11-2015

Effective Node Clustering and Data Dissemination In Large-Scale Wireless Sensor Networks

Mariam Saeed AI Nuaimi

UAEU

جامعة الإمارات العربيـة المتحدة
United Arab Emirates University

United Arab Emirates University

College of Information Technology

EFFECTIVE NODE CLUSTERING AND DATA DISSEMINATION IN LARGE-SCALE WIRELESS SENSOR NETWORKS

Mariam Saeed Al Nuaimi

This dissertation is submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

Under the Supervision of Professor Khaled Shuaib

November 2015

Declaration of Original Work

I, Mariam Alnuaimi, the undersigned, a graduate student at the United Arab Emirates University (UAEU), and the author of this dissertation entitled "*Effective Node Clustering and Data Dissemination in Large-Scale Wireless Sensor Networks*", hereby, solemnly declare that this dissertation is my own original research work that has been done and prepared by me under the supervision of Professor Khaled Shuaib, in the College of Information Technology at UAEU. This work has not previously been presented or published, or formed the basis for the award of any academic degree, diploma or a similar title at this or any other university. Any materials borrowed from other sources (whether published or unpublished) and relied upon or included in my dissertation have been properly cited and acknowledged in accordance with appropriate academic conventions. I further declare that there is no potential conflict of interest with respect to the research, data collection, authorship, presentation and/or publication of this dissertation.

Student's Signature: ______________ Date: 8th Dec 2016

Approval of the Doctorate Dissertation

This Doctorate Dissertation is approved by the following Examining Committee Members:

1) Advisor (Committee Chair): Khaled Shuaib
   Title: Professor
   Information Security Track
   College of Information Technology
   Signature ____________ Date 26/11/2015

2) Member: Farag Sallabi
   Title: Associate professor
   Networking Track
   College of Information Technology
   Signature ____________ Date 26/11/2015

3) Member: Mohammad Hayajneh
   Title: Associate professor
   Computer Systems Design Track
   College of Information Technology
   Signature ____________ Date 26-11-2015

4) Member (External Examiner): Ala Al-Fuqaha
   Title: Professor
   Department of Computer Science
   College of Engineering & Applied Sciences
   Institution: Western Michigan University
   Signature ____________ Date 11/26/2015

This Doctorate Dissertation is accepted by:

Dean of the College of Information Technology: Professor Omar El-Gayar

Signature ____________ Date 25/4/16

Dean of the College of the Graduate Studies: Professor Nagi T. Wakim

Signature Nagi Wakim Date 13|12|2016

Copy 4 of 7

Advisory Committee

1) Advisor: Khaled Shuaib

Title: Professor

Information Security Track

College of Information Technology

2) Co-advisor: Mohammed Abdel-Hafiz

Title: Associate Professor

Department of Electrical Engineering

College of Engineering

3) Member: Nader Mohamad

Title: Associate Professor

Networking Track

College of Information Technology

Abstract

The denseness and random distribution of large-scale WSNs makes it quite difficult to replace or recharge nodes. Energy efficiency and management is a major design goal in these networks. In addition, reliability and scalability are two other major goals that have been identified by researchers as necessary in order to further expand the deployment of such networks for their use in various applications. This thesis aims to provide an energy efficient and effective node clustering and data dissemination algorithm in large-scale wireless sensor networks. In the area of clustering, the proposed research prolongs the lifetime of the network by saving energy through the use of node ranking to elect cluster heads, contrary to other existing cluster-based work that selects a random node or the node with the highest energy at a particular time instance as the new cluster head. Moreover, a global knowledge strategy is used to maintain a level of universal awareness of existing nodes in the subject area and to avoid the problem of disconnected or forgotten nodes. In the area of data dissemination, the aim of this research is to effectively manage the data collection by developing an efficient data collection scheme using a ferry node and applying a selective duty cycle strategy to the sensor nodes. Depending on the application, mobile ferries can be used for collecting data in a WSN, especially those that are large in scale, with delay tolerant applications. Unlike data collection via multi-hop forwarding among the sensing nodes, ferries travel across the sensing field to collect data. A ferry-based approach thus eliminates, or minimizes, the need for the multi-hop forwarding of data, and as a result, energy consumption at the nodes will be significantly reduced. This is especially true for nodes that are near the base station as they are used by other nodes to forward data to the base station. MATLAB is used to design, simulate and evaluate the proposed work against the work that has already been done by others by using various performance criteria.

Keywords: Clustering protocols; wireless sensor networks; load balancing; routing protocols; energy efficiency protocols; ferry protocol.

Title and Abstract (in Arabic)

التقسيم ونقل البيانات الفعال في الشبكات الحساسة الضخمة

الملخص

التوزيع الكثيف والعشوائي لحساسات الشبكات الحساسة الضخمة يجعل من الصعب استبدال أو شحن الحساسات المتهالكة مما جعل ادارة الطاقة بكفاءة أحد أهم الأهداف في تصميم هذه الشبكا .. اضافة لذلك، لموثوقية وقابلية التوسع تعتبر كذلك من الأهداف المهمة التي مازالت مستهدفة من قبل الباحثين. لذلك في هذا البحث نهدف لتقديم نظام لتقسيم الأجهزة الحساسة ونقل المعلومات ذا فعالية و كفاءة في حفظ الطاقة في الشبكات الحساسة الضخمة. حيث أن هذا النظام المقترح سيقوم بحفظ الطاقة عن طريق تصنيف الأجهزة لإختيار الحساس الأصلح للقيام بتجميع المعلومات من باقي الحساسات في المجموعة على عكس الأنظمة الموجودة مسبقا التي تختار الحساس المجمع عشوائيا. بالإضافة لذلك فنظامنا يقوم باستخدام ميزة المعلومات العامة لتوفير قاعدة معلومات عن الحساسات الموجودة في الشبكة للتأكد من عدم وجود حساس معزول أو منسي ووجود تغطية كافية لكافة الحساسات الموجودة في المنطقة المغطاة. كذلك يوفر النظام حلول لتقليل المعلومات المرسلة لقاعدة البيانات وزيادة كفاءتها عن طريق استخدام نظام العبارة المجمعة المعاد شحنها و استخدام تقنية التنبيه الاختياري و تناوب الحساسات على تغطية منطقة معينة وذلك لتقليل ارسال معلومات مكررة. سنقوم باستخدام الماتلاب لتصميم وتنفيذ ومقارنة أداء نظامنا مع الانظمة الموجودة مسبقا من خلال استخدام عدة أوجه للإثبات كفاءة النظام.

مفاهيم البحث الرئيسية: أنظمة التقسيم، شبكة الأجهزة الحساسة، أنظمة حفظ الطاقة، أنظمة العبارة المجمعة.

Acknowledgements

First, all thanks be to God for giving me the strength, the patience, and the blessing in time and health to complete this dissertation. Then, I would like to express my deepest gratitude and appreciation to my advisor, Dr. Khaled Shuaib: you have been a great mentor and my role model during my undergraduate and graduate studies. I would like to thank you for your insightful comments, valuable advice, and continuous encouragement. Above all, thank you for believing in me. I would also like to thank my advisory committee members, Dr. Mohammed Abdel-Hafez, and Dr. Nader Mohamed for serving on this committee: thank you for your bright comments and suggestions which improved my dissertation tremendously. I also would like to thank the examiners of my thesis Prof. Ala Al-Fuqaha from Western Michigan University, Dr. Farag Sallabi and Dr. Mohammad Hayajneh from the College of IT, UAEU.

I would like to thank UAE University, the College of Information Technology, and the College of Graduate Studies for providing me with a scholarship and the opportunity to pursue and complete my PhD studies.

Finally, an especial thanks goes to my family: without you, I would not be who I am now.

Dedication

To the women of the world and to humanity

Table of Contents

List of Tables

List of Figures

List of Acronyms and Abbreviations

WSN	Wireless Sensor Networks
BS	Base Station
LEACH	Low Energy Adaptive Clustering Hierarchy
PEGASIS	Power-efficient Gathering in Sensor Information Systems
HEED	Hybrid Energy Efficient Distributed Clustering
SEP	Stable Election Protocol
MNUC	Mixed Unequal Clusters Size algorithm
HGMR	Hierarchical Geographic Multicast Routing
EADC	Energy-Aware Dynamic Clustering algorithm
ULEACH	Universal LEACH
T-LEACH	Threshold LEACH
MCH	Master Cluster Head
HABRP	Hierarchical Balanced Energy Efficient Routing Protocol
En	Energy Level
Dn	Distance from Base Station
CCH	Candidate Cluster Head
T	Energy Threshold
CH	Cluster Head
NRCA	Node Ranking Clustering Algorithm
TOA	Time of Arrival
UWB	Ultra-Wide Band
WAMR	Wireless Automatic Meter Reading
FNRCA	Ferry Node Ranking Clustering Algorithm
ST	Stopping Time
CP	Checkpoint

List of Symbols

X_i	X position of node i
Y_i	Y position of node i
$Closest_{CP}$	Closest checkpoint
$En\ (i)$	Residual Energy of node i
$D(i)$	Distance of node i
Average(Dn)	Average distance of all nodes from the base station
Average(En)	Average energy of nodes in the network
T(i)	Energy threshold on node i
r	Sensing radius
N	Total number of sensor nodes in the network
N_j	Total number of attached cluster heads to checkpoint j
n	Total number of cluster heads
m	Total number of checkpoints in the network
$BufSize$	Cluster head memory in bits
$NumberOfAttachedCHs$	Number of attached cluster heads associated
$TimeToTransmitAbit$	Time needed to transmit a bit of information
CHs$[CH_i, i = 1,..,n]$	A set of cluster-heads
CPs$[CP_j, j = 1,..,m]$	A set of checkpoints
CPS	A list of checkpoints
CHS	A list of cluster heads
N_T	Total number of cluster heads in the network
W_j	Weight of checkpoint j
E_o	Initial energy of each node
E_{elec}	Per bit energy consumption
E_{DA}	Energy for data aggregation
E_{amp}	Amplifier transmitting energy
Tmax	Time of the longest tour of the ferry
s	Round index of data collection
Tour_Time(s)	Total round trip =Traveling time + the stopping time
Travel_Time	Traveling time from the base station and coming back
l	Checkpoint index
k	Cluster head index
Ferry_Speed	Assumed fixed speed of the ferry

Chapter 1: Introduction

1.1. Overview

Recent improvements in electronic hardware technology have enabled manufacturers to develop low cost, low power, and small-sized motes [1, 2, 3, and 4]. Hundreds and thousands of these motes are deployed as wireless sensor networks (WSNs) serving many applications based on the specific requirements of each one [1, 5]. A diverse set of applications for sensor networks encompassing different fields has already emerged in areas including medicine, agriculture, environment, military, inventory monitoring, intrusion detection, motion tracking, machine malfunction, toys, and many others.

The denseness and random distribution of WSNs make it quite difficult to replace or recharge nodes' batteries, especially in applications such as: disaster recovery areas, environment monitoring, border monitoring, battlefields, underwater sensing, oil fields, and many others. Therefore, energy efficiency and management is a major design goal in these networks. In addition, reliability and scalability are two other major goals that have been identified by researchers in order to further expand the deployments of such networks for their use in applications requiring these features such as the military, environment and healthcare. Node clustering strategies and effective data collection and dissemination mechanisms within a WSN are considered major factors which affect the achievement of the main goal of prolonging the network lifetime while maintaining proper coverage and ensuring reliable data collection. The purpose of this study is to propose node clustering strategies and effective data collection and dissemination mechanisms within a WSN to prolong the network lifetime while maintaining proper coverage and reliable data collection.

1.2. Research Motivation

Wireless sensor networks, powered by batteries, are currently deployed for data gathering and application management in a wide range of areas. In most cases, the networks are dense, sometimes large-scale, and randomly distributed which makes it quite difficult to replace or recharge the batteries, especially when they are used in applications such as disaster recovery, environment monitoring, border monitoring, battlefields, underwater sensing, oil fields, and many others. Therefore, achieving the energy efficiency and management of WSNs is considered a major research goal. Providing efficient clustering, data gathering, and dissemination techniques to prolong the lifetime of WSNs implies better and less expensive management of such networks. In addition, reliability and scalability are two other major goals of researchers who aim to further expand the use of WSNs in applications requiring these features, such as the military and healthcare.

1.3. Research Problem

In WSNs, sensor nodes collect and aggregate data through the network to a repository system through the base station (sink) for further use and analysis. Data processing and wireless data transmission/reception are the two main energy-consuming tasks performed by the sensor nodes which have limited energy that is supplied by on-board batteries. Therefore, to increase the lifetime of a wireless sensor network, energy conservation is a key challenge that must be overcome, especially for large-scale and dense networks. This research investigates how to prolong the network lifetime of a large-scale WSN by conserving as much energy as possible in the deployed nodes while maintaining proper coverage and effective dissemination and collection of data for real-time and delay tolerant applications.

1.4. Research Objectives

The research proposed as a part of this work is composed of two main areas that are interconnected: clustering and data dissemination in large-scale wireless sensor networks. This research attempts to address the main problems of prolonging the network lifetime and maintaining proper coverage. In the area of clustering, our research aims to provide an effective and novel clustering scheme which is shown to improve on the existing approaches by providing longer lifetime and better area coverage. The other area of this research is focused on incorporating an efficient data dissemination technique to complement the clustering scheme and improve on the network lifetime while maintaining coverage for effectively collecting sensed data.

1.5. Research Contribution

The main contribution of the thesis can be summarized as follows:

- Research the existing WSN clustering algorithms, implement them in MATLAB and evaluate their performance using different criteria like network lifetime and consumed energy, and by varying the number of nodes and changing the placement of the base station.
- Propose and design a new energy efficient clustering algorithm to improve the network lifetime of WSNs by applying a new mechanism for cluster head selection and rotation which helps to reduce energy consumption and extend

node lifetime. Additionally, incorporate a duty-cycle technique in the design of the algorithm.

- Propose and design a new energy efficient data collection algorithm for delay tolerant applications through the use of a mobile ferry to collect data. Using a mobile ferry to collect data further preserves energy by reducing multi-hop forwarding. This in turn minimizes the energy consumed in the network when collecting and transferring data to the BS. In this algorithm, the area is divided into virtual grids and in each grid there is a checkpoint (stopping point) where the ferry stops and collects data from the cluster heads. In order to optimize the ferry's path, a weight is assigned to each checkpoint in order to choose the best sequence, the order of the checkpoints to be visited, and the required stopping time at each one. This eliminates a loss of messages due to incorrect predictions of the positions of the ferry or its movement.
- Implement the proposed algorithms in MATLAB, validate their performance through simulation, and compare their results to other well-known algorithms. Our work is shown to outperform other existing approaches in terms of the network lifetime and energy consumed. Moreover, our proposed algorithms achieved better reliability by incorporating effective data dissemination techniques which improve the performance further and help satisfy the requirements of certain applications of interest.

1.6. Thesis Structure

The rest of the thesis is organized as follows:

Chapter 2 provides an overview of WSNs, their characteristics, architectures, applications, and current research projects. Some parts of this chapter have previously been published in:

- Mariam Alnuaimi, Farag Sallabi, Khaled Shuaib, "A survey of Wireless Multimedia Sensor Networks challenges and solutions," Proceedings of the IEEE IIT'12, 25-27 April 2011, Abu-Dhabi, UAE.

Chapter 3 highlights the challenges in clustering a large-scale WSN, gives an in-depth literature review of the current existing algorithms in the area of clustering and classifies them based on the cluster's technique formation and the way that data is aggregated to the base station. Moreover, it shows the performance evaluation of these algorithms using different scenarios. The main contents of this chapter have been published in the following conference paper:

- Mariam Alnuaimi, Khaled Shuaib, Klaithem Alnuaimi, Mohammed Abdel-Hafez, "Performance analysis of clustering protocols in WSN," Proceedings of the Wireless and Mobile Networking Conference (WMNC), 2013 6th Joint IFIP, pp. 1-6, Dubai, UAE, April 22-24 2013.

Chapter 4 gives a detailed description of a proposed energy efficient clustering algorithm for WSNs using node ranking in electing cluster heads and thresholds to replace them. It provides a comparison of the performance of the proposed algorithm against two well-known algorithms in terms of network lifetime. The simulation demonstrates how the proposed algorithm outperformed other well-known algorithms in terms of the network lifetime and energy consumed. The contributions of this chapter have been previously published in the following publications:

- Mariam Alnuaimi, Khaled Shuaib, Klaithem Alnuaimi, Mohammed Abed-Hafez, "Clustering in Wireless Sensor Networks based on node ranking," in 2014 IEEE International Wireless Communications and Mobile Computing Conference (IWCMC), pp. 488-493, Nicosia, Cyprus, 2014.
- Mariam Alnuaimi, Khaled Shuaib, Klaithem Al Nuaimi, "Clustering in WSN using node ranking with hybrid nodes duty-cycle and energy threshold," in Proceedings of the 2014 IEEE 13th International Symposium on Network Computing and Applications, pp. 245-252. IEEE Computer Society, Cambridge, MA, USA, August 2014.
- Mariam Alnuaimi, Khaled Shuaib, Klaithem Alnuaimi, Mohammed Abed-Hafez, "An efficient clustering algorithm for wireless sensor networks," International Journal of Pervasive Computing and Communications 11, no. 3 (2015): 302-322, August 2015.

Chapter 5 surveys the recent progress made in using mobile ferries for data gathering in WSNs by addressing two areas: 1) determining the path of the ferry, and 2) the scheduling for dispatching the ferry to collect data from static sensors. It presents a classification of mobile ferries based on the role they play in addition to carrying information. Furthermore, it discusses the challenges in deploying mobile ferries in WSNs along with many of their possible applications. The main contents of this chapter have been published in the following conference paper:

- Mariam Alnuaimi, Khaled Shuaib, Klaithem Al Nuaimi, Mohammed Abdel-Hafez, "Data gathering in Wireless Sensor Networks with ferry nodes," 12th IEEE International Conference on Networking, Sensing and Control (ICNSC15), pp. 221-225, Taiwan, Taipei, April 2015

Chapter 6 gives a description of the newly proposed efficient data collection algorithm using ferry node to collect data from nodes of WSN based on a ferry's predetermined or fixed path selection. In this algorithm, the decision to select cluster heads is based on their residual energy and their distance from the ferry path. It also surveys the recent progress made by using mobile ferry nodes for data gathering in WSNs. It shows a simulation of the performance of the proposed algorithm in terms of network lifetime and the overall energy consumption of the network per round by using different ferry path scenarios and by changing the number of checkpoints. The models and results of this chapter have been published in the following conference paper:

- Mariam Alnuaimi, Khaled Shuaib, Klaithem Alnuaimi, Mohammed Abdel-Hafez, "Ferry-based data gathering in Wireless Sensor Networks with path selection," in the 6th International Conference on Ambient Systems, Networks and Technologies (ANT 2015), Procedia Computer Science 52 (2015): 286-293, London, UK, June 2015.

Chapter 7 gives a description of the proposed efficient data collection algorithm using ferry node to collect data from nodes of WSN based on a ferry's path selection. Two goals are set out in this algorithm: minimizing the overall round trip travel time of the ferry and minimizing the overall energy consumed in the whole network. The results of the simulation on the efficiency of the proposed algorithm compared to algorithms presented in the recent literature are provided in this chapter by using different evaluation criteria. The results of this chapter have been published in:

- Mariam Alnuaimi, Khaled Shuaib, Klaithem Alnuaimi, Mohammed Abdel-Hafez, “Data gathering in delay tolerant Wireless Sensor Networks using a ferry,” Sensors 15, no. 10 (2015): 25809-25830, October, 2015.

Finally, Chapter 8 concludes the thesis, and proposes some areas for further research and study.

Chapter 2: Introduction to WSNs

2.1. Overview

Recent improvements made in electronic hardware technology enabled manufacturers to develop low cost, low power and small size sensors. Hundreds and thousands of these sensors are deployed as wireless sensor networks (WSN) serving many applications based on the specific requirements of each one. A diverse set of applications for sensor networks encompassing different fields have already emerged, including medicine, agriculture, environment, military, inventory monitoring, intrusion detection, motion tracking, machine malfunction, toys, and many others.

In general, a wireless sensor network is a collection of nodes with sensing, computation, and wireless communication capabilities. These nodes, or motes, communicate with each other by forming a network of nodes and maintaining connectivity in a distributed way as shown in Figure 2-1. The distributed sensor nodes also communicate with the sink node through the gateway. There are two types of WSNs when it comes to deployments: structured WSN and unstructured or ad hoc WSNs. When deploying a structured WSN, the location and number of sensor nodes is planned beforehand. It is easy to control and maintain a structured WSN because the details of each sensor node are available. However, an unstructured WSN is composed of a number of sensor nodes that are deployed in an ad hoc manner into an area of choice. In such an environment, network maintenance, such as managing connectivity and detecting failures, might be difficult due to the large number of deployed nodes and the large coverage area. However, such deployments are critical to have in certain harsh environments where the deployment of pre-planned (structured) networks can be difficult, if not impossible. The advantage of a structured network is that fewer nodes can be deployed with less network maintenance and lower management cost.

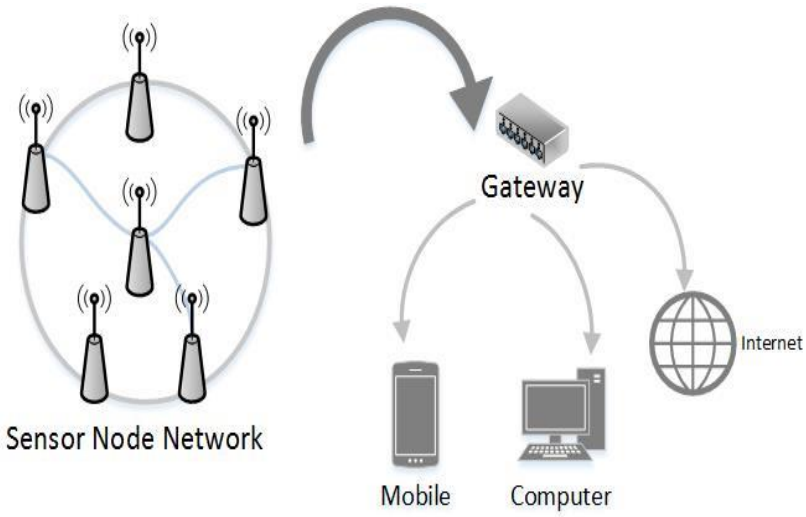

Figure 2-1: WSN overview

In this chapter, I provide an overview of wireless sensor networks in general. First, I discuss the main characteristics of sensor nodes within the WSN in Section 2.1. Then, in Section 2.2, I provide a brief discussion of the most well-known data transmission technologies within WSNs and compare them according to their transmission speed, frequency, bandwidth, and coverage. Section 2.3 highlights some examples of the applications of WSN. Section 2.4 demonstrates the three main architectures of a WSN. I also discuss some of the most recent research projects on wireless sensor networks and detail the areas of research in Section 2.5. Finally, Section 2.6 concludes this chapter.

2.2. Characteristics of WSNs

In this section, I will discuss the different characteristics of sensor nodes within a wireless sensor network. Sensor nodes must adapt to the environment in which they are deployed. Consequently, they have certain characteristics that ought to be considered when designing any sensor node. For example, since sensors cannot be recharged often, they must conserve their battery power for as long as possible. Moreover, they must organize themselves whenever a change occurs in their surroundings. The following sections provide more details on the important characteristics of the sensor nodes used in a WSN.

2.2.1. Self-Organized

When deployed in large quantities in a sensing field, sensors can automatically organize themselves to form an ad hoc multi-hop network to communicate with each other and with sink nodes. Typically, a WSN has one or more sinks (or base stations) that collect data from sensors within the WSN. These sinks are considered the gateways through which a WSN interacts with the outside world.

2.2.2. Energy and Memory Limitations

Sensor nodes have limited energy or battery life. This is due to the size of the sensor nodes as well as the environment into which they are deployed. Typically, a wide WSN will be difficult to maintain and therefore difficult to recharge by humans because it is deployed in areas to which they have limited access (for example, a battlefield, underground, or underwater). This means that energy consumption is an important aspect to consider when designing sensor nodes. Most WSNs are designed to conserve energy for as long as possible since their nodes will rarely be recharged or maintained, especially for ad hoc WSNs. The same applies to incorporated memory since the size of each sensor node does not allow for the inclusion of a large storage unit. Therefore, the amount of information stored is kept to a minimum and is relayed as soon as there is a chance to do so [1].

2.2.3. Heterogeneity of Nodes

In many applications, WSNs consist of different types of sensors. Sensor types, such as acoustic, proximity, position, pressure, optical, and many other types of sensors, are specific for sensing an input and communicating data to other sensors or to the base station for decision making purposes [2]. A WSN, in many scenarios, consists of multiple types of sensors as each one senses the attribute from which it is expected to collect data and all the data is collected by the base station. From there, the decision is carefully made based on the various data collected from the different types of deployed nodes.

2.2.4. Mobility of Nodes

Since sensor nodes are deployed in large quantities over a broad area, they may change their locations after their first deployment. This change may result from environmental variables, such as wind or water, or it can be due to the movement of the object to which the sensor nodes are attached or carried, such as a human or an airplane. Therefore, mobility can either be subsequent to an effect, or it can be a requirement of an application. Thus, sensor nodes usually have the ability to move from one location to another without affecting the data that is collected and communicated by the node itself [3].

2.2.5. Scalability

In simple terms, scalability usually refers to the ability to grow or expand without changing the original architecture or performance. As for a wireless sensor network, its scalability is demonstrated by its ability to grow or expand in terms of adding new nodes, new sensed data, and new methods of analyzing data without tremendously affecting the cost and without even the need to change the structure of the already deployed WSN. The scalability feature of wireless sensor networks enables them to adjust to the changes required by the application or the sensing field in the simplest manner possible [4].

2.2.6. Hard to Maintain or Manage in Case of Failure

The denseness and random distribution of WSNs make it quite difficult to replace or recharge nodes' batteries, especially in applications such as: disaster recovery areas, environment monitoring, border monitoring, battlefields, underwater sensing, oil fields, and many others. Therefore, energy efficiency and management are major design goals in these networks. In addition, reliability and scalability are two other major goals that have been identified by researchers as necessary in order to further expand the deployment of such networks in areas of application requiring these features, such as the military, environment, and healthcare.

2.3. Technology

The technology used in wireless sensor networks is usually associated with the sensor node itself. WSN technologies include, among others, Zigbee, Ultra-wideband, Bluetooth, Z-wave, and RFID. In this section, I will briefly discuss each technology, and compare the four technologies based on the range they cover, the frequency they use, their speed, and bandwidth.

2.3.1. Zigbee Technology

Zigbee is a multi-hop forwarding technology used by Zigbee Alliance which uses IEEE standard 802.15.4 [5]. Zigbee transfers data by forwarding packets from one node to another until reaching the target base station where the data must be collected. When using Zigbee, the nodes are either routers or leaf nodes. Router nodes transfer data from the children to their parents or to the destination, while leaf nodes can only transfer data to their parents. The benefits of Zigbee are its low cost, and the simplicity of its data transfer methodology.

2.3.2. Ultra-WideBand Technology

Ultra-wideband (UWB) technology sends very short pulses in a very short time and therefore requires large bandwidth for its transmission. Ultra-wideband technology is not easily blocked by obstacles such as walls and human bodies, which has led many solution providers to use it when building their wireless sensor networks. UWB sends short pulses in a short amount of time, therefore, estimating the Time of Arrival (TOA) of each pulse is more accurate than estimating the TOA for a large packet which might be lost during transmission. UWB is renowned for its accurate indoor positioning since it usually covers a short range of signal [6].

2.3.3. Bluetooth Technology

Bluetooth is a short-range transmission technology. Despite its short range of coverage, Bluetooth technology can easily be adopted by many systems, it is widely embedded in different devices such as mobile phones, laptops, sensors, and other devices that allow different types of devices to communicate [7]. When using Bluetooth technology, each sensor will have its own unique tag by which it can send and receive data. Bluetooth is not blocked by metallic objects as UWB is.

2.3.4. RFID Technology

RFID is an electromagnetic transfer of radio frequencies. Each node has an ID tag by which it can send and receive data through the network. RFID tags are usually recognized by RFID readers which can read the data transmitted by the tag. RFID tags are small and lightweight. However, they cover a very small range of only one to two meters [8]. RFID is a popular technology for use in applications such as tracking and identifying items.

2.3.5. Comparison of WSN Technologies

Table 2-1 shows a comparison between the WSN technologies mentioned earlier in this chapter. The table shows that Ultra-wideband has the highest speed and the highest bandwidth among the four technologies mentioned. However, Zigbee covers a wider range when communicating between sensors [9, 10]. Therefore, it requires fewer sensors to control the same area as any other technology. RFID, on the other hand, has the shortest range among the technologies under comparison. However, RFID is only used in certain cases, as I mentioned earlier, such as to track items within a building. IEEE provides a standard for each of the technologies under 802.15 and each technology uses a different standard depending on its needs.

Table 2-1: WSN technologies comparison

	Zigbee	**UWB**	**Bluetooth**	**RFID**
IEEE Standard	802.15.4	802.15.3a	802.15.1	802.15.4f
Speed (Mbps)	10	40-60	1-24	5
Bandwidth (Mhz)	1 – 2	>500	1	2
Frequency (Ghz)	2.4	3.1 – 10.6	2.4	2.45 – 5.8
Range (m)	10-20	10	1-100	1 – 2

2.4. Recent WSN Applications

WSN applications are evolving every day for the purposes of information gathering in order to better monitor and control the components that they manage. In the following section, I will discuss some of the recent applications of WSNs.

2.4.1. Smart Power Grid Systems

A smart power grid is an efficient and reliable automation service for electricity flow and is one of the recent applications of WSNs. WSNs are used to capture and analyze data related to power usage, power delivery, power generation, and power disturbances and outages. Sensors are used to identify energy usage frequency, phase angle, and the values of voltage to help utility companies manage electricity in an efficient way. Wireless automatic meter reading, or WAMR [11], is an example for such an application. WAMR collects customers’ real time energy consumption and provides them with archived readings. It can also control lights, air conditioners, heaters, and other devices within a building to help customers manage their electricity usage in an efficient way.

2.4.2. Smart Habitat Monitoring

Ecologists study the origins, migration patterns, behaviors, diseases, life processes and the environment inhabited by wildlife. Habitat monitoring applications provide ecologists with data on relevant environmental conditions, such as weather, that affect avian migration, for example. They are used to help settle large-scale land use disputes affecting animals, plants, and people [12]. The authors in [13] proposed an approach for monitoring the activities of birds in order to track 350 species of exotic birds migrating from Siberia to India overwinter. They implemented a habitat monitoring system in which sensors were attached to the bodies of the birds in order to track each bird's activity and make a record of it.

2.4.3. Smart Cloud

Cloud computing has gained a great deal of attention in recent years due to its wide deployment and the services that it offers. A cloud service implies the use of the Internet as a large repository or as a workspace. People can access the Internet anytime and anywhere. In [14], the authors proposed an intelligent smart cloud model. This model provides customized services to users by personalizing the content through smart processing based on the user's behavior. In this model, aspects of the users' behavior were collected by sensors mounted on their devices, such as mobile phones and tablets.

2.4.4. Smart Healthcare Delivery

Smart healthcare delivery applications are used for patient monitoring and care in remote sites. For example, images of a patient's facial expressions, respiratory conditions, or movements can be taken and forwarded to specialists at other hospitals that are far away in order for the remote doctor to make a better diagnosis. In [15], a healthcare sensor periodically captures information on vital signs (e.g., body temperature, blood pressure) and sends it to a gateway device. Once the information has been processed by the gateway, it is forwarded to doctors to help them make an initial diagnosis.

2.5. WSN Architectures

Different architectures were proposed to show how WSNs can be more scalable and more efficient, depending on the specific application Quality of Service (QoS) requirements and constraints [16]. Therefore, based on the designed network topology and architectures, the available resources in the network can be efficiently utilized and fairly distributed throughout the network, and the desired operations of the content can be handled. In general, network architectures for WSNs can be divided into three different groups, as mentioned in [17, 18, 19 and 20] and outlined below are composed of several components, which include: video and audio sensors, scalar sensors, multimedia processing hubs, storage hubs, sink, and the gateway.

2.5.1. Single-tier Flat Architecture

In this architecture, the network consists of homogeneous sensor nodes with the same capabilities and functionalities. All nodes can perform any function, such as sensing certain attributes, image capturing, multimedia processing, and transferring data to a sink over either a single-hop or a multi-hop path through transmission nodes, not cluster heads [21, 22], as shown in Figure 2-2.

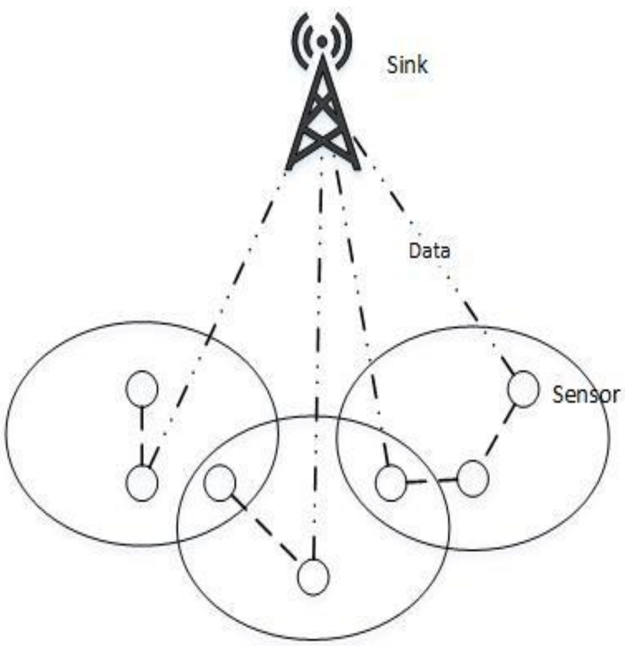

Figure 2-2: Single-tier flat architecture

2.5.2. Single-tier Clustered Architecture

Single-tier clustered architecture consists of heterogeneous sensors, such as camera, audio, and scalar sensors, that are grouped together to form a cluster. All heterogeneous sensors belonging to the same cluster send their sensed data to the cluster head, which has more resources and can perform complex data processing. The cluster head is connected either directly or indirectly to the sink or the gateway through a multi-hop path, as shown in Figure 2-3 [20, 21, 22 and 23].

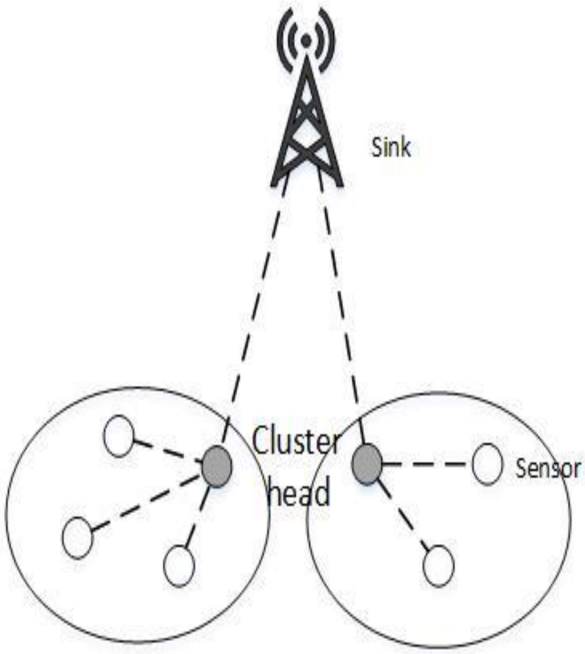

Figure 2-3: Single-tier cluster architecture

2.5.3. Multi-tier Architecture

In this architecture, the first tier consists of scalar sensors that perform simple tasks, like measuring scalar data from the surrounding environment (e.g., light, temperature, etc.), while the second tier consists of camera sensors that perform more complex tasks, such as image capturing or object recognition. The third tier consists of more powerful and higher resolution video camera sensors that are capable of performing more complex tasks, like video streaming or object tracking [24]. Each tier has a central hub for data processing and communicating with the upper tier. The third tier is connected with the sink or the gateway through a multi-hop path [25, 26], as shown in Figure 2-4.

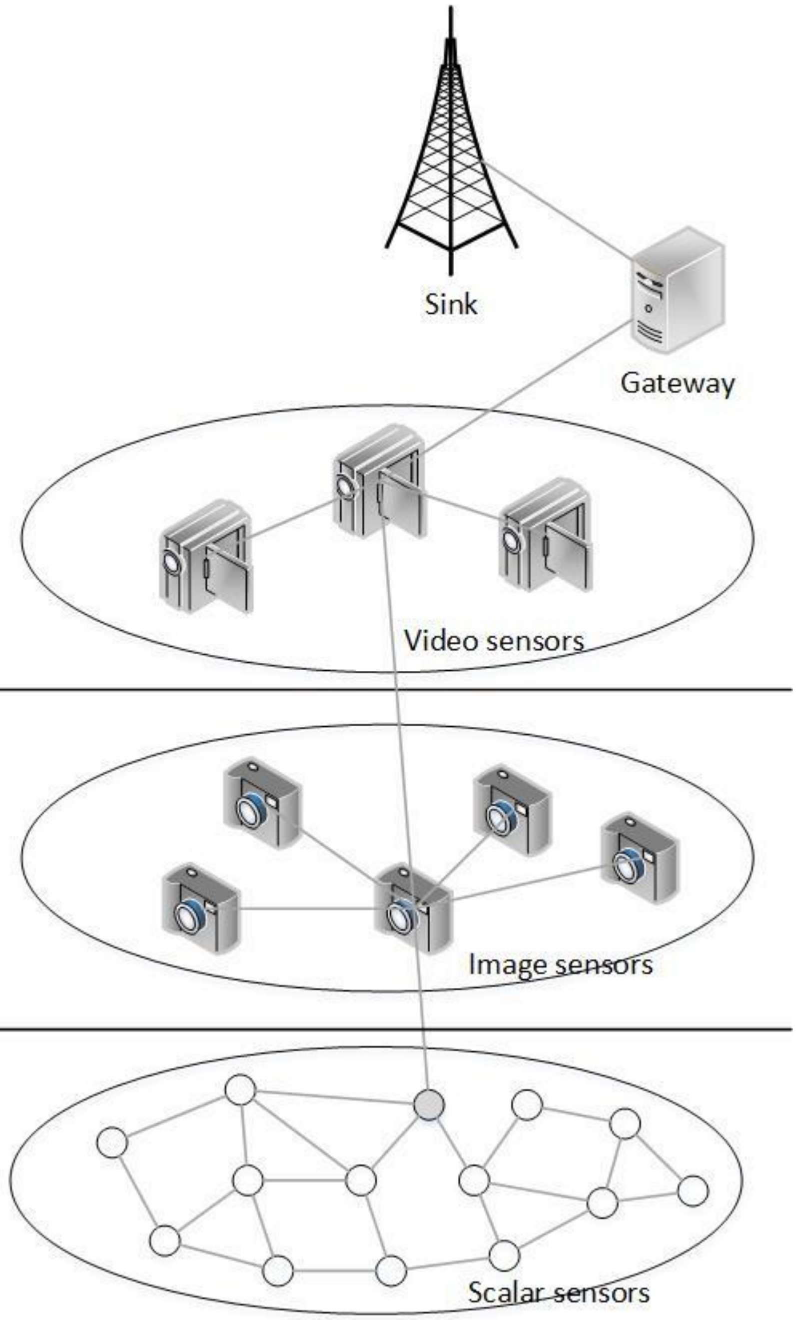

Figure 2-4: Multi-tier WSN architecture

2.6. Current Research Projects

In this section, I will discuss three current research projects which focus on wireless sensor networks.

First, in the area of visualization techniques in wireless sensor networks, many research papers have been published in order to stress the importance of this area [27, 28 and 29]. This research provides important benefits to the field, such as by being able to geospatially locate every sensor node within the sensor network and having better knowledge of its energy level, sensed data, and its location with respect to the collected data. In such cases, a Geographic Information System (GIS) is used in order to visualize the scattered sensors through the covered area. The sensors send their location data to the base station using multi-hop forwarding, in the case of outdoor sensors, while other techniques are used in the case of indoor localization, such as inertial measurements of building 2D maps of the location using the direction (angle) and speed from a starting point of movement of the sensor. Data, such as battery usage and sensed data, are sent along with the location data of the sensors to the controller station where the GIS displays a map of all the sensors and the details of each sensor are displayed once the user selects the sensor. The number of sensors usually ranges from 100 to 1000 sensors and the network size that is used is 600 × 400 m [27].

Second, seminal research has already been published on the issue of the security and efficiency of the data being transmitted through sensor networks [30, 31]. For example, data related to the military, disaster zones, and medicine must be efficiently and securely transmitted to the target control center where decisions must be made accurately and in a timely fashion. Therefore, any loss or delay of such data caused by malicious attacks would have a greatly negative effect on the decisions that need to be made. In such research, a Secure Efficient Data Transmission protocol (SET) is used, as well as an Identity-Based Signature (IBS) mechanism. The goal of the combined solution is to enhance the transmission security computation while not affecting the transmission speed. The solution was evaluated using a different number of sensor nodes, ranging from 100 to 1000 nodes, and by dynamically changing the cluster heads.

My final example is related to research focusing on efficient routing techniques in WSNs and the decision making process [32, 33]. This is an important research area for wireless sensor networks since an efficient algorithm for data dissemination would save nodes' energy, provide faster access to the data in case of emergency, and reduce the overall effort exerted by the whole network. The research proposes a solution that provides the aforementioned features by using a mathematical model that focuses on decisions such as nodes' positions, scheduling, data transmission routes, and paths to sinks. The solution is based on Period Iteration Heuristics (PIH) and Sequential Assignment Heuristics (SAH) approaches [33]. The attributes used for the evaluation were a network lifetime of up to 10 hours and up to 300 different sensor locations [33]. The solution was ultimately found to increase the lifetime of the network.

2.7. Summary

In this chapter, I have reviewed Wireless Sensor Networks' characteristics, technologies, applications, and architectures according to the most up-to-date research published in the area. I have elaborated on each characteristic of a WSN, such as its heterogeneity, scalability, mobility, and energy and memory limitations. I have also provided examples taken from the latest research published in the applications of WSN. Moreover, I have shown the architectures that are used when planning or even deploying a WSN, such as multi-tiered architecture. Finally, I have elaborated on a few of the research projects currently being conducted in the area of WSN research.

Chapter 3: Literature Review: Clustering Algorithms

3.1. Background

Research related to WSNs is not new and several problems related to them have been exposed and addressed within the last few years. The research carried out in this area can be divided into three main categories: clustering algorithms, data dissemination techniques, and routing protocols [34, 35]. In this chapter, the focus of our attention is on clustering algorithms, therefore, the literature reviewed will mainly address this research area.

In [36, 37], a hierarchical clustering algorithm for sensor networks, called the Low Energy Adaptive Clustering Hierarchy (LEACH), was introduced. The idea of this algorithm is to form clusters of sensor nodes based on the received wireless signal strength. Local cluster heads are used by members of the cluster as routers to the sink. The intent of this approach is to reduce node energy consumption as the transmissions of gathered data to the sink will only be done by cluster heads rather than by all the sensor nodes. LEACH randomly selects a number of sensor nodes as cluster heads and then rotates this role among the nodes in order to uniformly distribute the energy load among the sensors in the network. Each elected cluster head broadcasts an advertisement message to the rest of the nodes in the network, informing them of its new role as cluster head. All the non-cluster head nodes, after receiving this message, choose the cluster to which they want to belong. This decision is based on the signal strength of the received advertised message.

LEACH uses single-hop routing where each node can transmit directly to the cluster head, which in turn transmits directly to the sink, regardless of the distance. This technique might work well in dense WSNs, but not in large-scale networks with large distances between the nodes, due to a direct proportional energy consumption relationship with distance. LEACH elects cluster heads randomly regardless of their energy level and thus, it is not suitable for networks deployed at a large scale. Furthermore, the idea of dynamic clustering brings extra overhead, e.g. head changes, advertisements etc., which may diminish any gain realized in energy consumption. It also assumes that nodes always have data to send, and that nodes located close to one another have correlated data.

In addition, it is not obvious how the number of predetermined cluster heads (CHs) will be uniformly distributed throughout the network. Therefore, there is a possibility that the elected CHs will be more concentrated in one part of the network than in other parts. As a consequence of this, some nodes will not have any CHs in their neighborhood and will not be covered. Finally, the protocol assumes that being a CH consumes approximately the same amount of energy for each node. In order to mitigate some of these problems, multi-hop LEACH was proposed in [38]. Multi-hop LEACH is another extension of the LEACH routing protocol to increase energy efficiency through the use of multi-hop forwarding to reach the base station of the wireless sensor network. Cluster heads receive data from all nodes at a single-hop and send it to the base station through intermediate cluster heads. However, some of the abovementioned problems are still considered open research issues and have not yet been resolved.

A Power-Efficient Gathering in Sensor Information Systems (PEGASIS) grid scheme approach [39] is a chain-based algorithm showing an improvement over the LEACH protocol. PEGASIS forms chains from sensor nodes instead of forming multiple clusters. Only one node is selected from that chain to transmit to the base station or sink. Gathered data moves from one node to other neighboring nodes, is aggregated, and eventually sent to the base station. Each node uses the signal strength to measure the distance to all the neighboring nodes, and then adjusts the signal strength so that only one node can be heard. Therefore, the chain will consist of those nodes that are closest to each other and form a path to the base station. The aggregated data will be sent to the base station by any node in the chain and the nodes in the chain will take turns sending data to the base station.

Unlike LEACH, PEGASIS avoids cluster formation and uses only one node in a chain to transmit to the base station instead of using multiple nodes, thus saving energy consumed by the rest of the nodes within the network. However, PEGASIS creates more delay for distant nodes on the chain, especially if the wrong direction to the base station is taken. Moreover, the chain leader can become a bottleneck for the whole chain and the approach also assumes that all nodes in the network are able to reach the base station.

In addition to the two studies above, several other issues have been recently considered by researchers with regard to large-scale WSNs. The authors in [40] proposed a mixed unequal clusters size algorithm (MNUC) to prolong the life of the network. This study addresses the problem of a hot spot where nodes have to do more processing and transmission-related work when compared to other parts of the network. Therefore, their energy will be drained more quickly than that of the other nodes. The idea of the algorithm is to form clusters with unequal sizes. Nodes closer to the base stations will be gathered into smaller sized clusters and nodes that are far away from the base stations will have larger cluster sizes. Nodes that are closer to the base station will be used more, but the transmissions' ranges will be less.

Universal LEACH (ULEACH) was proposed in [41] as an improvement over LEACH. The selection of cluster heads in ULEACH is based upon the initial and residual energy of nodes. Data is sent using a multi-hop approach from the farthest node to the cluster heads and from the cluster heads to the master cluster heads (MCH). This algorithm incorporates some features of Hybrid Energy Efficient Distributed Clustering (HEED) [42] and PEGASIS into LEACH. Although it utilized the multi-hop data transmission approach, it does not take into account the distance of the master cluster heads from the base station. Therefore, there might be more delay in delivering the data if the master cluster heads are far from the base station, which will also result in an additional transmission cost.

Threshold LEACH (T-LEACH) was proposed in [43] as an improvement on LEACH. It is a threshold-based cluster head replacement scheme for clustering protocols of wireless sensor networks. T-LEACH minimizes the number of cluster head selections by using a threshold of residual energy. However, it still uses the random head selection process of LEACH without specifying any criteria with which to choose cluster heads.

Despite recent achievements in these three areas of research, hurdles must still be overcome and these have attracted the attention of many researchers who are working on areas such as quality of service, security, energy harvesting, and prolonging the network lifetime by conserving energy on deployed nodes.

This chapter is organized as follows. The current challenges of clustering algorithms are discussed in Section 3.2. In Section 3.3, a performance evaluation of some well-known algorithms is shown. Section 3.4 provides a discussion of the simulation results and Section 3.5 summarizes the chapter.

3.2. Current Challenges of Clustering Algorithms

Based on the literature review discussed in Section 3.1, there are several challenges which need to be considered while clustering a large-scale WSN. These issues are summarized below.

3.2.1. Selection of Cluster Heads

After dividing a WSN into clusters, it is important to choose the best cluster head for each one. The optimal selection of the cluster head is the one that is reachable by all member nodes in the cluster, and will increase the lifetime and reliability of the network. There are several approaches that can be used for cluster head selection, such as selecting the node with the maximum current energy among the cluster members. Another method is to select the node which can be reached by all nodes using the least amount of energy. Moreover, it is necessary to alternate the role of cluster heads among the nodes to avoid overloading a few nodes with more responsibility than others and, in so doing, deplete their energy too quickly.

There are several approaches for cluster head rotation. One approach is to use a time stamp to initiate the process of electing another cluster head. Another approach is to use the remaining energy level to initiate the process of electing another cluster head. For example, a cluster head might trigger a new cluster head election process if its remaining energy level goes below a specified threshold. Frequent cluster head rotation results in more clustering overhead and network interruption. On the other hand, less frequent rotation may cause some nodes to die faster than others. The study of the optimal selection and rotation of cluster heads is essential for prolonging the lifetime of the network and increasing its reliability [44].

3.2.2. Cluster Size

Most existing clustering protocols assume a fixed cluster communication range in distance, which implies that all clusters have the same physical size. This assumption results in unfair load balancing where cluster heads that are closer to the observed event will carry more traffic and their energy will be drained faster than distant cluster heads. In [45], a larger cluster size is suggested to cluster heads that have less data to forward to distribute the load evenly among the cluster heads. However, this requires the nodes to know their locations based on the position of the event that occurred and the location of the base station. Selecting appropriate cluster sizes to minimize energy consumption within a WSN, not just based on the communication range, but by considering other factors such as the denseness of the WSN, the location of the base station, the application requirement with respect to reliability and the frequency of the data collection is still an area of research that is open to further investigation.

3.2.3. Ensuring Connectivity

Maintaining connectivity is an important objective of clustering protocols. Every node in a network must be a member of a cluster. It is recommended, insofar as it is possible, that all nodes within a cluster are able to communicate with their cluster head directly to avoid multi-hop forwarding, which usually results in less energy consumption. However, in certain cases, where the cluster size is larger than the communication range of nodes or when nodes have died due to the depletion of their energy, multi-hop communication cannot be avoided. To strike a balance between choosing the most appropriate cluster size while maintaining proper connectivity within each cluster, intra-cluster communication is used to indicate the success of the cluster formation. There is another type of connectivity called inter-cluster communication which describes the communication that takes place between different clusters. Two main approaches were proposed in the literature: relaying data through cluster heads and relaying data through gateways. In [19, 46 and 47], the nodes on the clusters' boundaries are used as gateways to relay data among the cluster heads (shown in Figure 3-1). Network density has to be sufficiently high in order to ensure that enough gateways are present at the intersection areas between clusters. On the other hand, in [48, 49], the cluster head relays data only through cluster heads (shown in Figure 3-1 as a dotted line). An advantage of the second relay approach is that it enables all non-cluster nodes to sleep while not sensing or transmitting data. Selecting efficient intra-cluster and inter-cluster transmission ranges to ensure connectivity and prolong the network lifetime is an important issue in clustering which needs to be considered when designing a clustering algorithm.

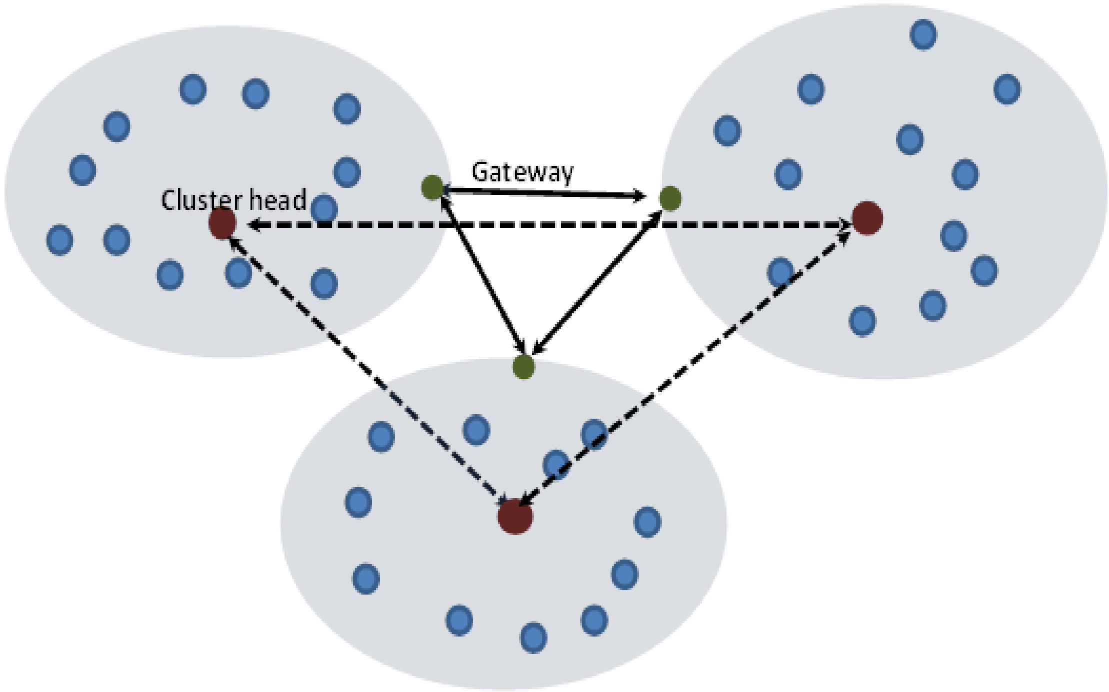

Figure 3-1: Routing via gateway nodes and cluster heads

3.2.4. Clustering the Network in the Presence of Duty-Cycle

Allowing sensors to sleep when they are not active contributes significantly to prolonging their battery lifetime. This is because listening consumes a great amount of energy that is comparable to reception. Therefore, a node's duty-cycle should be taken into consideration when designing clustering techniques. Incorporating a node's duty-cycle in the design of the clustering can be done in one of two ways, depending on the type of the application. In the first approach, non-cluster head nodes can be allowed to sleep when they are not sensing any data or when they are not communicating with their cluster heads. This approach is appropriate for applications where sensors are sending updates on a periodic basis at a predetermined time. The second approach is used if the application requires the sensors to continuously monitor the field for unexpected events, then a cluster head can determine which of its cluster members are sending redundant data and advise them to sleep [50].

3.3. Performance Evaluation

In this section, we conducted simulation studies to compare various clustering protocols. In our simulation studies, the application of border monitoring for intruder detection is considered. We simulated four clustering algorithms based on certain scenarios using MATLAB. We assume a rectangle shaped area instead of a square shaped area, as shown in Figure 3-2, which is commonly used in most research papers. This is because the segment of the belt region, which is usually the borderline between two countries or between any disputed areas and is where the sensors are randomly deployed, can be segmented into connected rectangles. I ran the simulation five times and took the average of the runs to present our results. In this work, we assume that the long borderline is divided into rectangular segments, each with one base station. The information gathered by the sensors within any one segment is communicated to the base station which is connected to either a wireless or wired backbone network delivering the information to a central database. The parameters used in our simulation are shown in Table 3-1.

We have considered the network lifetime as a performance metric, which is the time interval from the start of the operation of the sensor network until the death of the last node in the network. In our simulation, we consider several scenarios as discussed below.

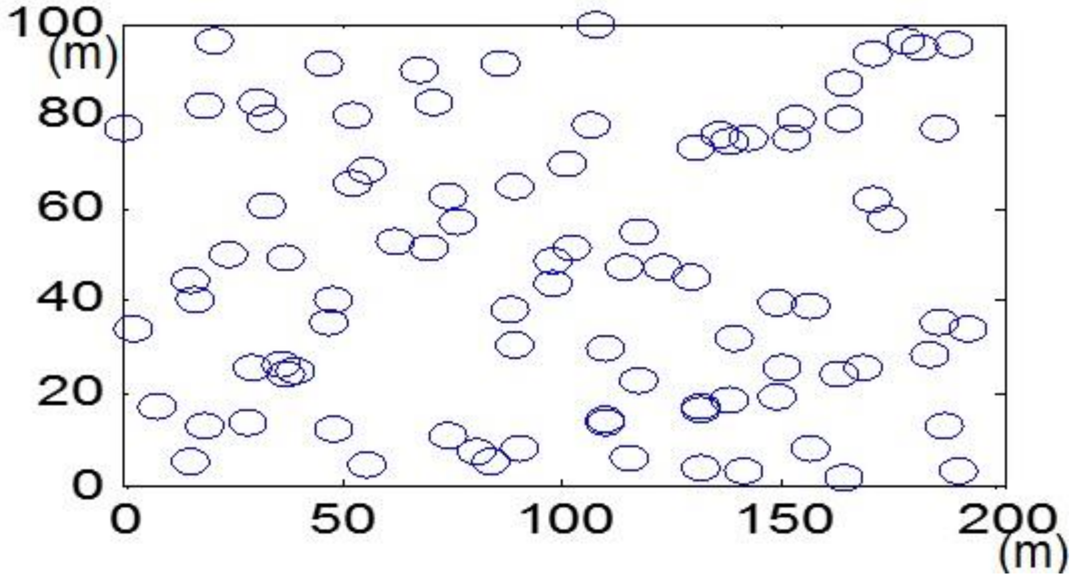

Figure 3-2: The simulated area with nodes placed randomly

Table 3-1: Parameters used in simulation

Notation	Description
N = 200	Total number of sensor nodes
E_o = 0.5J/node	Initial energy of each node
E_{elec} = 50nJ/bit	Per bit energy consumption
E_{DA} = 5nJ/bit	Energy for data aggregation
E_{amp} = 100 pJ/bit/m2	Amplifier transmitting energy
Maximum No. of rounds	5000
No. of bits (k)	2000
Area	200 x 100 (m)

3.3.1. First Scenario

In the first scenario, we placed the base station, or the sink, at the middle of the field segment (x = 100, y = 50), as shown as P1 in Figure 3-3. However, placing the sink in this position might not be desired for a border monitoring application assuming that the utilities will only be provided on the borderline.

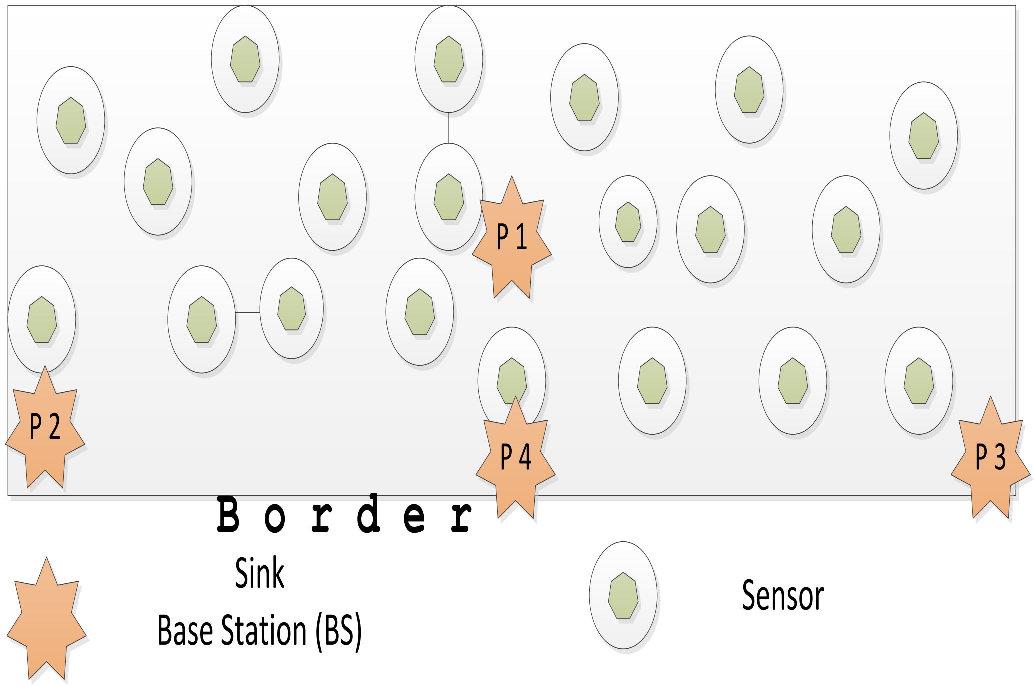

Figure 3-3: Illustration of the positions of the sink node

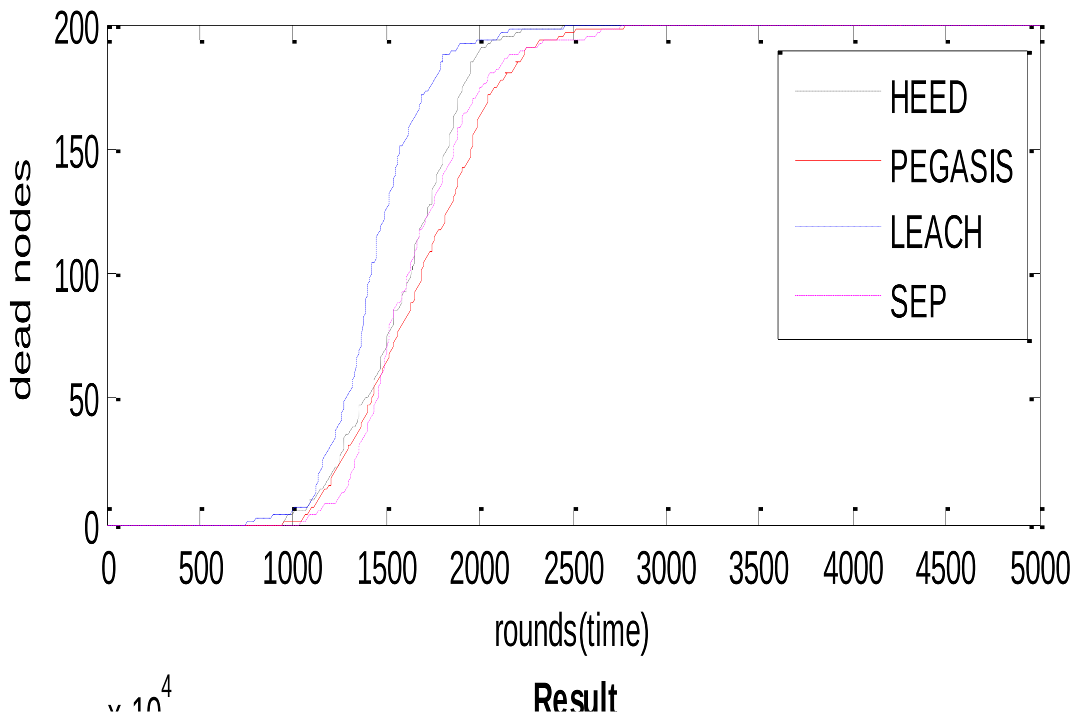

Figure 3-4: Results of the first scenario

Table 3-2: Simulation results of the first scenario

Protocols	**Measurements**	
	Round first node dies	**Round last node dies**
LEACH	821	2350
HEED	1024	2487
PEGASIS	1086	2674
SEP	1185	2829

From Figure 3-4 and Table 3-2, we can see that the last node died in LEACH at round 2350, which sets it apart as having the shortest network lifetime among the other protocols. HEED has the second shortest network lifetime after LEACH, as its last node died at round 2487. On the other hand, we can see that SEP has the longest network lifetime followed by PEGASIS, as their last nodes died at rounds 2829 and 2674, respectively.

3.3.2. Second Scenario

In the second scenario, we placed the base station at P2 (x = 0, y = 0) as shown in Figure 3-3. This will be suitable for providing the BS with a continuous power supply and more data storage capability as it will be directly connected to the borderline.

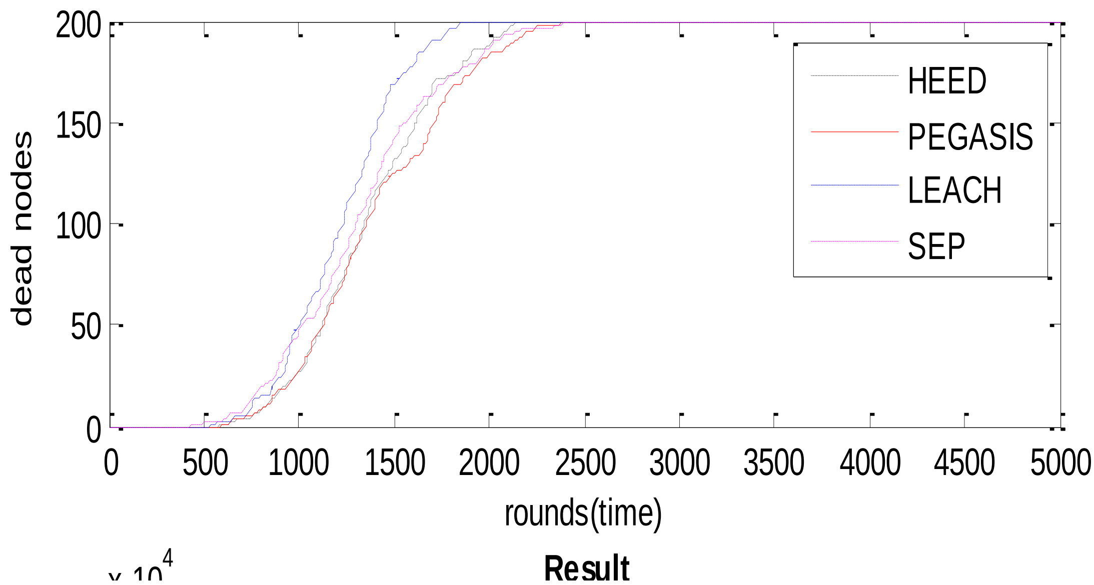

Figure 3-5: Results of the second scenario

Table 3-3: Results of the second scenario

Protocols	**Measurements**	
	Round first node dies	**Round last node dies**
LEACH	583	1820
HEED	585	2182
PEGASIS	599	2389
SEP	437	2410

From Figure 3-5 and Table 3-3, we can see degradation in the network lifetime of all protocols compared to the first scenario. The first node in LEACH died at round 583 and the last died at round 1820, whereas in the first scenario, the first died at 821 and the last died at 2350.

3.3.3. Third Scenario

In the third scenario, we placed the base station at P3 (x = 200, y = 0) as shown in Figure 3-3. The BS in this position has a similar connectivity as in the second scenario.

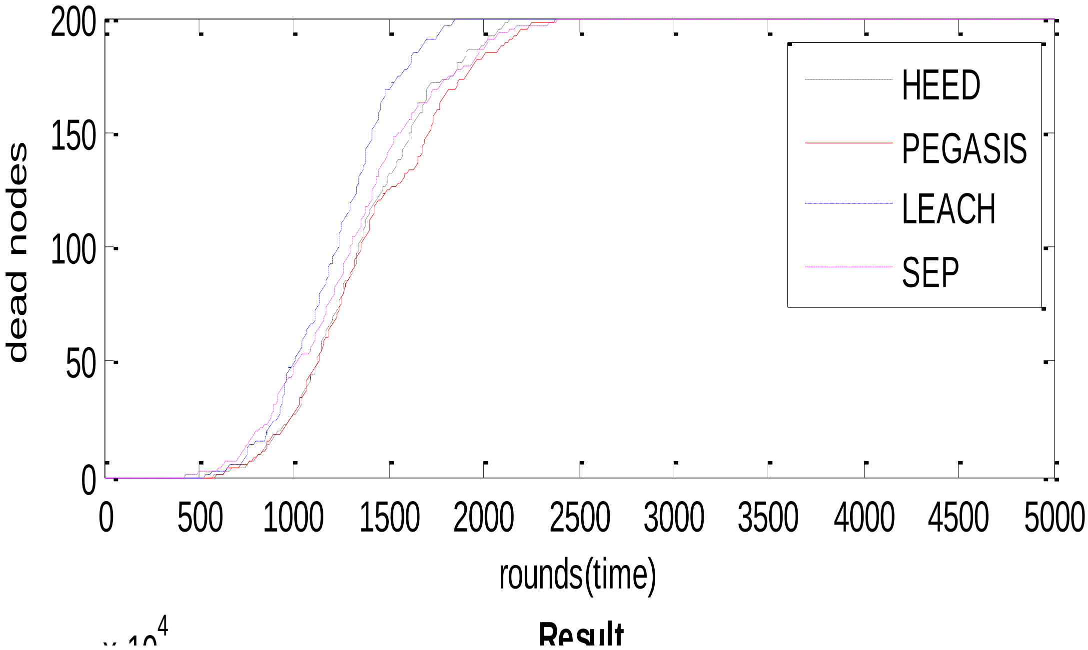

Figure 3-6: Results of the third scenario

Table 3-4: Results of the third scenario

Protocols	Measurements	
	Round first node dies	**Round last node dies**
LEACH	534	1920
HEED	633	2180
PEGASIS	590	2357
SEP	599	2404

From Figure 3-6 and Table 3-4, we can observe a similar performance in the network lifetime of all protocols as in the previous scenario. The first node in LEACH died at round 534 and the last died at round 1920.

3.3.4. Fourth Scenario

In the fourth scenario, we placed the base station at P4 (x = 100, y = 0) as shown in Figure 3-3. The BS in this position will have a similar connectivity as in the second and third scenarios.

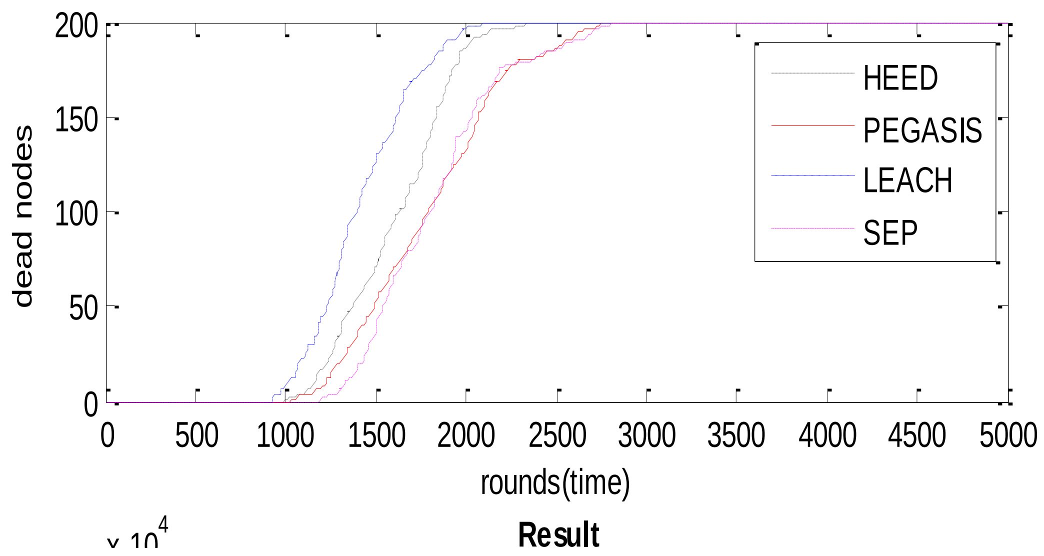

Figure 3-7: Results of the fourth scenario

Table 3-5: Results of the fourth scenario

Protocols	Measurements	
	Round first node dies	Round last node dies
LEACH	801	2303
HEED	1002	2427
PEGASIS	1022	2650
SEP	1001	2800

From Figure 3-7 and Table 3-5, we can see an improvement in the network lifetime of all protocols when compared to the second and third scenarios. The results are similar to the first scenario because the base station is placed in the middle. The first node in LEACH died at round 801 and the last died at round 2303.

3.3.5. Fifth Scenario

In this scenario, we vary the number of nodes to see if changing the number of nodes has any impact on the performance of the four protocols. The position of the base station will be fixed in the middle at P1 (x = 100, y = 50). We simulated 100 nodes, 200 nodes, and 300 nodes.

Table 3-6: Results of the fifth scenario

Protocols	**Measurements**					
	Round first node dies			**Round last node dies**		
Number of nodes	**100**	**200**	**300**	**100**	**200**	**300**
LEACH	655	821	1023	2235	2350	2474
HEED	927	1024	1130	2310	2478	2520
PEGASIS	1001	1086	1201	2469	2674	2805
SEP	940	1185	1300	2612	2829	2942

From Table 3-6, we can see an improvement in the network lifetime of all protocols as the number of nodes increases. The first node in LEACH died at round 655 when the number of nodes is 100, at 821 when the number of nodes is 200, and at round 1023 when the number of nodes is 300. Similar performance was observed for the other protocols. As the number of nodes increased, the density increased, thus making the transfer of data to the sink node less costly in most cases due to shorter transmission distances.

3.4. Discussion

From the results in the above section, we can assert that LEACH, SEP and HEED performed better when the base station was located in the middle of the field in the first and fourth scenarios. However, the network lifetime in the second and third scenarios decreased by around 15%. This can be justified because nodes on the edges consume more energy to reach the base station located at the opposite edge of the area compared to placing the base station in the middle of the field. On the other hand, the placement of the base station had the least effect on PEGASIS network lifetime across the three scenarios by around 10%. This can be explained as a result of using greedy chain aggregation from one node, to its closest neighbor, and all the way to the base station. Consequently, each node will lose less energy. However, this approach can cause a delay in receiving the sent data as it has to pass through many nodes on its way to the base station.

We can also notice that LEACH has the shortest network lifetime in all scenarios because LEACH treats all the nodes equally and randomly selects the cluster heads. HEED performed marginally better than LEACH because it uses the residual energy of each node to elect the clusters' heads. Moreover, we can see that SEP had the longest network lifetime of them all, because it uses advanced nodes that are equipped with more energy than the normal nodes.

3.5. Summary

In this chapter, the challenges in clustering a large-scale WSN were highlighted, some of the state-of-the-art clustering protocols presented in the current literature were discussed, and they were classified based on the techniques used to form their clusters and the way that their data is aggregated to the base station. I further considered the case of border monitoring and simulated these protocols to compare their performance results using different scenarios in terms of their network lifetime. Many aspects should be taken into consideration when designing clustering protocols, such as the optimal selection and rotation of the cluster heads, the cluster sizes, connectivity, the placement of the base station, and duty life cycle.

Chapter 4: Node Ranking Clustering Algorithm (NRCA)

4.1. Overview

Since data transmission can account for up to 70% of the power consumed in typical sensor nodes [41], substantial amounts of energy can be saved by reducing the distance traveled and the amount of data transmitted to the base station. The distance of the nodes from the base station and inter-node distances can make a big impact on saving nodes' energy and thus prolonging the network lifetime. This can be defined either as the time it takes for the first node to die, the time it takes for the last node to die, or the time it takes for a certain percentage of nodes in the WSN to die [51]. Moreover, in dense deployments of sensor nodes in a WSN, nodes can cooperate to send data and therefore distribute the consumption of energy between them.

In this chapter, we propose the use of a node ranking clustering algorithm (NRCA). The difference between this algorithm and other algorithms is that this algorithm uses a more efficient mechanism to select cluster heads. It is considered more efficient as it prolongs the network lifetime further by decreasing communication overheads caused by the frequent election of cluster heads which, as a result, decreases the energy consumed by nodes when compared to other algorithms. This is achieved by the proper election and replacement of cluster heads which involves measuring the distance and current energy level of nodes, using energy thresholds, and calculating the number of sensing rounds that cluster heads can serve before being replaced. In this algorithm, nodes are ranked based on their current energy level (En) and their positions (Dn) in reference to the BS. This ranking is used for choosing cluster heads which are also sorted into levels based on their position, or

Euclidean distance, from the BS. Therefore, each node is assigned a rank Rn (En, Dn) reflecting its likelihood of being elected as a cluster head. In the next section, I will introduce the proposed algorithm in more detail.

4.2. Description of NRCA

In most of the previously proposed clustering algorithms, a node is elected as a cluster head either randomly or based on it having the highest residual energy in a cluster. This selection might lead to inefficiencies [52]. For example, (and as was previously shown in [52]), node A in Figure 4-1 has higher residual energy than the other nodes, M and S, belonging to the same cluster as A. Thus, this node is typically elected as the new cluster head. As a result, this causes M and S in the same cluster to send data through A to the base station, thus taking a longer path as the location of A is in the opposite direction of the base station. The additional distance that the data needs to travel to arrive at the base station will result in more energy consumed. In addition, nodes can be forgotten or disconnected and are not covered by any of the cluster heads chosen, due to being far from any reachable cluster heads. Moreover, the frequent replacement of cluster heads in each round wastes more energy.

These three problems can be avoided in our proposed algorithm where data can be sent through the correct path or direction with respect to the BS, and by the BS, thus maintaining a global knowledge of all nodes in the WSN area to ensure that all live nodes are connected through the proper choice of cluster heads. Finally, I propose the use of an energy threshold technique in making decisions to replace cluster heads, which prolongs the lifetime of the nodes closer to the BS. This, in effect, prolongs the overall network lifetime as nodes closer to the BS are more critical than those far away nodes are in maintaining connectivity to the sink.

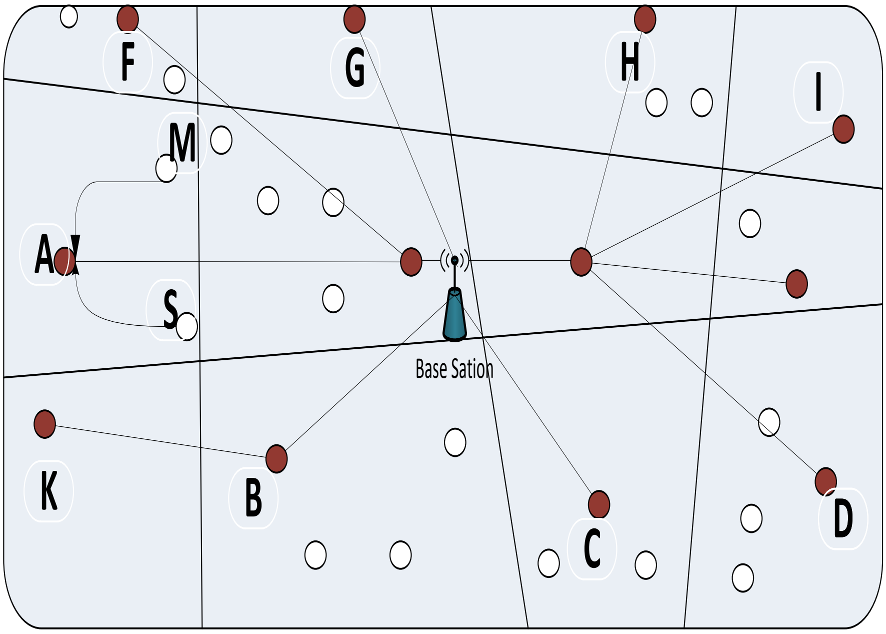

Figure 4-1: WSN clustering example of sending data to the BS in the wrong direction

In the proposed algorithm, the base station (BS) is placed in a fixed position and has unlimited energy. Thus, no constraints are assumed with regard to the power consumption as a result of data processing and communication. Through the initial step of the algorithm described below, the BS becomes aware of the locations of all sensor nodes either via collecting their GPS coordinates or any other mechanism [53].

The following steps give a description of the algorithm and cluster heads' selection process:

- Similar to the initial step taken in [36, 38, 39, 42 and 52], each node at the setting up phase broadcasts a message to its neighbors containing its energy level and location. Therefore, each node sets up a neighbor information table recording the energy levels and positions of its neighbors and broadcasts this information to its neighbors. This is conducted by all nodes in the network until the information about all the nodes in the network is received by the BS. This will provide the BS with a global knowledge of the network, and the pseudo code is shown in Figure 4-2.

- The BS divides the area into smaller partitions called clusters based on the assumed communication range of the nodes and their positions, i.e. geographical locations, by geographical partitioning, or by dividing nodes into groups. The size of, and distance between, any two of the farthest nodes within a cluster should be less than the pre-defined communication range. Therefore, no node will be out of coverage. The pseudo code is shown in Figure 4-2. Moreover, to ensure connectivity and save energy, we are assuming that nodes have a power control unit used to adjust the communication range based on a desired value other than the default one. This becomes useful when distances between nodes increase due to dead nodes. If an active node stops sending data to its cluster head, for a period of time equal to one round this node is considered dead or disconnected.
- Communication between nodes and their cluster heads, between cluster heads and between the base station and cluster heads are bi-directional.
- The BS calculates the number of rounds (a round is a time slot where the cluster head's election phase and the data transmission phase occur) cluster heads can serve based on their residual energy and on an initial pre-defined energy threshold, then relays this information to each cluster head.
- Cluster heads close to the BS will have higher energy threshold value, however, cluster heads that are farther from the BS will have a low energy threshold value.

- Cluster heads are replaced only when their energy level drops below the pre-defined or calculated energy threshold.
- Cluster heads, which are located closest to the network base station, are referred to as the first level cluster heads. The cluster heads that are located at more distant positions from the base station are considered second level, third level, etc.
- Higher-level cluster heads transmit to lower-level cluster heads in order to reach the BS using the least amount of energy.
- If there is a change in the network topology, due to nodes being considered dead or having residual energy below a certain threshold, the BS determines the next appropriate cluster head in each cluster while considering the changes.

```
If is_the_network_clustered = false
  for every node u ∈ Node-List do
      u advertise its position and its energy level to the BS
  end for
      For every node i ∈ Node-List do
            Sort nodes according to their geographical location
            //Partition sorted nodes into groups according to their communication range.
            If distance between i and i +1 < communication range then
               add i and i +1 to cluster_list
               else
               create new cluster_list
               add i + 1
            end if
            i+1
      End for
            for every node u Node-List do
            rank(u) = BS ranks u based on its energy  level (En) and Euclidean  position (Dn)
            from the BS
            end for
            for every node u ∈ Node-List_same _region do
                  if (rank(u) > rank(u + 1) ) then
                  canBeClusterHead = true
                  add node to Candidate_Cluster_heads_list
                  end if
                  u + 1
            end for
            for every cch ∈ Candidate_Cluster_heads_list  do
                  Candidate Cluster Head are ranked into levels based on their position from
                  the BS
            end for
            for every node ch  ∈ Cluster_heads_list  do
                  Calculate number of rounds cch can serve as a cluster head
                  Broadcast msgs that it is a cluster head
                  u joins the ch
            end for
   end if
u sends data to ch
```

Figure 4-2: The pseudo code for NRCA

4.3. Used Energy Model

In this chapter, the energy model adopted is the same used by [36, 39, 42, 52 and 54] and as shown in Table 4-1 where E_{elec} is the radio dissipated energy which is assigned a value of 50 nJ/bit to run the transmitter or receiver circuitry. E_{amp} is the used energy for the transmitting amplifier and assigned a value of 100 pJ/bit/m2. E_{Tx}(k, d) is the energy that a node dissipates for the radio transmission of a message of k bits over a distance d and expressed by equation (1).

Table 4-1 Parameters used in the simulation

Notation	**Description**
N = 100	Total number of sensor nodes
E_o = 0.5J / node	Initial energy of each node
E_{elec} = 50nJ / bit	Per bit energy consumption
E_{DA} = 5nJ / bit	Energy for data aggregation
E_{amp} = 100 pJ/bit/m2	Amplifier transmitting energy
Sensing field = 100 x 100 m	Area of the sensing field
Communication range	40 meters

$$E_{Tx}(k) = E_{elec} \times k + E_{amp} \times k \times d^2 \qquad (1)$$

In the same way, the equation of the energy dissipated by a node for the reception E_{Rx}(k) of a message of k bits which is due to running the receiver circuitry E_{elec} (k) can be expressed by equation (2):

$$E_{Rx}(k) = E_{elec} \times k \qquad (2)$$

4.4. Cluster Head Selection Process

After forming the clusters, the BS assigns a cluster head for each cluster based on the proposed NRCA. Nodes in each cluster are ranked based on their distance from the BS and their current energy level. Nodes with the maximum residual energy and minimum distance are chosen as cluster heads based on Equations 3 and 4.

$$\mathrm{NodeRanking}(\mathrm{En},\ \mathrm{Dn}) \tag{3}$$

where

$$(\mathrm{Dn}(\mathrm{i})) = \mathrm{Min}(\mathrm{D}(\mathrm{i},\ \mathrm{BS})),\ (\mathrm{En}(\mathrm{i})) = \mathrm{Max}(\mathrm{ResidualEnerg}) \tag{4}$$

$$|\mathrm{D}(\mathrm{i},\mathrm{BS})| = \sqrt{(\mathrm{X_i} - \mathrm{X_{bs}})^2 + (\mathrm{Y_i} - \mathrm{Y_{bs}})^2} \tag{5}$$

Residual (En) is the current energy level of the node i, D(i, BS) is the Euclidean distance of node i to the base station. Given a particular deployment region of interest, X_i and Y_i are the X and Y positions of node i. X_{bs} and Y_{bs} are the X and Y positions of the base station.

A cluster head in each cluster will be changed when its energy level reaches a pre-defined threshold or a calculated value and not every round. This will make it possible for a node, i, to continuously play the role of a cluster head for multiple rounds and thus save any energy that would have otherwise been wasted by the control used and messages exchanged in replacing it.

$$\mathrm{T}(\mathrm{i}) = \frac{\mathrm{Residual}(\mathrm{En}(\mathrm{i}))}{\mathrm{Average}(\mathrm{En})} \times \frac{\mathrm{Average}(\mathrm{Dn})}{\mathrm{D}(\mathrm{i},\mathrm{BS})} \tag{6}$$

$$\mathrm{Average}(\mathrm{Dn}) = \frac{\sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{D}(\mathrm{i},\mathrm{BS})}{\mathrm{n}} \tag{7}$$

$$\text{Average (En)} = \frac{\sum_{i=1}^{n} \text{Residual(En(i))}}{n} \tag{8}$$

Equation 6 shows how to calculate the energy threshold value used for all nodes. T(i), is calculated based on its residual energy, Residual (En(i)), is the average residual node energy within its cluster, the Euclidean distance between it and the BS D(i, BS), and the associated average, Dn. In the first round, all nodes have the same energy level. Consequently, ranking will depend solely on the distance. If a node is closer to the BS, it has a greater probability of becoming a CH. In the next rounds, the residual energy of each candidate node in the network is different. Therefore, the selection of CHs will depend both on their residual energy and Euclidean distance. According to Equation 6, nodes close to the BS will be changed more often as their threshold values will be higher. This is because they are critical to the network and depended on more to aggregate the data to the BS. However, nodes that are far from the BS will have a lower threshold and will be changed less frequently. The number of rounds a node, i, can stay as a CH, CountRound (i) is calculated based on the node residual energy and the calculated threshold value as shown in Equation 9.

$$\text{CountRound (i)} = \frac{\text{Residual(En(i))}}{\text{T(i)}} \tag{9}$$

4.5. Performance Evaluation

To evaluate the performance of the proposed algorithm against two other well-known algorithms (LEACH and PEGASIS), we used MATLAB to simulate the algorithms under consideration. Table 4-1 shows the parameters used in this simulation environment which are the standard parameters used by all researchers in this field. I ran the simulation five times and took the average of the runs to present our results.

The simulated area is 100 x 100 m. Every node was given an initial energy of 5 J. The energy for data aggregation is 5 nJ/bit. The energy to run the radio is 50 nJ/bit. The amplifier transmitting energy is 100 nJ/bit/m2. In our performance evaluation, we focused our attention on the main two algorithms, LEACH and PEGASIS, which were used as the baseline for all researchers in the field. In [34] we showed how PEGASIS outperformed HEED, therefore HEED was not selected. SEP was also not considered here as it uses heterogeneous nodes with different initial energy levels. Running the simulation, I considered several metrics to evaluate the performance of NRCA, as follows:

4.5.1. Cluster Formation and Cluster Head Selection

As we can see from Figure 4-3, PEGASIS forms a chain starting with the farthest node from the BS. A leader node is elected randomly in each round and it assumes all nodes can reach the BS. The leader node is the one responsible for transmitting all sensed data to the BS in each round. As shown, the leader node is far from the BS, so it consumes more energy to send the data to the BS, especially if it is the farthest node.

Figure 4-4 shows the cluster formation and cluster heads' election in LEACH. As can be seen, cluster heads are elected randomly in each round, so a cluster head can be the farthest node from the BS in its cluster (as shown in Cluster A) or it can be the node with the least energy. In both cases, the election leads to inefficiencies.

On the other hand, Figure 4-5 shows the NRCA cluster formation and cluster head selections. The nodes with the highest energy and closest to the BS in each cluster will be selected as cluster heads. For example, as shown in Figures 4-4 and 4-5, for Cluster A, the node closest to the base station was chosen as a cluster head, while in LEACH the farthest node in the same cluster was chosen. Therefore, the energy consumed to send data to the BS is reduced in NRCA. Moreover, there are no disconnected or forgotten nodes and thus no clusters are formed with only one node.

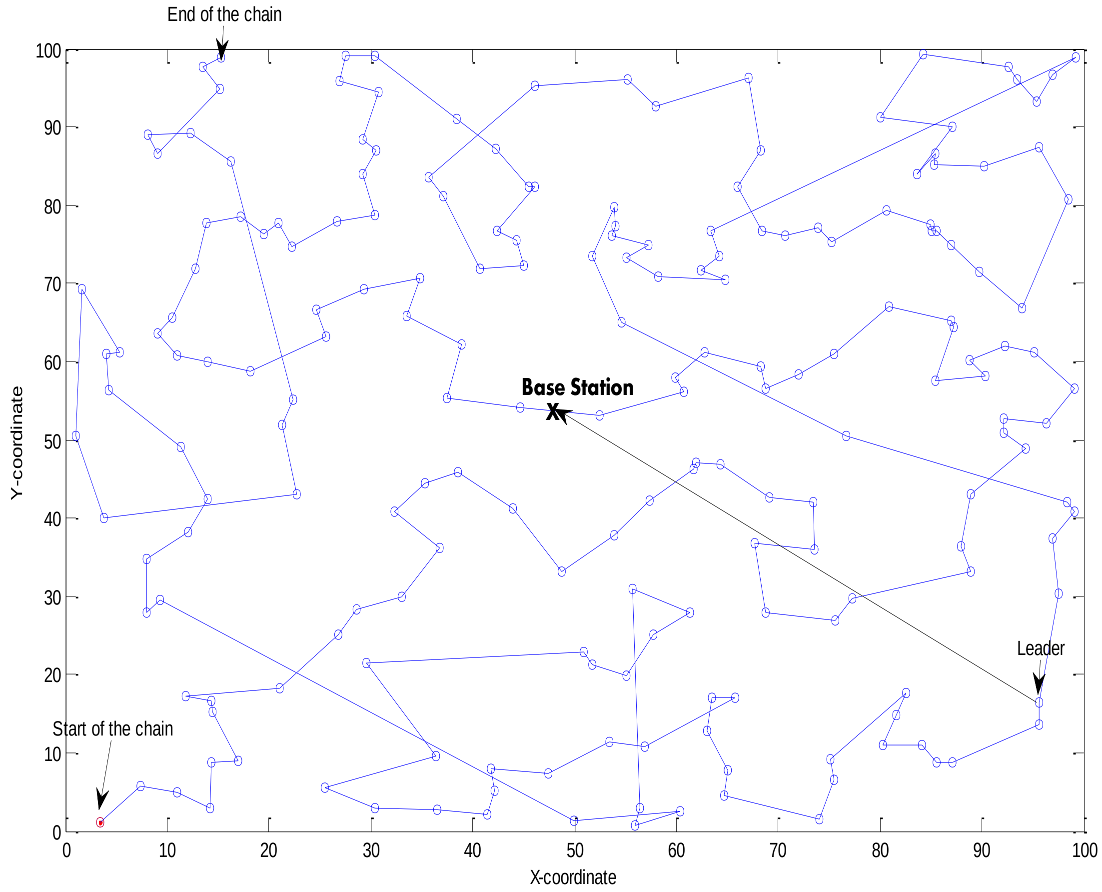

Figure 4-3: PEGASIS chain formation

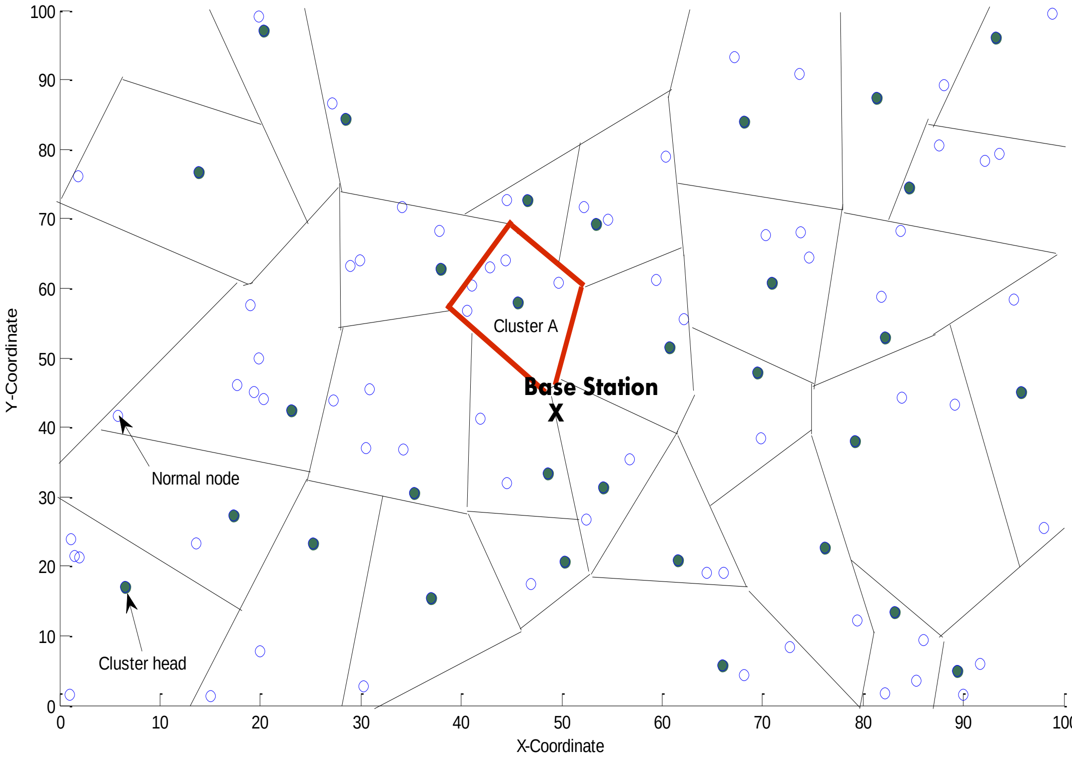

Figure 4-4: LEACH cluster formation and cluster head elections at first round

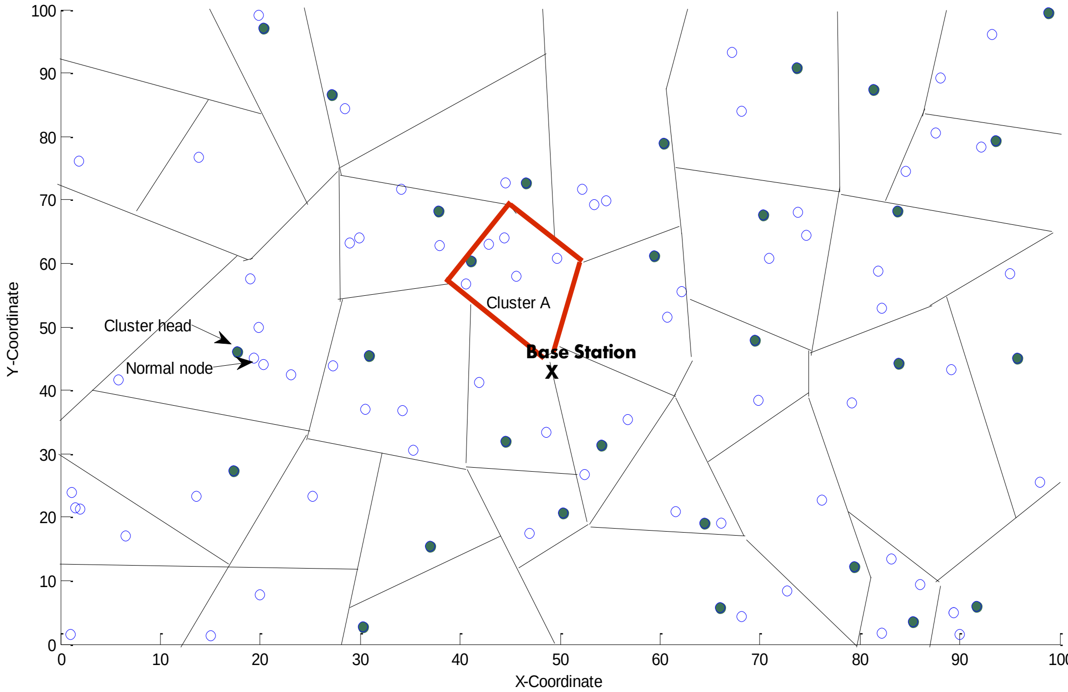

Figure 4-5: NRCA cluster formations and cluster head elections at first round

4.5.2. Network Lifetime

Network lifetime is defined here as the interval from the time the sensor network starts its operation until the death of the last node in the network. From Figure 4-6 and Table 4-2, we can see that the last node in the simulated WSN died in LEACH at round 2230, making it the lowest achiever with the shortest network lifetime among the other protocols considered. On the other hand, we can see that NRCA has the longest network lifetime, followed by PEGASIS, as their last nodes died at rounds 3200 and 2774, respectively. Table 4-2 and Figure 4-6 show how NRCA outperformed PEGASIS by 15% and LEACH by almost 70% for the network lifetime criterion. In this scenario, no threshold was chosen so heads will be changed every round.

Table 4-2: Simulation results for the network lifetime

Protocols	**Measurements**	
	Round first node dies	**Round last node dies**
NRCA	1179	3200
PEGASIS	1086	2774
LEACH	821	2230

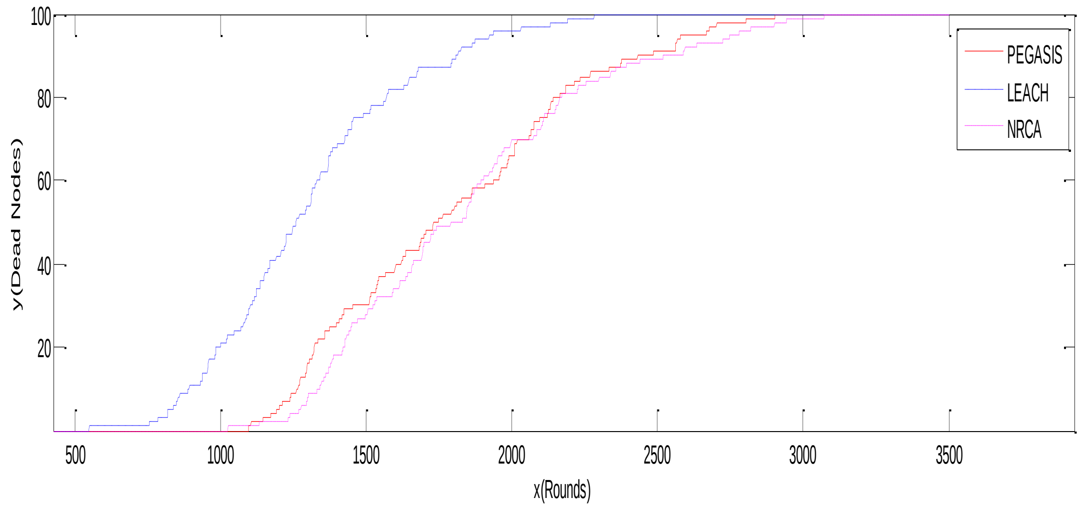

Figure 4-6: Simulation results for the network lifetime

4.5.3. Connectivity and Coverage

There exists a connectivity between the cluster head and nodes in the cluster if, and only if, the physical Euclidean distance between the cluster head and any node in the cluster is less than, or equal to, the transmission range of the cluster head. The more cluster head nodes there are, the better coverage or connectivity the network will have and the less distance and energy will be needed to send data. Better coverage also implies minimal or no forgotten or disconnected nodes. If an active node stops sending data to its cluster head, for a period of time equal to one round and its last known residual energy is greater than the average energy consumed by one round (5 micro Joules) then this node is considered disconnected. However, if its residual energy percentage was less than 5 micro Joules then it is considered dead.

At the startup phase, NRCA considers only nodes that report their energy levels and locations to the base station. From Table 4-3, we can see that NRCA has less disconnected nodes. This is due to the correct partitioning done by the base station based on the global knowledge it maintained in the setup phase and the assumed power control unit feature. As such, its performance with respect to connectivity and coverage is considered better.

Table 4-3: Number of disconnected nodes per selected rounds

Protocol	# of disconnected nodes in round 100	# of disconnected nodes in round 500	# of disconnected nodes in round 1000	# of disconnected nodes in round 2000
NRCA	3	8	18	35
PEGASIS	10	23	46	62
LEACH	18	38	56	89

4.5.4. Varying the Placement of the Base Station

In this simulation section, I changed the placement of the sink node or BS to see its effect on the performance of the algorithms. At position 1 (P1), I placed the BS at the center or middle of the WSN area, (x = 50, y = 50), and at position 2 (P2), I placed the BS on the borderline of the area where the WSN is being deployed, i.e. (x = 50, y = 0).

From the results in Table 4-4, we can notice that the change of the BS placement has the least effect on NRCA. PEGASIS follows with a minor effect. On the other hand, LEACH has been affected more by this change. It performed better when the BS was placed at the center of the WSN area. This is due to LEACH treating all the nodes without discrimination and randomly selecting the cluster heads.

Table 4-4: Simulation results of changing the placement of the BS

protocols	**Measurements**			
	Round first node dies		**Round last node dies**	
	Middle	Border	Middle	Border
NRCA	1185	1179	3302	3292
PEGASIS	1086	1022	2790	2574
LEACH	821	801	2350	2058

4.5.5. Varying the Number of Nodes

In this simulation, I varied the number of nodes, while keeping the deployment area fixed to see if changing the density of the nodes has any impact on the performance of the algorithms. The position of the base station was fixed at P1. I simulated 100 nodes, 200 nodes, and 500 nodes and looked at when the first and last nodes died as shown in Table 4-5.

Table 4-5: Simulation results for different number of nodes

Protocols	**Measurements**					
	Round first node dies			**Round last node dies**		
Number of Nodes	100	200	500	100	200	500
NRCA	1179	1185	1200	3220	3329	3442
PEGASIS	1086	1090	1109	2974	3174	3255
LEACH	821	830	846	2303	2303	2374

From Table 4-5, we can see an improvement in the network lifetime of all protocols as the number of nodes increases. The first node in LEACH died at round 821 when the number of nodes is 100, at 830 when the number of nodes is 200 and at round 846 when the number of nodes is 500. A similar performance was observed for the other protocols. As the number of nodes increased, the density increased, making the transfer of data to the sink node less costly in most cases due to shorter transmission distances.

4.5.6. Received Data by the BS

As shown in Figure 4-7, data received by the BS in NRCA was more than it was when using the other two algorithms. Data includes both control data sent in cluster head selections or network setup and the sensed data which is sent through sensors (control data was < 10%). This was due to NRCA choosing the most appropriate nodes as cluster heads based on both energy and the correct path to the BS. It is also due to minimizing the number of overhead messages needed for cluster head selection and replacement processes.

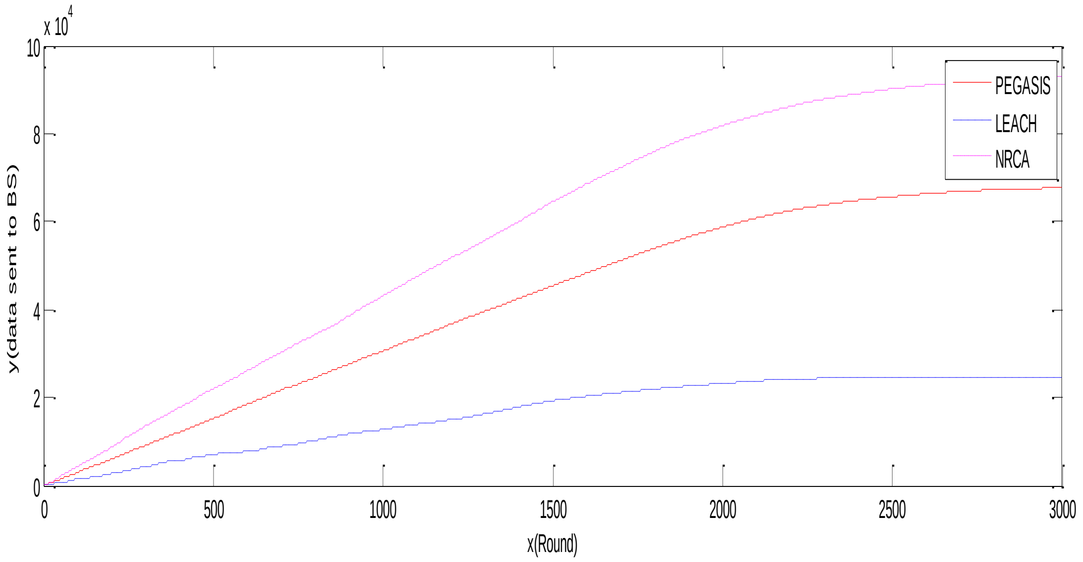

Figure 4-7: Received data by the BS per round

4.5.7. Energy Consumed

As shown in Figure 4-8, the energy consumed per round in NRCA is less than LEACH and PEGASIS with LEACH consuming the most. The amount of energy wasted on the frequent replacement of cluster head nodes by allowing them to serve as CHs in several rounds (as long as their energy did not drop below the specified threshold level) was the main factor in achieving this.

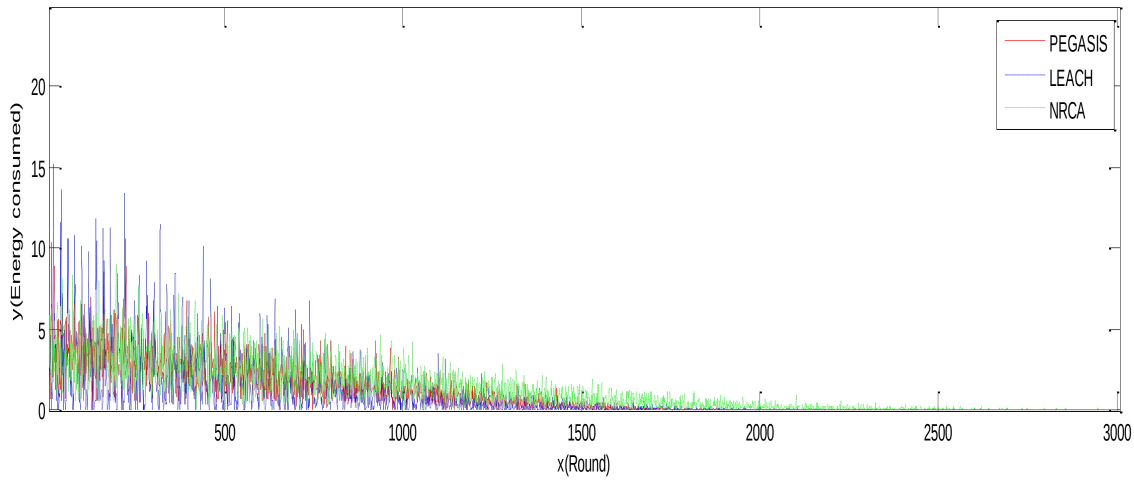

Figure 4-8: Energy consumed per round

4.5.8. Using Fixed Threshold Percentages of Remaining Energy to Replace CHs

In this simulation, I varied the fixed threshold values of the remaining energy to replace the CHs. I used 40%, 30%, 20% and 10% of the remaining energy. Table 4-6 shows the result. As we can see, at the beginning, using 40% as the energy threshold to replace the cluster heads performed better than the rest. When almost 50% of the nodes died, we achieved almost equal results for all threshold values. However, as the remaining nodes decrease the 10% threshold value performed better than the other values.

4.5.9. Using Dynamic Thresholding to Replace CHs (Variable Threshold)

Using the result achieved in the previous simulation section, I applied the formula as defined in Equation 6 to calculate the energy threshold when replacing CHs. In this formula, nodes closest to the sink are aimed to live the longest as they are critical to the network and are used by other nodes in the network to forward data to the base station. Using Equation 6 implies that the cluster heads close to the base station will have a higher replacement energy threshold value and they will be replaced more frequently than cluster head nodes that are farther from the BS and will have lower replacement energy threshold values.

Figure 4-9 shows the results obtained when simulation experiments were run using the variable energy thresholds calculated using Equation 6, versus using a fixed pre-defined threshold, and as opposed to replacing cluster heads in each round. As can be seen from Figure 4-9, the last node that died when using NRCA without threshold was at round 3200 and with a fixed threshold at round 4020, whereas with the use of the variable threshold values based on Equation 6, the last node died at round 4320. This shows how NRCA with variable and fixed threshold values outperformed NRCA without a using threshold in terms of network lifetime. We can also see that NRCA with variable threshold values outperformed NRCA with a fixed one in terms of network lifetime by almost 7%.

Table 4-6: Network lifetime using different threshold values

	Percentage of left energy to change CH (threshold of the left energy)			
Percentage of alive nodes	**10%**	**20%**	**30%**	**40%**
First died	843	850	858	861
90%	1230	1238	1245	1257
50%	1987	1982	1979	1978
10%	3580	3540	3522	3506
Last node died	4020	4010	3990	3970

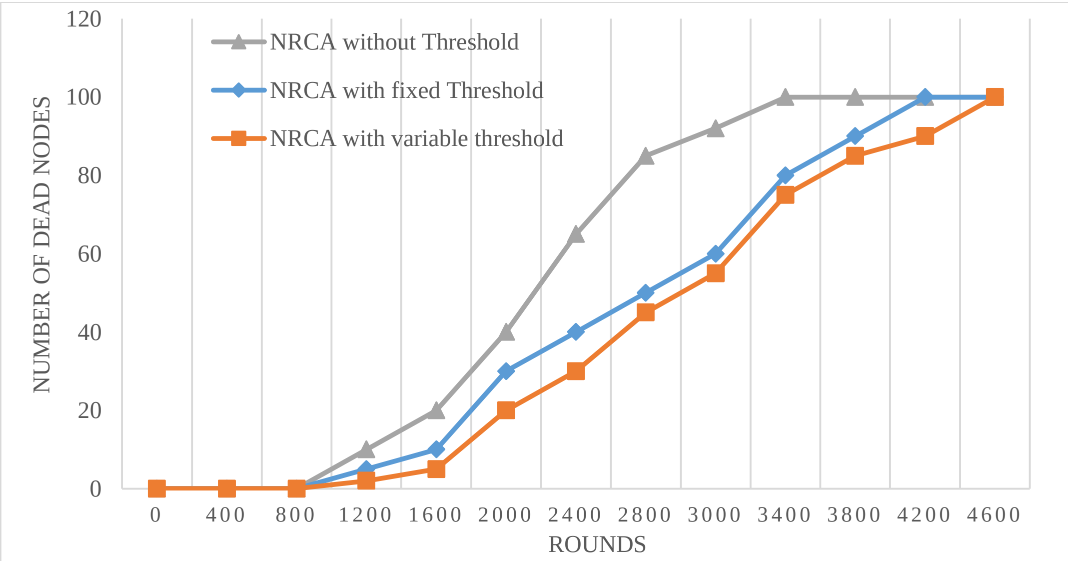

Figure 4-9: NCRA with fixed, variable and without threshold

4.5.10. Hybrid Node Duty-Cycle (Redundant Nodes Duty-Cycle Selection)

In large-scale dense wireless sensor networks, sensors are often deployed in large quantities to increase reliability and to extend the coverage [55]. As a result, there are many redundant sensor nodes collecting redundant data in such networks. However, to increase the network lifetime and distribute the load more equally among nodes, redundant nodes should take turns in covering the monitored area whenever possible. Initially, all nodes are in a working mode and for nodes monitoring the same coverage area, redundant data might be collected and communicated through the network, thus consuming energy. Therefore, I propose to apply hybrid node duty-cycles, where nodes take turns in monitoring a particular coverage area based on certain conditions.

I used a hybrid duty-cycle scheme where I combined both synchronous and asynchronous schemes. In order to determine which node should stay active or go to sleep within a cluster, each node will communicate with its direct neighbors and detect nodes that are within the same pre-defined detection range (sensing or coverage range). Nodes, covering the same detection range, will then agree on which node stays active based on its energy. If the energy of an awake working node is below a certain threshold, for example, 10% of the initial energy, the working node will send a broadcast message to wake up sleeping nodes within the detection range before it goes to sleep. For reliability purposes, sleeping nodes will wake up to enter into the detecting mode in the event that a period of time Ts has passed without it receiving any instructions from the awake node. This technique is efficient when monitoring a continuous event. From Figure 4-10, we can see that the last node in NCRA without using nodes duty-cycle died at round 4320, while with duty-cycle it was at 4660. This shows that using a duty-cycle strategy improved the performance by almost 8%.

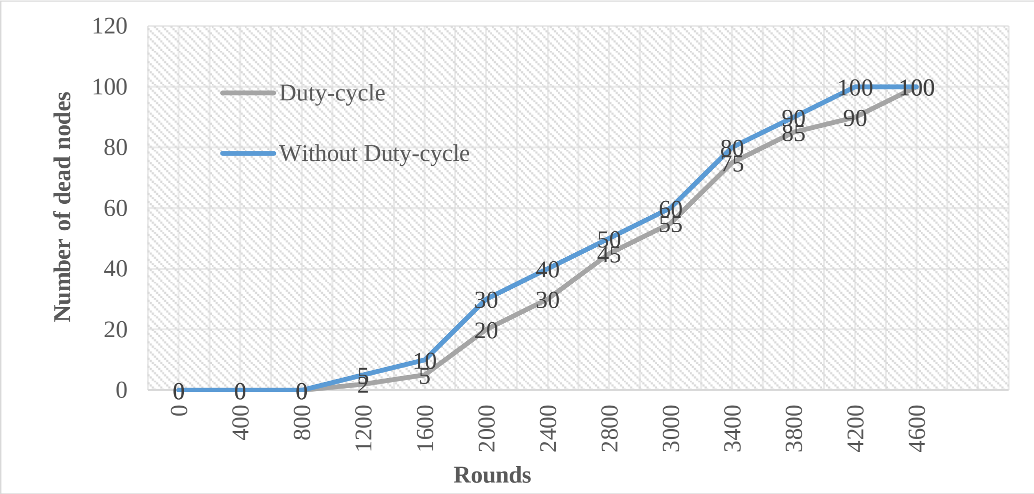

Figure 4-10: NRCA with and without duty-cycle

4.6. Summary

In this chapter, an energy efficient clustering algorithm for WSNs using node ranking in electing cluster heads was proposed. The performance of the proposed algorithm against two well-known algorithms in terms of network lifetime was compared. Through simulation, this chapter showed how the proposed algorithm outperformed PEGASIS by 15% and LEACH by almost 70% for the network lifetime criterion. However, NRCA required more computations than the other two algorithms due to computing of distances and the number of rounds the cluster heads can remain serving as such. Moreover, the performance of the algorithm using random cluster heads replacement and using threshold values to replace the cluster heads were compared and the simulation showed that using a threshold value outperformed the random replacement of cluster heads. Using an energy threshold to replace cluster heads improved the network lifetime by almost 15%. I also found that using variable energy threshold values to replace cluster heads improved the network lifetime even further, by almost 7% over the use of a fixed value. In addition to that, using a hybrid redundant node duty-cycle improved the network lifetime by 8%.

Chapter 5: Literature Review: Data Gathering using Mobile Nodes (Ferries)

5.1. Overview

Mobile ferries are an alternative way to collect data from dispersed sensor nodes, especially in large-scale networks and for delay tolerant applications. Unlike data collection via multi-hop forwarding among the nodes, ferries travel across the sensing field and collect data from the sensing nodes. The advantage of using a ferry-based approach is that it eliminates the need for multi-hop forwarding of data, and as a result, energy consumption at the nodes is significantly reduced. However, this increases data delivery latency and as such it might not be suitable for all applications. In this chapter, I survey the recent progress in using mobile ferry nodes for data gathering in WSNs by addressing two main areas: determining the path of the ferry and the scheduling of when to dispatch the ferry to collect data from sensors. I also highlight challenges facing the deployment of mobile ferries in wireless sensor networks.

5.2. Introduction

In general, a wireless sensor network is a collection of static nodes with sensing, computation, and wireless communication capabilities [56, 57, 58 59 and 60]. However, due to the nature of some applications such as disaster recovery, animals tracking and military applications, mobile nodes are needed [61]. Using mobile nodes to collect data from sensors in WSNs can improve the performance, such as the lifetime of a WSN and the maintained coverage area. Ferries are mobile elements that are used to carry data over distances to the base stations or to a data center. They are also used to connect isolated islands of WSNs. In addition, ferries can be used to resolve the issue of coverage for holes in a WSN resulting from the need to replace deployed fixed sensor nodes which have run out of energy. Mobile elements can be attached to people, animals, vehicles, robots, unmanned aerial vehicles or any movable object.

There are different types of ferries or mobile elements that are used in WSNs [61]. They can be classified according to the following subsections:

5.2.1. Ordinary Sensor Nodes

Ordinary sensor nodes are the source nodes that perform the sensing task as shown in Figure 5-1. A mobile ferry can be used as a scale sensor that senses data from the surrounding environment and sends the data (e.g., temperature, light, gas) to the cluster head or a collector. The advantage of these nodes is that they are moving, so they can track a movable event like an intruder detection in border monitoring applications.

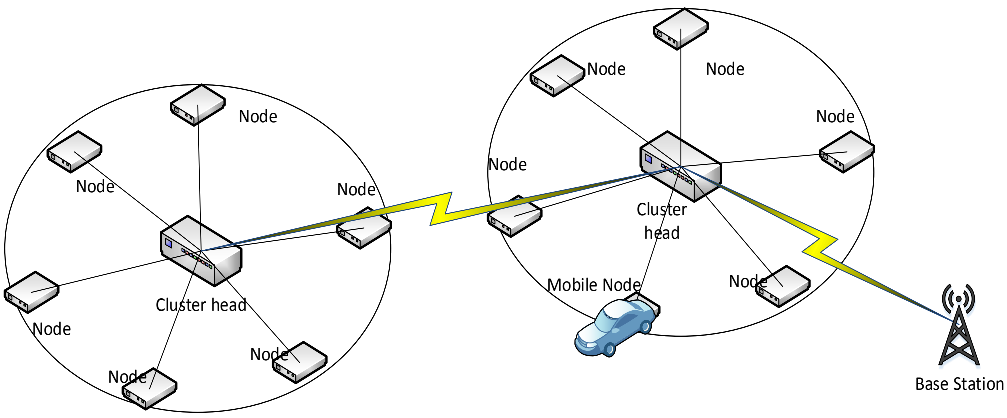

Figure 5-1: Mobile nodes are used as a part of WSNs

5.2.2. Mobile Sink or Base Station

The sink, or the base station (BS): destination where all data are gathered to be used by data centers or outside applications. The sink can be mobile and visit all nodes to collect data from them directly or through intermediate nodes as shown in Figure 5-2. Mobile sinks can increase the network lifetime, decrease delay, and decrease traffic. However, having a mobile sink requires full knowledge of, or control over, its movement and schedule.

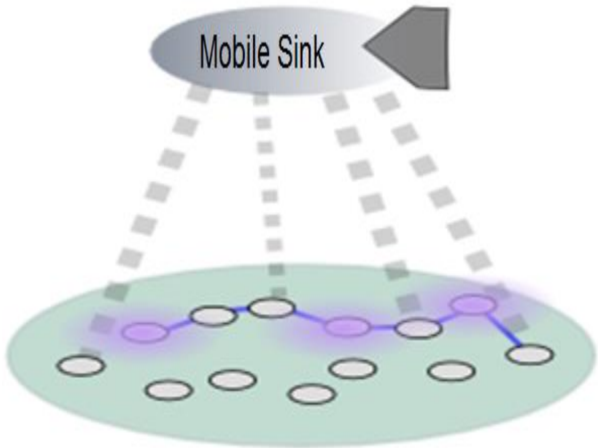

Figure 5-2: Mobile sink collects data from nodes

5.2.3. Mobile Support Nodes

Support nodes are the intermediate nodes that help the data to be transferred from the source (sensing nodes) to the destination (sink) as shown in Figure 5-3. A WSN might become partitioned into several islands for many reasons, which makes communication in the network impossible. In this case, mobile support nodes can be used to connect partitioned WSNs. Mobile support nodes can also be used to replace dead critical nodes in case of emergency. An example of this is sending a robot to cover a certain area when nodes on duty are dead. This strategy will help to provide greater coverage and increase the lifetime of the network.

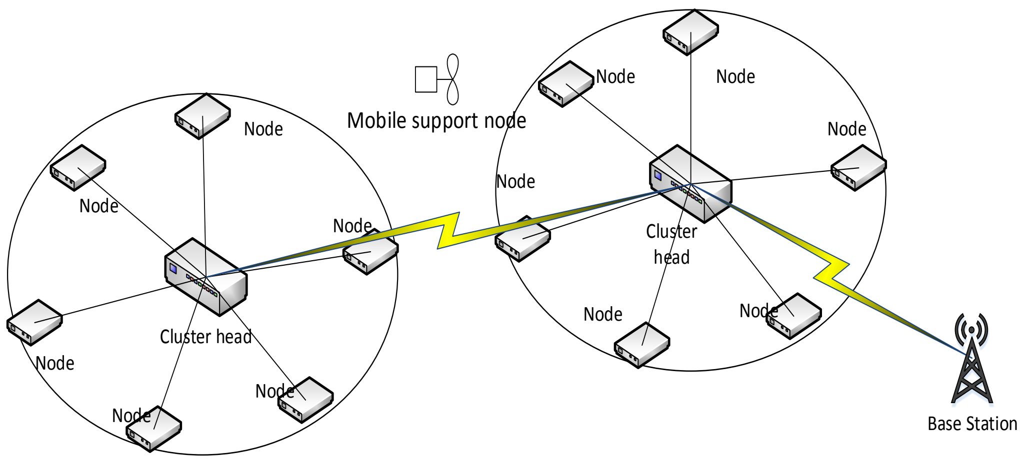

Figure 5-3: Mobile support node used to transfer data in WSNs

5.3. Applications Using Mobile Ferries in WSNs

Due to the nature of some applications of WSNs, mobile ferries are needed. Using ferries to collect data from sensors in WSNs can improve the performance of WSNs, such as their lifetime and their coverage. Below are some of the applications that can utilize the advantages of ferries.

5.3.1. Border Monitoring

Mobile nodes can be used in intrusion detection and border surveillance to collect data from sensors scattered along the border. BorderSense is an example of such an application, where mobile ferries such as Unmanned Aerial Vehicles (UAVs) are used to collect data from static sensors [62]. Mobile nodes can also be used as sensor nodes to provide additional coverage if needed. In addition to that, mobile nodes can track intruders based on information from static sensors and help to catch the intruders.

5.3.2. Disaster Recovery

During times of disaster, such as an earthquake or a tsunami, communication infrastructures are usually destroyed, which makes rescue and recovery efforts difficult. Therefore, there is a need for mobile nodes to be used in the collection of information from the surrounding environment and to aid in the rescue operation. In [63], multiple mobile sensors carried on vehicles are used across vast distances with minimal need for wired infrastructure to provide communication coverage for disaster recovery. In addition, static nodes can be deployed to monitor the disaster area and information can either be disseminated through multi-hop forwarding or by using mobile ferries carried by robots or by other means.

5.3.3. Environment Monitoring

Mobile elements can be attached to people, animals, vehicles or any movable object to continuously report environmental data for long periods of time. They can be used to detect air and water pollution, forest fires, and floods. CitiSense [64] uses wearable devices and mobile phones carried by users to collect environmental parameters (CO, NO2 and O3, temperature, humidity and barometric pressure) from static sensors to monitor air pollution in certain areas and correlate them to other events or aspects.

5.3.4. Military Applications

Military applications involve intrusion detection, battlefield surveillance, monitoring friendly forces, battlefield damage assessments, information gathering and smart logistics support in an unknown deployment area. In [65], static sensor nodes were deployed on the ground with the job of detecting and tracking vehicles passing through the area down a dirt road. The vehicle tracking information was collected from the sensors using a UAV in a flyover maneuver which was then sent to an observer at the base camp.

5.3.5. Intelligent Road Transportation

Applications of Intelligent Road Transportation (IRT) usually fall under navigation, traffic flow control (e.g., changing traffic lights) and the need to plan and build new infrastructure. Vehicles on the road are equipped with sensors that could act as mobile nodes on the road network and provide a rich source of data about traffic, the environment, and road conditions. This information assists traffic managers to regulate traffic effectively in order to maintain a good flow of traffic and minimize the risk of accidents and road congestion [66]. In addition, these mobile nodes can disseminate data to subscribed drivers who wish to avoid congestion and get reports on road and weather conditions in real-time at a low cost.

5.3.6. Animal Tracking

Sensors can be attached to animals to track them in support of wildlife research or simply to locate them. When sensors are attached to animals they became mobile sensor nodes. ZebraNet system is a WSN tracking system carried by animals across a large area that are being studied. The sensors send logged data on the animal's positions, their temperature, heart rate, and the frequency of their feeding to the base center to be used by wildlife researchers [67, 68]. These mobile sensors can also be used to collect data from scattered static sensors deployed in the farm area for various applications and communicate such data effectively to a base station for further transmission to a central database for processing.

5.3.7. Pipeline Monitoring

There are many applications for WSNs in monitoring water and oil pipelines. Mobile sensor nodes are used in pipeline monitoring because pipelines cover a large area and therefore it is costlier to deploy static nodes across them. TriopusNet is a mobile wireless sensor network system used for autonomous sensor deployment in pipeline monitoring. It releases sensor nodes from a centralized repository located at the source of the water pipeline and builds a wireless network of interconnected sensor nodes. When a node dies, or has a low battery level, the TriopusNet system sends a new node from the repository to replace the dead node [69].

5.4. Surveying Previous Research

Using mobile ferries in WSNs is a relatively new area of research which is gaining the attention of many scholars. Incorporating ferries in WSNs helps to eliminate the need for the multi-hop forwarding of data. It also reduces energy consumption at the node level. However, using ferries might cause delays in the collecting, disseminating, and processing of data and therefore it might not be suitable for all applications. The existing research in this field can be grouped into two main areas or categories, as listed in the sections below.

5.4.1. Determining the Path

The path that the ferry takes to collect the sensed data from the sensors can be categorized as either a random path or a planned path. Usually in case of the random path, the ferry is attached to people or animals moving randomly and collects sensed data whenever they are within the communication range of the static sensor nodes. In [70], mobile entities called mules were deployed in the environment. Mules picked up data from the sensors when they were in close range, buffered it, and dropped it off when they were within the communication range of the wired access points. They used a two-dimensional random walk to model the mobility of mules. Both the mules and the sensors were required to have memory capacities as they were both buffering data. In [71], mobile nodes were used in the sensor field as forwarding agents. When a mobile node entered within close proximity of the sensors, data was transferred to the mobile node to be deposited at the destination later. They used analytical models to understand key performance metrics such as data transfer, latency to the destination, and power consumption.

Due to the random mobility of the ferry, it is difficult to gather sensed data from all the deployed nodes. Unlike the random path approach, in the planned path approach, a path is determined before dispatching the ferry and thus the ferry is sent to cover a certain area near to the deployed sensors in order to collect data. In [72], an architecture of a wireless sensor network for a traffic surveillance application with mobile sinks was proposed. All sensor nodes in this architecture were assumed to be located within direct communication range of the mobile sink. All multi-hop transmissions of high-volume data over the network were converted into single-hop transmissions to further preserve the energy of the network. Therefore, nodes will transmit only in a single-hop fashion to the mobile sink.

In Mobi-Route [73], a routing protocol wherein the sink moves on a planned path to prolong the network lifetime in WSNs was proposed. In this protocol, the sink moved and stopped at certain points of interest. The stopping periods were designed to be long enough to allow for the collection of data. All the deployed static sensor nodes needed to be aware of the sink's movement and the location and time of the stops in order to send the sensed data to it.

The authors in [74] used a single ferry to collect data from a circular dense sensor network. They showed that the optimal mobility strategy of the ferry is achieved when moving at the border of the sensing area. They divided the area into circles starting from the source. The inner circles forwarded the data to the outer ones until the border was reached where the ferry was used to collect the sensed data.

5.4.2. Scheduling the Dispatch of the Ferry

The scheduling of when exactly to send the ferry to collect sensed data from nodes is a rather complicated task. In [75, 76], the researchers studied the scheduling problem when the path of the mobile sink was optimized to visit each node in the WSN before its buffer was full. Buffer overflow was used as a trigger to send the ferry to collect data to prevent data loss.

In [77, 78], the authors suggested that the mobile sink visit exact locations (rendezvous points) based on a predetermined schedule to collect data. The rendezvous points buffer and aggregate the data that originated from the source nodes through multi-hopping and transfer it to the mobile sink upon its arrival.

In [79] a ferry is used to help in collecting data in partitioned wireless sensor networks and transfer the collected data that is stored locally back to the base station. The authors classified the scheduling of ferry visit into three categories: time-based scheduling, location-based scheduling, and dynamic-based scheduling. Time-based scheduling occurs when a node dies and its death leads to partitioned WSNs. This node will have a higher priority for ferry visits. The location-based scheduling assigns the nodes closer to the base station a higher priority for the ferry's visit. The dynamic-based scheduling is based on calculating the distances between the current location of the ferry and the locations of the partitioned wireless sensor networks that have not yet been visited by the ferry, and selects the shortest distance for its next visit.

In [80] the authors considered on demand data collection. In this research, the sensor nodes broadcast data collection requests when their buffers are about to be full. Upon receiving such requests, the ferry moves toward the sensor nodes to collect the data and transfer it to the sink.

In [81] a mobile node was used to help in disseminating data to the sink. It was used to move back and forth along the linear network, and collect data from the individual sensors when it came within their communication range. The mobile node would then transfer the collected data to a base station. The mobile node was also used to perform other functions, such as data processing, aggregation, and could also transport messages from the sink to the sensor nodes.

Table 5-1 shows a literature review summary of the use of ferries in WSNs. The research is classified according to whether it uses a single-hop or multi-hop approach to forwarding data. Furthermore, the research is categorized according to whether all the nodes in the network are visited by the ferry or only a subset of nodes.

Table 5-1: Summary of related work

	Visits all nodes	**Visits subset of nodes**
Single-hop	[13], [14], [18], [10]	[16], [8]
Multi-hop	[9]	[15], [17], [11], [12]

5.5. Challenges in Using a Ferry in WSNs

Based on the previous background and literature review, several challenges can be identified in deploying ferries in WSNs. Below are some of these challenges which are still open to debate and can be tackled by future researchers.

5.5.1. Ferry Presence Detection

Detecting the presence of the ferry within the communication range is a very challenging issue, especially if the presence is brief and the path of the ferry is uncontrollable. Therefore, sensors need to be awake and in detection mode all the time to detect the presence of the ferry. This negates any efforts to conserve the nodes' energy.

5.5.2. Mobility of the Ferry

Ferries are rechargeable mobile elements that are used to carry data over distances to the base stations. Their mobility can be an issue when collecting the data from the sensors. Therefore, presence detection, speed, and the direction of the mobile ferry to enable nodes to send data and the ferry to collect data from the nodes in an efficient manner while preserving as much of the network nodes' energy as possible is an important challenge that needs to be met.

5.5.3. Efficient Energy Management Strategies

Energy management is an important issue in all networks. Efficient management can lead to prolonging the network lifetime of WSNs. Researchers in [82, 83, 84 and 85] surveyed the existing energy management schemes present in the literature for both static and mobile nodes. They found that keeping nodes in the awake mode consumes energy. A pre-defined policy for a mobile ferry to visit sensors in a WSN should be set when the motion of the ferry is controlled. As an example, the path, the speed, and the stopping periods of the ferry have to be defined in order to improve the performance of the network. This will allow nodes to sleep and wake up based on the ferry's schedule and proximity. As a result, the energy of the network will be preserved and the lifetime of the network will lengthen. If the path and schedule of the ferry is known or can be predicted, sensors can be awakened only when they expect the ferry to be within their communication range. This will further preserve the energy of the network. However, a further challenge would be to optimize the motion of the ferry in a controlled manner and thus efficiently manage the duty cycles of the sensing nodes depending on the deployed application.

5.5.4. Optimum Data Transfer

The communication time between static nodes and the mobile ferry might be short while a significant amount of data might need to be collected. Therefore, there is a need to provide coverage to the entire network and maximize the number of reliably sent messages to the ferry. In [86] the authors investigated how to efficiently collect data from stationary sensor nodes using multiple robotic vehicles, such as data ferries, under different circumstances. They proved that finding an optimum ferry path is an NP-hard problem. Therefore, finding a reliable and efficient path for the ferry to take to provide coverage for the entire network using the least energy and causing minimum latency is one of the most difficult challenges that needs to be investigated and addressed by researchers.

5.6. Summary

In this chapter, I have surveyed the recent progress made in using mobile ferries for data gathering in WSNs by addressing two areas: determining the path of the ferry and the scheduling of when to dispatch the ferry to collect data from static sensors. I presented a classification of mobile ferries based on the role they play in addition to carrying information. Furthermore, I surveyed the existing work on the path planning and scheduling of ferry dispatch. In addition, some of the common challenges in deploying mobile ferries in WSNs were discussed along with many of their possible applications.

Chapter 6: Ferry-based Gathering and Clustering Algorithm with Determined Path

6.1. Overview

Depending on the application, mobile ferries can be used for collecting data in a WSN, especially those on a large scale with delay tolerant applications. Unlike data collection via multi-hop forwarding among the sensing nodes, ferries travel across the sensing field to collect data. A ferry-based approach either eliminates or minimizes the need for the multi-hop forwarding of data, and as a result, energy consumption at the nodes can significantly reduced. This is especially true of nodes that are near the base station as they are used by other nodes to forward data. However, this increases data delivery latency and, as such, it might not be suitable for all applications.

In this chapter, an efficient data collection scheme using a ferry node is proposed with an emphasis on the effect of the ferry's path. In this scheme, the selection of cluster heads is based on their residual energy and their distance from the ferry's path. I simulated the proposed scheme in MATLAB using different scenarios to show their performance in terms of the network lifetime and total energy consumption in the network. I found that centered and diagonal fitted paths within the assumed sensing field performed better than the diagonal path in terms of the network lifetime and energy consumed. I also found that increasing the number of checkpoints increases the lifetime of the network but also increases delay.

6.2. Introduction

In this chapter, I propose a mobile ferry improved algorithm based on our previously published work on the node ranking clustering algorithm (NRCA) [87, 88]. Using the NRCA algorithm, the decision of selecting cluster heads is based on their residual energy and their distance from the base station where an energy threshold technique is used to replace cluster heads. In this chapter, the decision of selecting cluster heads is based on their residual energy and their distance from the planned ferry's path checkpoints. In addition, data is collected by the ferry instead of flooding the network with multi-hop forwarding. The network is divided into several clusters by the base station based on NRCA. Each cluster head collects data and sends it to the mobile ferry.

The rest of the chapter is organized as follows. My proposed data collection algorithm is described in Section 6.2. In Section 6.3, the performance evaluation in terms of the network lifetime is shown by using different criteria. Finally, Section 6.3 summarizes the chapter.

6.3. Ferry Node Ranking Clustering Algorithm (FNRCA)

In this chapter, I propose a ferry-based node ranking clustering algorithm (FNRCA) to collect data from the nodes. The difference between this algorithm and other algorithms is that it uses a more efficient mechanism to select cluster heads. This is achieved by measuring the distances, the current energy levels of the nodes, and calculating the number of rounds that each node can be a cluster head for, in order to maximize the network lifetime and decrease the excessive communication overheads used for electing new cluster heads. In this algorithm, nodes are ranked based on their current energy level (En) and their positions (Dn) with reference to the predetermined checkpoints on the ferry’s trajectory. This ranking is used to choose the cluster heads which are also sorted into levels based on their position, or Euclidean distance, from the checkpoints on the ferry’s trajectory. Therefore, each node is assigned a rank Rn (En, Dn) reflecting its candidacy for election as a cluster head.

The proposed algorithm is shown to be energy efficient because it minimizes the energy used by cluster heads to reach the BS by using a ferry. In the next subsection, I will introduce the proposed algorithm in more detail.

6.3.1. Assumptions

In the proposed algorithm, the base station (BS) is placed in a fixed position and has unlimited energy. Thus, no constraints are assumed with regard to power consumption due to data processing and communication. Moreover, it is assumed that the ferry dispatches from the base station and will return to it. In addition to that, it is assumed that there are no energy constraints on the ferry. Nodes are distributed randomly based on uniform distribution. Through the initial step of the algorithm described below, the BS becomes aware of the locations of all the sensor nodes either via collecting their GPS coordinates or by any other mechanism.

6.3.2. Description of the Algorithm

The proposed algorithm is an extension of our previously published work [87, 88] with node ranking being based on the planned path of the ferry. The following steps provide a description of the algorithm and the process of selecting the cluster heads:

- Similar to the initial step taken in [36, 39, 42, 45 and 52] each node at the set up phase broadcasts a message regarding its energy level and location to its neighbors. Therefore, each node sets up a neighbor information table recording the energy levels and positions of its neighbors and broadcasts this information to its neighbors. This is conducted by all nodes in the network until information about all the nodes in the network is received by the BS. This will provide the BS with a global knowledge of the network.
- The BS divides the area into smaller partitions called clusters based on the assumed minimum communication range of the nodes.
- The path of the ferry and checkpoints where the ferry will stop to collect data on its planned trajectory are predetermined by the BS and sent to the cluster heads.
- Nodes with the highest energy level (En) and least distance (Dn) from the closest checkpoint on the ferry's trajectory in each cluster become a cluster head (CH) after the first round is completed where cluster heads were chosen in reference to the BS using the NRCA.
- At each checkpoint, the ferry stops to collect the sensed data gathered from cluster heads associated with the checkpoint. Gathered data is collected either directly from the sensing nodes within these cluster heads' communication range or through multi-hop forwarding through other cluster heads for out-of-communication sensing nodes.
- Dissemination of data from cluster heads to the ferry is triggered by a control message communicated by the ferry to the cluster heads associated with each checkpoint. The length of time that the ferry will stay at each

checkpoint is determined based on several parameters as will be shown later.

- Cluster heads, which are located closer to the path of the ferry, are referred to as the first level cluster heads. The cluster heads that are located at more distant positions from the path are considered second level, third level, etc. Higher-level cluster heads transmit to lower-level cluster heads in order to reach the ferry with the least energy consumption.
- The used energy model for sensing and disseminating data in our simulation is the same used by [42, 52] as was described earlier in Section 4.1.1.

6.3.3. Cluster Head Selection Process

After the initial forming of clusters, the BS assigns a cluster head for each cluster based on NRCA. Nodes in each cluster are ranked based on how far they are from the path of the ferry and on their current energy level. Nodes with the maximum residual energy and minimum distance will be chosen as a cluster head based on NodeRanking (En, Dn) where

$$[\mathrm{Dn(i)} = \mathrm{Min(D(i,\ Closest_{CP}\ )\ )}\ ,(\mathrm{En(i)}) = \mathrm{Max(ResidualEnergy(i)}] \quad (1)$$

$$\left|\mathrm{D(i, Ferry\_path\_CP)}\right| = \sqrt{\left(\mathrm{X_i} - \mathrm{X_{cp}}\right)^2 + \left(\mathrm{Y_i} - \mathrm{Y_{cp}}\right)^2} \quad (2)$$

ResidualEnergy (En (i)) is the current energy level of the node i; D(i, Ferry_path_CP) is the Euclidean distance of node i to the closest checkpoint on the ferry's path. Given a particular deployment region of interest, X_i and Y_i are the X and Y positions of node i. X_{cp} and Y_{cp} are the X and Y positions of the closest checkpoint on the ferry's path

A cluster head in each cluster will be changed when its energy level reaches a pre-defined threshold or a calculated value and not at every round. This will make it possible for a node, i, to continuously play the role of a cluster head for multiple rounds and thus prevent wasting energy on control and exchanging messages to replace it.

6.3.4. Ferry's Stopping Time at Each Checkpoint

The stopping time (ST) is the period of time that the ferry will stay at each checkpoint, j, to allow the associated cluster heads to send their gathered data to the ferry. This time period depends on the number of associated cluster heads, their buffer sizes, and the transmission time of a bit.

$$\mathrm{ST}\ (\mathrm{j}) = \mathrm{BuffSize}(\mathrm{j}) \times \mathrm{numberOfAttachedCHs} \times \mathrm{timeToTransmitAbit} + \mathrm{T} \tag{3}$$

where BuffSize is the cluster head memory size in bits, the numberOfAttachedCHs is the number of cluster heads associated with the checkpoint j, timeToTransmitAbit is the time needed to transmit a bit of information to the checkpoint and T is an assumed constant delay added to account for propagation delay.

6.3.5. Problem Formulation

Given a set of cluster heads, n, in a multi-hop-based WSN, our aim is to use a ferry to collect gathered data from the cluster heads based on a pre-defined path while minimizing the overall energy consumed during such a process to prolong the network lifetime. I formulated our problem so that the ferry will take two paths. The first one will be a diagonal line across the middle of the field. The diagonal line can be adjusted to move closer to the cluster heads that have lower values of energy. In the second path, the ferry will move along a line in the center of the field as shown in orange in Figure 6-1. Along both paths, there will be checkpoints where the ferry will stop to collect data from the cluster heads.

Table 6-1: Parameters used in the simulation, values for the various energy parameters as per the energy model used by [25, 33, 35]

Notation	**Description**
N = 400	Total number of sensor nodes
E_o = 0.5J / node	Initial energy of each node
E_{elec} = 50nJ / bit	Per bit energy consumption
E_{DA} = 5nJ / bit	Energy for data aggregation
E_{amp} = 100 pJ/bit/m^2	Amplifier transmitting energy
Area = 200 x 200	Area used in the simulation in meters
# of cluster heads/ # Checkpoints = 10	Ratio of checkpoints to cluster heads
Packet size	256 bits
Data rate	256 Kbps
Cluster radius, coverage radius	30 m
Sensing radius	30 m
Buffer size	256 K Bytes

6.4. Performance Evaluation

To evaluate the performance of the proposed FNRCA algorithm, I used MATLAB to simulate four scenarios on a 200 x 200 m^2 sensing field. In the first scenario, I set the trajectory of the ferry to be diagonal while in the second scenario I fit the diagonal trajectory using curve fitting based on a one-degree polynomial function to move closer to the cluster heads with lower energy values. In the third scenario, I set the trajectory of the ferry to follow the center of the sensing field, while the fourth scenario represents our previous work, the NRCA algorithm, without using a ferry node for comparison. The base station was placed in three different locations: at the center of the field (x = 100, y = 100), at (0, 0) and at (0, 100). The ferry was dispatched from the base station along the planned trajectory path. I ran the simulation five times and took the average of the runs to present our results. Table 6-1 shows the parameters used in this simulation environment which are standard parameters used by all researchers in this field. Every node was given an initial energy of 5 J. The energy for data aggregation is 5 nJ/bit. The energy to run the radio is 50 nJ/bit. The amplifier transmitting the energy is 100 nJ/bit/m^2. Using a simulation, I considered the network lifetime metric to evaluate the performance of the four aforementioned scenarios.

6.4.1. Simulated Scenarios

As shown in Figure 6-1, the ferry will move along the diagonal path of the sensing field. It will move back and forth on this path while stopping at the checkpoints to collect the data from cluster heads then disseminate it to the BS. In the second scenario, the ferry will move back and forth on the path where the diagonal line is fitted to move closer to the cluster heads with lower values of energy. Curve fitting, using a one-degree polynomial function, was used to fit the line by assigning cluster heads residual energy values as a weight. The fitted line will move closer to the cluster heads with less energy. In the third scenario, the ferry will move on the horizontal line crossing the middle of the field. The fourth scenario is based on our previous NRCA algorithm without using any ferry nodes. When using a ferry, checkpoints are distributed along the path with a ratio in reference to the number of cluster heads, for instance, areas with more cluster heads will have a higher number of checkpoints.

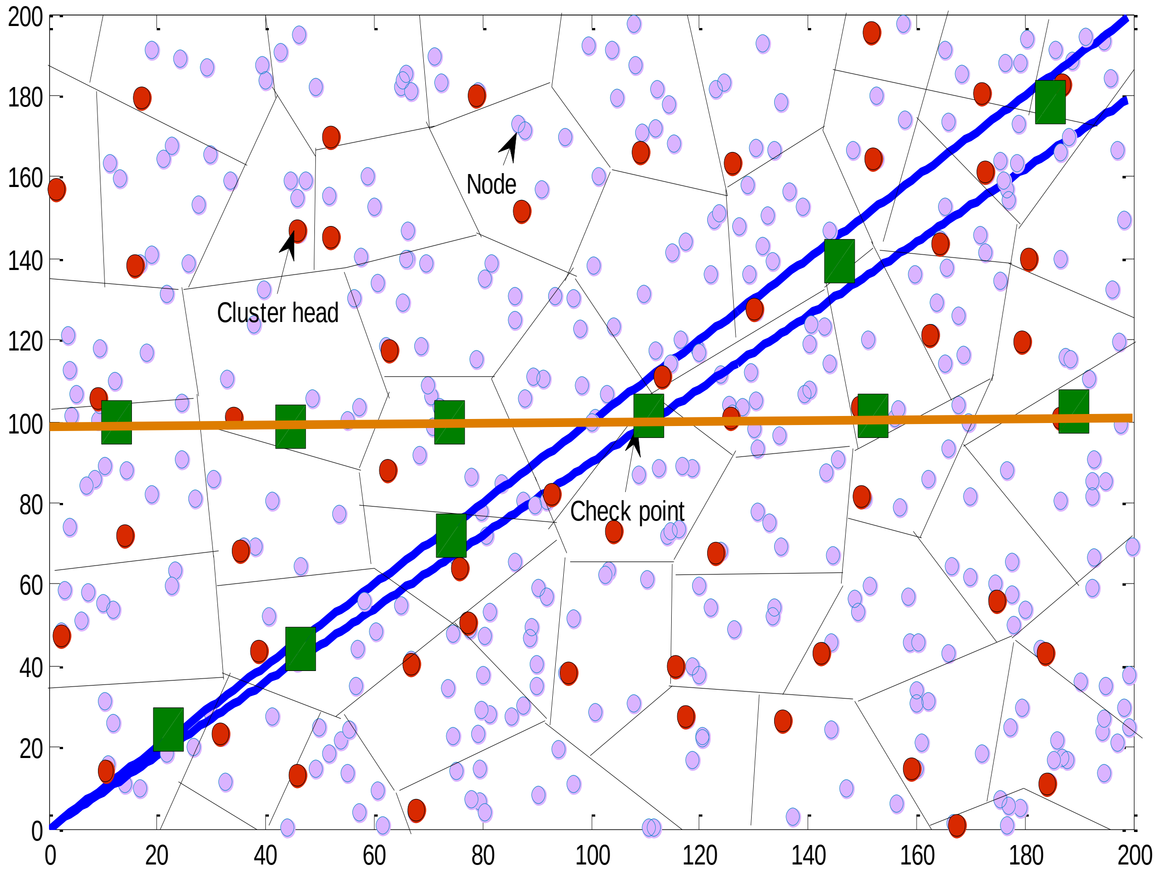

Figure 6-1: Paths of the ferry with checkpoints

6.4.2. Network Lifetime

Network lifetime is defined here as the time interval from the moment the sensor network begins its operation until the death of the last node in the network. From Table 6-2, we can see that the last node in NRCA died at round 3300, making it the lowest achiever with the shortest network lifetime when compared to the other scenarios. On the other hand, we can see that the centered and the diagonal fitted paths had longer network lifetimes as their last nodes died at rounds 3860 and 3837, respectively. This can be explained by the fact that the cluster heads on the opposite diagonal corners will be far from the path and their multi-hopping chain to reach the path will be longer. Also in the fitted diagonal, this result can be justified because the path will be closer to the cluster heads with less energy, which means that they will consume less energy to reach the checkpoints. The placement of the base station did not affect the result as I got the same result for the different placements of the base station.

Table 6-2: Simulation results for the network lifetime

Protocols	**Measurements**	
	Round first node died	**Round last node died**
Diagonal path	1479	3556
Fitted diagonal path	1760	3837
Center line	1810	3860
NRCA	1300	3300

6.4.3. Energy Consumed

As shown in Figure 6-2, the energy consumed per round in the fitted path is less than the diagonal unfitted one. Allowing the ferry to move closer to the cluster heads with lower energy values helps in reducing the energy consumption in these cluster heads and as a result, it prolongs the lifetime of these cluster heads and preserves the overall energy of the whole network.

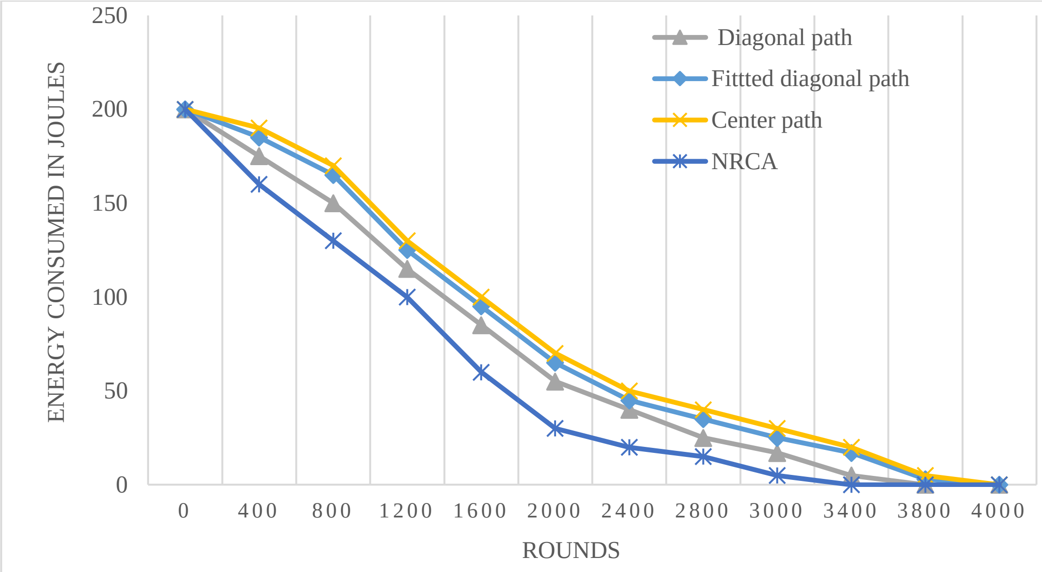

Figure: 6-2 Energy consumption in the network

6.4.4. Changing the Number of Checkpoints

I also simulated the second scenario while changing the ratio of the checkpoints to one checkpoint for every 20 cluster heads, one checkpoint for every 10 cluster heads, and one checkpoint for every five cluster heads. Table 6-3 shows the network performance based on changing the number of checkpoints. From Table 6-3 we can see that the network lifetime increases as the number of checkpoints increases. This is because the more checkpoints, the less distance the data will travel which saves the energy of the cluster heads and the overall energy of the network.

Table 6-3: Changing the number of checkpoints

# of Checkpoints	**Measurements**	
	Round first node died	**Round last node died**
# of cluster heads/ # Checkpoints = 15	1560	3600
# of cluster heads/ # Checkpoints = 10	1760	3817
# of cluster heads/ # Checkpoints = 5	1913	3910

6.5. Summary

In this chapter, an efficient data collection scheme using a ferry node was proposed with an emphasis on the effect of the predetermined ferry's path. In this scheme, the decision of which cluster heads to select is based on their residual energy and their distance from the ferry path. The proposed scheme was simulated in MATLAB using different scenarios to show their performance in terms of the network lifetime and total energy consumption in the network. I found that the centered and the diagonal fitted paths performed better than the diagonal path in terms of the network lifetime and energy consumed. I also found that increasing the number of checkpoints increases the lifetime of the network

Chapter 7: Ferry-based Gathering and Clustering with Undetermined Paths

7.1. Introduction

In this chapter, I propose a mobile ferry algorithm based on our previously published work, the node ranking clustering algorithm (NRCA) [87, 88]. Using NRCA, the decision of selecting cluster heads in a WSN is based on their residual energy, their distance from the base station, and an energy threshold that is used to replace the cluster heads. In this algorithm, the decision of selecting cluster heads is based on their residual energy and their distance from the ferry's path which is composed of checkpoints (CPs). The checkpoints' positions will initially be decided by deploying a virtual grid on the field and placing a checkpoint in the center of each grid. The checkpoints will then be changed based on its number of attached cluster heads. The Traveling Salesman Problem (TSP) will be used to find a Hamiltonian cycle to decide the path of the ferry. Checkpoints will represent the vertices and the distances between them will represent the edges. A cost function will be used to decide which vertices will be visited first so that the overall cost will be minimized. Since TSP is NP-hard [89, 90], when the number of stops to be made is greater than four, a genetic algorithm will be used to choose the sequence of checkpoints to be visited. The main contribution of this algorithm is in finding near optimal (in terms of consumed overall energy and round trip traveling time) random path for the ferry to follow to collect data from the sensor network.

The rest of the chapter is organized as follows. In Section 7.2, a summary of the current and closely related work is provided. Our proposed data collecting algorithm is described in Section 7.3. In Section 7.4, a performance evaluation in terms of the network lifetime of the proposed algorithm is shown by using different criteria. Finally, Section 7.5 summarizes the chapter.

7.2. Background Work

Using mobile ferries in WSNs is a relatively new area of research which is rapidly attracting the attention of many researchers. Incorporating ferries in WSNs helps to eliminate the need for the multi-hop forwarding of data [86]. It also reduces the energy consumption at the nodes. However, using ferries might add a delay in the collection, dissemination, and processing of data and thus might not be suitable for all applications.

In [91] the authors proposed path-planning algorithms for an autonomous underwater vehicle (AUV) which acts as a mobile sink node for the underwater sensor nodes. They used Value-of-Information (VoI) as the metric for choosing the path of the AUV. The VoI serves as a marker for evaluating the quality of information with respect to the collection time of that data.

The authors in [74] used a single ferry to collect data from a circular dense sensors network. They showed that the optimal mobility strategy of the ferry was achieved when moving at the border of the sensing area. They divided the area into circles starting from the source. The inner circles forward the data to the outer ones until the border was reached where the ferry was used to collect the sensed data. Thus multi-hop forwarding was used to finally reach the ferry. In [77, 78] the ferry visits exact rendezvous points to collect data. These points buffer and aggregate data to the ferry from the nodes though multi-hop forwarding. In [92] the WRP (weighted rendezvous points) algorithm was proposed where nodes are used as rendezvous points. Cluster heads and nodes send their collected data to these points through multi-hop forwarding. The tour path of the ferry to these points is built by assigning a weight to each one as represented by the distance in the number of hops from the path and the number of data packets each node is forwarding to the closest point. In [93] the authors chose cluster heads with the highest energy as rendezvous points and then built the tour of the mobile sink to these energy-rich cluster heads to collect data. In Section 7.4.6, I will compare our FNRCA against the WRP algorithm and the one used in [93].

Our proposed approach is different from previously published work insofar as the ferry does not have to visit each node in the network to collect information from it. Instead, the area will be divided into virtual grids and a checkpoint will be placed in each grid. The ferry will only visit these checkpoints to collect data. Our approach also uses TSP and a genetic algorithm to choose the optimum path of the ferry which consists of visiting a list of sequenced checkpoints. The sequence of checkpoints that will be visited will be decided by assigning a weight to each checkpoint and deciding which checkpoint will be visited first. Moreover, the NRCA algorithm will be applied in each virtual grid to decide on the best placement position for each checkpoint in order to preserve the energy of the whole network. Our aim is to minimize the overall round trip traveling time of the ferry and to minimize the energy consumed in the network. This is achieved by modifying NRCA to be applied in reference to the position of the checkpoint rather than the position of the base station (i.e., the sink). By doing this, each checkpoint will act like a virtual sink within each virtual grid. I referred to the new modified NRAC algorithm as ferry-based NRCA or FNRCA. In FNRCA distance used to rank the nodes is in reference to the checkpoint position rather than the position of the base station.

7.3. Ferry Node Ranking Clustering Algorithm (FNRCA)

In this chapter, I will propose a ferry-based node ranking clustering algorithm (FNRCA) to collect data from nodes. The difference between this algorithm and other algorithms is that this algorithm uses a more efficient mechanism to select cluster heads (CHs). This is done by measuring the distances, the current energy levels of nodes, and calculating the number of rounds for which each node can be a cluster head, in order to maximize the network lifetime and decrease excessive communication overheads used to elect new cluster heads. In this algorithm, nodes are ranked based on their current energy level (En) and their positions (Dn) with reference to the predetermined checkpoints on the mobile ferry's trajectory. This ranking is used for choosing cluster heads which are also sorted by levels based on their position, or the Euclidean distance from the ferry's checkpoints. Therefore, each node is assigned a rank Rn (En, Dn) reflecting its candidacy as a cluster head.

In our algorithm, once the ferry reaches a checkpoint, it broadcasts a notification message to all nodes in its communication range informing them of its presence at the respective checkpoint with which they are associated. Nodes within each cluster will then start sending any sensed data to their associate cluster heads to be transmitted to the base station. The number of cluster heads attached to that particular checkpoint, as will be demonstrated, determines the ferry's stopping time. Using this strategy, cluster heads will not have to worry about the speed or the direction of the ferry and energy that would otherwise be wasted by doing this will be preserved.

The algorithm also provides an efficient energy management strategy wherein cluster heads are only awakened when the ferry sends them a notification message to inform them of its presence. The idea of using the ferry's passing of cluster heads to collect data further preserves energy by reducing multi-hop forwarding which drains the cluster heads' energy throughout the network. To optimize the ferry's path, a weight is assigned to each checkpoint to be able to choose the best sequence, the order of checkpoints to be visited, and the necessary stopping time at each one. This eliminates the loss of messages due to any inaccurate prediction of the position of the ferry or its movement. Our algorithm uses three phases, as shown below in Figure 7-1.

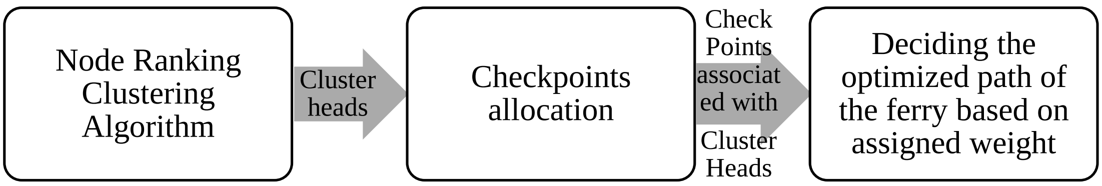

Figure 7-1: Illustration of phases used by FNRCA

The proposed algorithm is shown to be energy efficient because it aims to minimize the energy consumed in the network in the process of collecting and transferring data to the BS by using a mobile ferry. In the next section, I will introduce the proposed algorithm in more detail.

7.3.1. Node Ranking Clustering Algorithm

In this proposed algorithm, several assumptions are made: first, the base station (BS) is placed at a fixed position and has unlimited energy. Thus, no constraints are assumed with regard to power consumption due to data processing and communication. Second, all nodes are assumed to have the same energy level at the set up phase which is known to the BS. Third, the sensing field dimensions are also assumed to be provided to the BS. Fourth, it is assumed that the mobile ferry is dispatched from the base station and returns to it once its task is completed. In addition, it is assumed that there are no energy constraints with respect to the ferry which is assumed to be moving at a fixed speed. Nodes throughout the sensing field are randomly and uniformly distributed.

7.3.2. Description of the Algorithm

The proposed algorithm is an extension of our previously published work NRCA [87, 88] with node ranking being based on the planned path of the ferry rather than the location of the BS. The following steps provide a description of the algorithm and cluster heads' selection process:

- After clustering, the sensing field will be divided into virtual square grids based on the specified maximum sensing range. Each virtual grid will be of the size $r \times r$ where r is the maximum sensing range. Multiple clusters fall within one or more virtual grids.

- Initially, a ferry checkpoint (virtual base station) is placed at the center of each virtual grid.

- Initially, NRCA is used to choose CHs based on their location from the ferry's checkpoints.

- Nodes and cluster heads will associate themselves with the ferry's checkpoint based on their location within each virtual grid.

- Borderline nodes and cluster heads will be associated with cluster heads and checkpoints closer to them based on distance, respectively.

- After the initial phase, NRCA is applied in each virtual grid based on the position of the ferry's checkpoint and the energy values of the associated nodes. Therefore, the energy consumed per virtual grid will be minimized. This is explained below in subsequent sections.

- The ferry will be dispatched from the BS to visit all checkpoints and return to the BS using a Hamiltonian cycle, as will be demonstrated.

- At each checkpoint, the ferry stops to collect the gathered data from the cluster heads associated with it. Gathered data consists of sensed data and control information, like a node's energy values and a node's GPS location.

- Dissemination of data from cluster heads to the ferry is triggered by a control message communicated by the ferry to the cluster heads associated with each checkpoint. The time spent by the ferry at each checkpoint is determined based on several parameters as will be described.

- In the subsequent rounds of dispatching the ferry, the BS chooses the new locations of the checkpoints based on the collected information to minimize the energy of the overall sensing field, as will be shown. The BS will then determine the new path of the ferry by using the Hamiltonian cycle, as was carried out in the initial phase.

- The used energy model for sensing and disseminating data in our simulation is the same that is used by [42, 45 and 52], as was described in Section 4.1.1.

7.3.3. Cluster Head Selection Process

After the initial forming of clusters, and based on the information collected through the first dispatching round of the ferry, the nodes in each cluster are ranked by the BS based on their distance from the checkpoint to which they are attached and their current energy level. This information is dispatched back to the nodes through the next ferry visit. Nodes with the maximum residual energy and minimum distance will be chosen as a cluster head based on NodeRanking (En, Dn) where

$$[\mathrm{Dn(i)} = \mathrm{Min(D(i,\ Closest_{CP}\ )\ )}\ \ , \mathrm{(En(i))} = \mathrm{Max(ResidualEnergy(i)]} \quad (1)$$

where

$$\left|\mathrm{D(i, Closest_{CP})}\right| = \sqrt{\left(\mathrm{X_i} - \mathrm{X_{cp}}\right)^2 + \left(\mathrm{Y_i} - \mathrm{Y_{cp}}\right)^2} \quad (2)$$

and ResidualEnergy $(En\ (i))$ is the current energy level of node i, $D(i, Closest_{CP})$ is the Euclidean distance of node i to the closest checkpoint. Given a particular deployment region of interest, X_i and Y_i are the X and Y positions of node i. X_{cp} and Y_{cp} are the X and Y positions of the closest checkpoint on the sensing field.

A cluster head in each cluster will be changed when its energy level reaches a pre-defined threshold or a calculated value and not every sensing round. This will make it possible for ith node to continuously play the role of a cluster head for multiple sensing rounds and thus prevent wasting energy on the control and exchange messages that would otherwise be sent to replace it.

In the next two subsections, I will discuss the placement of the checkpoints and the amount of time the ferry will spend stopped at each checkpoint.

7.3.4. Ferry Checkpoint Locations

To decide the location of the initial ferry's checkpoints, I first create virtual grids based on the specified maximum sensing range. Each virtual grid will be of the size $r \times r$. A checkpoint will be initially placed in the center of each square in the virtual grid. Then NCRA will be applied to each square in the grid where nodes will be ranked according to their energy levels and their distance from the checkpoints. After the first round of the ferry, the checkpoints' positions will be changed in each grid by the BS based on the related information collected by the ferry in the first dispatched round. Each checkpoint in each virtual grid will be placed closer to the larger number of neighboring cluster heads. The checkpoint coordinates, $\mathrm{Xcp(j)}$ and $\mathrm{Ycp(j)}$, in each virtual grid are calculated by the following equations:

$$\mathrm{Xcp(j)} = \frac{1}{\mathrm{N_j}} \sum_{\mathrm{k=1}}^{N_{\mathrm{j}}} \mathrm{X_k^j} \qquad (3)$$

$$\mathrm{Ycp(j)} = \frac{1}{\mathrm{N_j}} \sum_{\mathrm{k=1}}^{N_{\mathrm{j}}} \mathrm{Y_k^j} \qquad (4)$$

where $\mathrm{N_j}$ is the total number of attached cluster heads associated with the same checkpoint.

The following is the pseudo code for choosing the checkpoint location in each grid:

- **Input**: a subset of cluster heads cpch in each virtual grid, the virtual grid dimensions and the sensing Range r between the ferry and clusters heads;
- **Output**: if the subset of all cluster heads which can be covered by a circle with a radius at most r, return to the circle's center (Eq. 3 and 4) or false otherwise and no change in the checkpoint position i.e. it will be its previous position.
- **if**
- radius > r **then**
  - **return** false; // no change in checkpoint position
- **else**
  - center(x,y)=(Eq.3 and 4)
  - **return** center. // checkpoint position will be the center (x , y)

- **end if**

7.3.5. Stopping Time of the Ferry at Each Checkpoint

The stopping time (ST) is the period of time the ferry will spend at each checkpoint, j, to allow the associated cluster heads to send their gathered data to the ferry. This time period depends on the number of associated cluster heads, their buffer sizes, and the transmission time of a bit based on the assumed medium physical characteristic.

$$\mathrm{ST}\ (\mathrm{j}) = \mathrm{BuffSize(CHj)} \times \mathrm{NumberOfAttachedCHs} \times \mathrm{TimeToTransmitAbit} + \mathrm{T} \qquad (5)$$

where $BuffSize$ is the cluster head memory size in bits, the $NumberOfAttachedCHs$ is the number of cluster heads associated with checkpoint j, $TimeToTransmitAbit$ is the time needed to transmit a bit of information to the checkpoint and T is an assumed constant delay added to account for propagation delay.

7.3.6. Problem Formulation

Given a set of cluster heads **CHs**$[CH_i, i = 1, .., n]$ and a set of checkpoints **CPs**$[CP_j, j = 1, .., m]$ in a multi-hop WSN, the dispatched ferry needs to move along a path to collect data from associated nodes when stopping at the checkpoints before returning to the base station while satisfying the following two main goals:

$$\mathrm{Tour\_Time(s)} = \mathrm{Travel\_Time\ (s)} + \textstyle\sum_{\mathrm{j}=1}^{m} \mathrm{ST_s(j)} \qquad (6\ \mathrm{a})$$

$$Tmax = \mathrm{Max}(\mathrm{Travel\_Time\ (s)} + \sum_{\mathrm{j}=1}^{m} \mathrm{ST_s(j)}\ ) \qquad (6\ \mathrm{b})$$

such that

$$\text{Tour\_Time}(s) < Tmax \tag{7}$$

where s = 1,2,3… corresponds to the round of data collection, Tour_Time is the total the round trip traveling time of the ferry from the base station to each checkpoint j plus the stopping time at each checkpoint of round s:

$$\text{Travel\_Time (s)} = \frac{\mathrm{D_s(BS, CP1)} + \sum_{j=1}^{m-1} \mathrm{D_s(CP(j), CP(j+1))} + \mathrm{D_s(CP(m), BS)}}{\text{Ferry\_Speed}} \tag{8}$$

where m is the total number of checkpoints, n is the number of cluster heads, Travel_Time is the round trip traveling time of the ferry from the base station to each checkpoint j plus the stopping time at each checkpoint. $D(BS, CP1)$ is the distance from the base station to the first checkpoint and $D(CP_m, BS)$ is the distance from the last checkpoint visited by the ferry to the base station. $Ferry\_Speed$ is the assumed fixed speed of the ferry.

The second goal is to minimize the overall energy consumption of the network by applying our FNRCA in each virtual grid, i.e. finding the

$$Min \sum_{i=1}^{n} \mathrm{E(i)} \tag{9}$$

and by minimizing the sum distance from the checkpoints and their associated cluster heads as,

$$Min \sum_{cp=1}^{m} \sum_{ch=1}^{n} \mathrm{D(ch, cp)} \tag{10}$$

such that

$$\forall\, ch \in CHS,\ \exists\, cp \in CPS : D(ch, cp) \leq r \tag{11}$$

where n is total number of cluster heads, ch is a cluster head, CHS is the cluster heads list, cp is a checkpoint CPS is the list of checkpoints and r is the max sensing Radius.

In order to choose the optimum path of the ferry and achieve the above goals and constraints, a weighting scheme is used to order the checkpoints in the sequence in which they will be visited by the ferry.

7.3.7. Checkpoint Weighting Scheme

To determine the path of the ferry or which checkpoints to visit first, a weighting scheme is used based on determining the following weights:

- Checkpoints with a larger number of attached cluster heads:

Checkpoints with a larger number of attached cluster heads will contribute more to the amount of data collected and in order to reduce data loss, they will be prioritized and visited first by the ferry. The weight for such a CP, j, is calculated as:

$$\mathrm{W}_1(\mathrm{j}) = \frac{\mathrm{Attached}\,clusterheads(j)}{N_T} \qquad (12)$$

where N_T is the total number of cluster heads in the network.

- Checkpoints closer to the base station:

Checkpoints closer to the base Station will be given a higher weight in order for the ferry to start the collection process there first and then move to ones that are further away. Their weight will be calculated as:

$$W_2(j) = \frac{1}{D(CP(j), BS)} \tag{13}$$

- Checkpoints closer to each other:

Checkpoint closer to each other will reduce the travel time and distance covered by the ferry; therefore, they will have a higher priority when it comes to being visited first by the ferry and their weight will be calculated as:

$$W_3(j) = \frac{1}{\min\left[D\big(CP(j), CP(l)\big)\right]}, l = 1,2,..m, l \neq j \tag{14}$$

The overall weight (W) is computed as:

$$\max W = \sum_{j=1}^{m}(W_1(j) + W_2(j) + W_3(j)) \tag{15}$$

Following is the pseudo code for ordering the checkpoint in the Traveling Salesman Problem sequence to be visited by the ferry according to the weight given to them:

- **Input**: a set of checkpoints, their attached cluster heads.
- **Output**: A sequence of checkpoint for the ferry to follow.
- //Optimal Traveling_Salesman_Problem_tour
- **while** there exist checkpoints **do**
    - **for all** CPj (j = 1, 2, ..., m − 1) **do**
      find the weight W from Eq. 15.
    - **End for**
- Select CP with maximum weight
- Add it to the TSPtour list {CPj,CPj+1…}
- Remove it from the set
- **end while**
- **return** TSPtour

In our work, TSP will be used to find a Hamiltonian cycle to decide the path of the ferry, where checkpoints represent vertices and the distance between them will represent the edges. The above weight will be used to choose which vertices will be visited first such that the overall consumed energy and round trip traveling time will be a minimum.

Assuming a directed graph (G) with weights on the edges where $G = (Vertex, Edge)$ we will find a Hamiltonian cycle where the cycle covers all the vertices only once and seeks a minimal weight subset of edges.

The abovementioned problem can be solved easily in a short time if the number of checkpoints are four or less by trying all possible paths (4!) and finding the minimum weight among them. However, if the number of checkpoints is more than four, there will be permutations of possible paths which will take a much longer time and require greater processing capabilities. Therefore, I used a genetic algorithm to find the best path based on our own fitness function, goals, and weights as shown in the subsections below.

7.3.8. Applying a Genetic Algorithm to Elect a Path

Since TSP is an NP-hard problem [89, 90, 94, 95, 96, 97 and 98] I used a genetic algorithm to find the optimum sequence of checkpoints to be visited by the ferry. Genetic algorithms are heuristic approaches which can be used to solve the TSP. They use simple chromosomes to encode solutions of data and apply crossover and mutation operators to these chromosomes to find an optimum solution. Good solutions will be selected by the fitness function and reproduced to produce a better solution, while the bad ones will be removed. After several generations, the genetic algorithms will produce an optimum solution to the problem.

In this work, I represented the ferry's path as a list of genes or chromosomes where each checkpoint will have a number to identify it, e.g. $(1, 2, 3, 4 \ldots\ldots CPm)$ and the path or the solutions will be represented by the ordered sequence of checkpoints. Zero is used to represent the base station. The path will start and stop at the base station, so each path will contain 0 at the start and end of its sequence. An example of a path representation will be [0, 3, 1, 2, 4, 0]. Below is the pseudo code for the genetic algorithm used:

- **Input:** p(t) and c(t) are parent paths and offspring candidate paths in current generation t.
- // Input will be taken from the previous pseudo code
- **Output:** The optimum solution TSP.
- T←0;
- Initialize p(t);
- Evaluate p(t);
- **While** (there exist p(t)) **do**
  - Perform crossover and mutation p(t) to get c(t);
  - Evaluate c(t) with the fitness function(c(t));
  - Select p(t+1) from p(t) and c(t);
  - T ← t+1;
- **End While**
- **End**

7.3.8.1. Crossover Operation

I used an ordered crossover (OX) in our genetic implementation which was used in BERLIN52, which the best known program for TSP so far [90, 94, 96 and 98]. Given two parent chromosomes, two random crossover points are selected, thus partitioning them into left, middle, and right portions. The child inherits its left and right portions from Parent 1, and its middle section is determined by their order and position from the Parent 2. An example of ordered crossover is shown below:

Given the following two paths:
1st Path = (1 2 3 4 5 6 7 8 9)
2nd Path = (4 5 2 1 8 7 6 9 3)
Based on the used OX genetic implementation [98], the following steps are performed:

1. Partition each path into three segments (left, middle, right)

1st Path = (1 2 3 | 4 5 6 7 | 8 9)
2nd Path = (4 5 2 | 1 8 7 6 | 9 3)

2. Copy the middle segment of both paths, the two candidate paths become as follow:
   1st candidate path = (- - - | 4 5 6 7 | - -)
   2nd candidate path = (- - - | 1 8 7 6 | - -)
3. Reorder each of the sequences starting from the right segments according to their order in the second path without repeating the already copied numbers
4. Generate new candidate paths as:
   1st candidate path = (2 1 8 | 4 5 6 7 | 9 3)
   2nd candidate path = (3 4 5 | 1 8 7 6 | 9 2)

7.3.8.2. Mutation Operator

The resulting children from an ordered crossover operation will now be subjected to the mutation operator in the final step to form a new generation. This operator randomly flips or alters one or more bit values at randomly selected locations in a chromosome. An example is shown below where 8 has been altered to 9:

Path 1	=	(1 2 3 4 5 6 7 [8])
Candidate path 1	=	(1 2 3 4 5 6 7 [9])

For implementing mutation in MATLAB I used the "*MutationFcn*" command.

7.3.8.3. Fitness Function

The fitness function is used to measure the goodness of the produced children in terms of pre-defined goals where bad solutions are eliminated and good solution are kept. Our two goals, as shown in Equations 7 and 9, are first, to evaluate the total traveling time of the ferry, and second, to evaluate the total energy consumed in the whole network subject to the constraint in Equation 11. Based on the first goal, the fitness function "*Time_Fitness_Fun*" in Equation 16 will evaluate the traveling time where the shorter the traveling time the better the path will be, however if the traveling time is greater than Tmax, is eliminated by assigning a negative value to the function represented by, $-\infty$, to exclude this solution from the solutions set. From the second goal, the smaller the total energy consumed that the path gives, the fitter the solution will be. Such paths will be preserved to be used to produce a better solution.

For the implementation of the fitness functions, I used the MATLAB "fitnessfcn" command given by:

$$Time\_fitness\_Fun = \begin{cases} Tour\_Time(s), & Tour\_Time(s) < Tmax \\ -\infty\,, & Tour\_Time(s) \geq Tmax \end{cases} \quad (16)$$

$$Energy\_fitness\_Fun = \sum_{i=1}^{n} E(i) \quad (17)$$

7.4. Performance Evaluation

To evaluate the performance of the proposed FNRCA algorithm, I used MATLAB to simulate the algorithm on a 200 meter x 200 meter sensing field. Table 7-1 shows the parameters used in this simulation environment which are standard parameters used by all researchers in this field. Every node was given an initial energy of 5 J. The energy for data aggregation is 5 nJ/bit. The energy to run the radio is 50 nJ/bit. The amplifier transmitting energy is 100 nJ/bit/m^2. The packet size is 256 bits. The data rate is 256 Kbps. Ferry speed is 100 meters/min which represents a fast walk. Using the simulation, I considered the network lifetime, energy consumed, and the total time of one tour of the ferry metrics to evaluate the performance.

Table 7-2: Parameters used in the simulation, values for the various energy parameters as per the energy model used by [25, 33, 35, 38]

Notation	Description
N = 400	Total number of sensor nodes
E_o = 0.5J / node	Initial energy of each node
E_{elec} = 50nJ / bit	Per bit energy consumption
E_{DA} = 5nJ / bit	Energy for data aggregation
E_{amp} = 100 pJ/bit/m^2	Amplifier transmitting energy
Area = 200 x 200	Area used in the simulation in meters
# Checkpoints	Varies according to the sensing range and the area : Area/sensing Raduis r
Packet size	256 bits
Data Rate	256 Kbps
Max sensing Radius : r	60-100 m
Buffer size	256 K Bytes
Tmax	Time of the longest tour of the ferry
Ferry_speed	100 m/min

7.4.1. Simulated Scenarios

As shown in Figure 7-2 (a and b), the ferry will follow a nonlinear path leaving from the base station, which will be across the center of the sensing field. It visits each checkpoint only once per round to collect the data from the cluster heads and carry this data back to the BS. I showed in our previously published work [99] that a centered predetermined path outperformed the diagonal path in terms of the network lifetime and energy consumed. In the figure, four checkpoints in Figure 7-2(a) and nine checkpoints in Figure 7-2(b) are used.

To evaluate the performance, I looked at the network lifetime, energy consumed, and the duration of the overall round trip as will be shown in the subsections below. I ran the simulation five times and took the average of the runs to present our results.

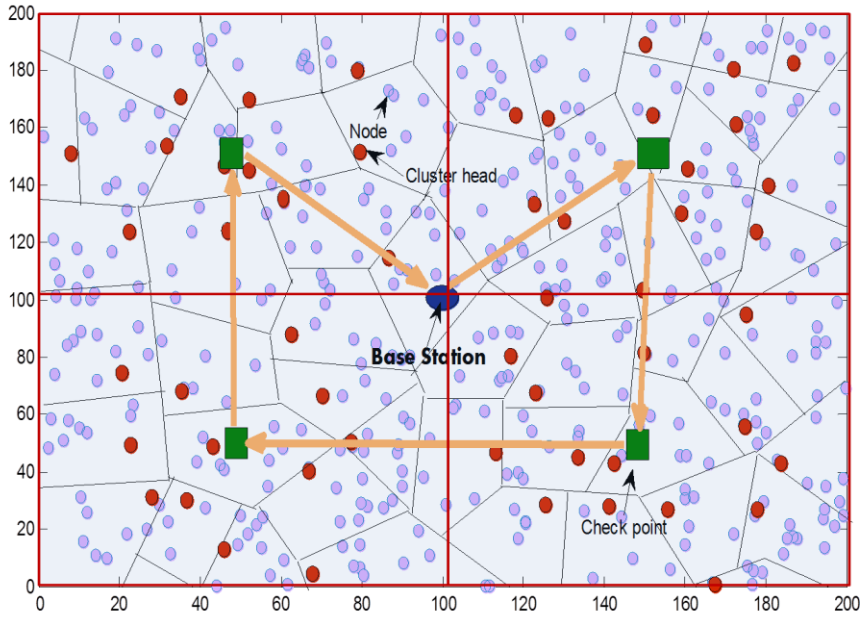

Figure 7-2 (a): Path of the ferry where 4 checkpoints are used

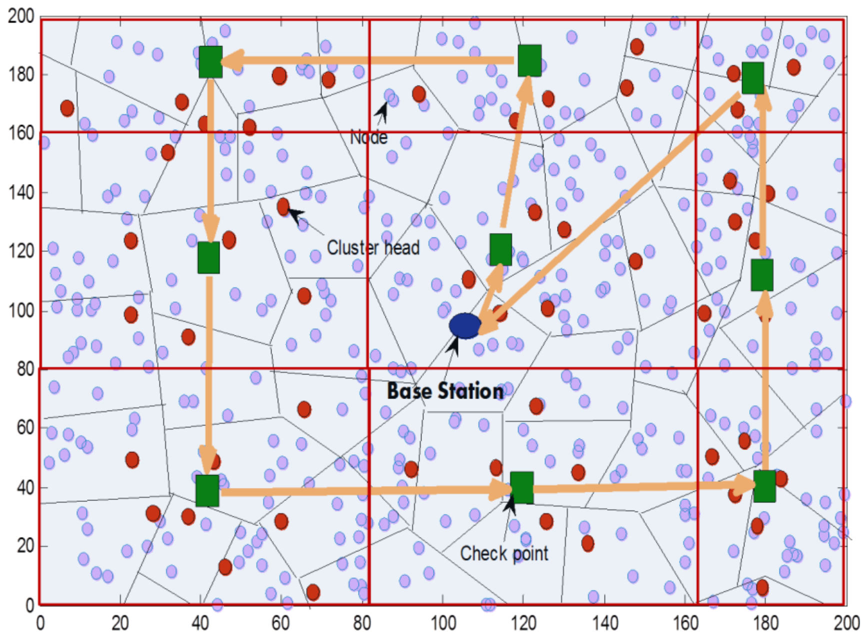

Figure 7-2 (b): Paths of the ferry with checkpoints, where 9 checkpoints are used

7.4.2. Performance Based on Network Lifetime

Network lifetime is defined here as the interval from the time the sensor network starts its operation until the death of the last node in the network. I compared the performance of four checkpoints TSP with a genetic algorithm, referred to as the optimized path, to the case of using a predetermined fixed path in the center of the sensing field and the multi-hop NRCA without the use of a ferry. From Table 7-2, we can see that the last node died in NRCA at round 3311, making it the lowest achiever with the shortest network lifetime when compared to the other two. On the other hand, we can see that the optimized nonlinear path based on TSP with a genetic algorithm had the longest network lifetime as its last nodes died at round 4003, compared to the predetermined path where the first node died at round 1763 and the last at round 3830.

Table 7-2 Simulation Results for the network lifetime based on Figure 7-2 (a)

Protocols	**Measurements**	
	Round first node dies	**Round last node dies**
Optimized path_TSP_Genetic	2010	4003
Predetermined fixed path	1763	3830
NRCA	1300	3311

7.4.3. Performance Based on Energy Consumed

As shown in Figure 7-3, the energy consumed per round in the optimized path case is less than the predetermined one and NRCA. Dividing the region into virtual grids with a checkpoint in each helps in reducing the energy consumption in these grids and, as a result, prolongs the lifetime of the cluster heads and preserves the overall energy of the whole network.

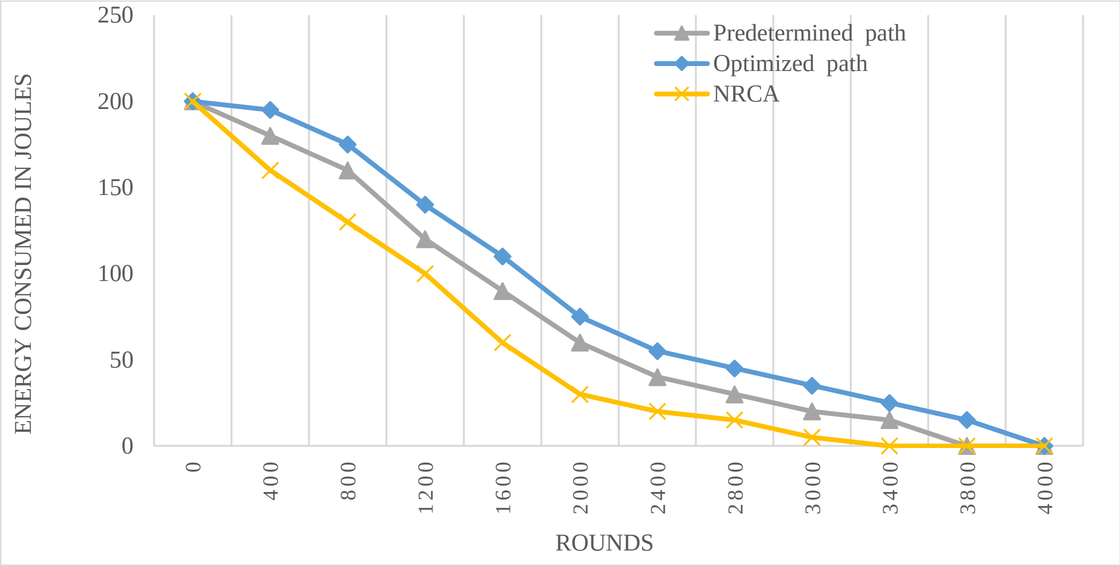

Figure 7-3: Energy consumption in the network

7.4.4. The overall Time of One Round Trip of the Ferry

The total overall time of one round trip of the ferry is defined as the overall traveling time of the ferry from the base station to each checkpoint and its return to the BS, plus the time spent stopped at each checkpoint, once, in order to collect data. As shown in Table 7-3, the predetermined path with four checkpoints took around 5.40 minutes per one round of data collection whereas four minutes were recorded for the optimized path.

Table 7-3: Simulation results for one round collection

	Predetermined path	Optimized path_TSP_Genetic
Time in minutes	5.40	4

7.4.5. Changing the Number of Checkpoints

By changing the sensing range for the optimized path, the number of checkpoints will be changed as well. I changed the sensing range to 20 meters and the number of checkpoints to be 25, 40 meters and the number of checkpoints to be 9, 50 meters and the number of checkpoints to be 4, and finally, 100 meters and the number of checkpoints to be 1. Table 7-4 shows the network performance as a result of changing the sensing range and the number of checkpoints. From Table 7-4 we can see that the network lifetime increases as the number of checkpoints increases. This is because the more checkpoints we have, the less distance that the data will have to travel, which in return saves the cluster heads' energy and the overall energy of the network. However, looking into the overall time that it takes the ferry to undertake one round of data collection, we can see from Table 7-5 and Figure 7-4 that it increases as the number of checkpoints increases. Thus, round trip traveling time has a direct relationship to the number of checkpoints. This is due to the increase in the length of the traveling path plus the increase in the amount of time spent stopped at each checkpoint. Given a particular application, the number of checkpoints can be chosen for a particular scenario based on the maximum tolerable delay.

Table 7-4: Simulation results for the network lifetime using different numbers of checkpoints

#Checkpoints	**Measurements**	
	Round first node dies	**Round last node dies**
Sensing range 20 #Checkpoint 25	2460	4433
Sensing range 40 #Checkpoint 9	2111	4120
Sensing range 50 #Checkpoint 4	2010	4003
Sensing range 100 #Checkpoint 1	1400	3500

Table 7-5: Simulation results for one round collection

#Checkpoints	Time in minutes
Sensing range 20 #Checkpoint 25	18.60
Sensing range 40 #Checkpoint 9	9.60
Sensing range 50 #Checkpoint 4	5.40
Sensing range 100 #Checkpoint 1 which is the base station	3

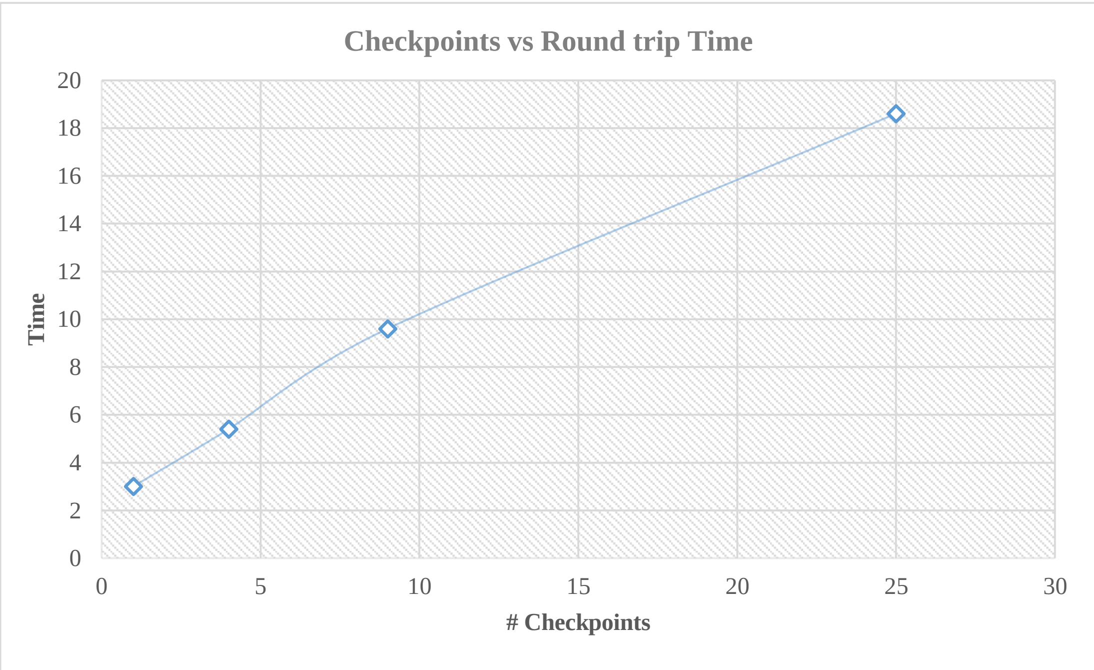

Figure 7-4: Number of checkpoints vs round trip time

7.4.6. Performance Evaluation of FNRCA Against Other Algorithms

I looked into comparing our algorithm with other state-of-the-art algorithms under not exact, but similar, operating conditions. I use a hybrid path, but in the literature, some of the algorithms (such as that found in [100, 101 and 102]) use controlled paths for the ferry to follow. In addition, there are other algorithms which use multiple mobile sinks (such as in [77, 78, 100 and 102]) while I only use one mobile object – which is the ferry – as a temporary sink. Moreover, I limit the multi-hopping in my algorithm to one hop count, while the other algorithms (especially the ones using rendezvous approaches [77, 78, 92, 93 and 101]) use multi-hop forwarding (one or more hop count) combined with the use of mobile elements to collect data. Nevertheless, I have considered further analysis and comparisons against two recently developed algorithms [92 and 93]. In these two recent algorithms, the authors proved that their proposed algorithms outperformed other existing ones.

In order to compare our algorithm with these two, I adapted their used parameters shown in Table 7-6 in our simulation. In Figure 7-5, that FNRCA outperformed WRP [92] and Charalampos et al. [93] in terms of network lifetime as the last node in FRNCA died after 1200 seconds compared to 1000 in WRP and 1050 in Charalampos et al. is shown. In WRP, 50% of the nodes died after 4500 seconds, in Charalampos et al. they died after 4800 seconds, while in FRNCA 50% of its nodes died after 6000 seconds. This can mostly be attributed to the use of multi-hop communication in WRP and Charalampos et al. which consumes more energy in general and results in a faster depletion of energy in the cluster heads that are closer to the ferry path.

However, in our algorithm, checkpoints are just locations where the ferry will stop to collect data from cluster heads that belong to its virtual grid, where each cluster head is just one hop count from the checkpoint position with which they are associated. Charalampos et al. achieved slightly better results than WRP since it selects cluster heads with higher energy as rendezvous points when using nodes with different initial energy values. However, as can be seen in Figure 7-6, similar results are achieved by both WRP and Charalampos et al. when using the same energy value to begin with. In both cases, FNRCA outperformed the two algorithms when using the same or different initial energy values. In addition, comparing the graph for FNRCA in Figure 7-5 and Figure 7-6 reveals minimal changes in its performance regardless of whether the same initial energy value was used by all nodes or uniformly distributed ones were used. This is mainly due to the fact that FNRCA incorporates current energy values in selecting and rotating cluster heads and minimizes multi-hop communication.

Table 7-6: Parameters used in the simulation to compare FNRCA to [92, 93]

Notation
N = 200
Initial node energy, E_o = uniformly selected from the nodes from 50-100 J / node
Area = 200 x 200
Checkpoints = 25
Packet size = 30 Bytes
Data Rate = 40 Kbps
Max sensing Radius : r = 50 m
Ferry_speed = 1 m/sec

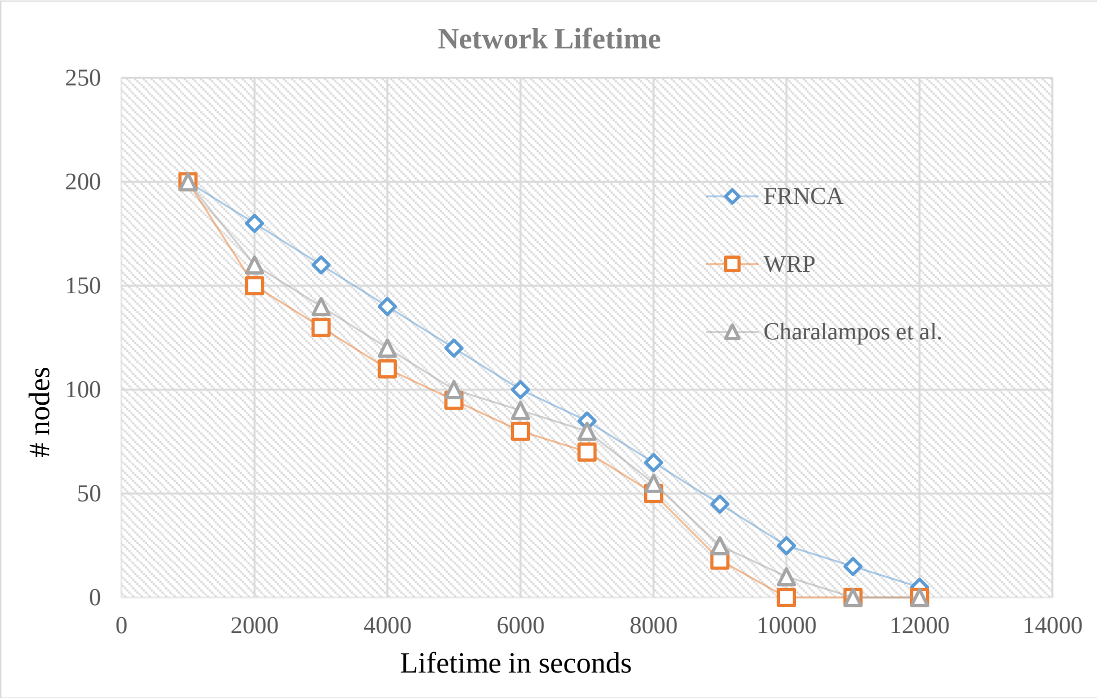

Figure 7-5: Network lifetime for FNRCA WRP and Charalampos et al. using different initial energy values

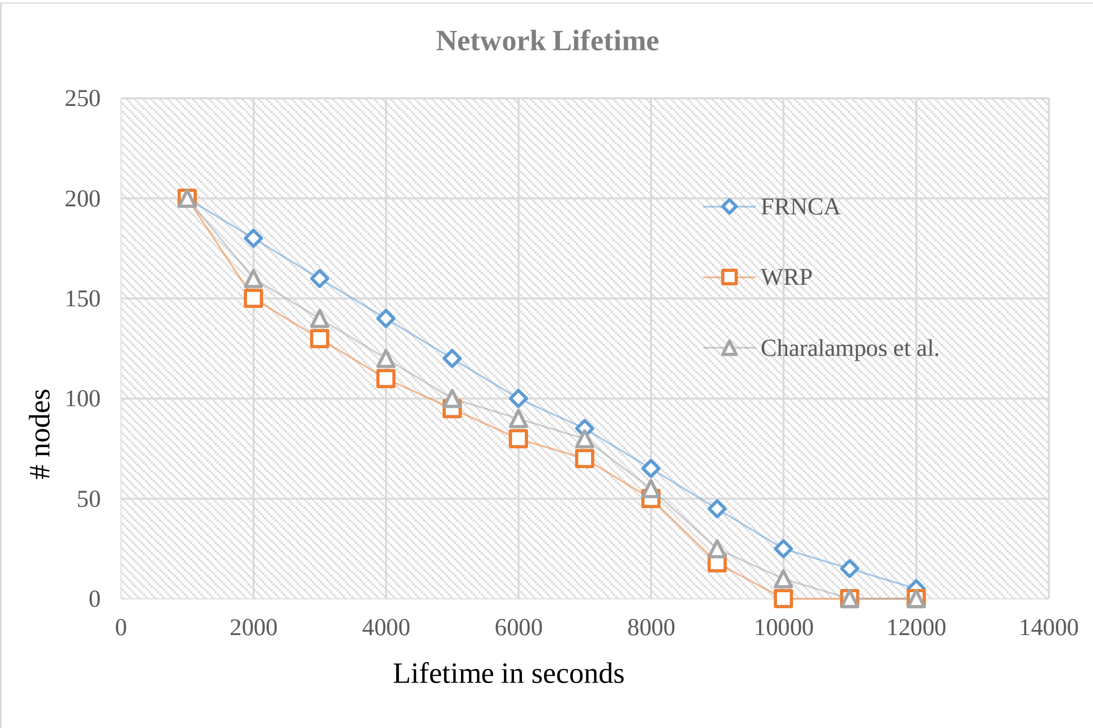

Figure 7-6: Network lifetime for FNRCA WRP and Charalampos et al. using the same initial energy values

7.5. Summary

In this chapter, an efficient data collection algorithm using a ferry node was proposed while considering the overall ferry round trip travel time and the overall consumed energy in the network. To minimize the overall round trip travel time, I divided the sensing field area into virtual grids based on the assumed sensing range and assigned a checkpoint in each one. A genetic algorithm with weight metrics was used to solve the Traveling Salesman Problem (TSP) and decide on an optimum path for the ferry to collect data. I utilized my previously published node ranking clustering algorithm (NRCA) in each virtual grid and when choosing the location for placing the ferry's checkpoints. I simulated the proposed algorithm in MATLAB and showed its performance in terms of the network lifetime, total energy consumption, and the total travel time.

Through simulation, I demonstrated the efficiency of the proposed algorithm when compared to using a traditional multi-hopping method to collect data and using fixed predetermined paths. Moreover, through simulation I showed that a nonlinear trajectory achieves a better optimization in terms of network lifetime, overall energy consumed, and the round trip travel time of the ferry when compared to a linear predetermined trajectory. The results of the simulation also showed that using a greater number of checkpoints increases the network lifetime, however, it increases the round trip travel time of the ferry as well. In addition, I compared my proposed algorithm against two other recently developed algorithms that were used by their authors to prove that they outperformed the previous algorithms. By doing so, I showed through my results that my proposed algorithm was able to outperform these other two algorithms in terms of network lifetime.

Chapter 8: Conclusions and Future Work

8.1. Conclusions

In this thesis, an energy efficient clustering algorithm for WSNs using node ranking in electing cluster heads was proposed. The performance of the proposed algorithm against two well-known algorithms was compared by using extensive simulation. Through simulation, I showed how the proposed algorithm outperformed some well-known algorithms like PEGASIS and LEACH. The performance of the algorithm using random cluster head replacement and using threshold values to replace the cluster heads were compared, and simulation showed that using threshold values outperformed the random replacement of cluster heads. Using an energy threshold to replace cluster heads improved the network lifetime as well. I also found that using variable energy threshold values to replace cluster heads improved the network lifetime even more over the use of a fixed value. In addition to that, using a hybrid redundant node duty-cycle has improved the network lifetime further.

Moreover, an efficient data collection algorithm using a ferry node is proposed while considering the overall ferry round trip travel time and the overall consumed energy in the network. To minimize the overall round trip travel time, I divided the sensing field area into virtual grids based on the assumed sensing range and assigned a checkpoint to each one. A genetic algorithm with weight metrics to solve the Traveling Salesman Problem (TSP) and decide on an optimum path for the ferry to collect data was used. I utilized my previously published node ranking clustering algorithm (NRCA) in each virtual grid and in choosing the location for placing the ferry's checkpoints. I simulated the proposed algorithm in MATLAB and showed its performance in terms of the network lifetime, total energy consumption, and the total travel time. Through simulation, I demonstrated the efficiency of the proposed algorithm when compared to using a traditional multi-hopping method to collect data and using fixed predetermined paths. Moreover, I showed through simulation that a nonlinear trajectory achieves a better optimization in terms of network lifetime, overall energy consumed, and the round trip travel time of the ferry when compared to a linear predetermined trajectory. The results of the simulation also showed that using a greater number of checkpoints increases the network lifetime, however, it increases the round trip travel time of the ferry as well. In addition to that, I compared the proposed algorithm two of the most recent algorithms in the field and showed how it outperformed them in network lifetime.

8.2. Future Work

In the near future, I plan to simulate more of the ferry algorithms and compare their performance to my proposed algorithm by using different criteria. I looked into comparing the FNRCA algorithm to other state-of-the-art ones, however, I was unable to conduct a fair comparison under the same constraints and conditions. In FNRCA, a random uncontrolled path is used, but in the literature, I found some algorithms that use controlled paths for the ferry to follow and comparing them with the proposed algorithm would be unfair. Furthermore, some of the algorithms use multiple mobile sinks while only one mobile object (which is the ferry) is used in FNRCA. Moreover, in the FNRCA, the multi-hop count is limited to one hop count while some of the other algorithms, especially the ones using rendezvous approaches, still use multi-hop forwarding combined with the use of mobile elements to collect data. Nevertheless, I am still considering undertaking further analysis and making possible comparisons as work in the future. I also plan to test NRCA for worst case scenario when all nodes have the same energy and the same distance from the base station and see how will it perform compared to other algorithms. Moreover, I am planning to consider using limits rather than a single threshold for changing cluster heads.

In addition, I plan to consider some of the physical characteristics of the medium, such as considering channel fading and radio interference as they are considered to be two of the challenges that must be overcome when designing energy efficient protocols for WSNs. Moreover, adding several ferries to collect data can also be an improvement over the current proposed algorithm and can decrease the time delay in case of emergency or in non-delay tolerant applications. I am also considering changing the algorithm to have a speed-controlled flyover ferry – instead of stopping at each checkpoint, the ferry can decrease its speed while flying over checkpoints to collect the data. I also plan to use particle swarm optimization (PSO) which starts with random particles (solutions) and then searches for optima by updating generations. Moreover, I plan to consider the area of joint decision making for selecting cluster heads, checkpoints and the path of the ferry.

Bibliography

1. Akyildiz, I., Su, W., Sankarasubramaniam, Y., & Cayirci, E. “A survey on sensor networks,” Communications Magazine, IEEE 40, no. 8, 102-114, 2002.
2. Römer, K., & Mattern, F. “The design space of wireless sensor networks,” Wireless Communications, IEEE, 11(6), 54-61, 2004.
3. Ekici, E., Gu, Y., & Bozdag, D. “Mobility-based communication in wireless sensor networks, ” IEEE Communications Magazine, 44(7), 56, 2006.
4. Akkaya, K., & Younis, M. “A survey on routing protocols for wireless sensor networks,” Ad Hoc Networks, 3(3), 325-349, 2005.
5. Baronti, P., Pillai, P., Chook, V. W., Chessa, S., Gotta, A., & Hu, Y. F. “Wireless sensor networks: A survey on the state of the art and the 802.15. 4 and ZigBee standards,” Computer Communications, 30(7), 1655-1695, 2007.
6. Gezici, S., Tian, Z., Giannakis, G. B., Kobayashi, H., Molisch, A. F., Poor, H. V., & Sahinoglu, Z. “Localization via ultra-wideband radios: a look at positioning aspects for future sensor networks,” Signal Processing Magazine, IEEE, 22(4), 70-84, 2005.
7. Buratti, C., Conti, A., Dardari, D., & Verdone, R. “An overview on wireless sensor networks technology and evolution,” Sensors, 9(9), 6869-6896, 2009.
8. Yick, J., Mukherjee, B., & Ghosal, D. “Wireless sensor network survey,” Computer Networks, 52(12), 2292-2330, 2008.
9. Vieira, M. A. M., Coelho Jr, C. N., da Silva, D. C., & da Mata, J. M. (2003, September). “Survey on wireless sensor network devices,” In Emerging Technologies and Factory Automation, Proceedings. ETFA'03. IEEE Conference, 1: 537-544, 2003.
10. Sohraby, K., Minoli, D., & Znati, T. “Wireless sensor networks: technology,” protocols, and applications. John Wiley & Sons, 2007.
11. Depuru, S. S. S. R., Wang, L., & Devabhaktuni, V. “Smart meters for power grid: challenges, issues, advantages and status,” Renewable and Sustainable Energy Reviews, 15(6): 2736–2742, 2011.

12. Szewczyk, R., Osterweil, E., Polastre, J., Hamilton, M., Mainwaring, A., & Estrin, D. “Habitat monitoring with sensor networks, ” Communications of the ACM 47(6) no. 6 (2004): 34-40, 2004.

13. Kaur, R., Sharma, M. “An approach to design habitat monitoring system using sensor networks,” International Journal of Soft Computing and Engineering (IJSCE), NCAI2011: 5-8, 2011.

14. Song, S., & Yoon, Y. “Intelligent smart cloud computing for smart service,” In Grid and Distributed Computing, Control and Automation, 64-73. Springer Berlin Heidelberg, 2010.

15. Chao, S. H. A., Wang, R., Huange, H., & Sun, L. “A type of healthcare system based on intelligent wireless sensor networks,” The Journal of China Universities of Posts and Telecommunications 17 (2010): 30-39, 2010.

16. Sexena, R. N., Roy, A., & Shin, J. “Cross-layer algorithms for QoS enhancement in wireless multimedia sensor networks,” In IEICE Trans. Commun. E91-B: 2716–2719, 2008.

17. Almalkawi, I. T., Guerrero Zapata, M., Al-Karaki, J. N., Morillo-Pozo, J. “Wireless Multimedia Sensor Networks: current trends and future directions,” Sensors 10: 6662-717, 2010.

18. McCurdy, N. J. & Griswold, W. “A system architecture for ubiquitous video,” In Proceedings of the 3rd Annual International Conference on Mobile Systems, Applications, and Services (Mobisys '05), pp. 1-14, 2005.

19. Amis, A. D. et al. “Max-min d-cluster formation in wireless ad hoc networks,” Proc. IEEE INFOCOM, Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE (Vol. 1, pp. 32-41), 2000.

20. Hengstler, S., & Aghajan, H. “Application-oriented design of smart camera networks,” In Proceedings of the 1st ACM/IEEE International Conference on Distributed Smart Cameras (ICDSC '07): 12–19, 2007.

21. Barton-Sweeney, A., Lymberopoulos, D., & Savvides, A. “Sensor localization and camera calibration in distributed camera sensor networks,” In Proceedings of the 3rd International Conference on Broadband Communications, Networks and Systems (BROADNETS '06), 2006.

22. Akyildiz, I. F., Melodia, T. & Chowdhury, K. R. "Wireless Multimedia Sensor Networks: Applications and testbeds," Proc. IEEE. Vol. 96, No. 10: 1588–1605, 2008.

23. Akyildiz, I.F., Melodia, T., & Chowdhury, K. R. "A survey on wireless multimedia sensor networks," Comput. Netw. 51: 921–960, 2007.

24. Soro, S., & Heinzelman, W. B. "On the coverage problem in video-based wireless sensor networks," in Proceedings of the 2nd International Conference on Broadband Networks (BROADNETS '05): 9–16, 2005.

25. Tezcan, N., & Wang, W. "Self-orienting Wireless Multimedia Sensor Networks for maximizing multimedia coverage," In Proceedings of IEEE International Conference on Communications, ICC '08, Beijing, China: 2206–2210, 2008.

26. Margi, C., Petkov, V., Obraczka, K., Manduchi, R. "Characterizing energy consumption in a visual sensor network testbed," In Proceedings of 2nd International Conference on Testbeds and Research Infrastructures for the Development of Networks and Communities, TRIDENTCOM 2006, Barcelona, Spain, 2006.

27. Sun, B., Shufean, M. A., & Sun, F. "A geographic information system framework for data visualisation in wireless sensor networks," International Journal of Sensor Networks, 19(1): 51-61, 2015.

28. Petrioli, C., Nati, M., Casari, P., Zorzi, M., & Basagni, S. "ALBA-R: Load-balancing geographic routing around connectivity holes in wireless sensor networks," IEEE Transactions in Parallel and Distributed Systems, 25(3): 529-539, 2014.

29. Cerotti, D., Gribaudo, M., & Bobbio, A. "Markovian agents models for wireless sensor networks deployed in environmental protection," Reliability Engineering & System Safety, 130: 149-158, 2014.

30. Lu, H., Li, J., & Guizani, M. "Secure and efficient data transmission for cluster-based wireless sensor networks," IEEE Transactions in Parallel and Distributed Systems, 25(3): 750-761, 2014.

31. Wang, D., & Wang, P. "Understanding security failures of two-factor authentication schemes for real-time applications in hierarchical wireless sensor networks," Ad Hoc Networks, 20: 1-15, 2014.

32. Tunca, C., Isik, S., Donmez, M. Y., & Ersoy, C. “Distributed mobile sink routing for wireless sensor networks: A survey,” Communications Surveys & Tutorials, IEEE, 16(2): 877-897, 2014.
33. Keskin, M. E., Altınel, İ. K., Aras, N., & Ersoy, C. “Wireless sensor network lifetime maximization by optimal sensor deployment, activity scheduling, data routing and sink mobility,” Ad Hoc Networks, 17: 18-36, 2014.
34. Alnuaimi, M., Shuaib, K., Nuaimi, K., & Abdel-Hafez, M. “Performance analysis of clustering protocols in WSN.” In Wireless and Mobile Networking Conference (WMNC), 2013 6th Joint IFIP: 1-6: pp. 488-493. IEEE, Dubai, 2013.
35. Radi, M., et al. “Multipath routing in wireless sensor networks: Survey and research challenges,” Sensors 12, No. 1: 650-685, 2012.
36. Heinzelman, W. B., Chandrakasan, A. P., & Balakrishnan, H. “Energy-efficient communication protocol for wireless microsensor networks,” Proceedings of the 33rd Hawaii International Conference on System Sciences, 8: 3005–3014, 2000.
37. Heinzelman,W. B. “An application-specific protocol architectures for wireless networks, ”, Ph.D. Dissertation, MIT, Cambridge, 2000.
38. Biradar, R. V., Sawant, S. R., Mudholkar, R. R., & Patil, V. C.. “Multi-hop routing in self-organizing Wireless Sensor Networks,” IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 1: 155-164, 2011.
39. Lindsey, S., & Raghavendra, C. S. “PEGASIS: power efficient gathering in sensor information systems,” In Proceedings of the IEEE Aerospace Conference, Vol. 3, pp. 3-1125, Big Sky, Montana, 2002.
40. Du, G., Shi, Q., Tang, Y., & Sun, X. “A mixed non-uniform clustering algorithm for wireless sensor networks,” In 2011 IEEE 13th International conference on communication Technology (ICCT): pp. 661-665. IEEE, 2011.
41. Kumar, N., Bhutani, P., & Mishra, P. “U-LEACH: A novel routing protocol for heterogeneous wireless sensor networks,” In International Conference on Communication, Information & Computing Technology (ICCICT): pp. 1-4, Mumbai, India, 2012.

42. Younis, O., & Fahmy, S. "HEED: A hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks," IEEE Transactions on Mobile Computing 3, No. 4 (2004): 366-379, 2004.
43. Hong, J., Kook, J., Lee, S., & Kwon, D. Yi, S. "T-LEACH: The method of threshold-based cluster head replacement for wireless sensor networks," Inf. Syst. Front 11: 513–521, 2011.
44. Younis, O., Krunz, M., & Ramasubramanian, S. "Node clustering in wireless sensor networks: recent developments and deployment challenges," IEEE Network, Vol. 20, No. 3: 20-25, 2006.
45. Shu, T., Krunz, M., & Vrudhula, S. "Power balanced coverage time optimization for clustered wireless sensor networks," In Proceedings of the 6th ACM International Symposium on Mobile Ad Hoc Networking and Computing: 111-120. ACM, 2005.
46. Banerjee, S., & Khuller, S. "A clustering scheme for hierarchical control in multihop wireless networks," Proc. IEEE INFOCOM, Twentieth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE (Vol. 2, pp. 1028-1037). IEEE, 2001.
47. Basagni, S. "Distributed clustering algorithm for ad hoc networks," Proc. Int'l. Symp. Parallel Architectures, Algorithms, and Networks, pp. 310-315. IEEE, 1999.
48. Heinzelman, W., Chandrakasan, A., & Balakrishnan, H. "An application-specific protocol architecture for wireless microsensor networks," IEEE. Wireless Commun. Vol. 1, No. 4: 660-670, 2002.
49. Younis, O. and Fahmy, S. "Distributed Clustering in Ad Hoc Sensor Networks: A Hybrid, Energy-Efficient Approach," Proc. IEEE INFOCOM, Hong Kong, Mar. 2004; an extended version appeared in IEEE Trans. Mobile Comp. , vol. 3, no. 4, Oct.-Dec. 2004.
50. Younis, O., Krunz, M., and Ramasubramanian, S. "Node Clustering in Wireless Sensor Networks: Recent Developments and Deployment Challenges," IEEE Network, vol. 20, no. 3, pp. 20-25, 2006.
51. Blough, D. M., & Santi, P. "Investigating upper bounds on network lifetime extension for cell-based energy conservation techniques in stationary ad hoc

networks," In Proceedings of the 8th annual international conference on mobile computing and networking: 183-192. ACM, 22-26 September, 2002.

52. Nikolidakis, S. A., Kandris, D., Vergados, D. D., & Douligeris, C. "Energy efficient routing in wireless sensor networks through balanced clustering," Algorithms 6.1: 29-42, 2013.
53. Koutsonikolas, D., Das, S. M., Hu, Y., & Stojmenovic, I. "Hierarchical geographic multicast routing for wireless sensor networks," Wireless Networks 16, no. 2: 449-466, 2010.
54. Tyagi, S., & Kumar, N. "A systematic review on clustering and routing techniques based upon LEACH protocol for wireless sensor networks," Journal of Network and Computer Applications. 36: 623–645, 2013.
55. Zhou, Z., Xang, X., Wang, X., & Pan, J. "An Energy-efficient data-dissemination protocol inwireless sensor networks," In IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM), Buffalo, NY, USA, 2006: 10-19.
56. Baronti, P., Pillai, P., Chook, V., Chessa, V., & Gotta, A. "Wireless sensor networks: a survey on the state of the art and the 802.15.4 and ZigBee standards," Computer communications C30, No. 7: 1655-1695, 2007.
57. Alnuaimi, K., Alnuaimi, M., Mohamed, N., Jawhar, I., & Shuaib, K. "Web-based wireless sensor networks: a survey of architectures and applications," The 6th International Conference on Ubiquitous Information Management and Communication, Malaysia, 20-22February, 2012.
58. Alnuaimi, M., Sallabi, F., & Shuaib, K. "A survey of wireless multimedia sensor networks challenges and solutions," IEEE International Conference on Innovations in Information Technology (IIT'11): 191-196, Abu Dhabi, UAE, 2011.
59. Alnuaimi, M., Shuaib, K., & Jawhar, I. "Performance evaluation of IEEE 802.15.4 physical layer using MatLab/Simulink," IEEE International Conference on Innovations in Information Technology: 1-5, 19-21 November, 2006.

60. García Villalba, L. J., Orozco, A. L. S., Triviño Cabrera, A., & Barenco Abbas, C. J. "Routing protocols in wireless sensor networks," Sensors 9, No. 11: 8399-8421, 2009.

61. Di Francesco, M., Das, S. K., & Anastasi, G. "Data collection in wireless sensor networks with mobile elements: A survey," ACM Transactions on Sensor Networks (TOSN) 8:1, 2011.

62. Sun, Z. et al. "BorderSense: border patrol through advanced wireless sensor networks," Ad Hoc Networks 9, No. 3: 468-477, 2011.

63. Lu, W., Seah, W. K. G., Peh, E. W. C., & Ge, Y. "Communications support for disaster recovery operations using hybrid mobile ad-hoc networks," In 32nd IEEE Conference on Local Computer Networks (LCN 2007), IEEE, pp. 763-770, 2007.

64. Nikzad, N. et al. "CitiSense: improving geospatial environmental assessment of air quality using a wireless personal exposure monitoring system," In Proceedings of the conference on Wireless Health, p. 11, 2012.

65. Arora, A., et al. "A line in the sand: a wireless sensor network for target detection, classification, and tracking," Computer Networks, Vol. 46, No. 5: 605-634, 2004.

66. Koppanyi, Z., Lovas, T., Barsi, A., Demeter, H., Beeharee, A., & Berenyi, A. "Tracking vehicle in gsm network to support intelligent transportation systems," Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., vol. XXXIX-B2: 139-144, 2012.

67. Juang, P., Oki, H., Wang, Y., Maronosi, M., Peh, L., & Rubenstein, D. "Energy-efficient computing for wildlife tracking: design tradeoffs and early experiences with ZebraNet," ACM Sigplan Notices Vol 37, No. 10: 96-107, 2002.

68. Kaur, R., & Sharma, M. "An approach to design habitat monitoring system using sensor networks," International Journal of Soft Computing and Engineering (IJSCE), NCAI2011: 5-8, 2011.

69. Lai, T. T. T., Chen, W. J., Li, K. H., Huang, P., & Chu, H.H. "Computing the impact of cyber attacks on complex missions," In 2011 IEEE International Systems Conference (SysCon), pp. 46-51. IEEE, 2012.

70. Shah, R.C., Roy, S., Jain, S., & Brunette, W. "Data MULEs: modeling a three-tier architecture for sparse sensor networks," Proc. First IEEE Int'l Workshop Sensor Network Protocols and Applications: 30-41, 2003.
71. Jain, S., Shah, R. C., Brunette, W., Borriello, G., & Roy, S. "Exploiting mobility for energy efficient data collection in sensor networks," Mobile Networks and Applications Vol. 11, No. 3: 327-339, 2006.
72. Song, L., & Hatzinakos, D. "Architecture of wireless sensor networks with mobile sinks: sparsely deployed sensors," IEEE Trans. Vehicular Technology, July 2007, Vol. 56, No. 4: 1826-1836, 2007.
73. Luo, J., Panchard, J., Piorkowski, M., Grossglauser, M., & Hubaux, J. "MobiRoute: routing towards a mobile sink for improving lifetime in sensor networks," Proc. Second IEEE/ACM Int'l Conf. Distributed Computing in Sensor Systems (DCOSS): 480-497, 2006.
74. Luo, J., & Hubaux, J. P. "Joint mobility and routing for lifetime elongation in wireless sensor networks," Proceedings of the 24th IEEE Conference on Computer Communications (INFOCOM 2005), Vol. 3: 1735, 2005.
75. Somasundara, A., Ramamoorthy, A., & Srivastava, M. "Mobile element scheduling for efficient data collection in wireless sensor networks with dynamic deadlines," Proc. 25th IEEE Int'l Real-Time Systems Symp. (RTSS): 296-305, 2004
76. Gu, Y., Bozdag, D., Brewer, R., & Ekici, E. "Data harvesting with mobile elements in wireless sensor networks," Computer Networks, Vol. 50, No. 17: 3449-3465, 2006.
77. Xing, G., Wang, T., Xie, Z., & Jia, W. "Rendezvous design algorithms for wireless sensor networks with a mobile base station," Proc. ACM MobiHoc: 231-240, 2008.
78. Xing, G., Wang, T., Xie, Z., & Jia, W. "Rendezvous planning in wireless sensor networks with mobile elements," IEEE Trans. Mobile Computing, Vol. 7, No. 11:1-14, 2008.
79. Chen, T. C., Chen, T. S., & Wu, P. W. "Data collection in wireless sensor networks assisted by mobile collector," In Wireless Days, 1st IFIP: 1-5. IEEE, 2008.

80. He, L., Zhuang, Y., Pan, J., & Xu, J. “Evaluating on-demand data collection with mobile elements in wireless sensor networks,” In Vehicular Technology Conference Fall (VTC 2010-Fall), IEEE 72nd: 1-5. IEEE, 2010.

81. Jawhar, I., Ammar, M., Zhang, S., Wu, J., & Mohamed, N. “Ferry-based linear wireless sensor networks,” In Global Communications Conference (GLOBECOM), 2013 IEEE:304-309. IEEE, 2013.

82. Anastasi, G., Conti, M., Di Francesco, M., & Passarella, A. “Energy conservation in wireless sensor networks: A survey,” Ad Hoc Networks 7, No. 3: 537-568, 2009.

83. Raghunathan, V., Ganeriwal, S., & Srivastava, M. “Emerging techniques for long lived wireless sensor networks,” IEEE Communications Magazine, April 2006, Vol. 44, No. 4: 108-114, 2006.

84. Yu, J. G., Qi, Y. Y., Wang, G. H., & Gu, X. “A cluster-based routing protocol for wireless sensor networks with nonuniform node distribution,” AEU-International Journal of Electronics and Communications 66, No. 1: 54-61, 2012.

85. Du, G., Shi, Q., Tang, Y., & Sun, X. “A mixed non-uniform clustering algorithm for wireless sensor networks,” In 2011 IEEE 13th International Conference Communication Technology (ICCT), 661-665. IEEE, 25-28, 2011.

86. Xue, L. et al. “Multiple heterogeneous data ferry trajectory planning in wireless sensor networks,” In INFOCOM, 2014 Proceedings IEEE: 2274-2282. IEEE, 2014

87. Alnuaimi, M., Shuaib, K., Alnuaimi, K., & Abed-Hafez, K. “Clustering in wireless sensor networks based on node ranking,” In Wireless Communications and Mobile Computing Conference (IWCMC), 2014 International: 488-493. IEEE, 2014.

88. Alnuaimi, M., Shuaib, K., & Alnuaimi, K. “Clustering in WSN using node ranking with hybrid nodes duty-cycle and energy threshold,” In 2014 IEEE 13th International Symposium on Network Computing and Applications (NCA): 245-252. IEEE, 2014.

89. Introduction to Travel Salesman problem [Online]. Available at: http://www.oocities.org/gopal_mba/tsp/modi.html [Accessed: March 06, 2015].

90. Abdoun, O., Abouchabaka, J., & Tajani, C. “Analyzing the performance of mutation operators to solve the Travelling Salesman Problem,” arXiv preprint arXiv:1203.3099, 2012.

91. Khan, F., Khan, S. A., Turgut, D., & Boloni, L. “Greedy path planning for maximizing value of information in underwater sensor networks,” Proceeding of the 39th Annual IEEE Conference on Local Computer Networks, 8-11 September, 2014, Edmonton, Canada, 2014.

92. Salarian, H., Chin, K., & Naghdy, F. “An energy efficient mobile sink path selection strategy for wireless sensor networks,” IEEE Transactions on Vehicular Technology, PP (99):1, 2013.

93. Konstantopoulos, C., Pantziou, G., Vathis, N., Nakos, V., & Gavalas, D. “Efficient mobile sink-based data gathering in wireless sensor networks with guaranteed delay,” In Proceedings of the 12th ACM international symposium on mobility management and wireless access (MOBIWAC 2014): pp. 47-54, 2014.

94. Woeginger, G. J. “Exact algorithms for NP-hard problems: A survey,” In Combinatorial Optimization—Eureka, You Shrink!, 185-207. Springer Berlin Heidelberg, 2003.

95. Gutin, G., & Punnen, A. P. (eds.). “The traveling salesman problem and its variations,”,Vol. 12. Springer Science & Business Media, 2002.

96. Albayrak, M., & Allahverdi, N. “Development of a new mutation operator to solve the traveling salesman problem by aid of genetic algorithms,” Expert Systems with Applications 38, No. 3: 1313-1320, 2011.

97. Pang, W., Wang, K. P., Zhou, C. G., Dong, L. J., Liu, M., Zhang, H. Y., & Wang, J. Y. “Modified particle swarm optimization based on space transformation for solving traveling salesman problem,” In Proceedings of 2004 International Conference Machine Learning and Cybernetics, 2004, Vol. 4: 2342-2346. IEEE, 2004.

98. Ahmed, Z. H. “Genetic algorithm for the traveling salesman problem using sequential constructive crossover operator,” International Journal of Biometrics & Bioinformatics (IJBB) 3, no. 6: 96, 2010.

99. Alnuaimi, M., Shuaib, K., Al Nuaimi, K., Abdel-Hafez, M. “Ferry-based data gathering in Wireless Sensor Networks with path selection,” The 6th

International Conference on Ambient Systems, Networks and Technologies (ANT 2015), Procedia Computer Science 52 (2015): 286-293, London, UK, Jun 2015.

100. Almi’ani, K., Viglas, A., & Libman, L. “Energy-efficient data gathering with tour length-constrained mobile elements in wireless sensor networks,” In 35th IEEE Conf. on Local Computer Networks (LCN): 582–589. IEEE, 2010.

101. Mai, L., Shangguan, L., Lang, C., Du, J., Liu, H., & Li, Z. “Load balanced rendezvous data collection in wireless sensor networks,” In 8th IEEE Intl.l Conference on Mobile Ad hoc and Sensor Systems (MASS): 282–291. IEEE, 2011.

102. Gao, S., Zhang, H., & Das, S. K. “Efficient data collection in wireless sensor networks with path-constrained mobile sinks, on mobile computing,” IEEE Transactions, Vol.10, No. 4: 592, 608, 2011.

List of Publications

1. Alnuaimi, M., Sallabi, F., Shuaib, K. “A survey of Wireless Multimedia Sensor Networks challenges and solutions,” Proceedings of the IEEE IIT12. Abu-Dhabi, UAE, 25-27 April, 2011.
2. Alnuaimi, K., Alnuaimi, M., Mohamed, N., Jawhar, I., Shuaib, K. “Web-based wireless sensor networks: a survey of architectures and applications,” Proceedings of the 6th International Conference on Ubiquitous Information Management and Communication. ACM, Kuala Lumpur, Malaysia, 20-22 February, 2012.
3. Alnuaimi, M., Shuaib, K., Alnuaimi, K., Abdel-Hafez, M. “Performance analysis of clustering protocols in WSN,” Proceedings of IFIP/IEEE WMNC2013. Dubai, UAE, 22-24April, 2013.
4. Alnuaimi, M., Shuaib, K., Alnuaimi, K., Abed-Hafez, M. “Clustering in Wireless Sensor Networks based on node ranking,” In 2014 IEEE International Wireless Communications and Mobile Computing Conference (IWCMC), pp. 488-493, Nicosia, Cyprus, 2014.
5. Alnuaimi, M., Shuaib, K., Al Nuaimi, K. “Clustering in WSN using node ranking with hybrid nodes duty-cycle and energy threshold,” In Proceedings of the 2014 IEEE 13th International Symposium on Network Computing and Applications, pp. 245-252. IEEE Computer Society, Cambridge, MA, August 2014.
6. Alnuaimi, M., Shuaib, K., Al Nuaimi, K., Abdel-Hafez, M. “Data gathering in Wireless Sensor Networks with ferry nodes,” 12th IEEE International Conference on Networking, Sensing and Control (ICNSC15), pp. 221-225, Taiwan, Taipei, April 2015.
7. Alnuaimi, M., Shuaib, K., Al Nuaimi, K., Abdel-Hafez, M. “Ferry-based data gathering in Wireless Sensor Networks with path selection,” The 6th International Conference on Ambient Systems, Networks and Technologies (ANT 2015), Procedia Computer Science 52 (2015): 286-293, London, UK, Jun 2015.
8. Alnuaimi, M., Shuaib, K., Alnuaimi, K., Abed-Hafez, M. “An efficient clustering algorithm for wireless sensor networks,” International Journal of Pervasive Computing and Communications 11, no. 3 (2015): 302-322, August, 2015.

9. Alnuaimi, M., Shuaib, K., Alnuaimi, K., Abed-Hafez, M. "Data gathering in delay tolerant Wireless Sensor Networks using a ferry," Sensors 15, no. 10 (2015): 25809-25830, October, 2015.