmissions. Also, some of these transmissions can be redundant since the change in energy may not be significant and thereby, does not affect the network configuration. Unnecessary transmissions of nodes information increase the total energy consumption, another issue to be considered.

More importantly, the existing cluster-based algorithms employ TDMA schedules and CDMA codes for data transmission between CMs, CHs and the BS [12] [13]. In practice, it is hard to create such schedules and codes since they require synchronization between the nodes in the network, which is difficult to achieve in ad-hoc WSNs. For this reason, many WSNs do not use TDMA or CDMA schemes. Instead, they employ CSMA for the simplicity of access control, i.e., it is asynchronous and the transmission can be done in a distributed manner. This means that there is no need to assign a time slot for each node that needs to transmit. Also, the main standard for WSNs is IEEE 802.15.4 which employs CSMA for data transmission [29]. Therefore, it is more practical to have a routing algorithm that utilizes only CSMA during data transmission.

2.6 Summary

This chapter presented the basic concepts in clustering on which the rest of the thesis is built. It described representative cluster-based routing algorithms, from the first proposed one to the most recent ones that deal with energy efficiency, covering their key features in selecting CHs, forming clusters and inter-cluster routing, as well as their shortcomings. This motivates us to propose a new cluster-based routing algorithm, to be presented in Chapter 3.

Chapter 3

Energy-efficient and Load-balancing Cluster-based Routing Algorithm

Based on our comments presented in Section 2.5 regarding the issues in the existing cluster-based algorithms, we have created an Energy-efficient and Load-balancing Cluster-based, ELC, routing algorithm for CSMA-based WSNs. The main objective of ELC is to reduce the total energy consumption in the network and extend network lifetime. This chapter presents a detailed description of the ELC algorithm as well as its anticipated advantages. Section 3.1 first presents the clustering in ELC which includes the CH selection and cluster forming. Section 3.2 then introduces the inter-cluster communication in ELC. Section 3.3 describes the operation of ELC which includes the interaction between the BS and the nodes in a WSN. Section 3.4 presents a case where no network re-configuration is required and how ELC handles it.

3.1 Clustering

Clustering includes selecting CHs and determining the CMs associated with each CH in order to form the different clusters. The following subsections describe the CH selection and cluster forming in ELC.

3.1.1 Cluster Head Selection

We adopt the idea of comparing the residual energy of each node with that of its neighbors to decide whether this node is a CH or not [21]. Determining the neighbors of each node decides on the nodes that would participate in these residual energy comparisons. This can affect the decision whether or not the corresponding node is a CH. Eventually the number of selected CHs in the network can be affected. The number of selected CHs influences the network's energy dissipation as was described in Subsection 2.3.1. This shows the importance of the approach used in determining the neighbors of each node in the network. As a result, we propose a new method for the BS to determine the neighboring relationship between nodes in the network as follows. Node j is considered to be a neighbor of node i if the physical distance between them, $d_{i,j}$, is less than the distance limit, d_{limit}. The value of d_{limit} affects the number of neighbors of each node and, consequently, the number of CHs to be selected and the network's energy dissipation. Hence, the value of d_{limit} that produces a network with the lowest energy dissipation is selected. Before describing the details of the procedure used to determine the neighbors of each node and to select CHs, the criterion for choosing the value of d_{limit} is further discussed next.

3.1.1.1 Distance Limit Determination

Appropriate value of d_{limit} should neither be too low nor too high. If this value is too low, then each node would be associated with a small number of neighbors. In this case, the node is more likely to have the highest level of residual energy compared to all of its neighbors and thus it is more likely to be elected as a CH. This could eventually result in a large number of CHs, where some of these CHs can be close to each other. This can cause collisions and redundancy when CHs aggregate the collected data and forward it towards the BS. Moreover, a network with a large number of CHs can increase its energy dissipation since CHs tend to consume more energy. On the other hand, if d_{limit} is too high, then each node would be associated with many neighbors and this gives the node a lower chance to be elected as a CH. This is because if one of the neighbors of the node is a CH, then the node cannot be a CH. In other words, there would be a small number of CHs distributed sparsely in the network deployment area and as a result CMs need to consume more energy to reach their CHs. Therefore, the value of d_{limit} that produces a desired number of CHs is selected. The desired number of CHs as defined in [17] is in the range of $1-6\%$ of the total number of nodes, according to the experiments conducted in [17] and described in Subsection 2.4.1. In this way, we can establish a good distribution of CHs by employing a deterministic CH selection, as was observed from the existing algorithms, and by providing the desired number of CHs, as described above.

To verify that the desired number of CHs for our CSMA-based WSNs is indeed in the range of $1-6\%$ of the total number of nodes, we run our ELC algorithm on the simulation platform NS-2and observe the total energy dissipation over time for three ranges of number of CH's: less than 1%, $1-6\%$, and more than 6%. The sensing area used in the simulation test consists of nodes randomly distributed in $100 \times 100m^2$ with a constant packet rate of 1 pkt/sec in a two-ray ground propagation environment, and the BS is located at $(50m, 175m)$.

For each of the three possible ranges of number of CHs, the simulation is repeated ten times since the nodes are randomly distributed. In each simulation run, the total energy consumption in the network over the simulation time is determined. Then, its average is taken over the ten simulation runs. The variation of the total energy consumption from this average is approximately 2%. This average total energy consumption for the three possible ranges of number of CHs, illustrated in Fig. 3.1, clearly shows that a network with a number of CHs between 1% and 6% has the lowest energy consumption.

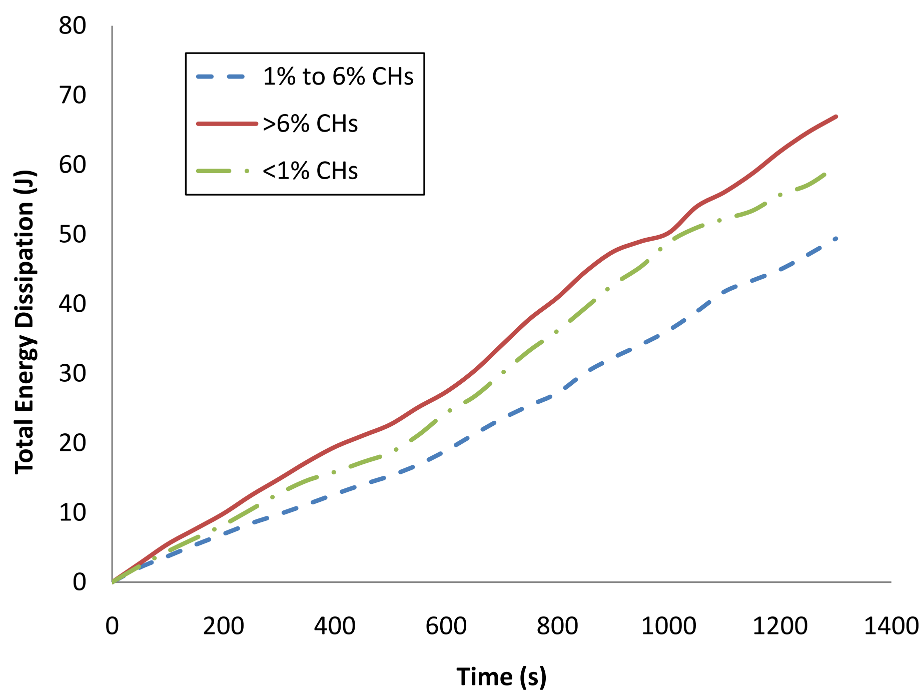


Fig. 3.1: Total energy dissipation over time for different numbers of CHs.

Using a curve fitting technique, we create an equation that can determine the value of d_{limit} that produces the desired number of CHs for different values of number of nodes in a fixed-size sensing area with randomly distributed nodes as follows

1. Run the ELC algorithm on the simulation platform while varying the total number of nodes N from 50 to 250.

2. For a given number of nodes, choose one value of d_{limit} that produces a number of CHs in the range of $1-6\%$. This is done by trying out different values until the required range of the number of CHs is obtained. As a result, there is one data point for each (N, d_{limit}).

3. Use curve fitting to construct a curve that has the best fit for these data points. The regression tool in Microsoft Excel that employs the least-squares method can be used to find the best fit curve.

By following those steps for a $100 \times 100m^2$, the constructed curve is shown in Fig. 3.2 and we have found that the following equation best represents this curve

$$d_{limit} = 152.12N^{-0.392} \tag{3.1}$$

where d_{limit} is in meters. Using Equation (3.1) one can determine d_{limit} that produces the desired number of CHs for any number of nodes in the considered sensing area. This equation can be used as long as the sensing area is the same as the one we have considered. If another sensing area is employed, then the same procedure can be followed to determine the new d_{limit} equation.

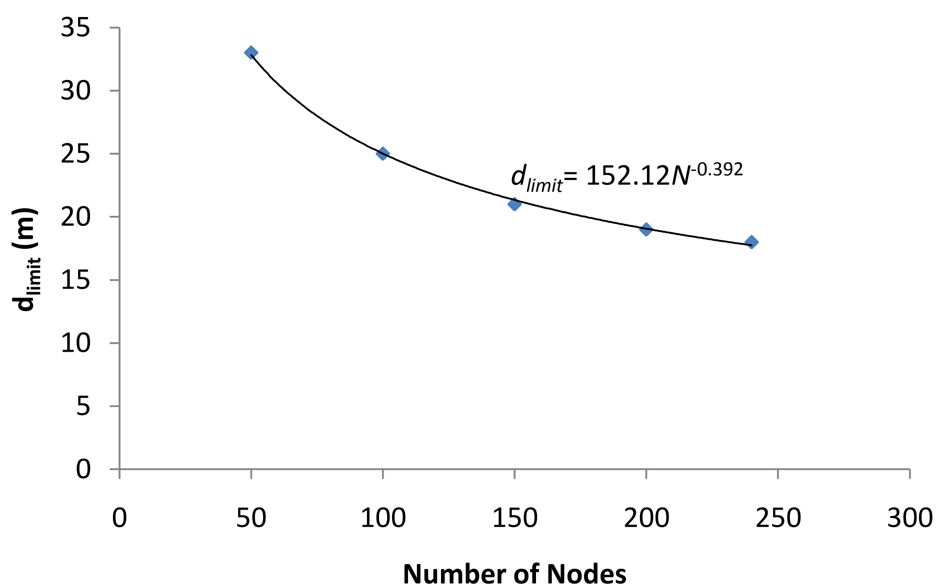


Fig. 3.2: d_{limit} vs. the total number of nodes.

3.1.1.2 CH Selection Procedure

First, the BS determines the neighbors of each node in the network. Let $\mathbf{N}_i$ denote the set of neighbors of node i. $\mathbf{N}_i$ is determined based on d_{limit} as follows: $\mathbf{N}_i = \{j : j \in \mathbf{N}, d_{i,j} < d_{limit}\}$, where $\mathbf{N}$ is the set of all nodes in the network.

Next, for each node $i \in \mathbf{N}$, its normalized residual energy E_i/E_i^{max} is compared against those of its neighbors, where E_i is the current energy of node i and E_i^{max} is the initial maximum energy of node i. If $(E_i/E_i^{max}) > \max_{j\in\mathbf{N}_i}(E_j/E_j^{max})$, then node i is selected as a CH. If $(E_i/E_i^{max}) = \max_{j\in\mathbf{N}_i}(E_j/E_j^{max})$ but the neighbors having the highest value of normalized residual energy are not CHs, then node i is selected as a CH. Otherwise, node i is not elected as a CH. In this way, the BS selects all the CHs.

The pseudocode of the CH selection procedure is given in Pseudocode 1.

Pseudocode 1 CH Selection

```
for each node i ∈ N do
  for each node j ∈ N do
    if d_{i,j} < d_{limit} then
      add node j to N_i
    end if
  end for
end for
for each node i ∈ N do
  if E_i/E_i^{max} > max_{j∈N_i}(E_j/E_j^{max}) then
    node i is a CH
  else if E_i/E_i^{max} == max_{j∈N_i}(E_j/E_j^{max}) and neighbor j is not a CH then
    node i is a CH
  else
    node i is not a CH
  end if
end for
```

3.1.2 Cluster Forming

The CM-CH distance is used as a parameter in cluster forming to reduce CMs' transmission power and, consequently, their energy consumption to reach the corresponding CHs. To balance the load generated by the CMs among the different CHs, we propose in the ELC algorithm to consider the number of members associated with a CH as one of the parameters in cluster forming. By incorporating the number of members, we aim to balance the size of clusters. Different clusters would have similar number of CMs. Thus, each of the CHs would be associated with similar number of CMs. As a result, the load generated by the CMs can be distributed among the different CHs. In other words, the load and energy consumption can be balanced among CHs.

Taking into account the CM-CH distance and the number of members parameters in the cluster forming of the proposed ELC algorithm, we propose a cost when CM j is associated with CH h as follows

$$CH_{h,j} = \sigma \times \left(\frac{d_{h,j}}{d_{max}}\right)^f + \psi \times \left(\frac{m_h}{N}\right) \tag{3.2}$$

where $h = 1, ..., H$, H is the number of CHs, $j = 1, ..., P$, P is the number of CMs, $d_{h,j}$ is the distance between CH h and CM j, d_{max} is the maximum distance between any two nodes in the network, m_h is the current number of members associated with CH h, N is the total number of nodes, σ and ψ are the weighting coefficients that sum to 1 and f is the path loss exponent. The first component of Equation (3.2) represents the CM-CH distance parameter between CH h and CM j, where by prioritizing a smaller value, we aim to reduce the energy consumed by CMs to communicate with their CHs and, consequently, reduce the network's energy dissipation. As for the second component of Equation (3.2), it represents the number of members associated with CH h which indicates its load, where preferring a CH with a smaller load in the association decision of every CM can lead to balancing the load among different CHs. In this case, employing only the first component by adjusting the weighting coefficients contributes to only reducing the network's energy consumption, while employing only the second component contributes to load balancing and hence, increasing network lifetime. On the other hand, employing both of the components aims to reduce network's energy consumption as well as increase network lifetime.

For a given CM j, it is associated with CH h that has the lowest cost which is calculated using Equation (3.2). In other words, a CM is associated with the CH that is the closest and has the lowest cluster population. At first, all the clusters have no members and every time a CM is added to a cluster, the number of members of the corresponding CH is updated accordingly. Thereby, the cost of that CH changes as well, which affects the association decision of the next CM to be added to a cluster. In this manner, we aim to balance the size of clusters and, consequently, their load.

The pseudocode of the cluster forming procedure is given in Pseudocode 2.

Pseudocode 2 Cluster Forming

for $each\ CM\ j$ **do**
  if $CH_{h^*,j} == \min_{h=1,\ldots,H}(CH_{h,j})$ **then**
    $node\ j\ is\ a\ CM\ of\ CH\ h^*$
    $next\ hop\ of\ node\ j\ is\ CH\ h^*$
    $m_{h^*} \leftarrow m_{h^*} + 1$
  end if
end for

3.2 Inter-Cluster Communication

From our discussion in Section 2.5, it is difficult to say whether the multi-hop transmission is more efficient than the one-hop transmission in the inter-cluster communication. To address this issue in the proposed ELC algorithm, besides the possibility of one-hop transmission, we allow the possibility of multi-hop transmission for any of the CHs to deliver its data to the BS. To decide which type of transmission is more efficient for any of the CHs, a cost for each of the possible paths is calculated and the path with the lowest cost is chosen. In order to determine the cost of each of the paths, the parameters that contribute to the cost between the CHs as well as the BS, need to be determined. The chosen parameters in the proposed ELC algorithm are presented next. Afterwards, we discuss the method used in calculating the routing paths.

3.2.1 Inter-Cluster Routing Parameters

To ensure that CHs do not consume a large amount of energy in their data transmission to other CHs, the CH-CH distance is used as a parameter. Moreover, to make sure that CHs do not deplete their energy fast, the residual energy of CHs is also used as a parameter in calculating the cost of the routing paths. Considering the residual energy of CHs can be used to guarantee that only the CHs with high remaining energy participate in the routing paths, which can help in balancing the CHs' energy consumption. This can extend network lifetime.

The load of a CH can be generated from its own cluster as well as from relaying data of other CHs. Taking into consideration the load of a CH in forwarding data between CHs, a new parameter which is the number of members associated with a CH is also employed in determining the cost of the multi-hop routing paths in the proposed ELC algorithm. This enables a CH with many already connected CMs, i.e., with higher load, to avoid relaying many data messages from other CHs. As a result, the load of CHs and their energy consumption can be balanced. Even though in the cluster forming we aim to balance the size of clusters by using the number of members parameter, Equation (3.2) includes as well the CM-CH distance parameter in calculating the cost of a CH. This can affect the chosen CH for a given CM and, consequently, the formed clusters. Therefore, the number of members is also included in calculating the cost of the paths in the inter-cluster communication to help in balancing the load among CHs.

3.2.2 Inter-Cluster Routing Calculation

Using the three variables, namely, CH-CH distance, CH's residual energy and number of members associated with a CH, we propose a shortest-path approach with a new cost definition to construct the multi-hop routing paths between CHs and the BS. The shortest path between any of the CHs and the BS, which is the path with the lowest cost, is determined using a graph search algorithm.

A graph consists of a set of vertices **V**, that includes all the CHs and the BS, linked with a set of edges **E**, that includes the directed edges between the vertices. The edge between vertex V_i and vertex V_j is denoted as $e_{i,j}$. The cost of the directed edge $e_{i,j}$ between V_i and V_j indicates the cost in the direction from V_i to V_j and it is defined as

$$W_{e_{i,j}} = \alpha \times \left(\frac{d_{i,j}}{d_m}\right)^f + \beta \times \left(1 - \frac{E_j}{E_j^{max}}\right) + \gamma \times \left(\frac{m_j}{N}\right) \tag{3.3}$$

where $i \neq j$,$(i, j) \in$ **E**. In Equation (3.3),

- $d_{i,j}$ is the distance between V_i and V_j while d_m is the maximum distance between any two nodes in the network including the BS and f is the path loss exponent.
- E_j is the energy of V_j and it is normalized by its initial maximum energy (E_j^{max}). The normalized residual energy is subtracted from 1 to determine the amount of energy already consumed, where a vertex with low consumed energy and, consequently, high residual energy is preferred in determining the shortest path.
- m_j is the number of CMs associated with V_j. It is normalized by the total number of nodes(N).
- α, β, and γ are the weighting coefficients that sum to 1.

The three parameters considered in calculating the routing paths in the proposed ELC algorithm are represented by the three components of Equation (3.3). Equation (3.3) is used to represent the cost of the directed edge between any two CHs. Note that the cost of the directed edge from any of the CHs to the BS is based on the distance between the CH and the BS only, since the other remaining parameters are not applicable for the BS. In this way, the chosen paths between CHs and the BS can be multi-hop or one-hop depending on the cost of the different routing paths, where the shortest path is selected.

The graph search algorithm used to determine the shortest path in the proposed ELC algorithm is the Dijkstra algorithm [30]. For a given source vertex in a graph, the Dijkstra algorithm can be used to determine the shortest path between that vertex and a destination vertex. The shortest path is the path with the lowest cost, where the cost of a path P_k is the sum of the cost of its edges ($W_{e_{i,j}}$) and it is represented by the following C_k

$$C_k = \sum_{e_{i,j} \in P_k} W_{e_{i,j}} \tag{3.4}$$

where $i \neq j$, $k = 1, 2, ..., T$ and T is the number of paths for one source-destination (s, d) pair. Thus, the shortest path P_k^* is defined as

$$P_k^* = \arg\min_{P_k} \left(\sum_{e_{i,j} \in P_k} W_{e_{i,j}} \right). \tag{3.5}$$

3.3 Operation of ELC

The network activity is organized into rounds, where the network routing configuration is periodically updated at the beginning of every round. Every node in the network has a timer that triggers the event of starting a new round. If this event is not triggered, then the nodes continue the data transmission, where CMs transmit to their CHs, and CHs aggregate the data and forward it towards the BS. The main events of ELC that occur in one round are summarized in Fig. 3.3. At the beginning of every round, nodes send their status information, if needed, to the BS which performs CH selection, executes cluster forming and determines the inter-cluster communication. As a result, the BS broadcasts the relevant information for configuring the network to all the nodes to be used for data transmission. The following subsections describe the different events in more details.

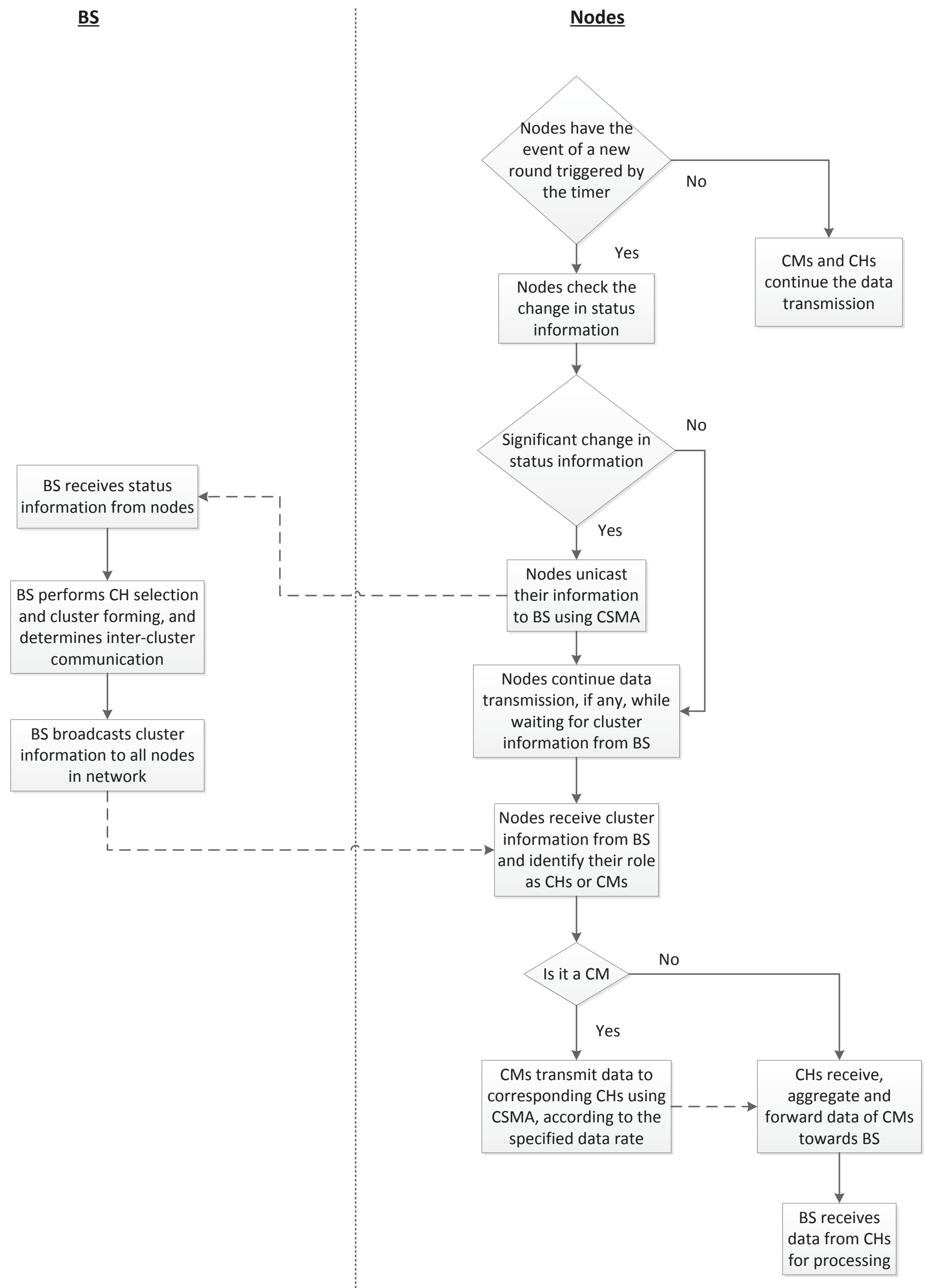


Fig. 3.3: Main events of ELC in one round.

3.3.1 Updating the BS with Nodes' Status Information

In the first round, each node in the network unicasts its location (x, y) and its residual energy in a one-hop manner to the BS using CSMA. The packet used in this case allocates 6 bytes for the data portion, where 2 bytes are used for the x value, 2 bytes for the y value and 2 bytes for the energy value. This packet includes a field in its header indicating its type which is a message carrying the status information about the corresponding node. As a result, the size of this packet is small and hence, one-hop transmissions are used to deliver the nodes' status information. The packet format of the status update message is illustrated in Appendix B.

As for the other rounds, we propose in ELC the idea of having only some of the nodes sending their status information in every round, which can reduce the total energy consumption in the network. This is done as follows. Each node calculates the amount of change in its energy value relative to that of the previous round. If this change in percentages exceeds a pre-determined threshold, the node decides to send its energy value to the BS. Also, in the case that the location of the node has changed, the node would decide to send its new location to the BS. Based on these decisions, the node unicasts the required information to the BS using CSMA. If the node's location has not changed and its change in energy has not exceeded the pre-determined threshold, then the node would not send its status information to the BS.

The value of the threshold used to decide whether a node should update the BS with its energy value is based on the nodes' energy consumption throughout the network operation. This threshold is a percentage value that can be determined using the following steps.

1. Consider the nodes' energy consumption in one round and determine the change in each node's energy in percentages.
2. Calculate the average of these percentage values over all the nodes.
3. Perform steps 1 and 2 for all the rounds of the network operation, where eventually each round has an average value for the percentage change in a node's energy.
4. Calculate one average value over all the rounds, which represents the required threshold value.

Determining the threshold value can be done in the testing process before the deployment of the WSN. In our case, simulation is used to determine the threshold value.

3.3.2 Configuring the Network by the BS

The BS employs the status information of the nodes to configure the network. In the case of some nodes deciding not to send their status information to the BS as discussed in Subsection 3.3.1, the BS would determine if these nodes are alive with the same status information as in the previous round or they are dead. This is done using the nodes' old energy value that can be used to anticipate if a node is dead in the current round or it is alive with the same status information. After the BS determines the nodes' status information, it executes CH selection and cluster forming as described in Section 3.1.

After forming the clusters, the BS broadcasts a message in a one-hop manner to all the nodes in the network to inform them about their CHs. This message includes a field in its header indicating its type which is a message carrying the information about the CH of each node in the network. The broadcasted message includes an array, where the index of the array represents node IDs and the corresponding values represent the nodes' CH IDs. In our implementation, each ID is represented with one byte. This message is used by the nodes to identify their role (CHs or CMs) and their next hop (in case of CMs). This would cover the intra-cluster communication. The packet format of the clustering information message is illustrated in Appendix B.

Finally, the BS constructs the multi-hop routing paths between all the CHs and the BS as described in Section 3.2. The information about the produced routing paths is converted into hop-by-hop information and stored in a table. Each row in this table belongs to one of the CHs while each column represents a CH that generates a message to be forwarded. The value at a specific row and column represents the ID of the next-hop-CH for forwarding the corresponding message. The BS unicasts to each CH the corresponding row in a one-hop manner, where each ID is represented with one byte. The unicasted message includes a field in its header indicating its type which is a message carrying the information about the inter-cluster communication. Using the information in this message, a CH can determine its next-hop-CH based on the CH that generated the message to be relayed. The packet format of the inter-cluster routing information is illustrated in Appendix B.

3.3.3 Data Transmission

After each node receives the clustering information from the BS, as described in Subsection 3.3.2, it identifies its role as a CH or a CM. More specifically, if the ID of a node's CH matches its own ID, then this node is a CH. Otherwise, this node is a CM of the corresponding CH. As for the information about the multi-hop routing paths, each CH receives the corresponding message and stores it in the array $next\_hop\_info$. The index i of this array represents the ID of a CH that generates a message to be forwarded and the value at $next\_hop\_info[i]$ represents the ID of the next-hop-CH.

As a result, the data transmission starts, where CMs that have data start sending it to the corresponding CH using CSMA, according to their data rate. As for the CMs with no data, they go into sleeping mode. CHs periodically aggregate the data messages received from their CMs and forward it towards the BS using CSMA. Each CH forwards its aggregated data towards the BS using its own array $next\_hop\_info$. For instance, a CH i transmits its data to $next\_hop\_info[i]$, where index i is used since CH i is the source of the data to be relayed. The ID at $next\_hop\_info[i]$ can belong to the BS or to another CH j. If CH j receives the data of CH i, then CH j would forward the data to $next\_hop\_info[i]$ using its own array $next\_hop\_info$, where index i is used since CH i is the source of the data to be relayed by CH j. This is done until the data has reached the BS. The received data at the BS is processed and analyzed depending on the WSN application.

3.4 Handling a Case with No Re-Configuration

In some cases of a network operation, there may exist some rounds where the change in any node's status information is relatively the same as that of any of the other nodes. Thereby, the decisions in terms of CH selection and cluster forming may not be affected. As a consequence, the configured network by ELC may have the same configuration as in the previous round. Hence, there is no need to re-configure the network. Re-configuring the network consumes energy, by receiving the large messages containing the new configuration, as well as increases latency since the reception and the processing of these messages introduces time delays. Therefore, to avoid unnecessary re-configurations, ELC algorithm includes a method to handle such a case.

In ELC, after the BS determines the new configuration, it compares the new configuration with that of the previous round. If it is the same, then the BS broadcasts a small message with zero bytes for the data portion to all the nodes to continue their operation as in the previous round. The header of this message includes a field indicating the type of the message which is a message informing the nodes to continue as in the previous round. This enables the nodes to avoid the reception of the longer messages containing the new information about the clustering and the routing paths. The packet format of the no re-configuration message is illustrated in Appendix B. On the other hand, if the new configuration is different, then the BS sends the new information about the clustering and the routing paths to the nodes in the network as described in Subsection 3.3.2.

3.5 Summary

This chapter introduced the proposed algorithm ELC. It presented a detailed description of ELC and its different components. Also, the operation of ELC in a WSN is described as well as a case with no required network re-configuration. In order to verify the anticipated improvements that ELC can achieve, ELC is tested and compared against some of the existing cluster-based algorithms in the following chapter.

Chapter 4

Performance Study of ELC

This chapter presents extensive simulations carried out by Network Simulator version 2.35 (NS-2) in order to verify the expected improvements that ELC can achieve as compared to other existing cluster-based algorithms. The benefits of employing a centralized algorithm in ELC are examined by comparing its performance against Modified-LEACH which is a recent modification of the distributed LEACH. In addition, in order to investigate the potential improvements that can be achieved by the different components of ELC, various versions of ELC as well as the proposed ELC are compared against LEACH-C and CBCDACP. LEACH-C and CBCDACP are two of the existing centralized cluster-based algorithms that have potential in reducing energy consumption and thus, they are used in the evaluation of our proposed ELC algorithm. Also, ELC is compared against LEACH-C and CBCDACP under three network scenarios where the BS is at three different locations, according to the considered WSN application. This enables to verify the improvements achieved by ELC under different applications. From the results of the simulation tests, the effect of varying the location of the BS is studied, in order to determine the scenarios for which ELC can be best suitable.

The rest of this chapter is structured as follows. Section 4.1 introduces the models and parameters used in the simulations. Section 4.2 describes the different simulation scenarios considered in our study. Section 4.3 presents and discusses the simulation results.

4.1 Simulation Models and Parameters

In this section, we describe the wireless channel propagation models and the radio energy dissipation models used in our simulations. Afterwards, the network topology considered in our simulations is presented as well as the parameters of ELC and the performance metrics.

4.1.1 Wireless Channel Propagation Models

In some applications where sensing nodes are distributed in an outdoor environment (for example in a farm), there are few determined objects that can reflect the transmitted signal. The reflection off the ground usually dominates the other reflections. Hence, the channel propagation can be modeled using the two-ray ground model, which considers the transmitted signal as well as its reflection off the ground. In other applications where the sensing nodes are distributed in an indoor environment, there are a random number of obstacles. These obstacles can block the signal path and produce random variations to the power of the received signal. Therefore, for these scenarios, the shadowing model is used for the channel propagation. The shadowing model is a statistical model which considers the received power at a certain distance as a random variable that depends on the fading effects. The following subsections discuss these two models.

4.1.1.1 Two-ray Ground Model

A transmitted signal can encounter multiple objects that reflect the signal. In the case where the signal reflected off the ground dominates other reflected signals, the two-ray ground model [31] is employed as the channel propagation model. The two-ray ground model considers the line-of-sight path as well as the ground reflection path. At the receiver, the transmitted signal and its reflected version are summed together to produce the received signal, which has distortions as compared to the original one.

In the two-ray ground model, the distance between the transmitter and the receiver plays a major role in determining the power of the received signal. For distances d smaller than a critical distance, denoted as the crossover distance d_c, the original signal and the reflected signal are combined either constructively or destructively. This depends on the phase difference between the two signals. As a result, the received power experiences a sequence of maxima and minima. By averaging out these maxima and minima, the received power falls at a rate proportional to d^2. Thus, the received power is represented using the following equation

$$P_r = \frac{P_t G_t G_r \lambda^2}{(4\pi)^2 d^2 L} \tag{4.1}$$

where P_t is the transmitted power, G_t is the transmitter gain, G_r is the receiver gain, λ is the channel wavelength and L is the system loss, which is usually set to 1.

As for distances d larger than the crossover distance, the original signal and the reflected signal are combined destructively. This is because the two signals are out of phase. As a result, the received power falls off rapidly. In average, the received power falls at a rate proportional to d^4. Thus, the received power is represented using the following equation

$$P_r = \frac{P_t G_t G_r h_t^2 h_r^2}{d^4 L} \tag{4.2}$$

where h_t is the height of the transmitter antenna and h_r is the height of the receiver antenna.

The crossover distance is calculated using the following equation

$$d_c = \frac{4\pi h_t h_r}{\lambda}. \tag{4.3}$$

4.1.1.2 Shadowing Model

In some cases, the transmitted signal experiences random variations in its power due to the obstacles in its path. These obstacles can block as well as scatter the transmitted signal. This results in random attenuations in the transmitted signal. Hence, a statistical model for the received power at a certain distance is required to represent these attenuations. The shadowing model [31] considers the fading effects by including a random variable in the calculations of the received power. The shadowing model is based on a path loss model and a log-normal model, where the received power at distance d, $P_r(d)$, is represented using the following equation

$$\left[\frac{P_r(d)}{P_r(d_0)}\right]_{dB} = -10 f \log\left(\frac{d}{d_0}\right) + X_{dB} \tag{4.4}$$

where $P_r(d_0)$ is the received power at a reference distance d_0 that is set to 1m, f is the path loss exponent and X_{dB} is a Gaussian random variable with zero mean and a standard deviation denoted as the shadowing deviation.

In the simulations, the shadowing parameters which include the path loss exponent and the shadowing deviation are set to 3.6 and 7.4, respectively. This corresponds to an indoor environment with many obstacles [32].

4.1.2 Radio Energy Dissipation Models

There are two energy models that can be used in simulations, depending on the wireless channel propagation model used. The first energy model is associated with the two-ray ground model and the second one is associated with the shadowing model. The two energy models are discussed next.

4.1.2.1 Energy Model Associated with Two-ray Ground Model

In this energy model, the transmitter dissipates energy to run the radio electronics and the power amplifier while the receiver dissipates energy to run the radio electronics as illustrated in Fig. 4.1 [17]. This model is the same as the one considered in LEACH-C [17]. The power amplifier dissipation model is based on the two-ray ground model. In other words, power loss is proportional to d^2 or d^4 for distance $d < d_c$ or $d > d_c$, respectively. To transmit an l-bit message to a node at a distance d, the radio consumes the following energy

$$\begin{aligned} E_{Tx}(l,d) &= E_{Tx,elec}(l) + E_{Tx,amp}(l,d) \\ &= \begin{cases} lE_{elec} + l\epsilon_{fs}d^2, d < d_c \\ lE_{elec} + l\epsilon_{mp}d^4, d \geq d_c. \end{cases} \end{aligned} \tag{4.5}$$

To receive an l-bit message, the radio consumes the following energy

$$E_{Rx}(l) = E_{Rx,elec}(l) = lE_{elec} \tag{4.6}$$

where E_{elec} is the electronics energy, d_c is the crossover distance and the amplifier energy for one bit is $\epsilon_{fs}d^2$ or $\epsilon_{mp}d^4$. The parameter ϵ_{fs} is deduced from Equation (4.1) by first solving the equation for P_t (amplifier power), which gives $P_t = \frac{P_r(4\pi)^2 d^2}{G_t G_r \lambda^2}$. Then it is divided by the ratio of the channel bandwidth R_b and the size of the message l, in order to convert it to Joules to represent the amplifier energy. The result of this is equivalent to $l\epsilon_{fs}d^2$ that is used in the energy model as the amplifier energy and thus, the parameter ϵ_{fs} is expressed using the following equation

$$\epsilon_{fs} = \frac{P_r(4\pi)^2}{R_b G_t G_r \lambda^2}. \tag{4.7}$$

Similarly, the parameter ϵ_{mp} is deduced from Equation (4.2). As a result, ϵ_{mp} is expressed using the following equation

$$\epsilon_{mp} = \frac{P_r}{R_b G_t G_r h_t^2 h_r^2} \tag{4.8}$$

where P_r is the received power threshold, R_b is the channel bandwidth, G_t is the transmitter gain, G_r is the receiver gain, λ is the channel wavelength, h_t is the height of the transmitter antenna and h_r is the height of the receiver antenna. Also, the crossover distance can be approximated using the two parameters ϵ_{fs} and ϵ_{mp} as follows

$$d_c = \sqrt{\frac{\epsilon_{fs}}{\epsilon_{mp}}}. \tag{4.9}$$

Nodes also consume energy while they are not in sleeping mode and they consume E_{DA} for data aggregation. The values of the parameters used in [17] are adopted as a reference of comparison. However, since the standard IEEE $802.15.4$ is employed in the simulations, the channel bandwidth is set to 250 kbps. Thus, the two parameters ϵ_{fs} and ϵ_{mp} are set accordingly. The parameters and their values are summarized in Table 4.1.

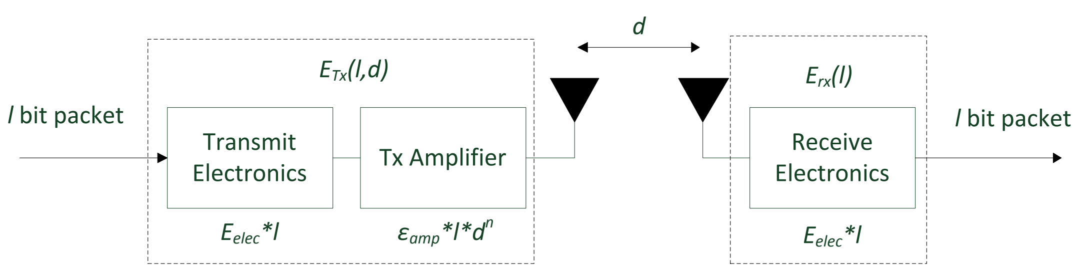


Fig. 4.1: Radio energy dissipation model.

Table 4.1: Simulation parameters.

Parameter	Value
E_{elec}	50 nJ/bit
ϵ_{fs}	38.69 pJ/bit/m^2
ϵ_{mp}	0.0052 pJ/bit/m^4
d_c	86.2 m
R_b	250 kbps
E_{DA}	5 nJ/bit/signal
Frequency	914 MHz

4.1.2.2 Energy Model Associated with Shadowing Model

In this case, the radio energy dissipation model is similar to the one used with the two-ray ground model. However, there is no distinction between short distances and long distances. More specifically, the power loss is only proportional to d^f where f is the path loss exponent. This is because in the shadowing model, the power loss can be averaged out over all distances, where the rate of falloff is proportional to d^f. In other words, to transmit an l-bit message to a node at a distance d, the radio consumes the following energy

$$E_{Tx}(l,d) = E_{Tx,elec}(l) + E_{Tx,amp}(l,d) = lE_{elec} + l\epsilon_{fs}d^f. \tag{4.10}$$

To receive an l-bit message, the radio consumes the following energy

$$E_{Rx}(l) = E_{Rx,elec}(l) = lE_{elec} \tag{4.11}$$

where E_{elec} is the electronics energy as before, f is the path loss exponent, and the amplifier energy $\epsilon_{fs} = 0.000325$ pJ/bit/m^2 according to the chosen shadowing parameters. The shadowing parameters affect the value of the received power threshold which in turn affects the value of ϵ_{fs} according to Equation (4.7). Also, the frequency used in this model is 2.4 GHz which is the typical frequency used for WSN $802.15.4$. For outdoor environment with the two-ray ground model, the frequency is set to 914 MHz, which is commonly used for remote control and short data aquisition applications.

4.1.3 Network Topology

In this chapter, the simulated network has 100 nodes randomly distributed between $x = 0$ and $x = 100$m, and between $y = 0$ and $y = 100$m, which represents the sensing area. The BS is placed at three different locations to represent three typical network scenarios. In the first scenario, it is placed at $(x = 50, y = 175)$ which corresponds to having the BS outside the sensing area as illustrated in Fig. 4.2. This scenario represents the case where WSN is used to monitor an outdoor environment such as a farm where the BS cannot be placed inside the sensing area. BS is integrated with a personal computer responsible for data processing and thus it has to be placed in the farmhouse far away from the sensing area. In the second scenario, the BS is placed at the boundary of the sensing area which is chosen to be at $(x = 50, y = 100)$ as illustrated in Fig. 4.3. And in the third network scenario, it is placed at $(x = 50, y = 50)$ which is the central point of the sensing area, as illustrated in Fig. 4.4. These two scenarios represent the cases where WSN is used, for example, for home alarm, monitoring and health care applications and therefore, the BS is placed within the sensing area.

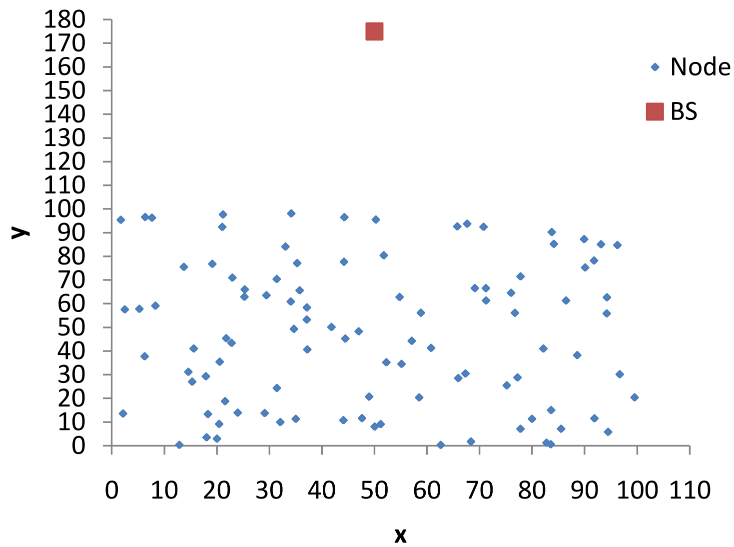


Fig. 4.2: Simulated network scenario with BS at (50,175).

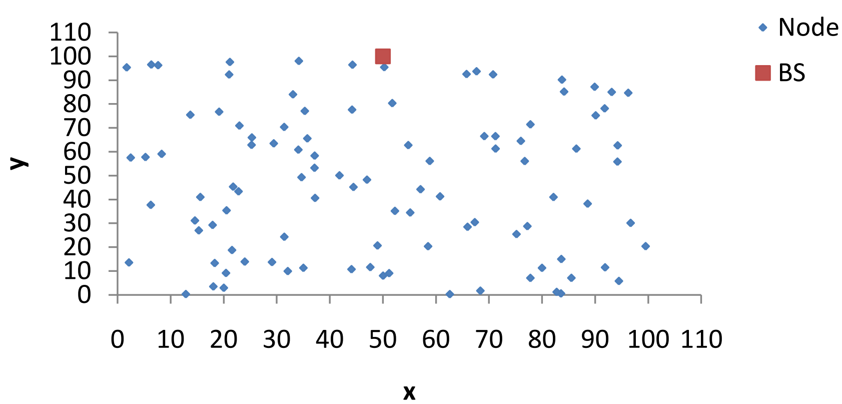


Fig. 4.3: Simulated network scenario with the BS at (50,100).

4.1.4 Parameters of ELC

The distance parameter in Equations (3.2) and (3.3) has a major effect on the network's energy consumption that ELC is aiming to reduce. This is because transmitting to a very far node consumes a large amount of energy and thus, increases the network's energy consumption. Reducing

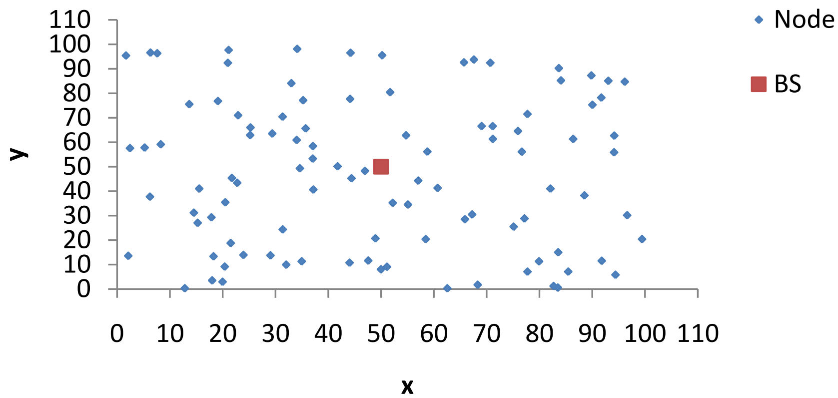


Fig. 4.4: Simulated network scenario with BS at (50,50).

energy consumption can also lead to extending network lifetime, which is another objective of ELC. The other parameters in these equations can provide energy balancing in the network that can lead to network lifetime extension. However, balancing energy without reducing energy consumption cannot effectively extend network lifetime. Thus, the distance parameter which reduces energy consumption requires more weight in these equations than the other parameters. Furthermore, the energy parameter in these equations has a more direct effect on balancing energy among CHs than the ”number of members” parameter that balances the load which in turn balances the energy consumption. Hence, the energy parameter has more weight than the number of members parameter. As a result, instead of setting the coefficients σ and ψ in Equation 3.2 to 0.5 and 0.5, respectively, σ is increased to 0.6 and thereby, ψ is set to 0.4. As for the coefficients α, β and γ in Equation 3.3, instead of setting each one of them to 0.333, α should be larger than β and β should be larger than γ, according to the above discussion. Also, they should sum up to 1. Hence, α, β and γ are set to 0.5, 0.3 and 0.2, respectively.

In the simulations of ELC, the threshold used to decide if a node should update the BS with its status information is set to 3%. This value is obtained from simulation as explained in Subsection 3.3.1.

4.1.5 Performance Metrics

In order to effectively compare the performance of the proposed ELC algorithm against the existing cluster-based algorithms, we use the following performance metrics:

- **Data Received at the BS:** This metric is the number of data messages received successfully at the BS. It is true that in ELC, we do not provide improvements in terms of link quality and the amount of delivered data, but at the same time, the anticipated improvements in terms of energy efficiency should not affect the amount of data delivered to the BS. To verify this, the number of data messages received at the BS is calculated.

- **Total Energy Dissipation:** This metric is the total amount of energy consumed by the nodes in the network which can show the energy efficiency of the algorithm under test. One of the main objectives of the proposed ELC algorithm is to reduce the total energy consumption in the network and that is why we measure the total energy dissipation. It can be used to show the effectiveness of incorporating multi-hop inter-cluster routing and the consideration that only some nodes update the BS with their status information.

- **Network Lifetime:** This metric is indicated by the number of nodes that have higher than 50% of their initial energy at the end of simulation (at 1300 sec). More nodes with higher than 50% of their energy means longer network lifetime is anticipated. By observing the variation in network lifetime between ELC and other algorithms, the effectiveness of balancing load and energy consumption among nodes in ELC can be determined. More specifically, the effectiveness of employing the residual energy and the number of members parameters in ELC can be studied.

4.2 Simulation Scenarios

Table 4.2 summarizes the scenarios considered in our simulations. In each scenario, the simulation is repeated ten times since the nodes are randomly distributed and in every simulation run the nodes have new locations in the same sensing area. Then, the average of the results obtained from the ten simulations is determined and presented for discussion. In the simulations, each node sends data towards the BS at a constant packet rate which could be either 1 or 2 packets per second depending on the simulation scenario. The packet format of the data packet is illustrated in Appendix B, where the size of the data payload is 107 bytes. It is assumed that each node has the ability to control its transmission power and to reach the BS in one hop. The initial energy of each node is 2 J and the round time is 20 s, which is adopted from LEACH-C [17] as well as other centralized cluster based protocols [23] [24]. It is also noted that the IEEE $802.15.4$ standard for Physical (PHY) layer and CSMA-based Media Access Control (MAC) layer is employed in all simulation scenarios for all algorithms. The PHY layer offers a theoretical maximum data rate of around 250 kbps. The MAC layer does not enable packet acknowledgment and re-transmission mechanisms. This is because we want to verify that the improvements offered by ELC do not decrease the amount of data received at the BS as compared to other algorithms. If re-transmissions are enabled, then it would not be possible to identify whether the successful delivery of data is because of the re-transmissions or because of the effectiveness of the algorithm.

Table 4.2: Simulation scenarios

BS Location	**Test Number**	**Channel Propagation Model**	**Algorithms under Test**	**Data Rate & Network Load**	**Purpose**
(50,175)	Test 1	Two-ray Ground	Modified-LEACH, ELC	1 pkt/sec resulting in 100 kbps	Investigating the benefits of centralizing the algorithm.
	Test 2	Two-ray Ground	LEACH-C, CBCDACP, ELC and its versions	1 pkt/sec resulting in 100 kbps	Evaluating the contribution of each component in ELC to the anticipated improvements.
	Test 3	Two-ray Ground	LEACH-C, CBCDACP, ELC	2 pkt/sec resulting in 200 kbps	Investigating the effect of increasing the load on the anticipated improvements of ELC.
(50,100)	-	Shadowing	LEACH-C, CBCDACP, ELC	1 pkt/sec resulting in 100 kbps	Study the effect of varying BS location.
(50,50)	-	Shadowing	LEACH-C, CBCDACP, ELC	1 pkt/sec resulting in 100 kbps	Study the effect of varying BS location.

4.3 Simulation Results and Discussions

As mentioned earlier, in each scenario, the simulation is repeated ten times and the average of the results obtained from these simulations is determined whether in terms of delivered data, energy consumption or network lifetime. These average values are presented in the figures and tables of the different tests. The variations of the results from the average values are approximately 2% and they are not shown in the figures for readability. In all the considered simulation scenarios, we have observed that ELC does not result in any pair of CHs that are neighbors. In other words, CHs are well distributed in ELC. Also, clusters in ELC have similar number of members as compared to LEACH-C and CBCDACP. This is determined by finding out how much the number of members of the different clusters is spread out around the mean. This is done by calculating the standard deviation of the number of members in the different clusters in each round, and then averaging it out over all the simulation rounds. The average standard deviations in ELC, LEACH-C and CBCDACP are 3.74, 9.26, and 7.02, respectively. Balancing the cluster populations can help to balance the load among CHs.

The following subsections present and discuss the results of each of the simulation scenarios.

4.3.1 The Performance of ELC with BS at $(50, 175)$

Here, we present and discuss the results of the three tests conducted to evaluate the performance of ELC while the BS is at $(50, 175)$.

4.3.1.1 Test 1: Benefits of the Centralized Algorithm in ELC

In this test, the benefits of employing a centralized algorithm in ELC are investigated. In ELC, the number of members is employed to provide load and energy balancing among CHs, which affects network lifetime, and using this parameter requires a centralized algorithm. Moreover, in ELC, the BS constructs the routing paths between CHs and the BS, which can control the entire paths to produce better decisions in terms of balancing load and energy. Therefore, network lifetime is used to verify the advantages of the centralized algorithm. Also, Modified-LEACH has a number of advertisement messages between the nodes to form the clusters, which can increase the possibility of collisions between different types of messages. This can result in leaving some nodes without CHs and losing data. Consequently, the amount of data delivered at the BS can be affected, and hence, it is calculated in this test for the two algorithms. Next, the results and their discussion are presented.

Data Received at BS. Fig. 4.5 shows the number of data messages received successfully at the BS over time, for ELC and Modified-LEACH. Also, the number of data messages sent by all the sources over time is shown in the same figure, as a reference. In this way, the figure illustrates the amount of data sent from the sources and the amount of data received at the destination. From Fig. 4.5, it can be seen that ELC delivers more data to the BS than Modified-LEACH. In

Modified-LEACH, CHs are selected probabilistically which results in CHs that are not well distributed. This leads to nodes that are far away from their CHs, which increases the probability of dropping data messages. Moreover, since CHs are not well distributed, Modified-LEACH has more collisions between the data messages that are transmitted from CMs to CHs and then to the BS, which reduces the amount of data received at the BS. Since retransmissions are not enabled in the MAC layer, the data messages that suffer from collisions and packet drops will not be retrieved. Also, Modified-LEACH suffers from collisions between advertisement and join-request messages. This can result in nodes without CHs, which reduces the amount of data received at the BS. On the other hand, ELC selects CHs deterministically using the global knowledge of the network, where each neighborhood has one CH and hence, CHs are well distributed. In this case, CMs are close to their CHs and there is less chance of dropping data messages. Furthermore, there is less chance of having collisions between data messages. Moreover, by centralizing the algorithm, there is no need for any advertisement or join-request messages and as a consequence there is less collision in the network. Therefore, ELC delivers more data to the BS.

Network Lifetime. Table 4.3 shows that ELC has 592.30% more nodes with higher than 50% of their energy at the end of simulation as compared to Modified-LEACH. In Modified-LEACH, clusters are formed without balancing their size, which results in unbalanced load distribution among CHs. On the other hand, in ELC, the number of members is employed as a parameter in cluster forming at the BS, which helps in balancing the size of clusters and, hence, the load among CHs. Moreover, the BS employs the global knowledge of the network to produce better decisions in terms of constructing multi-hop routing paths between CHs and the BS. The parameters used in constructing the multi-hop routing paths, i.e., residual energy and number of members, contribute to balancing the energy and load of CHs. Furthermore, Modified-LEACH is a distributed algorithm that employs a random CH-selection which results in CHs being very close to each other and thus, it results in some CMs being very far away from their CHs. This means that CMs consume more energy than the case in ELC where CHs are selected deterministically at the BS and are better distributed throughout the sensing area. As a result, the nodes in Modified-LEACH have less remaining energy. From the above discussion, ELC achieves longer network lifetime than Modified-LEACH.

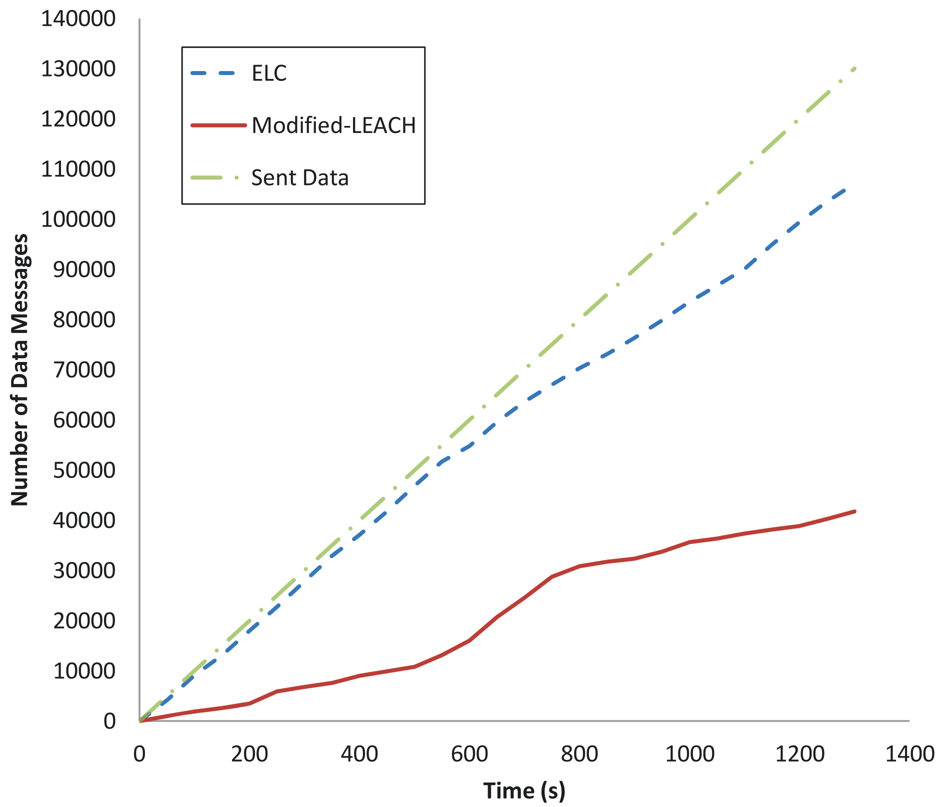


Fig. 4.5: Data received at the BS as a function of time, for Modified-LEACH and ELC.

Table 4.3: The number of nodes that have more than 50% of their energy, for Modified-LEACH and ELC

Algorithm	Number of nodes with more than 50% of their energy
Modified-LEACH	13
ELC	90

4.3.1.2 Test 2: Evaluating Different Components of ELC

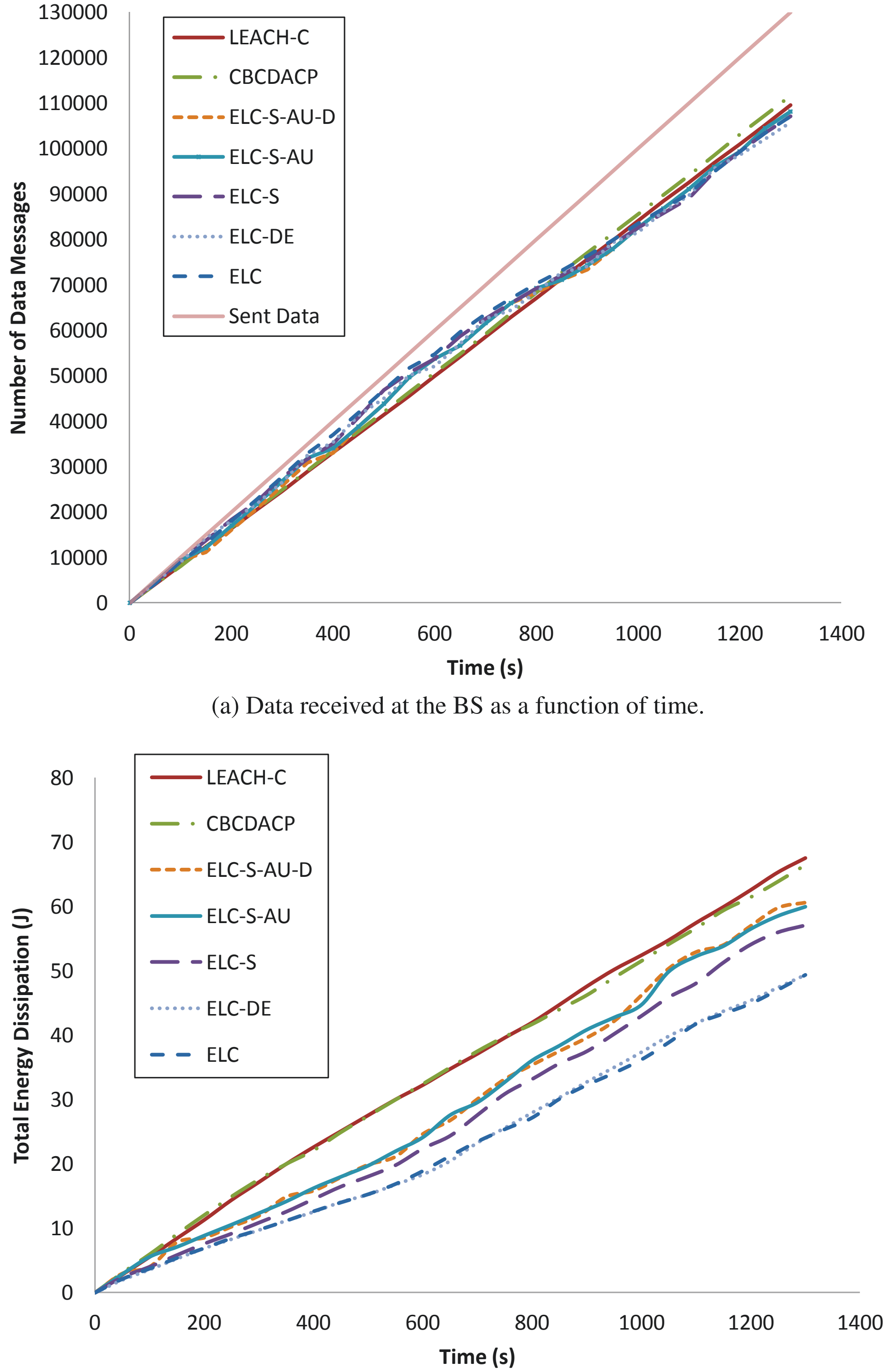
In this test, the contribution of each component of ELC is evaluated in terms of reducing energy consumption and extending network lifetime. This is done by considering five versions of ELC as follows. *ELC-S-AU-D* is a version of ELC that considers only the distance parameter in cluster forming, has all the nodes updating the BS with their status information in every round, and employs only one-hop communication between CHs and the BS. This version is used to evaluate the CH selection and cluster forming of ELC, without the number of members parameter. *ELC-S-AU* is similar to ELC-S-AU-D except that it employs the number of members parameter in cluster forming. This enables to evaluate the number of members parameter in cluster forming. *ELC-S* is similar to ELC-S-AU except that it includes the consideration that only the nodes with significant change update the BS with their status information. This helps to determine the contribution of the consideration that only some nodes update the BS with their information. *ELC-DE* is similar to ELC-S except that it incorporates the multi-hop inter-cluster routing using only residual energy and distance parameters. These two parameters are the ones considered in the inter-cluster communication of CPMS. The purpose of using ELC-DE is to assess the contribution of the multi-hop transmission. Finally, *ELC* is similar to ELC-DE except that it includes the number of members parameter in inter-cluster routing and it is the complete version. This is to assess the contribution of the number of members parameter in inter-cluster communication. Next, the results and their discussion are presented.

Data Received at BS. The number of data messages received successfully at the BS is shown in Fig. 4.6a as a function of time, for LEACH-C, CBCDACP and all versions of the proposed algorithm. Also, this figure shows the number of data messages sent from the nodes towards the BS, as a point of reference. LEACH-C, CBCDACP and all versions of the proposed algorithm deliver almost the same amount of data which verifies that employing ELC does not affect the amount of delivered data. The loss in delivered data as compared to the sent data is due to the collisions during the CSMA mechanism and due to the channel propagation.

Energy Consumption. From Fig. 4.6b, *ELC-S-AU-D* consumes less energy than LEACH-C and CBCDACP, where the average energy reduction is $5.58\text{J}(7.70\%)$ as compared to LEACH-C and $3.95\text{J}(5.90\%)$ as compared to CBCDACP. In the proposed algorithm, by using the distance parameter in cluster forming, CMs are closer to their CHs than the case in LEACH-C and CBC-

DACP, where some CMs can be far away from their CHs since there is a fixed number of CHs. In *ELC-S-AU*, the energy reduction as compared to LEACH-C and CBCDACP is almost the same as that of ELC-S-AU-D. This is because ELC-S-AU includes, besides the distance parameter, the number of members as a parameter in cluster forming which affects the load and energy balancing of CHs and, consequently, affects the network lifetime more than the total energy reduction. As for *ELC-S*, the average energy reduction is $7.87\text{J}(17.99\%)$ as compared to LEACH-C and $6.32\text{J}(15.14\%)$ as compared to CBCDACP. This increase in the energy reduction as compared to the case of ELC-S-AU is due to the consideration that only the nodes with significant change in their status information update the BS, which is included in ELC-S. Thus, this consideration contributes to network's energy reduction by approximately 9.76%.

In *ELC-DE*, the average energy reduction is $18.10\text{J}(32.60\%)$ as compared to LEACH-C and $16.50\text{J}(29.60\%)$ as compared to CBCDACP. The increase in energy reduction as compared to the case of ELC-S is due to the incorporation of the multi-hop routing paths between CHs and the BS. On the other hand, LEACH-C and CBCDACP depend only on the one-hop communication between CHs and the BS, which requires a large amount of energy. This will be further verified later. Employing multi-hop inter-cluster communication in ELC contributes to networks' energy reduction by approximately 14.53%. As for ELC, the energy reduction as compared to LEACH-C and CBCDACP is almost the same as that of ELC-DE. This is because ELC differs from ELC-DE in that it includes the number of members as a parameter in the inter-cluster communication, which affects the load and energy balancing of CHs and, consequently, affects the network lifetime more than the total energy reduction.



(a) Data received at the BS as a function of time.

(b) Total energy dissipation as a function of time.

Fig. 4.6: The results of evaluating the different components of ELC.

Verifying the Benefits of the Multi-hop Transmission. From the results above, the incorporation of the multi-hop routing paths introduces reductions in the total energy consumption. Here, we present an example to verify the fact that employing multi-hop routing paths as opposed to one-hop transmission can indeed reduce energy consumption. This example considers the CHs in each algorithm transmitting their data towards the BS.

In LEACH-C, in one of the rounds, CHs are transmitting their data to the BS in a one-hop manner as shown in Fig. 4.7. Based on the distance of these one-hop transmissions, Equation (4.5) is used to calculate the consumed energy in transmission. The total energy consumed in transmitting one data message from every CH to the BS is 7.69×10^{-3}J. For CBCDACP, it is similar to LEACH-C since it is using one-hop communication as well.

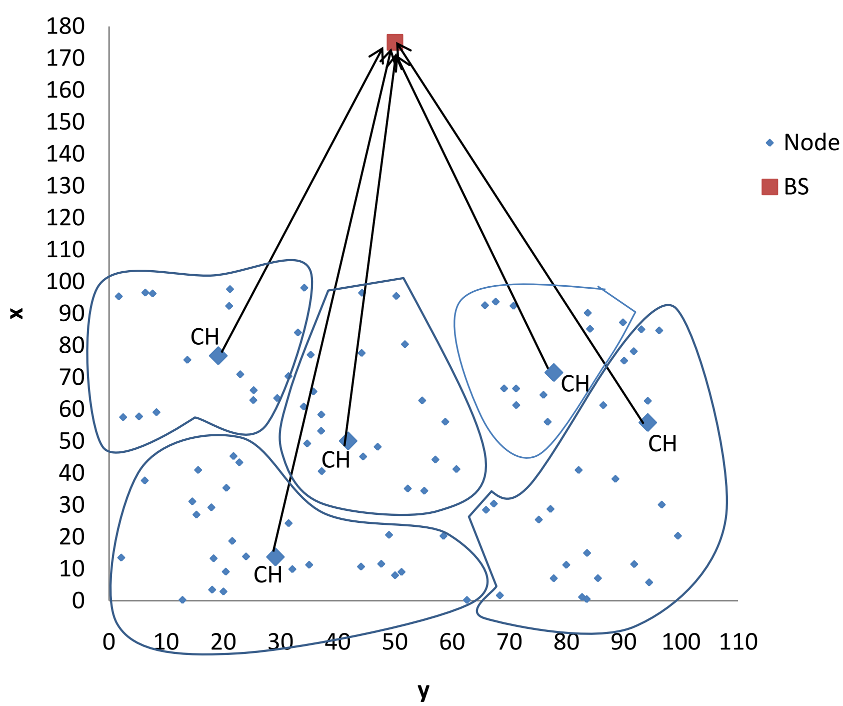


Fig. 4.7: One-hop inter-cluster communication in LEACH-C.

In ELC, in one of the rounds, CHs are transmitting their data to the BS as shown in Fig. 4.8. According to Equation (4.5), the consumed energy in transmitting one data message from each of the CHs to the BS along the appropriate hops is 2.79×10^{-3}J. The total consumed energy in receiving the data messages at the intermediate hops is 0.15×10^{-3}J, according to Equation 4.6. As a result, the total energy consumed is 2.94×10^{-3}J. This value is less than that in LEACH-C.

Therefore, the multi-hop routing paths in ELC can indeed reduce the total energy consumption in the network as compared to the one-hop communication in LEACH-C and CBCDACP.

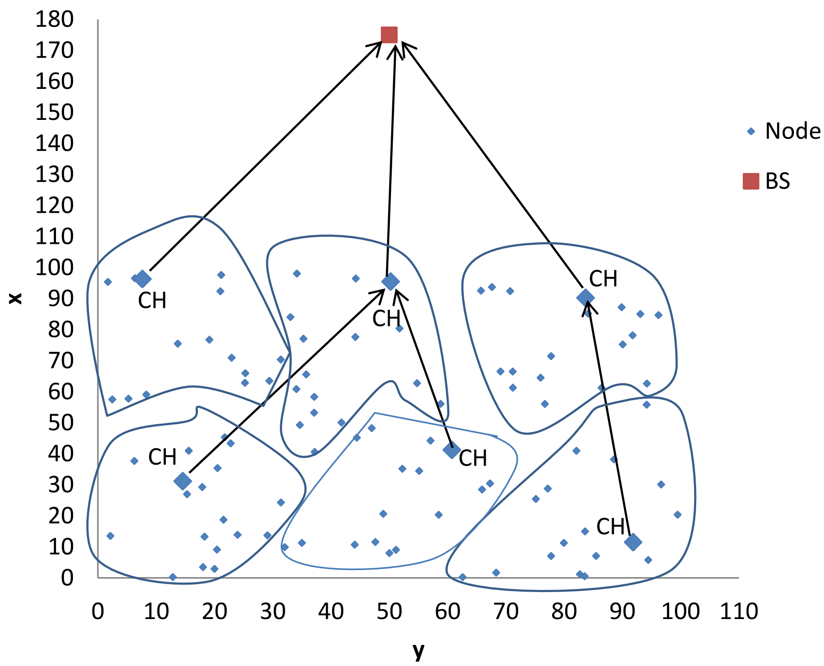


Fig. 4.8: Multi-hop inter-cluster communication in ELC.

Network Lifetime. The number of nodes with higher than 50% of their energy at the end of simulation is presented in Table 4.4, as an indication of network lifetime. *ELC-S-AU-D* has 90.00% more of such nodes than LEACH-C and 35.71% more than CBCDACP. In LEACH-C and CBCDACP, CHs selection and cluster forming consider distance between the nodes and the density factor of CHs. However, they do not look into balancing energy consumption among CHs, which affects network lifetime. On the other hand, in ELC-S-AU-D, the nodes' residual energy is used as a parameter in CH selection and thus, the energy consumption can be balanced among the nodes in the network, since CHs tend to consume more energy. As for *ELC-S-AU*, it has 155% more nodes with higher than 50% of their energy than LEACH-C and 82.14% more than CBCDACP. The increase in the number of nodes with more energy is due to the inclusion of the number of members as a parameter in cluster forming, which helps in balancing the size of clusters as well as the load among CHs. The number of members used in cluster forming contributes to increasing network lifetime by approximately 55.71%. In *ELC-S*, there is 205.00% more nodes with higher than 50% of their energy than LEACH-C and 117.85% more than CBC-DACP. The increase in number of nodes as compared to the case of ELC-S-AU is due to the consideration that only the nodes with significant change in their status information update the BS, which is included in ELC-S. This consideration reduces network's energy consumption as explained earlier, which leads to more nodes with higher energy and longer network lifetime. The contribution of this consideration to increasing network lifetime is approximately 42.85%.

In *ELC-DE*, there is 285.00% more nodes with higher than 50% of their energy than LEACH-C and 175.00% more than CBCDACP. The increase in number of nodes as compared to the case of ELC-S is due to employing the multi-hop inter-cluster communication that reduces network's energy consumption as explained earlier and hence, it results in more nodes with higher energy and longer network lifetime. In addition, employing the residual energy parameter in the multi-hop inter-cluster communication helps in balancing the energy consumption among CHs. As a result, the multi-hop inter-cluster communication in ELC contributes to increasing network lifetime by approximately 68.57%. As for *ELC*, it has 350.00% more nodes with higher than 50% of their energy than LEACH-C and 221.42% more than CBCDACP. The increase in number of nodes as compared to the case of ELC-DE is due to the inclusion of the number of members as a parameter in constructing the multi-hop paths in the inter-cluster communication of ELC, which balances the load among CHs. The number of members used in the inter-cluster communication contributes to increasing network lifetime by approximately 55.71%.

Table 4.4: The number of nodes that have more than 50% of their energy, for LEACH-C, CBC-DACP and the different versions of ELC.

Algorithm	**Number of nodes with more than 50% of their energy**
LEACH-C	20
CBCDACP	28
ELC-S-AU-D	38
ELC-S-AU	50
ELC-S	61
ELC-DE	77
ELC	90

4.3.1.3 Test 3: Increasing Load

In this test, the energy consumption and network lifetime are studied while increasing the load. The results are summarized in Table 4.5. This table also includes the results of ELC from the previous test where the network load is lower.

While increasing the load, the percentage of average energy reduction does not change, where it is 32.60% as compared to LEACH-C and 29.60% as compared to CBCDACP. As for the increase in number of nodes with $> 50\%$ of energy, the value has changed to 457.14% as compared to LEACH-C and 290.00% as compared to CBCDACP. The reason for the increase is as follows. In the case where clusters have unbalanced sizes, the load of the corresponding CHs would be different and unbalanced, as in LEACH-C and CBCDACP. After the increase in load generated by CMs, the load difference between CHs would increase as well. This means that by increasing the network load, balancing the load among CHs would play a more important role in increasing network lifetime. This is achieved in ELC by employing the number of members as a parameter in cluster forming as well as in the inter-cluster routing. In addition, increasing the network load results in increasing the gap between the energy consumption of CHs and that of CMs. This means that balancing energy among nodes becomes more critical, which can be achieved in ELC by considering the nodes' residual energy in CH selection.

Table 4.5: Performance comparison while increasing load.

Network Load	**Average Energy Reduction**		**Increase in #Nodes with >50% Energy**	
	Compared to LEACH-C (J)	**Compared to CBCDACP (J)**	**Compared to LEACH-C**	**Compared to CBCDACP**
100 kbps	18.10	16.50	70	62
200 kbps	17.96	16.20	32	29

4.3.2 The Performance of ELC with BS at $(50, 100)$ and $(50, 50)$

In this subsection, we present the results of the test for the case of the BS at $(50, 100)$ as well as that for the case of the BS at $(50, 50)$. The specifics of these scenarios were presented previously in Table 4.2. In order to compare the results of these tests to the case where the BS is at $(50, 175)$, another simulation test is required for the $(50, 175)$ case. More specifically, a simulation test to compare ELC against LEACH-C and CBCDACP is run under the same conditions of the $(50, 100)$ test and the $(50, 50)$ test except that the BS is at $(50, 175)$. Mainly, this test is similar to the comparison done between ELC, LEACH-C and CBCDACP in Test 2 above, except that the shadowing model is used instead.

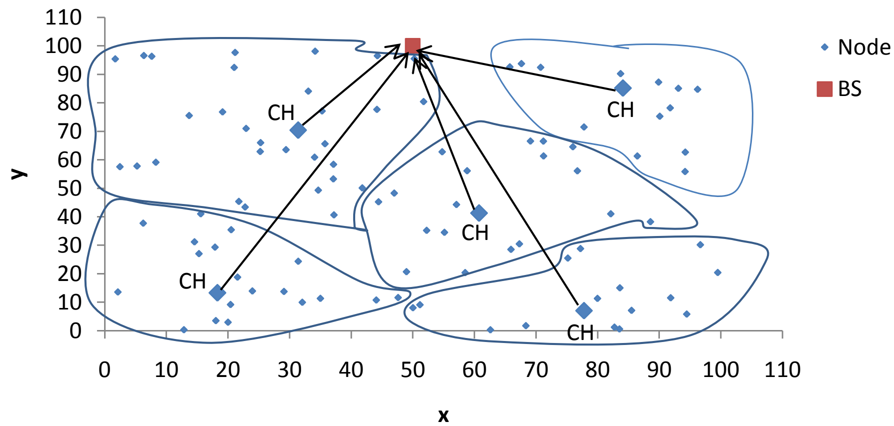
The average reduction in the total energy consumption and the increase in the number of nodes with more than 50% of their energy, as compared to LEACH-C and CBCDACP, are presented in Table 4.6 for the three cases of BS location.

Table 4.6: Performance comparison while varying BS location.

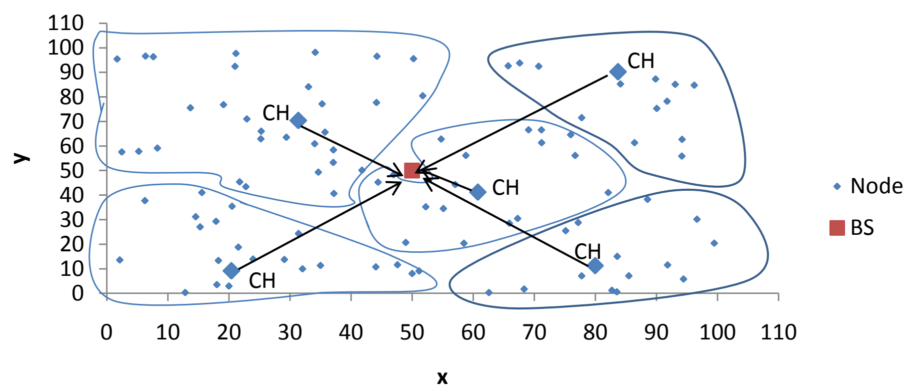
BS Location	**Average Energy Reduction**		**Increase in #Nodes with >50% Energy**	
	Compared to LEACH-C (J)	**Compared to CBCDACP (J)**	**Compared to LEACH-C**	**Compared to CBCDACP**
50,175	7.13 (27.12%)	6.31 (24.74%)	40.90%	32.85%
50,100	5.66 (26.56%)	4.51 (23.41%)	33.33%	28.0%
50,50	4.41 (22.05%)	3.57 (20.78%)	25.00%	23.37%

4.3.2.1 Comments

From Table 4.6, the average energy reduction by ELC becomes smaller as the BS gets closer to the sensing area. When the BS gets closer to the sensing area, the CHs in LEACH-C and CBCDACP transmit in shorter-distance one-hop communications to the BS as illustrated in Fig. 4.7, Fig. 4.9a and Fig. 4.9b. This reduces the energy consumption of one-hop communications since energy consumption depends on the distance between the transmitter and the receiver. For ELC, the path with the lowest cost between each CH and the BS is selected. By having the BS closer to the nodes, the cost of the direct paths between CHs and the BS in ELC (which depend on the distance) become smaller than the cost of the multi-hop paths that depend on Equation 3.4, where the distance parameter has the highest weight. As a result, most of the CHs in ELC deliver their data in one-hop transmissions to the BS as illustrated in Fig. 4.10a and Fig. 4.10b. This results in almost the same energy consumption of the one-hop communication in LEACH-C and CBCDACP. However, there are still few CHs in ELC that deliver their data in multi-hop paths as can be seen from Fig. 4.10a and Fig. 4.10b, since their multi-hop communication has lower cost than the one-hop communication. In this case, ELC consumes less energy than LEACH-C and CBCDACP. Therefore, as the BS gets closer to the sensing area, there is less difference in the energy consumed in the routing paths of CHs between LEACH-C, CBCDACP and ELC. This indicates that ELC provides less reduction in energy consumption when the BS gets closer to the sensing area.

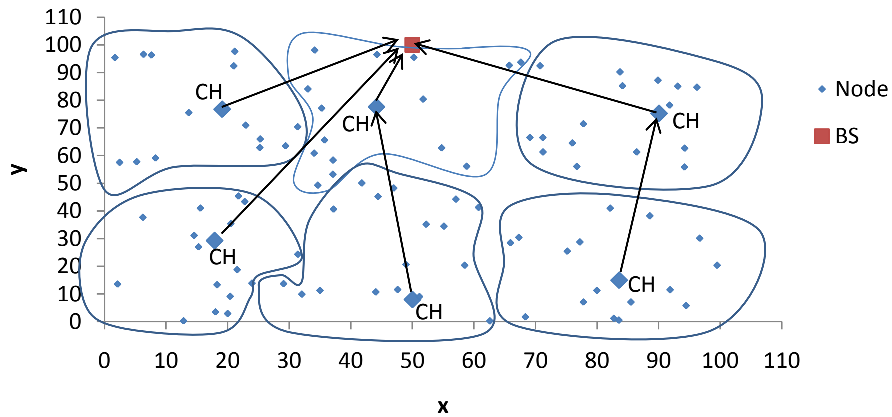


(a) Routing configuration in LEACH-C in the case of BS at $(50, 100)$

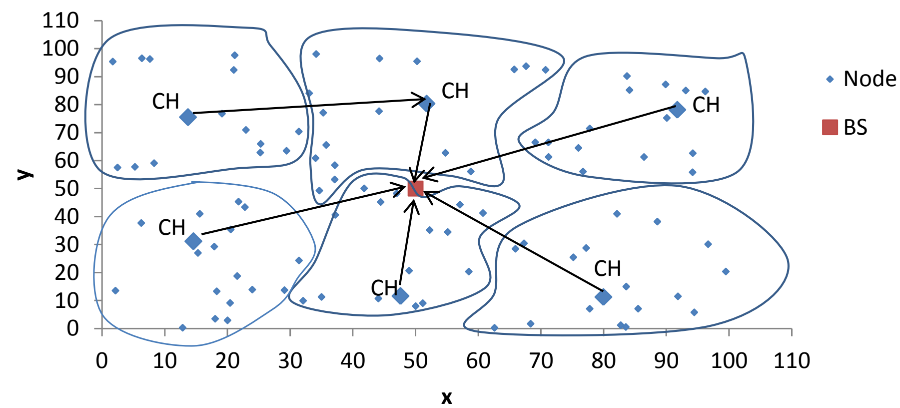


(b) Routing configuration in LEACH-C in the case of BS at $(50, 50)$

Fig. 4.9: Routing configurations in LEACH-C.



(a) Routing configuration in ELC in the case of BS at $(50, 100)$



(b) Routing configuration in ELC in the case of BS at $(50, 50)$

Fig. 4.10: Routing configurations in ELC.

Also, as a result of having less energy reduction, the increase in number of nodes with more energy is reduced as well. This is because if the difference in total energy consumption between the three algorithms is getting smaller, the difference in the number of nodes with more energy is getting smaller as well. However, ELC can still achieve energy balancing and network lifetime extension in the three cases of BS location. This is achieved by considering residual energy in CH selection, and employing the number of members and energy parameters in cluster forming and in inter-cluster routing.

ELC can achieve improvements at a lesser degree over LEACH-C and CBCDACP in terms of energy consumption and network lifetime as the BS location is moved, as shown in Table 4.6. In addition, the results in this test demonstrate that the proposed ELC algorithm is smart enough to identify the efficient method to deliver data from CHs to the BS, whether it is one-hop or multi-hop type of transmission.

By studying the results of the $(50, 175)$ test with the shadowing model, the reduction in energy and the increase in number of nodes with more energy are less than the case of the two-ray ground model (presented previously). In the two-ray ground model, most of the transmissions in LEACH-C and CBCDACP fall under the d^4 power loss whereas most of the transmissions in ELC belong to the d^2 power loss corresponding to short-distance transmissions. This contributes to a larger gap between LEACH-C, CBCDACP and ELC. However, in the shadowing model, the transmissions in the three algorithms fall under the d^f power loss where f is the path loss exponent (3.6 in our simulations). As a result, ELC provides less reduction in the energy consumption as compared to the case of the two-ray ground model. Moreover, in the shadowing model test, the shadowing parameters affect the received power threshold which in turn affects the calculation of the amplifier energy ϵ_{fs}, according to Equation 4.7. Smaller amplifier energy in the shadowing model yields to smaller nodes' energy consumption according to Equation 4.10, which results in more number of nodes with more than 50% of their energy. Thus, there are small differences in the number of nodes with more energy between the three algorithms. However, ELC can still achieve improvements in both cases, two-ray ground and shadowing channel models, according to Tables 4.5 and 4.6.

4.4 Summary

This chapter presented a simulation study of ELC. The performance evaluation of ELC against other cluster-based routing algorithms was presented and discussed. Through different simulation tests, the expected improvements that each component of ELC can achieve were verified. It was demonstrated that ELC can reduce the total energy dissipation and extend network lifetime while sustaining the performance in terms of the amount of data delivered to the BS. In addition, three cases of BS location were used to determine the case in which ELC can achieve more improvements. Even though ELC achieved improvements in all three cases, ELC showed more improvements in terms of reducing energy consumption and extending network lifetime in the case where the BS is outside of the sensing area. This case corresponds to those applications that have the BS outside the area generating the data to be collected, such as a farm generating and delivering data to a BS (personal computer) inside the owner's house that is next to the farm.

Chapter 5

Conclusion

This thesis introduces an Energy-efficient and Load-balancing Cluster-based, ELC, routing algorithm. ELC employs CSMA in data transmission and hence, it is more practical than existing cluster-based algorithms that assume the use of both TDMA and CDMA which requires synchronization between nodes. In ELC, we introduce a new method to determine the neighbors of each node in the network to be used in the energy comparisons of the CH selection. This is done while ensuring that the network has desired number of CHs by employing the $distance\ limit$ in identifying the neighbors of each node. In addition, a new parameter is used in the cluster forming of ELC, which is the number of members associated with a CH. Employing this parameter helps in balancing the size of clusters and thus, balances the load among CHs. Moreover, in ELC, we incorporate multi-hop routing paths to forward data from CHs to the BS, instead of depending only on the one-hop communications that can consume a large amount of energy. These multi-hop paths are based on CHs' residual energy, distance between CHs, and the number of members associated with a CH. In this way, we aim to reduce energy consumption, and balance the load and energy consumption among CHs. Furthermore, we introduce a method where only the nodes with significant change in their status information would send an update to the BS, as opposed to all the nodes updating the BS, which also reduces the energy consumption in the network.

The advantages of ELC are verified by comparing its performance against that of existing cluster-based algorithms using NS-2. In order to demonstrate the benefits of employing a centralized algorithm, ELC is compared against Modified-LEACH, which is a distributed algorithm. The simulation results show that the centralized algorithm can produce well distributed CHs that deliver more data to the BS. Also, the centralized algorithm is able to balance load and energy among CHs by producing better decisions at the BS instead of at the nodes and hence, extend the network lifetime. In addition, the different components of ELC are evaluated against LEACH-C and CBCDACP, where the contribution of each component in reducing energy consumption and extending network lifetime is determined. Based on the analysis of the different simulation tests, it was concluded that ELC is able to extend network lifetime by balancing the load and energy among the nodes, and reduce the total energy dissipation in the network. This is observed while ELC sustains the performance in terms of the amount of data received at the BS. Moreover, the applications where the BS is outside the sensing area are identified as the most suitable scenarios for employing ELC. In these scenarios, ELC demonstrates the best improvements in terms of energy reduction and network lifetime extension.

5.1 Future Work

In this thesis, ELC is tested and evaluated against other cluster-based routing algorithms using simulations. Experimental testing can be beneficial to demonstrate the improvements achieved by ELC on a real testbed. This can also show the practicality of ELC.

There are different applications, such as smart grid, that can benefit from the advantages of cluster-based algorithms. Smart grid is an electric power-grid infrastructure used to provide effective communications, monitoring and automation [33]. In smart grid, each house can be equipped with a smart meter that collects information about the electricity usage in the corresponding house. A group of smart meters can form a Neighborhood Area Network (NAN) that can be established using a WSN because of the low cost, flexibility and intelligence of WSNs [34]. In this way, power disturbances and outages can be avoided by monitoring the power conditions and providing the required diagnostics and protection. This can also help in the issue of increasing electricity demand and network congestion [33]. Employing cluster-based algorithms in NANs can help in the issue of latency and limited bandwidth. Moreover, a NAN consists of smart meters, routers and a collector. The responsibilities of a router is similar to a CH in a cluster-based algorithm, where a router collects data from the corresponding smart meters and forwards it to the collector [34]. The collector here is similar to the BS used in cluster-based algorithms. Also, since the smart meters and routers are not moving, a cluster-based routing would be suitable for this kind of application because of the movement limitations. Therefore, it would be interesting as a future work to adapt our proposed ELC algorithm to be used in NANs, according to the requirements of smart grid applications.

Appendix A

Multiple Access Techniques

A.1 TDMA

Time Division Multiple Access, TDMA, is one of the multiple access techniques used in wireless communication networks to allow more than one user to share the same frequency channel. In TDMA, the transmission channel is divided into time slots where each node has its assigned time slot to be used for transmitting in each TDMA cycle. Each node employs the entire frequency band during its transmission [35]. This is illustrated in Fig. A.1, where N is the number of nodes and B is the frequency band.

Guard times are used to ensure that there is no overlap between the transmissions of different nodes, in case there is mistiming in their transmissions. At the beginning of each bit stream that represents a data message, there is a header that allows the receiver to synchronize to the transmitted bit stream. Synchronization between the transmitting nodes as well as between the transmitter and the receiver increases the overhead in TDMA [35]. Also, it is difficult to achieve in ad-hoc WSNs.

A.2 CDMA

Code Division Multiple Access, CDMA, is another channel access method that employs spread-spectrum technology as well as a coding scheme. More than one node can transmit at the same time using the entire frequency band. Different codes are used by the nodes to produce the transmitted data messages. An example of coding a data signal is shown in Fig. A.2, where T_d is the pulse duration of the data signal and T_c is the pulse duration of the code signal. The receiver

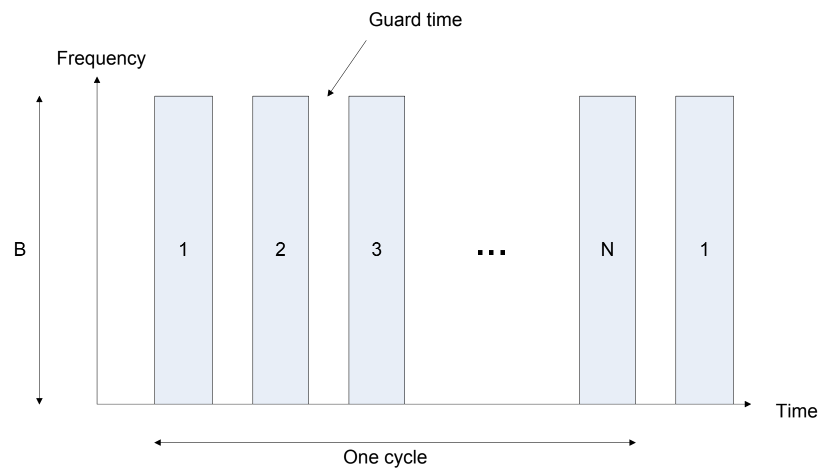


Fig. A.1: Time Division Multiple Access

recovers the data message from the desired node using these codes. Moreover, in CDMA, the data messages have a higher bandwidth than what the actual data requires [35].

Usually, the codes used in CDMA are orthogonal, which helps in eliminating interference. As a result, synchronization between the nodes is required which is difficult to achieve in ad-hoc WSNs. Moreover, the number of codes should not exceed a certain limit since they have to be orthogonal. Thus, the number of active transmitters is limited. In addition, CDMA requires that all the data messages at the receiver should have approximately the same power. Otherwise, signals with high power would overcome those with low power [35]. In WSNs, channel fading produces nodes messages that have different power levels and hence, employing CDMA complicates the operation.

A.3 CSMA

Carrier Sense Multiple Access, CSMA, is a random access technique that is able to avoid collisions by sensing the channel before transmitting. Any node can transmit at any time using the entire frequency band. However, before transmitting a message, each node senses the channel, if it is busy, the transmission is deferred. Otherwise, the node transmits the message. This is illustrated in Fig. A.3. There are different extensions to CSMA; the most important ones are CSMA/Collision Avoidance (CA) and CSMA/Collision Detection (CD). CSMA/CA includes collision avoidance where a node defers its transmission for a random interval in case the chan-

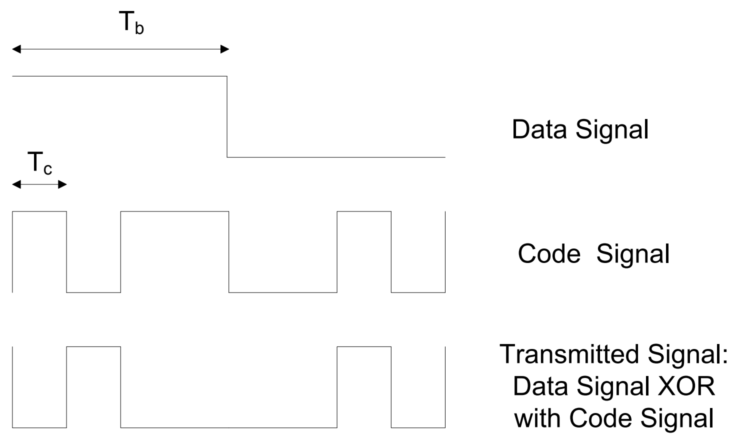


Fig. A.2: Code Division Multiple Access

nel is busy. This reduces the probability of collisions on the channel. CSMA/CD improves the performance of CSMA by terminating the transmission as soon as a collision is detected in order to avoid wasting bandwidth. However, detecting a collision in wireless channels is a challenging problem [35].

In CSMA, no synchronization is required between the nodes, this is beneficial for WSNs. Also, nodes do not have to wait its turn to transmit data, which can reduce latency. Therefore, the standard used in WSNs that is IEEE $802.15.4$ employs CSMA in transmitting data [29].

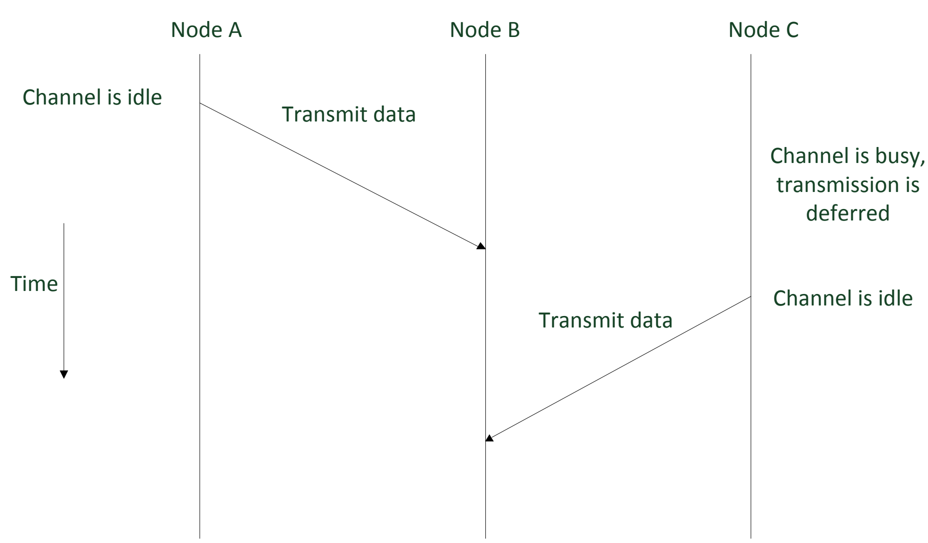


Fig. A.3: Carrier Sense Multiple Access

Appendix B

ELC Packet Format

The packet format of all the messages introduced in the proposed ELC algorithm is illustrated in Fig. B.1. This figure also includes the headers and footers that belong to the PHY and MAC layers. Fig. B.1 shows the hierarchical breakdown of the packet fields. The packet in the physical layer, which is denoted as PHY Protocol Data Unit (PPDU), contains a header and MAC Protocol Data Unit (MPDU) which is encapsulated into the PPDU. The MPDU consists of a MAC Header (MHR), a footer, and the MAC Service Data Unit (MSDU). The MSDU contains the packet fields that are employed specifically by our proposed ELC algorithm. Depending on the type of the message, the fields are set accordingly as illustrated in the figure.

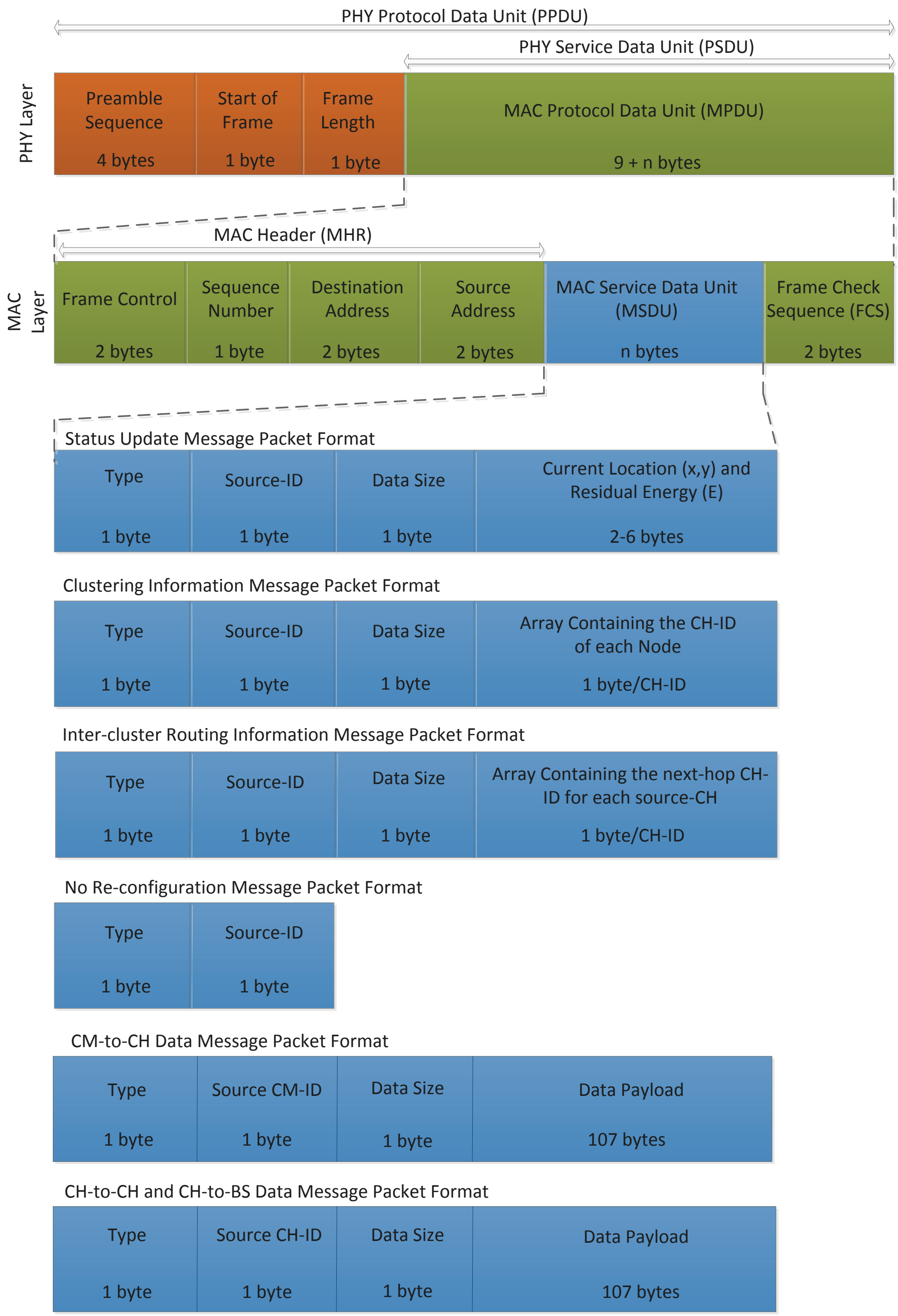


Fig. B.1: ELC packet format.

References

[1] I. Khemapech, I. Duncan, and A. Miller, “A survey of wireless sensor networks technology,” *Sensors Peterborough NH*, 2005.

[2] J. Yick, B. Mukherjee, and D. Ghosal, “Wireless sensor network survey,” *Computer Networks*, vol. 52, no. 12, pp. 2292 – 2330, 2008.

[3] K. Akkaya and M. Younis, “A survey on routing protocols for wireless sensor networks,” *Ad Hoc Networks*, vol. 3, no. 3, pp. 325 – 349, 2005.

[4] M. A. Al-Qutayri and J. S. Jeedella, *Integrated Wireless Technologies for Smart Homes Applications, Smart Home Systems*. InTech, 2010.

[5] H. Huang, S. Xiao, X. Meng, and Y. Xiong, “A remote home security system based on wireless sensor network and gsm technology,” in *Proc. 2010 Second International Conference on Networks Security Wireless Communications and Trusted Computing (NSWCTC’10)*, vol. 1, pp. 535 –538, Apr. 2010.

[6] M. Shu-guang, “Construction of wireless fire alarm system based on Zigbee technology,” *Procedia Engineering*, vol. 11, pp. 308 – 313, May 2011.

[7] T. Defeng, L. Shixing, X. Wujun, and Z. Yongming, “A fire monitoring system in zigbee wireless network,” in *Proc. 2010 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC’10)*, pp. 48–51, Oct. 2010.

[8] A. Barbato, L. Borsani, A. Capone, and S. Melzi, “Home energy saving through a user profiling system based on wireless sensors,” in *Proc. First ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings*, pp. 49–54, 2009.

[9] K. Shinghal, A. Noor, N. Srivastava, and R. Singh, “Wireless sensor networks in agriculture: for potato farming,” *International Journal of Engineering Science and Technology*, vol. 2, no. 8, pp. 3955–3963, 2010.

[10] D. D. Chaudhary, S. P. Nayse, and L. M. Waghmare, “Application of wireless sensor networks for greenhouse parameter control in precision agriculture,” *International Journal of Wireless and Mobile Networks (IJWMN)*, vol. 3, pp. 140–149, Feb. 2011.

[11] N. Suryadevara and S. Mukhopadhyay, "Wireless sensor network based home monitoring system for wellness determination of elderly," *IEEE Sensors Journal*, vol. 12, pp. 1965–1972, Jun. 2012.

[12] S. K. Singh, M. P. Singh, and D. K. Singh, "A survey of energy-efficient hierarchical cluster-based routing in wireless sensor networks," *International Journal of Advanced Networking and Applications*, vol. 2, no. 2, pp. 570–580, 2010.

[13] J. Lotf, M. Hosseinzadeh, and R. Alguliev, "Hierarchical routing in wireless sensor networks: a survey," in *Proc. 2nd International Conference on Computer Engineering and Technology (ICCET'10)*, pp. V3(650–654), Apr. 2010.

[14] A. A. Abbasi and M. Younis, "A survey on clustering algorithms for wireless sensor networks," *Computer Communications*, vol. 30, no. 14-15, pp. 2826 – 2841, 2007.

[15] J. Yu and P. Chong, "A survey of clustering schemes for mobile ad hoc networks," *IEEE Communications Surveys & Tutorials*, vol. 7, pp. 32–48, 1st Qtr. 2005.

[16] D. Wei and H. Chan, "A survey on cluster schemes in ad hoc wireless networks," in *2nd International Conference on Mobile Technology, Applications and Systems*, pp. 1–8, Nov. 2005.

[17] W. B. Heinzelman, *Application-Specific Protocol Architectures for Wireless Networks*. PhD thesis, MIT, Jun. 2000.

[18] W. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "An application-specific protocol architecture for wireless microsensor networks," *IEEE Transactions on Wireless Communications*, vol. 1, pp. 660 – 670, Oct. 2002.

[19] C.Sha, R.C.Wang, H.P.Huang, and L.J.Sun, "Energy efficient clustering algorithm for data aggregation in wireless sensor networks," *The Journal of China Universities of Posts and Telecommunications*, vol. 2, pp. 104–109, Dec. 2010.

[20] K. T. Kim, H. K. Yoo, B. H. Son, and H. Y. Youn, "An energy efficient clustering scheme for self-organizing distributed wireless sensor networks," in *Proc. IEEE Tenth International Conference on Computer and Information Technology (CIT'10)*, pp. 1468–1473, Jul. 2010.

[21] S. Moad, M. A. Azim, and N. Bouabdallah, "ADEEC: A novel adaptive energy efficient clustering protocol in wireless sensor network," in *Proc. IEEE Global Communications Conference (GLOBECOM'10)*, pp. 1–6, Dec. 2010.

[22] W. Wang, B. Wang, Z. Liu, L. Guo, and W. Xiong, "A cluster-based and tree-based power efficient data collection and aggregation protocol for wireless sensor networks," *Information Technology Journal*, vol. 10, no. 3, pp. 557–564, 2011.

[23] J. Ferdous, M. J. Ferdous, and T. Dey, "A comprehensive analysis of CBCDACP in wireless sensor networks.," *Journal of Communications*, vol. 5, pp. 627–636, Aug. 2010.

[24] F. Bajaber and I. Awan, "Adaptive decentralized re-clustering protocol for wireless sensor networks," *Journal of Computer and System Sciences*, vol. 77, pp. 282–292, Mar. 2011.

[25] S. Muruganathan, D. Ma, R. Bhasin, and A. Fapojuwo, "A centralized energy-efficient routing protocol for wireless sensor networks," *IEEE Communications Magazine*, vol. 43, pp. S8 – 13, Mar. 2005.

[26] S. Varma, N. Nigam, and U. Tiwary, "Base station initiated dynamic routing protocol for heterogeneous wireless sensor network using clustering," in *Proc. Fourth International Conference on Wireless Communications and Sensor Networks (WCSN)*, pp. 1–6, Dec. 2008.

[27] A. Kusdaryono and K.-O. Lee, "A clustering protocol with mode selection for wireless sensor network," *Journal of Information Processing Systems*, vol. 7, pp. 29–42, Mar. 2011.

[28] A. F. Salami, F. Anwar, A. M. Aibinu, H. Bello-Salau, and A. H. Abdalla, "Investigative analysis of clustering routing protocols for scalable sensor networks," in *Proc. Fourth International Conference On Mechatronics (ICOM'11)*, pp. 1–5, May 2011.

[29] C. Buratti, A. Conti, D. Dardari, and R. Verdone, "An overview on wireless sensor networks technology and evolution," *Sensors*, vol. 9, pp. 6869–6896, Aug. 2009.

[30] J. Kurose and K. Ross, *Computer Networking: A Top-Down Approach*. Addison-Wesley, fifth ed., 2009.

[31] A. Goldsmith, *Wireless Communications*. Cambridge University Press, 2005.

[32] T. Henderson, "Radio propagation models." Available at http://www.isi.edu/nsnam/ns/doc/node216.html, Nov. 2011.

[33] V. Gungor, B. Lu, and G. Hancke, "Opportunities and challenges of wireless sensor networks in smart grid," *IEEE Transactions on Industrial Electronics*, vol. 57, pp. 3557–3564, Oct. 2010.

[34] B. Lichtensteiger, B. Bjelajac, C. Muller, and C. Wietfeld, "RF mesh systems for smart metering: System architecture and performance," in *Proc. 2010 First IEEE International Conference on Smart Grid Communications (SmartGridComm '10)*, pp. 379–384, Oct. 2010.

[35] A. Leon-Garcia and I. Widjaja, *Communications Networks: Fundamental Concepts and Key Architectures*. McGraw-Hill, 2nd ed., 2006.