POLITEHNICA UNIVERSITY
OF BUCHAREST

Doctoral School of Electronics, Telecommunications and Information Technology

Decision No. 939 from 21-10-2022

Ph.D. THESIS SUMMARY

Mohammed Kareem Hussein

DATA REDUCTION ALGORITHMS TO IMPROVE THE ENERGY EFFICIENCY OF SINGLE HOP OR CLUSTER BASED WIRELESS SENSOR NETWORKS

ALGORITMI PENTRU REDUCEREA VOLUMULUI DE DATE ÎN VEDEREA AMELIORĂRII EFICIENȚEI ENERGETICE A REȚELELOR WIRELESS DE SENZORI CU TRANSMISIE DIRECTĂ SAU ORGANIZATE ÎN CLUSTERE

THESIS COMMITTEE

Chairman	Prof. Dr. Eng. Gheorghe Brezeanu	from	Univ. Politehnica of Bucharest
Doctoral supervisor	Prof. Dr. Eng. Ion Marghescu	from	Univ. Politehnica of Bucharest
Referee	Prof. Dr. Eng. Ioan Nicolaescu	from	Military Technical University Ferdinand I Bucharest
Referee	Prof. Dr. Eng. Ioan Bogdan	from	Technical University “Gheorghe Asachi”
Referee	Prof. Dr. Eng. Calin Vladeanu	from	Univ. Politehnica of Bucharest

BUCHAREST 2022

Table of Contents

1. Introduction……………………………………………………………………………..1

1.1 Research Problem, Objectives, and Limitations 2

1.2 The Thesis Content 3

2. An Overview of Data Reduction Methods used in Wireless Sensor Networks…………...4

2.1 Challenges of wireless sensor networks 6

2.2 Efficient Energy use in WSN 7

2.3 Sensor data compression 7

2.3.1 Overview Sensor Data Compression 8

2.3.2 Methods for data segmentation 8

2.4 Related Work 9

2.4.1 Data Reduction Based on Coding.. 9

2.4.2 Data Reduction Based on Forecasting.. 9

2.4.3 Data Reduction Based on Multiple sensors 9

2.5 Conclusion 10

3. Performance of Data Reduction Algorithms for WSN using different Real-Time Dataset……………………………………………………………………………………..11

3.1.4 Data Reduction based – An Efficient Data Collection and Dissemination (EDCD2) Algorithm 10

3.2 Real-Time Datasets 13

3.2.1 Data 1-AirQ.. 13

3.4 Simulation and Results.. 14

3.5 Conclusion 15

4. Impact of various data packet sizes on the performance of cluster-based WSN………..16

4.1 The LEACH Routing Algorithm.. 16

4.3 The effect of data packet size on the performance of cluster-based WSN - Scenario 1 ... 17

4.5 Conclusion 20

5. A Hybrid Data Reduction Algorithm (HDRA)…………………………………………21

5.1 System Scenario 21

5.2 The HDRA Algorithm presentation 22

5.2.1 Implementation and performance evaluation 24

5.2.2 Artificial Dataset 24

5.3 Analysis of the Results... 25

5.3.1 Scenario 1 25

5.4 Conclusion 27

6. Conclusions and Future Works……………………………………………………………28

6.1 Conclusions 28
6.2 Original contributions 30
6.3 Future Works.. 31
6.4 List of publications.. 31
Bibliography……………………………………………………………………………33

List of Tables

Table 3. 1 Some data samples of node1 – Data1- AirQ………………………………..14
Tabel 4. 3 Various data packet size 17
Table 5. 1 Simulation parameters………………………………………………………22

List of Figures

Figure 1. 1 Architecture of WSN……………………………………………………..1
Figure 2. 1 The architecture of WSN……………………………………………………5
Figure 3. 4 General block diagram of EDCD2 algorithm... 12
Figure 3. 6 Structure of Data1- AirQ 13
Figure 3. 11 Accuracy of Applying Various Algorithms for all Selected Nodes from DATA2-ARHO.. 14
Figure 4. 1 Network design for LEACH protocol...17
Figure 4. 2 Timeline of LEACH operation: the set-up and steady state phase………...17
Figure 4. 4 Sensor field area 50m x 50m...18
Figure 4. 5 The number of nodes alive per round by applying various packet sizes.......18
Figure 4. 6 The number of nodes alive per transmissions by applying various packet sizes...19
Figure 4. 7 Energy consumed per transmission...19
Figure 5. 1 Sensor field area 200m x 200m……………………………………………21
Figure 5. 5 Sample of real-time data…………………………………………………...24
Figure 5. 6 artificial dataset……………………………………………………………25
Figure 5. 7 Sample artificial dataset……………………………………………………25
Figure 5. 8 Residual energy……………………………………………………………26
Figure 5. 9 Alive Nodes Per Transsmissions…………………………………………...26
Figure 5. 10 Alive Nodes Per Rounds…………………………………………………27

Lista de abrevieri

WSN	Wirless Sensor Network
IOT	Internet of Things
BS	Base station
Qos	Quality of service
CPU	Central processing unit
CH	Cluster head
CS-RD	Coding schemes are the based relative difference
CS-FP	Coding schemes are the based factor of precision
BLUE	Best linear unbiased estimation
DPS	Dual Prediction System
GW	Gateway
FC	Fusion centre
LEACH	Low-energy adaptive clustering hierarchy
EDCD2	An Efficient Data Collection and Dissemination
FICA	Fast Independent Component Analysis
NNF	Neural Network Fitting
NNTS	Neural Network Time Series
LRMV	Linear Regression with Multiple Variables
AirQ	Air Quality
ARHO	American River Hydrologic Observatory
GSB	Grand St. Bernard
IBRL	Intel Berkeley Research Lab
HDRA	Hybrid Data Reduction Algorithm
S	Sensor
N	Node
T	Temperature
H	Humidity

Chapter 1
Introduction

The increased interest in wireless sensor networks (WSN) has resulted in numerous technical advances and hundreds of peer-reviewed papers over the last few years. Significant advances in this field have enabled a wide range of commercial and military applications [1].

WSNs are made up of a large number of tiny, low-cost, low-power, multipurpose sensor nodes that communicate wirelessly across short distances. Sensor nodes in WSNs are typically distributed at random in the region of interest and are extensively used for tracking and monitoring duties. Such a network's sensing device may detect data in the surroundings and transmit it to the base station (as shown in Figure 1.1).

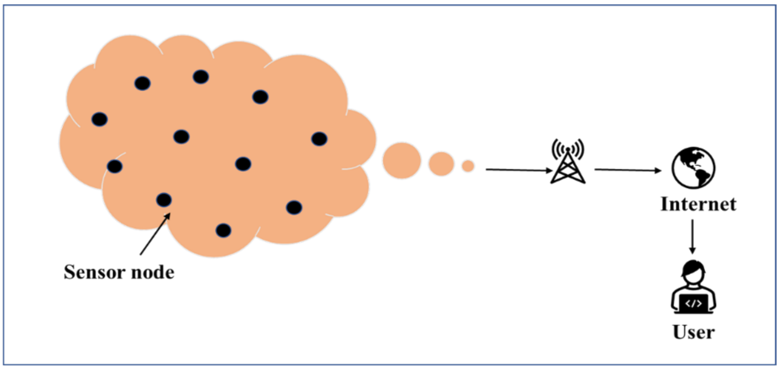

Figure 1. 1 *Architecture of WSN*

With the evolution of technology, various practical uses of WSNs have been developed. WSN applications can be divided into two broad categories: surveillance and tracking. Surveillance is used to analyse, manage, and comprehensively control the operation of a system in real-time. Tracking is commonly used to follow the evolution of an event, an individual, or an animal. Indoor/outdoor environmental monitoring, commercial surveillance, precision agriculture, biomedical or health monitoring, electric network monitoring, and army location monitoring are all examples of existing monitoring uses. Tracking applications include environmental traffic tracking, military target tracking [2], and so on .WSN has various properties that set it apart from other types of networks. Some of the most important WSN characteristics are briefly explained below [3] [4]:

- **Battery energy.** As a result of the power running out, several of the sensor nodes become invalid and are left unoccupied. It is recommended that pre-planned processes and algorithms for the conservation of battery energy be addressed to reduce the impact of this effect. In addition, the amount of energy used by the node for data transmission far surpasses the amount of energy used for the actual implementation of computing. The amount of electricity that a sensor node consumes in order to send a relatively tiny amount of data can be used to implement thousands of lines of code. Therefore, one of the most important issues that experts in communications need to look at is how to reduce the amount of power consumed while a network is in operation to make the most of the wireless sensor network's lifespan.

- **Dynamic.** Some sensor nodes will eventually disconnect from the network, either because the batteries in their devices have run out of power or because of some other issue. Because of the requirements of the tasks, it is possible that certain sensor nodes may become active at the same time, or that the network will receive brand new nodes. Because of this, it is quite likely that the topology of the network will need to be altered; hence, the WSN topology should contain features such as reconfiguration, self-healing, self-adjustment, and so on.
- **Multi-hop communications.** In the WSN, sensor nodes can only communicate with their direct neighbours. If one node needs to connect to another node that is outside the range of the node's radio transmitter, it must employ a multi-hop path that involves intermediate nodes.

1.1 Research Problem, Objectives, and Limitations

WSNs acquire multivariable environmental data. WSNs have limited resources and energy. Radio communication between the node and base station is a major energy consumer. Also, not all WSN applications can use data reduction methods. These techniques save sensor node power in different ways. Designing unique data reduction strategies for WSN nodes with numerous sensors is crucial.

Consequently, the main objectives of my thesis are as follows:

- Provide a critical evaluation of current data reduction methods and routing protocols to increase the WSN energy efficiency.
- Investigate the impact of different data packet sizes on the performance of cluster based WSN.
- Develop and analyse the performance of different approaches of data reduction in WSN.
- Develop a new hybrid data reduction algorithm for cluster based WSN.
- Evaluation of the suggested algorithm's performance and comparing it to that of other algorithms on a variety of performance metrics.

This study has the following assumptions and limitations:

i. The basic features of the system model considered in this study are:

- Each node transmits its data to the cluster head node.
- The cluster head is a normal node that means it has limited energy in the battery.
- The cluster head aggregates the data from the cluster members.
- The cluster head node has limitations in the size of its packets.

ii. The proposed algorithm and cluster-based protocol applies for different artificial dataset with different scenarios (see chapter 5).

iii. The performance metrics to be used in this study are the accuracy of transmitted data, the energy consumption rate, the lifetime of the sensor nodes, etc.

iv. MATLAB is used to simulate the algorithms and to evaluate their performance.

1.2 The Thesis Content

In C**hapter 2** an overview of the data reduction algorithms used in WSN will be presented. I will begin with an overview of the WSN which includes: the main challenges of WSNs; efficient energy use in WSNs; and sensor data compression. Furthermore, we will discuss previous works that focused on data reduction and routing algorithms, with the goal of reducing sensor node energy consumption. Besides, it concludes the limitations of current studies and further research opportunities as well.

In chapter 3 I will investigate different data reduction methods for WSN. These methods are data reduction based EDCD2, FICA, NNF, NNTS, and LRMV. I will use datasets existing on the Internet: Data1-AirQ, Data2-ARHO, Data3-GSB, and Data4_Intel and the parameters used Accuracy, Data redaction ratio and Total energy consumption.

In chapter 4 I will investigate the impact of different data packet sizes on the performance of cluster-based WSN. According to the findings of this study, the size of the sent data packet influences the life cycle of cluster-based WSN. MATLAB was used for the simulation trials. Energy consumption per transmission, alive node each round, and alive node per transmission are the performance indicators measured.

In Chapter 5 I will present a new Hybrid Data Reduction Algorithm (HDRA) for cluster-based WSN developed in course of my PhD stage. The developed HDRA algorithm is explained in detail, HDRA considering the data reduction and routing processes as well. The overall goal of the HDRA algorithm is to reduce the amount of energy that the sensor node consumes. Both the proposed HDRA algorithm and the cluster-based protocol were evaluated by using an artificial dataset with a variety of situations.

Chapter 6 will be dedicated to the general conclusions regarding the results, the author contributions, the dissemination activity, and the future work.

Chapter 2

An Overview of Data Reduction Methods used in Wireless Sensor Networks

Over the last few years, there is a noticed growing interest in the wireless sensor network (WSN) that has led to a lot of technical achievements and a large number of peer-reviewed publications that showed significant results for civil and also military applications [6].WSN is made of different autonomous devices that are distributed spatially in order to provide accurate monitoring for physical and environmental conditions through sensors. WSN has a wireless connection to the Internet through an integrated gateway. In terms of Internet of Things (IoT) the integration of Software Defined Networking (SDN) and IoT to make network control simpler, easier, and less stringy. SDN is a promising new paradigm shift capable of enabling a simplified and robust wireless programmable network to serve a range of physical objects and applications [7].

it is an integrated section of the Internet infrastructure, and it is considered to be a paradigm of communication, with a holistic perspective of connectivity with all objectives available in daily life activities. This makes IoT a concept with the purpose of promoting the Internet immersivity and pervasively in order to make the Internet access easier and to make interaction with the other devices smooth [8].WSN include several amounts of compact, cost-effective, low-power, multifunctional sensor nodes that connect over short ranges without wire inputs. WSN architecture is shown in Figure.2.1 Which also shows the different components for WSN, including sensor nodes, a gateway, and a fusion station [9].

The sensor nodes: they are electronic autonomous devices and can sense and compute the storage as well as the storage of communication. The sensor node has two main components, namely the board and mote. The first one is the sensor that acquires information from the surrounding environment such as light, chemical level and temperature, while the mote is used to integrate the microcontroller and the transceiver of the radio. The base station: it is the concentrator of the data, that were sent by the sensor nodes, and that includes other sensor nodes in the networks. Besides, base station (BS) might remotely setup or even reconfigure the motes.

The gateway: it connects external networks with WSN if there is a need. Also, it is the component that detects information from the surrounding environment, and its corresponding transmission is carried out in conjunction with all the other nodes. In terms of data messages, they are sent to the destination hop by hop in order to reduce power of transmission required by the source node during the direct transmission to the destination.

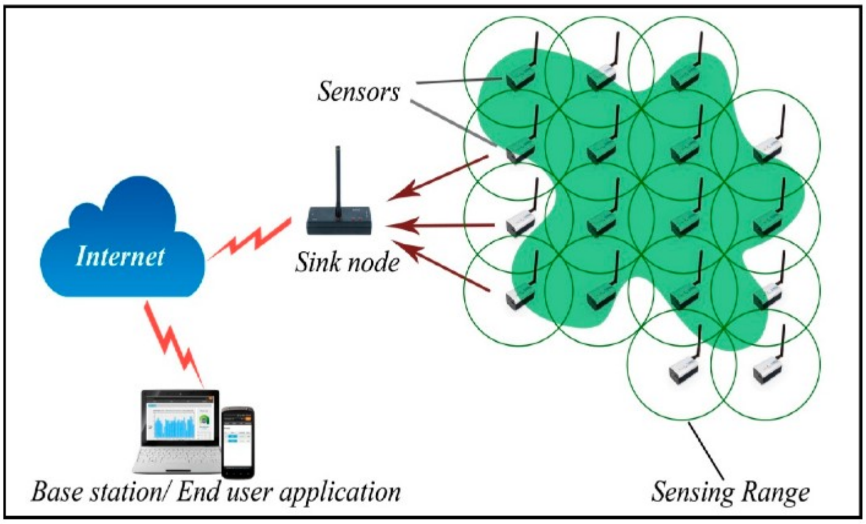

Figure 2. 1 *The architecture of WSN [9].*

1- Design factors

There are many factors that influence designing a sensor network such as programmability, maintainability, resources, fault tolerance, and scalability [9].

- Programmability: there might be a need to program the nodes in some cases such as the addition of a new feature for the environment to monitor. Also, some tasks must be done remotely; for example, the existence of WSN modes in places that are not accessible.
- Maintainability: the adaptation of WNSs to the typological changes is a must such as adding a new node, which also includes any general change in the typology of the network.
- Resources: in the cases, when the resources of energy are limited, the optimization of energy consumption is a major issue. These issues include the inability to change batteries, which has a negative effect on the nodes' lifetime.

2- Data collection and processing

WSN is adaptable in terms of untethered sensing, but it has possible hurdles in terms of obtaining a prolonged lifetime due to the limited energy budget, which is often based on batteries. Although WSN is adaptable in terms of untethered sensing, it has possible hurdles in terms of obtaining a prolonged lifetime. The ability of wireless sensor networks (WSNs) to give sensed data is the fundamental motivation for employing WSNs in many computer applications. This is since communication is an energy-intensive process. One strategy for cutting down on communication while maintaining high data quality is to make predictions based on patterns found in the data that has been sensed. This technique is known as model-driven data gathering, and it is best applicable to the practice of data reporting frequently, which is common in many applications that make use of pervasive computing. In situations like these, the data trend model may be developed locally on one of the nodes, and it would then represent the information that was reported to the data collecting sink. As long as the perceived data and the model forecast are consistent with one another, there is no requirement for any additional communication. As a consequence of this, the perversion state is the only one that calls for the data associated with the model to be modified and then sent back to the sink node.

Because of current challenges such as power consumption, routing algorithms, and efficiency, WSN is an important study topic. Despite the open issues in WSN, there are currently a large number of apps available. In any event, one of the keys aims for the design of any application is to retain WSN if possible, in terms of life cycle and function. [10].

2.1 Challenges of wireless sensor networks

- **Energy:** Sensors require power to perform a variety of tasks. Data collecting, information handling, and data communication all cost energy; also, continuous monitor to the info intended for accurate process requires a substantial expanse of the energy from the sensor nodes component's central processing unit when they are idle. After they have been depleted, power batteries must be replaced or recharged. Because of demographic factors, it can be difficult to recharge or replace batteries [11].
- **Security:** WSN is not only used on the battlefield, but also for surveillance, building monitoring, and burglar alarms, as well as in critical systems like airports and hospitals. This makes security a very hard problem to solve. In WSN, data sending between the sensor nodes or between the sensors and the BS needs to be kept private. If it isn't, someone could listen in on the conversation. In sensor networks, individually node and the BS need to check that the data they received from a reliable transmitter and not from an attacker who deceived genuine nodes into tolerant untruthful data. Untruthful information can make it hard to predict how a network will work. The data should be kept in good shape. Data shouldn't change, and users should get the correct data.
- **Architecture:** Architecture defined as a set of regulations for achieving functionality, as well as a collection of interfaces, functional components, protocols, and physical hardware. WSN structural design should be both long-lasting and inaccessible. As if the number of nodes were to be raised, since WSNs do not have a stable set of communication procedures to which they must comply, the quality of service (QoS) must be adaptable to fulfil the vast range of intended application situations. Since there is a one-to-one correspondence between. Computational speed and transmission bit rates might result in suboptimal energy performance.

2.2 Efficient Energy use in WSN

This paragraph is focusing on a synthesis of Data reduction algorithms. Anyone can remark that for reducing the consumed energy there are many different algorithms. We can identify five main classes of energy efficient techniques, namely:

- Data reduction.
- Radio optimization.
- Energy efficient routing.
- Sleep/wakeup schemes.
- Charging solution.

In this summary I will briefly dwell on the first three.

1. **Data Reduction:** To limit the quantity of data delivered to the sink, nodes along the channel execute data fusion. Data compression is a method of encoding information that reduces the number of bits required to represent the initial message. It saves energy by shortening transmission times as packet sizes decrease. Existing compression algorithms, however, are inapplicable to sensor nodes due to resource constraints. As a result, particular strategies for adapting to the processing and power capacities of wireless motes have been devised [14] [15] [13]. The authors of [16], devised an adaptive data aggregation system with the goal of utilizing the spatial correlation between sensor nodes. It is a mechanism used on the cluster head node level with unidirectional data, and each cluster member node supports a single sensor. In [17], the authors suggested two approaches to minimizing the CH packet size by considering the accuracy of prediction of sensed data at the base station. The proposed coding schemes are the based relative difference (CS- RD) and the based factor of precision (CS-FP).
2. **Radio optimization:** The radio module is the main component that causes battery depletion of sensor nodes. To reduce energy dissipation due to wireless communications, researchers have tried to optimize radio parameters such as coding and modulation schemes, power transmission and antenna direction.
3. **Modulation optimization:** aims to find the optimal modulation parameters that result in the minimum energy consumption of the radio [13].

2.3 Sensor data compression

There are a variety of engineering applications where the accuracy of estimation can be improved by using data from distributed sensors. However, in order to obtain an accurate estimate, the data must first be compressed. This is necessary due to the fact that the fusion centre only has a limited number of connections and computing capacity. One method for achieving this goal that does not involve the loss of information is to lessen the number of data dimensions. It is possible to utilize best linear unbiased estimation, often known as BLUE, to compress data from local sensors to a size that is acceptable.

2.3.1 Overview Sensor Data Compression

A wireless sensor network (WSN) is made up of many separate sensor nodes that work together to complete a specific task. Each wireless sensor node has a number of sensors, a central processing unit (CPU), a communication module, and a battery-powered energy source.

WSNs pose a number of problems, the most important of which is how well they use power. This is because the battery life has a big effect on how long the network will last. The energy stored in the sensor is used up during processing and transmission. It has been found that around 80% of the energy used is for moving data around in the network through the connection. To make sensor nodes last longer. There have been many different suggestions for how to make wireless sensor networks use less energy [24]. the performances of different algorithms that can reduce the number of data packets transmitted by sensor nodes within the WSN, the performance comparison shows that the based-on reduction in the data packet transmissions from the source to the sink [25].

2.3.2 Methods for data segmentation

In partitioning methods, there are two basic clustering methods, which are inherently different from each other. Partitioning methods were very common in the past, especially in market segmentation research. Nevertheless, this kind of partitioning has two drawbacks. First, the use of diverse categories increased the difficulty of finding homogeneous market segments. Second, it is a must for the investigator to decide on the number of clusters to be formed in advance. In typical methods, an investigator in a segmentation study starts blindly, because the investigators might not even know of the existence of clusters on the data.

There are also hierarchical methods, recognized commonly as bind methods, and many other an essential method, and most of them start through the identification of guest pairs that match each other according to the selection of identical categories. Hierarchical methods, however, give a beginning point without any provision for produced clusters, which are balanced and homogenous. Another recent alternate method is the two-phase cluster analysis method. That is the investigator starts with the hierarchical method, and then uses the iterative partitioning method. The hierarchical method produces the incipient clusters, and then the reiterated partitioning process links several cases to other clusters through a series of detach passes, generated by the computer. During the iteration cycles, it changes cases between the clusters to improve the clusters’ homogeneity, which makes the two-stage method able to give tighter clusters [15].

2.4 Related Work

We have identified current works aimed at improving network performance by focusing on data processing, distribution, and scheduling. In the following, I will give some considerations on three methods of data reduction: data reduction based on coding, data reduction based on prediction, data reduction based on multiple sensors.

2.4.1 Data Reduction Based on Coding

In [29], the authors presented a proposal for a configurable, safe, and clandestine distributed data stockpiling system that is capable of processing encrypted data and controlling the calculation output based on multi-cloud technology, based on the properties following characteristics: homomorphic cipher, weighted secret share scheme and error-correcting codes. We use error corrections codes on the excrescent residue number System.

In [32], the authors suggested the Lasso (least absolute shrinkage and selection operator) approach, and the aggregation of multivariable sensor data in various co-ordinate spaces, utilizing the form and characteristics of sensor data, to minimize the overall energy usage.

2.4.2 Data Reduction Based on Forecasting

In [37], the author suggested the Dual Prediction System (DPS) for the explanation of the expected effect of DPS's an accumulation of data on WSN's by the transmitting the predictive model from either sensor nodes to gateways (GW).

A previous study [38], aimed to values prediction for other indicators, referred to as Dependent Indicator Sets (DS), the authors suggested a connexion-based on the reduction method for streaming the data derives from the quantitative formulas between correlated indicator sets, known as Regress or Indicators Set (RS).

2.4.3 Data Reduction Based on Multiple sensors

The study [6] suggested an approach to provide more effective real time data collection model for multivariate sensor network wireless and IoT called RDCM applications and to lower the payload bits until it is submitted to the fusion centre (FC).

the author [14] suggested a new approach for assessing the efficiency of different multivariate WSN data reduction models by measuring the model threshold during the training process.

2.5 Conclusions

WSN makes it possible to do flexible and untethered sensing, but it faces the challenge of extending its lifetime while working within a constrained energy budget, a challenge that is typically associated with batteries. As a consequence of this, it is frequently the case that the primary source of energy depletion is communication. Because of this, highlights some of the previous efforts that have been made on data reduction methods for WSN.

In addition to this, it provides a concise summary of the limitations of the already available studies as well as potential directions for future research. The majority of the work that needs to be done with the most recent data reduction technologies can be done by sensor nodes being able to convey data directly (in a single hop) to the base station (BS). Before sending data to the BS, the majority of currently used routing protocols (multi-hop) do not take into consideration the possibility of data reduction. Therefore, this work proposed new algorithm with considering the data reduction and routing as well, which aims to save energy consumption by the sensor node.

Based on the research described in this chapter I published the next papers:

[70] M. K. Hussein, "*Data Reduction Algorithms for Wireless Sensor Networks Applications: Review,*" 2021 3rd International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA), 2021, pp. 1-7, Doi: 10.1109/HORA52670.2021.9461309.

[7] H. Mohammed, R. M. KHALEEFAH, M. k. Hussein and I. Amjad Abdulateef, *"A Review Software Defined Networking for Internet of Things,"* 2020 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA), 2020, pp. 1-8, Doi: 10.1109/HORA49412.2020.9152862.

Chapter 3

Performance of Data Reduction Algorithms for WSN using different Real-Time Datasets

The energy efficiency of a sensor node is affected by the process of data packet transmission from the sensor node to the base station (BS) and the amount of the transmitted data. Therefore, this section uses a diversity of datasets to examine how data reduction techniques affect WSN performance. The used data reduction techniques are “Neural Network Fitting”, “Network Time Series”, “Linear Regression with Multiple Variables”, “an efficient data collection and dissemination”, and “Fast Independent Component Analysis” which are named NNF, NNTS, LRMV, EDCD2, and FICA. Respectively: Data 1-AirQ, Data 2-ARHO, Data 3- GSB, Data 4- Intel. The performance metrics measured are energy consumption, data accuracy, and percentage of data reduction.

3.1.4 Data Reduction based – An Efficient Data Collection and Dissemination (EDCD2) Algorithm

EDCD2 is a scheme to bring up-to-date measured data to the BS [54]. EDCD2 was used to decrease the number of transferred packets from nodes (multiple sensors) It should be noted that there are two versions of EDCD, EDCD1, and EDCD2 for sensor nodes with one and multiple sensors, respectively. For IoT sensor node supported single and multiple sensors, the described methods are marked as EDCD1 and EDCD2, respectively. The sensor data in a WSN might just have a single characteristic (a univariate), or it might contain numerous characteristics (multivariate). The data collected by the sensor board is referred to as univariate data. This is because the sensor board is only meant to collect one type of data (light, temperature, or humidity). In a similar vein, some WSN systems have each sensor board supplied with multivariate sensors so that it may satisfy the requirements of a wide variety of applications.[14].

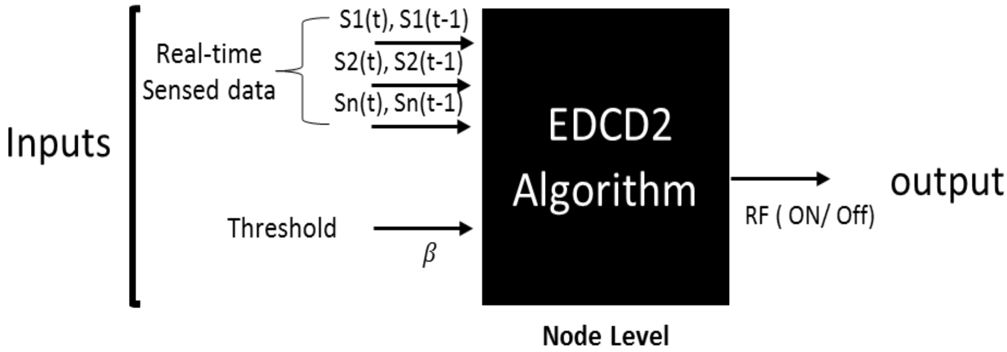

Figure 3. 4 *General block diagram of EDCD2 algorithm*

In this section, the application of the EDCD2 to reduce the size of data transferred is described. Figure 3.4 shows the block diagram of WSN data reduction based on the EDCD2 algorithm. The basic idea of EDCD2 is to avoid transmitting the sensed data if the value of the relative difference between the currently sensed data S(t) and the last transmitted data S(t-1) is smaller than the threshold value β for all sensors of the same node, otherwise, the sensed data S(t) will be transmitted to the BS. The detailed description of the EDCD2 algorithm based on data reduction is given in the following pseudocode [54].

//EDCD2//

Inputs:
$S_i(t), S_i(t-1)$ for each sensor $S_i$and β.

Output: D_s
Begin:
For $i = 1{:}n$ Do // i=1, 2, n ;
Set $S_i(t-1)$← last measuring value transmitted by the sensor S_i
Read: the sensor value (SV_i) at t time
Set$S_i(t) \leftarrow (SV_i)$
//Calculate the relative differences (R_f)
R_f=Abs $(S(t) - S(t-1))$ / $(S(t) + S(t-1)) \times 0.5)$
If $R_f > \beta$ Then
Set $SS_i \leftarrow 1$
Else: Set $SS_i \leftarrow 0$
End if
End For
// Recalculate the node data size (D_s)
Set $D_s \leftarrow 0$;
For $i = 1{:}n$ Do
$D_s = (D_s + (SV_i \times SS_i\ )\ )$
End For
// The decision to send data
If $D_s = 0$ Then
RFtransmit (Off) // no update / no send
Else: RFtransmit (On) // update(send)
End If
End Algorithm

3.2 Real-Time Datasets

The considered algorithms are evaluated on different benchmark real-time datasets, as described in the following subsections. It's important to note that, usually only part of the data from specific nodes of these datasets are used to assess the performance of current data reduction methods in WSN [57][58][59][6][60][61][62]. The reason is that most data reduction methods focus on reducing the amount of transferred data without considering how this data is forwarded to the CH /BS. In other words, they assume that the sensor nodes can directly transmit the sensed data to the CH /BS. The selected algorithms NNF, NNST, EDCD2, LRMV, and FICA are evaluated on real-time datasets as shown below:

3.2.1 Data 1-AirQ

Data 1- Air Quality (AirQ) is a WSN data set, including air pressure, humidity, and temperature sensors. These sensor data have been collected by 56 sensor nodes in year 2017 at Krakow, Poland. The number of the sample is 3500. For more information, see the main source [63]. Figure 3.6 shows the structure of Data1- AirQ. In addition, some samples of sensors value provided in Tables 3.1

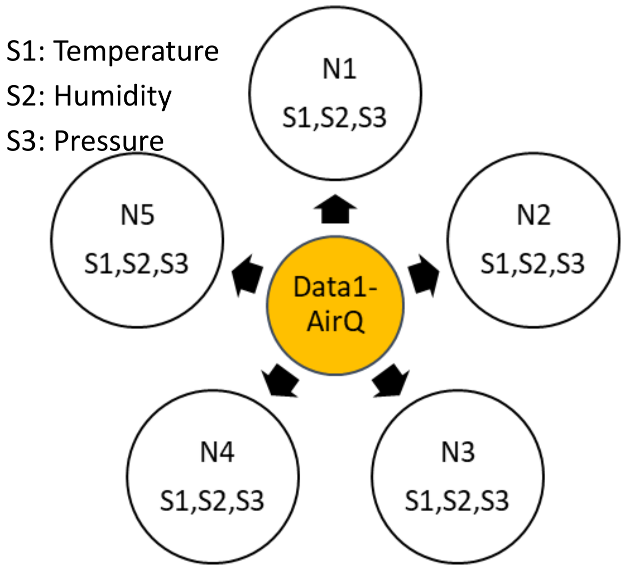

Figure 3. 6 *Structure of Data1- AirQ*

Table 3. 1 *Some data samples of node1 – Data1- AirQ*

Sample	Temperature[°C]	Humidity [%]	Pressure [Pa]
1	6	92	101906
2	6	92	101869
3	5	94	101837
4	5	92	101834
5	4	94	101832
6	5	94	101833
7	9	78	101842
8	11	66	101831
9	15	50	101798
10	17	42	101745

3.4 Simulation and Results

Figure 3.11 shows the accuracy of the applied algorithms EDCD2, FICA, NNF, NNTS, and LRMV for all selected nodes N1, N2, N3, N4, N5 in case of DATA2-ARHO dataset. From the results, the EDCD2 algorithm has the best accuracy compared to the other algorithms FICA, NNF, NNTS, and LRMV. The reason for this is the average total absolute error which has the lowest value of 0.199 when EDCD2 is used for all nodes. EDCD2 is not based on data reduction training methods like FICA, NNF, NNTS, and LRMV. The key idea of EDCD2 is to avoid transmitting the sensed data if the value of the relative difference between the currently sensed data S(t) and the last transmitted data S(t-1) is smaller than the threshold value β for all sensors of the same node, otherwise, the sensed data S(t) will be transmitted to the BS

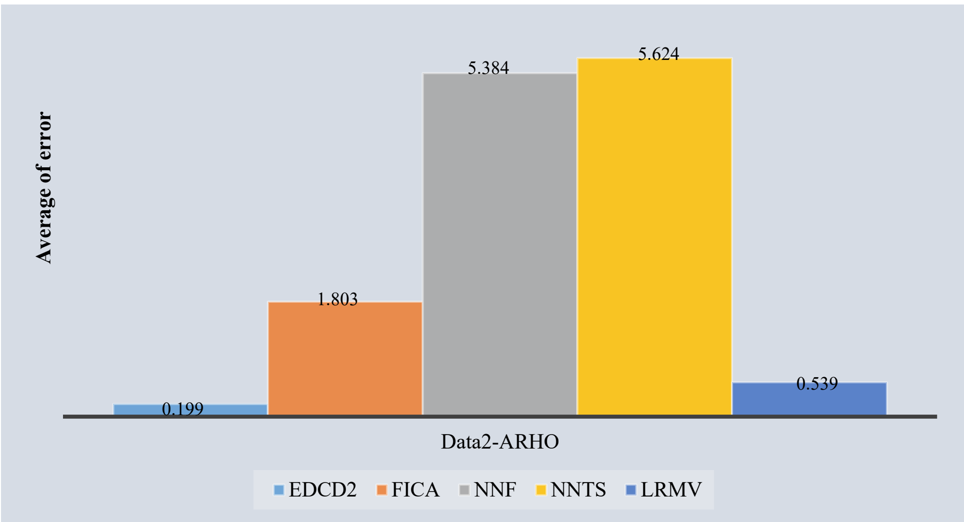

Figure 3. 11 *Accuracy of Applying Various Algorithms for all Selected Nodes from DATA2-ARHO*

In comparison with FICA, NNF, NNTS, and LRMV, as described in the pseudocode, these methods require reference parameters to be determined during the training phase, and the validation of the model will be affected by the value of these parameters and the number of samples. Furthermore, the accuracy of the data reduction models that is dependent on training decreases over time due to the increment in the approximation error. Moreover, NNTS and NNF algorithms have the worst performance in terms of accuracy, and the average absolute errors are 5.38 and 5.62, respectively. In summary, EDCD2 is a threshold-based data reduction algorithm. EDCD2 transmits measurement data only when the relative difference between the current measurement data and the last transmitted data is larger than the threshold value.

3.5 Conclusion

This chapter investigates the influence of data reduction techniques on the functionality of WSN by means of different datasets. In MATLAB, experiments are run for various strategies to decrease the quantity of data supplied. Real-time data sets are used to test the assigned approaches EDCD2, NNST, FICA, LRMV, and NNF. Data accuracy, energy usage, and percentage of data reduction are the performance measures measured. The study's findings reveal that the chosen algorithm aids in the reduction of sent data and energy usage, and the effectiveness of every algorithm differs depending on the data that is employed. EDCD2, for example, is not reliant on data reduction training approaches like FICA, NNF, NNTS, and LRMV.

The main idea behind EDCD2 is to avoid transmitting sensed data if the relative difference between the currently sensed data S(t) and the last transmitted data S(t-1) for all sensors in the same node is less than a certain threshold value; otherwise, the sensed data S(t) will be transmitted to the BS. As described in the pseudocode, FICA, NNF, NNTS, and LRMV require the determination of reference parameters during the training phase, and the validation of the model is reliant on the value of these parameters' dataset and the number of samples.

The results presented in this chapter was published in the papers:

[72] Hussein, M K; Marghescu, Ion; Nayef. A.M. Alduais: *Performance of Data Reduction Algorithms for Wireless Sensor Network (WSN) using Different Real-Time Datasets: Analysis Study,* 2022 International Journal of Advanced Computer Science and Applications (IJACSA).

[25] M. I. Husni, M. K. Hussein, N. A. M. Alduais, J. Abdullah and I. Marghescu, *"Performance of Various Algorithms to Reduce the Number of Transmitted Packets by Sensor Nodes in Wireless Sensor Network,"* 2019 11th International Conference on Electronics, Computers and Artificial Intelligence (ECAI), 2019, pp. 1-7, doi: 10.1109/ECAI46879.2019.9042081, WOS:000569985400

Chapter 4

Impact of various data packet sizes on the performance of cluster-based WSN

4.1 The LEACH Routing Algorithm

Low-energy adaptive clustering hierarchy (LEACH) is a pioneer in the field of WSN routing. LEACH's main purpose is to enhance energy efficiency by randomly counting, using a rotary cluster head (CH). Figure 4.1 shows the concept of the LEACH protocol. The LEACH operation consists of various runs, each phase is separated into two steps: the set-up phase and the continuity period as seen in Figure 4.2.CH selection, classification, and allocation of CH programs to members' nodes via the Time Division Multiple Access during the installation phase (TDMA). Each node participates with a changed priority value of 0 to 1 in the CH election process. If the random number generated for the sensor node is lower than T(n), then this node is CH. Equation 1. is the basis for T(n) value.

$$T(n) = \begin{cases} \dfrac{p}{1 - p * (r \, mod \, \frac{1}{p})} & : IF \; n \in G \\ 0: & otherwise \end{cases} \tag{4.1}$$

When P shows the intended number of sensor nodes to be converted into CHs, including all the sensor nodes, r represents the present round and G refers to the number of sensor nodes which did not take place during the preceding 1/P round of the CH elections. A node that becomes the CH in round r will not be able to participate in the next 1/P rounds.

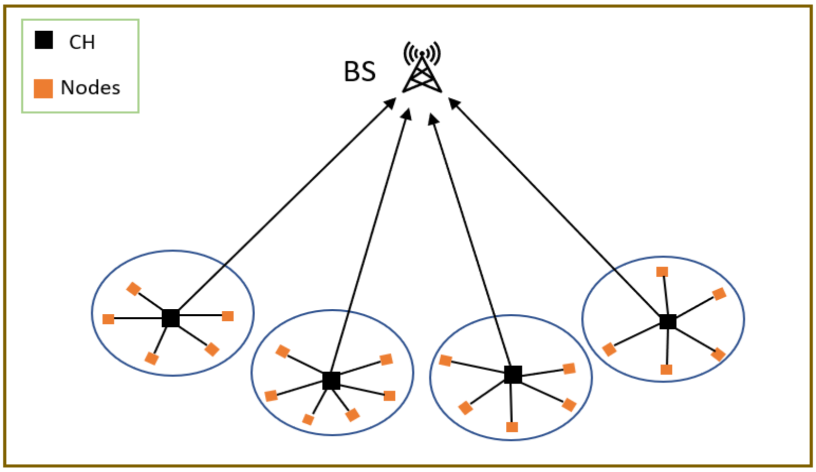

Figure 4. 1 *Network design for LEACH protocol*

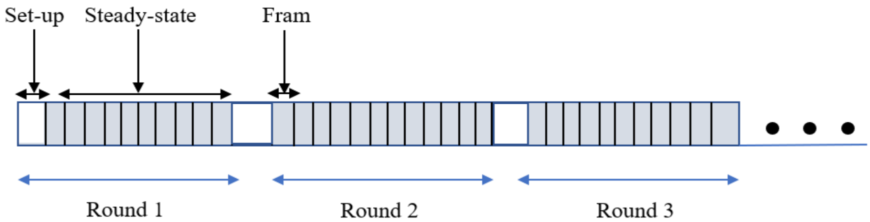

Figure 4. 2 *Timeline of LEACH operation: the set-up and steady state phase*

4.3 The effect of data packet size on the performance of cluster-based WSN - Scenario 1

To evaluate the effect of the data packet size on the performance of the cluster-based WSN networks. In this scenario, BS-location (20, 40) and sensor filed area (50m x 50m) are fixed. Data packet size are set in Table 4.3 for frame size. The number of nodes is 54 and the initial energy of each node begins at 2 Jules. The nodes are placed in the field of the sensor as shown in Figure 4.4

Table 4. 1 *Various data packet size*

	Frame size	data packet size
Packet 1	XBee - ZigBee-sensors 1	656 bits
Packet 2	XBee - ZigBee-sensors 2	696 bits
Packet 3	XBee – 802.15.4-sensors 1	704 bits
Packet 4	XBee – 802.15.4-sensors 2	744 bits
Packet 5	XBee - ZigBee-sensors 4	776 bits
Packet 6	XBee – 802.15.4-sensors 4	824 bits

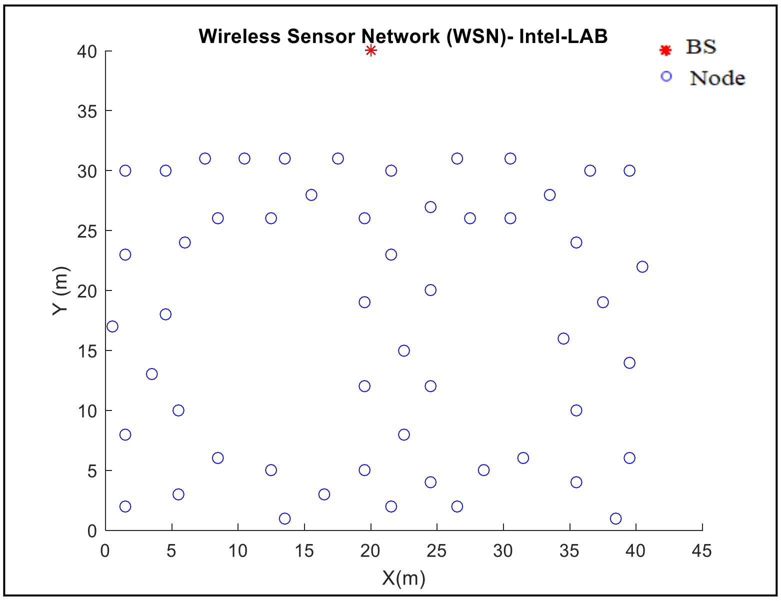

Figure 4. 4 *Sensor field area 50m x50m*

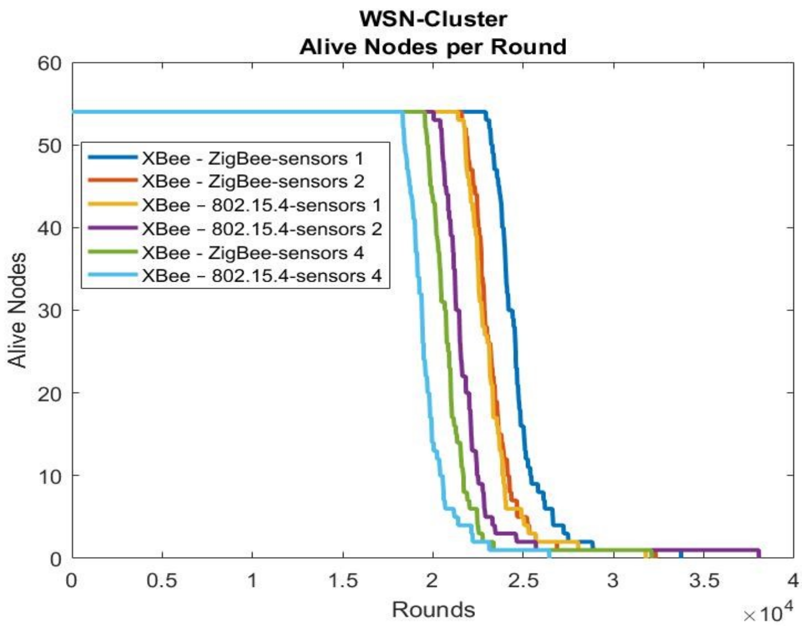

Figure 4. 5 *The number of nodes alive per round by applying various packet sizes*

Figure 4.5 illustrates the number of nodes each round by using a cluster-based protocol for various data packet sizes as described before. According to the results, the maximum number of alive nodes observed when using the XBee - ZigBee-sensors1 (656 bit) indicates that the networks' lifetime remains until 2 rounds. In other words, the minimum energy consumed by the network was at XBee - ZigBee-sensors 1. Conversely, when the data packet size is 824 bit (802.15.4-sensors 4), the number of active nodes was the least, which means that the lifetime of the network merely will last up to 2.3 rounds.

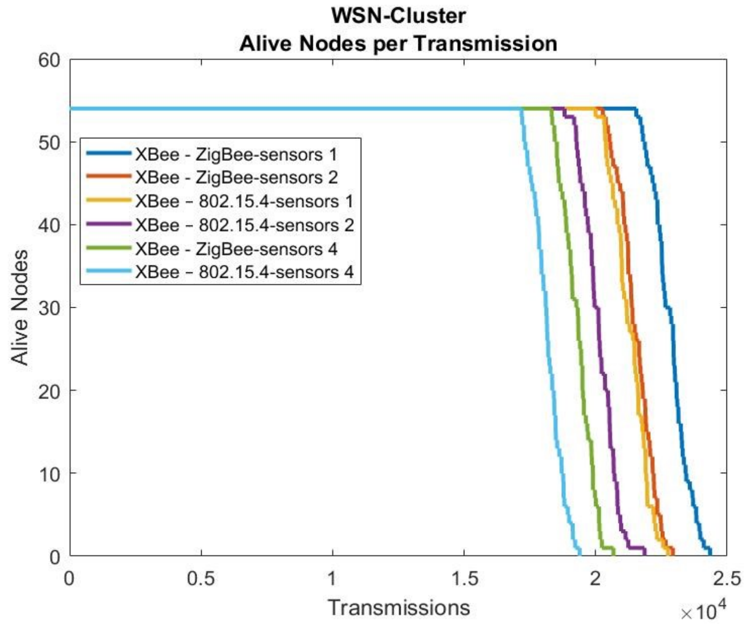

Figure 4. 6 *The number of nodes alive per transmissions by applying various packet*

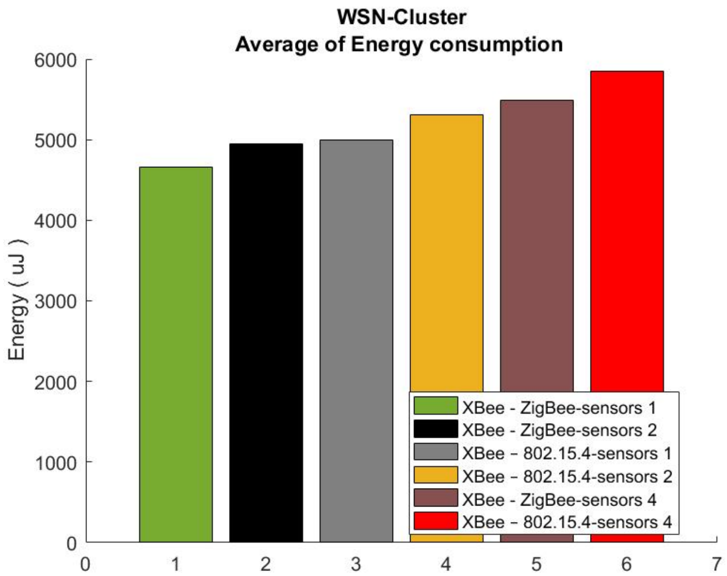

Figure 4. *7 Average Energy consumed per transmission*

In addition, Figure 4.6 and Figure 4.7 showed the number of alive nodes and average energy consumption per transmission by applying cluster-based protocol for different data packet sizes. we can observe from the maximum numbers for alive nodes when using the XBee - ZigBee-sensors1 (656 bit), which that is means the number of transmissions reached up. Conversely, the minimum number of alive nodes occurred when the data packet size is 824 bit (802.15.4-sensors 4) and the number of transmissions reached up only. In short, the data packet sizes affect the performance of the network. When the data packet size is 656, 696, 704, 744, 776, and 824, number of the active nodes is significantly reduced according to the data packet size.

4.5 Conclusion

The effect of different packet sizes on the performance of WSN-based clusters is investigated in this chapter. MATLAB was used for the simulation trials. Energy consumption per transmission, alive node each round, and alive node per transmission are the performance indicators measured. According to the findings of this study, the size of the sent data packet influences the life cycle of cluster-based WSN.

Based on the results of using the cluster protocol in different situations, such as indoors (54 nodes, 50 x 50 m2) and outdoors (200 nodes, 700 x 700 m2), the protocol works better indoors than outdoors because the cluster heads and nodes are closer together. But in this thesis, the challenge is to improve network performance over a wide area. It is advised that routing protocols be integrated with data reduction strategies to obtain high performance. The research developed and analysed in this chapter was disseminated in paper:

[71] M. K. Hussein, *"Impact of various data packet sizes on the performance of WSN-based clusters: Study,"* 2021 5th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT), 2021, pp. 424-428, Doi: 10.1109/ISMSIT52890.2021.9604692.

Chapter 5

A Hybrid Data Reduction Algorithm (HDRA)

Nowadays, wireless sensor networks (WSNs) appear as an active research area due to challenging issues, including power consumption, routing algorithms, efficiency, and so on. Notwithstanding the open problems in WSN, there is already a high number of applications available. In any case for the design of any application, one of the main goals is to keep WSN if possible, in terms of the life cycle and function. In most of the work of the latest data reduction methods, sensor nodes can send data directly (single hop) to the base station (BS).

Another issue, the most of the current routing protocols (multi-hop) didn't consider the problem of data reduction before forwarding the data to the BS. Therefore, this work proposed new algorithm with considering the data reduction and routing as well, which aims to save energy consumption by the sensor node. Table 5.1 shows the Simulation parameters. The proposed algorithm and cluster-based protocol applies for different artificial dataset and scenarios as described as following:

5.1 System Scenario

- Scenario #1: Generate an Artificial dataset that includes one type of sensor which is a Temperature sensor. Then set the generated data for 100 nodes randomly deployed in filed area size 200 ×200 m^2.

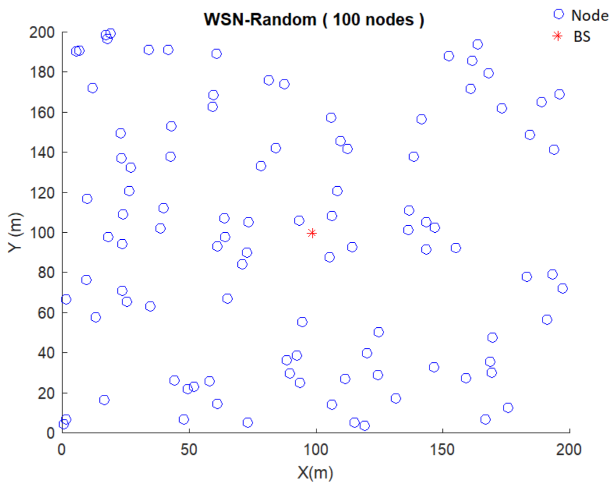

Figure 5. 1 *Sensor field area 200m x 200m*

Scenario #2: It is generated Generate an Artificial dataset that includes one type of sensor which is Temperature. Then set the generated data for 200 nodes randomly deployed in filed area size (200 ×200) m^2.

Scenario #3: It is generated Generate an Artificial dataset that includes one type of sensor which is Temperature. Then set the generated data for 200 nodes randomly deployed in filed area size (400 ×400) m^2.

Scenario #4: It is generated Generate an Artificial dataset that includes one type of sensor which is Temperature. Then set the generated data for 400 nodes randomly deployed in filed area size (800 ×800) m^2.

Repeat Scenarios #1-4 with change the type of sensors data to Humidity.

Case 2: Generate an Artificial dataset that includes two types of sensors which are Temperature and Humidity. Then set the generated data for all scenarios #1-4.

Table 5. 1 *Simulation parameters*

PARAMETERS	VALUE
Initial Energy of a Node (E_o)	1 JOULE
E_{TX}	50×10^{-9}
E_{RX}	50×10^{-9}
TRANSMIT AMPLIFIER (E_{AMP})	100×10^{-12}
DATA AGGREGATION ENERGY (E_{DA})	5×10^{-9}
THE PERCENTAGE OF CH (P)	0.5%
NUMBER OF NODES	SCENARIO#1-4
FIELD AREA	SCENARIO#1-4
BS LOCATION	SCENARIO#1-4
PACKET SIZE(BITS)	500

5.2 The HDRA Algorithm presentation

In this section I will present the HDRA (Hybrid Data Reduction Algorithm) which was developed during the PhD stage. The main purpose of HDRA is to extend the lifetime of the entire cluster based WSN. It aims to reduce the number of packets transmitted by sensor nodes if the value of the sensor reports does not change significantly. In other words, the proposed HDRA aims to make decisions before forwarding sensor node data to the cluster head. If no change is required, the update data will be changed to zero, otherwise, the node needs to update its data to the cluster head.

A detailed description of the proposed algorithm is illustrated in the pseudocode below.

// Proposed Algorithm

1. ***Set*** $(\boldsymbol{X}, \boldsymbol{Y})$ ***// Field*** *Dimensions in meters* $x\ \mathrm{m}\ X\ x\ \mathrm{m}$
2. ***Set*** *"Number of Nodes in the field"*
3. ***Set*** *Initial _Energy ←1//units in Joules //*
4. ***Set ENERGY MODEL*** *// "Energy required to run circuity (both for transmitter and receiver) See Table X"*
5. *Set NCH←p*Number of Nodes; // "p=0.05 Number of Clusters" According to Ref [17]*
6. ***Set*** *Round ←1 // Round of Operation//*
7. ***Set*** *Alive _nodes ←* $Number\ of\ Nodes$ *;*
8. ***Set*** *Number of Transmissions←0;*
9. **FOR** each NODE (i) **do**
10. **Call** NOED(i)Parameters ← **Node Structure(*i*)** *// Call Node Structure ()*
11. ***SET*** NOED(i). update Data ← 1 *// To Send only first value for Node i at Round 1*
12. ***SET*** *NODE(i). L_ Transmitted Data= [NODE(i). Sensor Data (1))]// Set the first Raw sensors value [T, H] to the last transmitted data at time (t-1)*
13. **END** FOR
14. **WHILE** *Alive _nodes* > 0 **do**
15. **IF** Round > 1 // *When Round =1 non-reduction mod // Only one time to transmit first value//*
16. **FOR** each NODE (i) **do**
17. ***// Phase Updating data***
18. **Set** $\boldsymbol{S(t-1)}$← NODE (*i*). *L_ Transmitted Data*
19. **Read:** *the current sensor value at* t *time (Round)*
20. **Set** $S(t)$ ← NODE (i). Data (Round)
21. **//Calculate the relative differences ($\boldsymbol{R_f}$)**
22. $\boldsymbol{R_f}$=Abs $(S(t) - S(t-1))$ / $(S(t) + S(t-1)) \times 0.5)$
23. **If** $R_f > Thrd$ **Then**
24. **Set:** NOED(i). update Data ← 1*;*
    NODE (i). *Transmitted Data* ← NODE (i). Data (Round)
25. **Else:** **Set** NOED(i). update Data ← 0
26. **End if**
27. **Set Round** ← $(Round$ *+1)*
28. **END FOR**
29. **FOR** each NODE (i) **do**
30. **If** NODE(i). update Data==1
31. **CALL** SELECT_CH () //
32. **CALL** STEADY-STATE PHASE //
33. **Determine** *// Energy Dissipation for Normal Nodes & Cluster Head //*
34. **IF** NODE(i). E<=**0**
35. Dead nodes← Deadnodes +1
36. *Alive _nodes←Alive_nodes* – 1
37. **END FOR**
38. **END WHILE**
39. **END**

5.2.1 Implementation and performance evaluation

This section studies the performance of the proposed algorithm and cluster-based protocol (See section 5.2). The value of threshold is set to 0.01, 0.03, and 0. 05. The proposed algorithm and cluster-based protocol applies for different artificial dataset as described in the next subsection.

5.2.2 Artificial Dataset

To evaluate the proposed algorithm, an artificial dataset based on real-time dataset have been generated. To do so, the dataset selected based on the sensor node area.

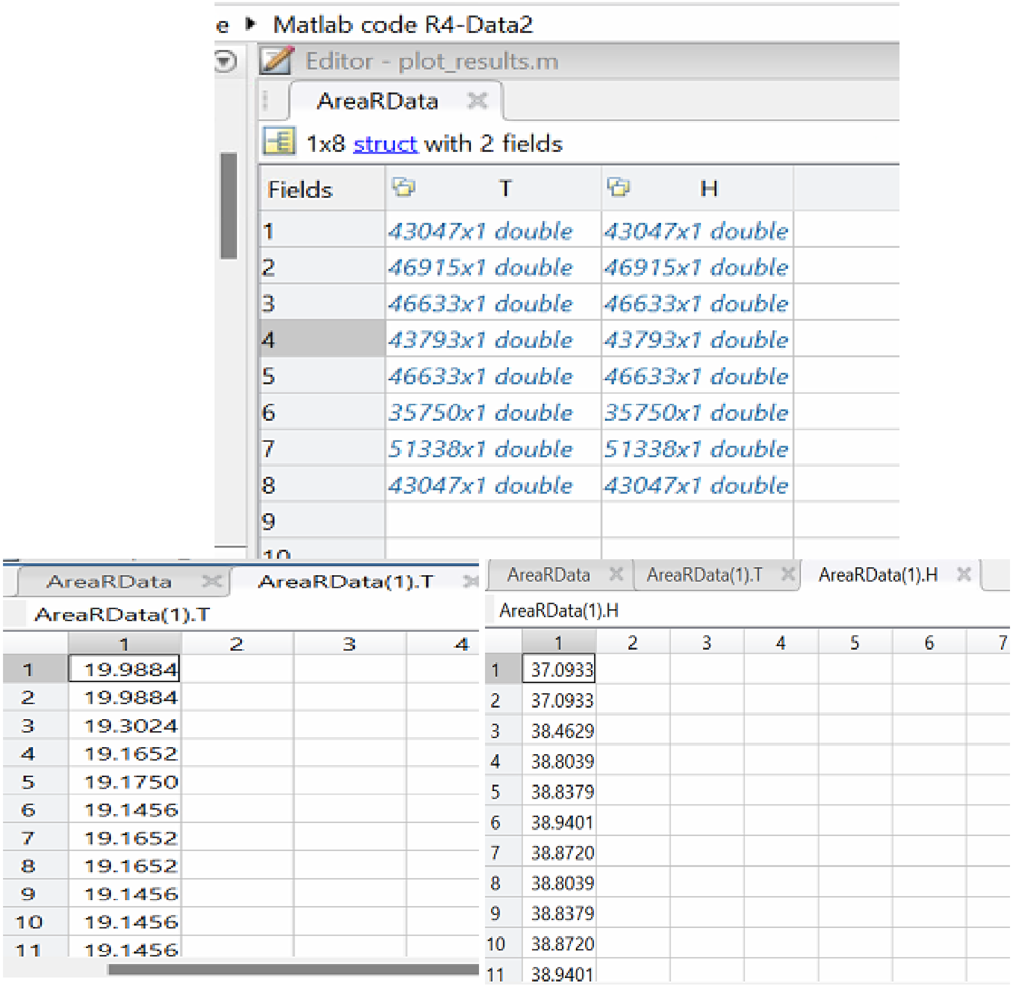

***Figure 5. 2** Sample of real-time data*

NODEData

1x200 struct with 3 fields

Fields	Area	DataT	DataH
1	5	200000x1 double	200000x1 double
2	7	200000x1 double	200000x1 double
3	5	200000x1 double	200000x1 double
4	3	200000x1 double	200000x1 double
5	4	200000x1 double	200000x1 double
6	4	200000x1 double	200000x1 double
7	5	200000x1 double	200000x1 double
8	6	200000x1 double	200000x1 double
9	1	200000x1 double	200000x1 double
10	5	200000x1 double	200000x1 double

***Figure 5. 3** artificial dataset*

NODEData(2).DataT

	1	2	3	4
1	22.3600			
2	20.3706			
3	21.2232			
4	24.0358			
5	24.4474			
6	25.8194			
7	23.3596			
8	24.4474			
9	18.1460			
10	23.7614			
11	18.5478			

NODEData(2).DataH

	1	2	3	4	5	6	7	8
1	32.9308							
2	36.6118							
3	37.0933							
4	37.0246							
5	41.5789							
6	32.8257							
7	41.5116							
8	34.5360							
9	41.7133							
10	38.7357							
11	34.8139							

***Figure 5. 4** Sample artificial dataset*

5.3 Analysis of the Results

5.3.1 Scenario 1

Figure 5.8 and Figure 5.9 show the residual energy and the number of live nodes per transmission when using the cluster-based routing protocol alone as well as this protocol combined with three variants of the data reduction algorithm (p=0.01, 0.03, 0.05) considering several types of sensors (Temperature, Humidity, and Temperature& Humidity), the analysis is performed on all scenarios. Here I will only refer to the results obtained for scenario 1. From the results, it's clear that the maximum number of alive nodes observed when applying the proposed algorithm, which means that the number of transmissions reached up to 21000, 26000, 16000 for T, H and TH respectively.

Conversely, the minimum number of alive nodes occurred by a cluster-based protocol without considering the sensor node updating data status and the number of transmissions reached up to 11000. In addition, Figure 5.10 shows the number of alive nodes per round by applying cluster-based protocol and the proposed algorithm for Scenario 1 with sensor type set to T, H and TH. According to the results, the proposed algorithm improves the network lifetime having in mind that the maximum number of rounds was 83000 rounds compared to 12000 rounds for cluster-based protocol. In short, the proposed algorithm improved the performance from the point of view of the network lifetime and saved energy.

Temperature & Humidity | Temperature (T) | Humidity(H)

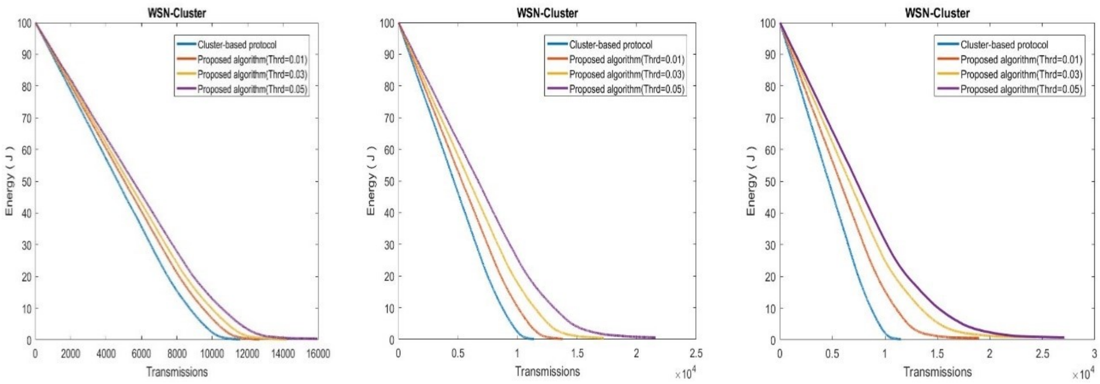

Figure 5. 8 *Residual energy*

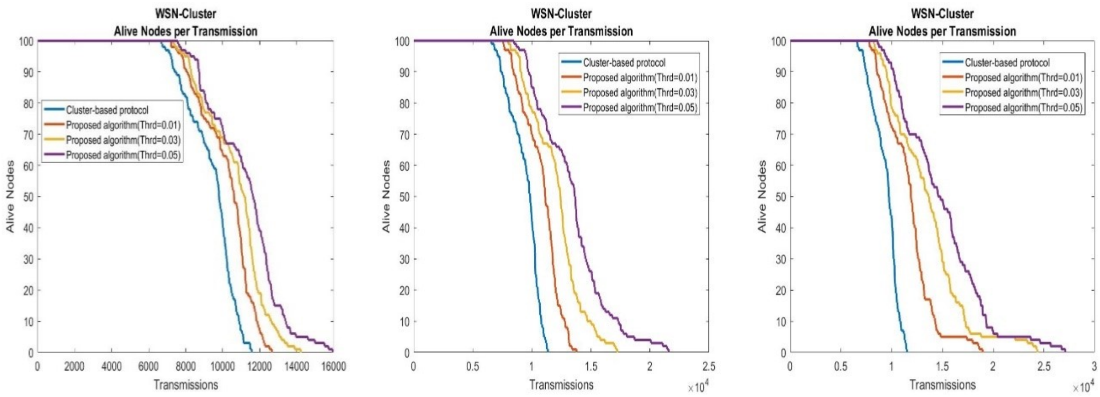

Figure 5. 9 *Alive Nodes Per Transmissions*

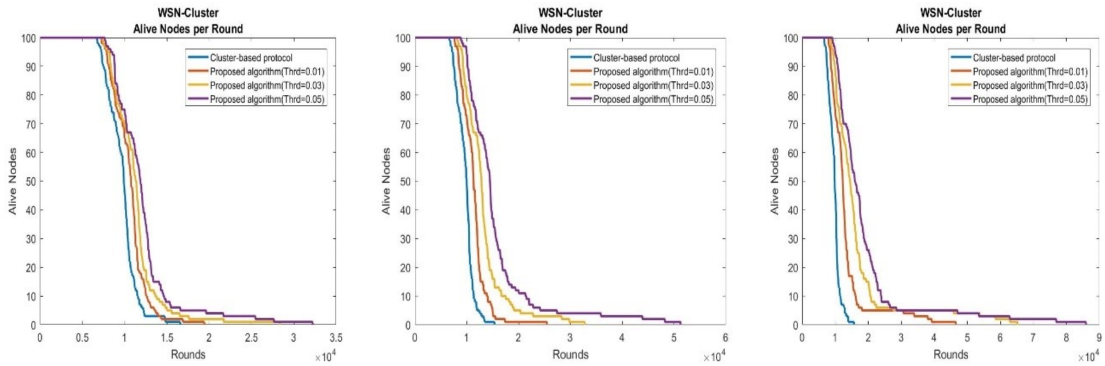

Figure 5. 10 *Alive Nodes Per Rounds*

5.4 Conclusions

In this chapter, the developed HDRA algorithm is explained considering the data reduction and routing processes as well. The overall goal of the HDRA algorithm is to reduce the amount of energy that the sensor node must consume. Both the proposed HDRA algorithm and the cluster-based protocol were evaluated by using an artificial dataset with a variety of situations. From the results, it's clear that the maximum number of alive nodes was observed when applied proposed algorithm for Scenario 1,2 ,3 and 4 with sensor type set to T, H and TH, which means that the average number of transmissions for all scenarios reached up to 17550,21250,13700 for T, H and TH respectively.

Conversely, the number of alive nodes occurred by a cluster-based protocol without considering the sensor node updating data status and average number of transmission packets for all scenarios reached up to 10300. According to the results, the proposed algorithm improves the network lifetime where the average of the maximum number of rounds that network nodes still alive for all scenarios was 48550 rounds compared to 10750 rounds for cluster-based protocol. In short, the proposed algorithm saved energy and improved the performance of the network from the point of view of the network lifetime.

The algorithm analysed in this paper is also presented in a paper:

Hussein, M K; Marghescu, Ion: *Hybrid Data Reduction Algorithm,* International Journal of Advanced Computer Science and Applications (**submitted)**.

Chapter 6

Conclusions and Future Works

6.1 Conclusions

WSNs are currently a hot study topic due to difficult concerns like as power consumption, routing algorithms, efficiency, and so on. Despite the open issues in WSN, there are currently a large number of apps available. In any event, one of the keys aims for the design of any application is to preserve WSN as long as feasible in terms of life cycle and usefulness. Sensor nodes can transport data directly (single hop) to the base station in the majority of the work of the latest data reduction technologies (BS). Another issue is that most existing routing protocols (multi-hop) do not take data reduction into account before forwarding data to the BS.

As a result, in this thesis, we designed and studied the performance of numerous data reduction techniques for cluster-based WSN with the goal of lowering energy consumption and increasing longevity. As a result, the created algorithm takes into account both the data reduction and routing operations. The suggested algorithm's overall purpose is to lower the amount of energy that the sensor node needs consume. The new algorithm was compared to the original LEACH cluster-based protocol. This is due to the fact that the LEACH protocol is a popular protocol that is regularly used in cluster based WSNs and has shown acceptable results. Previous research on data reduction and routing algorithms have been analysed in order to meet the goal of this thesis. Using a range of real-time datasets, we created and examined the influence of data reduction approaches on WSN performance. NNF, NNTS, LRMV, EDCD2, and FICA are data reduction methods. Following that, we evaluated the effect of various data packet sizes on the performance of a cluster-based WSN. Finally, we dubbed HDRA a new hybrid data reduction approach for cluster-based WSN. The HDRA algorithm's overall purpose is to lower the amount of energy that the sensor node needs consume. Using an artificial dataset with a range of circumstances, both the suggested HDRA method and the cluster-based protocol were assessed. The thesis structure has been separated into six chapters in order to present the proposed data reduction algorithms.

In Chapter one gives a brief introduction in this thesis. The overall definition of the WSN and some of its key features have been covered in this chapter. The research topic and goals have also been addressed.

In Chapter 2 an overview of WSNs was given. This chapter also discussed about previous work that focused on data reduction and routing algorithms to save energy on the sensor nodes. Also, concludes the limitations with the studies that have been done so far and what more research could be done.

In Chapter three I have developed and investigated different data reduction methods for WSN. These methods are data reduction based EDCD2, FICA, NNF, NNTS, and LRMV. The selected datasets are Data1-AirQ, Data2-ARHO, Data3-GSB, and Data4_Intel. From these results, the average data reduction percentage for applied EDCD2, FICA, NNF, NNTS, and LRMV algorithms through a real-time dataset named Data1-AirQ is 33%, 33%, 67%, 67%, and 33%, respectively. The NNF and NNTS algorithms achieve the greatest data reduction. According to Figure 3.10, both algorithms, NNF and NNTS, have adequate accuracy, with EDCD2 exhibiting the lowest error. Similarly, the average data reduction percentage for NNF, NNTS, EDCD2, LRMV, and FICA algorithms used to use a real-time dataset termed Data2-ARHO is 67%, 67%, 56%, 33%, and 67%, respectively. Although the NNF and NNTS algorithms obtained the best data reduction ratio, they also have the biggest error and the worst accuracy performance, as shown in Figure 3.11. The average data reduction percentage for NNF, NNTS, EDCD2, LRMV, and FICA algorithms applied through a real-time dataset named Data3-GSB is 67%, 67%, 67%, 67%, 33%, and 33%, respectively. The NNF, NNTS, and EDCD2 algorithms have the maximum data reduction, as shown in Figure 3.12. The EDCD2 algorithm has the lowest error. FICA had the weakest accuracy performance, with the highest mistakes. The average data reduction percentage for NNF, NNTS, EDCD2, LRMV, and FICA algorithms applied through a real-time dataset named Data4 Intel is 50%, 50%, 83 percent, 25%, and 75%, respectively. It should be noted that the EDCD2 algorithm produces the greatest data reduction. According to Figure 3.13 and Tables 3.21 and 3.22, both NNF, NNTS, EDCD2, and LRMV algorithms have reasonable accuracy, with FICA exhibiting the largest inaccuracy.

In Chapter four I have investigated the impact of different data packet sizes on the performance of cluster-based WSN. According to the findings of this study, the size of the sent data packet influences the life cycle of cluster-based WSN. It is advised that routing protocols be integrated with data reduction strategies to obtain high performance.

In Chapter five I have presented a new hybrid data reduction algorithm for cluster-based WSN developed during my PhD stage. The proposed HDRA algorithm is explained in this chapter, with HDRA considering the data reduction and routing procedures as well. The HDRA algorithm's overall purpose is to lower the amount of energy that the sensor node needs consume. Using an artificial dataset with a range of circumstances, both the suggested HDRA method and the cluster-based protocol were assessed. According to the results, the greatest number of alive nodes observed when applying the proposed algorithm for Scenarios 1,2,3, and 4 with sensor type set to T, H, and TH, which implies the average number of transmissions for all scenarios reached up to 17550,21250,13700 for T, H, and TH, respectively. In contrast, a cluster-based protocol's minimal number of living nodes happened without taking into account the sensor node updating data status, and the average number of transmission packets for all situations reached up to 10300. The suggested algorithm improves network longevity, with the average of the greatest number of rounds that network nodes remained alive for all situations being 48550 rounds, compared to 10750 rounds for the cluster-based protocol. In a nutshell, the proposed technique enhanced network life and saved energy.

6.2 Original contributions

The major contributions of this thesis are summarized as follows:

- Provided a critical evaluation of current data reduction methods and routing protocols for WSN energy efficiency. The outcome of this contribution has been concluded the limitations in current studies and further research opportunities as well. ***(I have published the outcome in two papers in the proceedings of IEEE Conferences*** [7] [70]***)*** .
- Investigated the impact of different data packet sizes on the performance of WSN-based cluster. In the application of WSN, sensor nodes rely on batteries with a limited lifespan to function. In addition, the data packet size of sensor nodes affects battery life. This is because energy efficiency is affected by the size of data transmissions from the node to the base station (BS). Therefore, this contribution evaluates the impact of various packet sizes on the performance of WSN-based clusters. The results of this study concluded that the size of the transmitted data packet affects the life cycle of WSN-based clusters. To achieve high performance, it is recommended to integrate routing protocols with data reduction methods. ***(I have published the outcome as a paper in the proceedings of an IEEE Conference*** [71]***)*** .
- I have analysed the performance of different approaches to data reduction for WSN. These approaches are data reduction based EDCD2, FICA, NNF, NNTS, and LRMV. The selected datasets are Data1-AirQ, Data2-ARHO, Data3-GSB, and Data4_Intel. (See Chapter 3). ***(I have published the outcome as a paper in International Journal of Advanced Computer Science and Applications a WOS indexed journal*** [72]***)*** .
- I developed a new hybrid data reduction algorithm for WSN-based clusters. The developed HDRA algorithm considering the data reduction and routing processes as well. The overall goal of the HDRA algorithm is to reduce the

amount of energy that the sensor node must consume. The developed algorithm improved the performance of the network life and saved energy. ***(I have submitted the outcome as a paper in International Journal of Advanced Computer Science and Applications a WOS indexed journal).***

6.3 Future Works

From this work, several lines of research can be taken. In this section, suggestions for future works are made, such as:

- Develop Information-centric networking-based adaptive duty-cycle for energy-efficient wireless sensor networks.
- Develop a new data reduction method for WSN-based clusters that considers the mobility of the nodes.
- Develop a new technique to extending network lifespan by considering compressing gathered data at the cluster head level before transferring it to the BS.
- Examine the performance of the suggested algorithm for the most common chain-based routing protocol known as PEGASIS.

6.4 List of publications

A. Papers published during the PhD stage

1. M. K. Hussein, "*Data Reduction Algorithms for Wireless Sensor Networks Applications: Review*," 2021 3rd International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA), 2021, pp. 1-7, doi: 10.1109/HORA52670.2021.9461309,

2. Hussein, M K; Marghescu, Ion; Nayef. A.M. Alduais, *Performance of Data Reduction Algorithms for Wireless Sensor Network (WSN) using Different Real-Time Datasets: Analysis Study*, International Journal of Advanced Computer Science and Applica tions;West Yorkshire, Vol. 13, Iss. 1, (2022). DOI:10.14569/IJACSA.2022.0130178, WOS:000754703200001
3. M. K. Hussein, "*Impact of various data packet sizes on the performance of WSN-based clusters: Study*," 2021 5th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT), 2021, pp. 424-428, doi: 10.1109/ISMSIT52890.2021.9604692.

4. M. I. Husni, M. K. Hussein, N. A. M. Alduais, J. Abdullah and I. Marghescu, "*Performance of Various Algorithms to Reduce the Number of Transmitted Packets by Sensor Nodes in Wireless Sensor Network*," 2019 11th International Conference on Electronics, Computers and Artificial Intelligence (ECAI), 2019, pp. 1-7, doi: 10.1109/ECAI46879.2019.9042081, WOS:000569985400109

5. H. Mohammed, R. M. KHALEEFAH, M. k. Hussein and I. Amjad Abdulateef, "*A Review Software Defined Networking for Internet of Things," 2020 International Congress* on Human-Computer Interaction, Optimization and Robotic Applications (HORA), 2020, pp. 1-8, doi: 10.1109/HORA49412.2020.9152862.

6. Hussein, M K; Marghescu, Ion: *Hybrid Data Reduction Algorithm: International Journal of Advanced Computer Science and Applica,* **submitted.**

B. Papers below the PhD stage with other related topics

1. J. Abdullah, M. K. Hussien, N. A. M. Alduais, M. I. Husni and A. Jamil, "*Data Reduction Algorithms based on Computational Intelligence for Wireless Sensor Networks Applications*," 2019 IEEE 9th Symposium on Computer Applications & Industrial Electronics (ISCAIE), 2019, pp. 166-171, doi: 10.1109/ISCAIE.2019.8743665.

2. A. Hamid Mohammed, R. M. KHALEEFAH, M. k. Hussein and A. Hatem AlMarzoogee, "*The Method of Calibration Compensation for Fiber Nonlinearity- A Review*," 2020 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA), 2020, pp. 1-8, doi: 10.1109/HORA49412.2020.9152914.

Bibliography

[1] I. F. Akyildiz, T. Melodia, and K. R. Chowdury, *"Wireless multimedia sensor networks: A survey,"* IEEE Wirel. Commun., vol. 14, no. 6, pp. 32–39, 2007.
[2] Z. Fei, B. Li, S. Yang, C. Xing, H. Chen, and L. Hanzo, *"A survey of multi-objective optimization in wireless sensor networks: Metrics, algorithms, and open problems,"* IEEE Commun. Surv. Tutorials, vol. 19, no. 1, pp. 550–586, 2016.
[3] L. Yong-Min, W. Shu-Ci, and N. Xiao-Hong, *"The architecture and characteristics of wireless sensor network,"* in 2009 International Conference on Computer Technology and Development, 2009, vol. 1, pp. 561–565.
[4] C. Buratti, A. Conti, D. Dardari, and R. Verdone, *"An overview on wireless sensor networks technology and evolution,"* Sensors, vol. 9, no. 9, pp. 6869–6896, 2009.
[5] D. Bhattacharyya, T. Kim, and S. Pal, *"A comparative study of wireless sensor networks and their routing protocols,"* Sensors, vol. 10, no. 12, pp. 10506–10523, 2010.
[6] N. A. M. ALDUAIS and A. A. J. JIWA ABDULLAH, *"RDCM: An Efficient Real-Time Data Collection Model for IoT / WSN Edge With Multivariate Sensors,"* IEEE Access, vol. 7, pp. 89063–89082, 2019.
[7] A. Hamid Mohammed, R. M. KHALEEFAH, M. k. Hussein and A. Hatem AlMarzoogee, *"The Method of Calibration Compensation for Fiber Nonlinearity- A Review,"* 2020 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA), 2020, pp. 1-8, doi: 10.1109/HORA49412.2020.9152914.
[8] A. Camacho Martínez, "Implementation and testing of LOADng: *A routing protocol for WSN,"* 2012 - upcommons.upc.edu, visited on January -2019.
[9] Y. El Khamlichi, A. Tahiri, A. Abtoy, I. Medina-Bulo, and F. Palomo-Lozano, *"A hybrid algorithm for optimal wireless sensor network deployment with the minimum number of sensor nodes,"* Algorithms, vol. 10, no. 3, 2017.
[10] S. Ozdemir and Y. Xiao, *"Secure Data Aggregation in Wireless Sensor Networks: A comprehensive overview,"* Computer Networks, vol. 53, no. 12, pp. 2022–2037, 2009.
[11] E. Troubleyn, I. Moerman and P. Demeester, *"QoS Challenges in Wireless Sensor Networked Robotics",* Springer Wireless Personal Communications, vol. 70, no. 3, June 2013, pp. 1059-1075.
[13] T. Rault, A. Bouabdallah, and Y. Challal, "*Energy efficiency in wireless sensor networks: A top-down survey,*" Comput. Networks, vol. 67, pp. 104–122, 2014.
[14] N. A. M. Alduais, J. Abdullah, A. Jamil, and H. Heidari, *"Performance Evaluation of Real-Time Multivariate Data Reduction Models for Adaptive-Threshold in Wireless Sensor Networks,"* IEEE Sensors Lett., vol. 1, no. 6, pp. 1–4, 2017.
[15] N. A. M. Alduais, J. Abdullah, and A. Jamil, "*APRS: adaptive real-time payload data reduction scheme for IoT/WSN sensor board with multivariate sensors*," Int. J. Sens. Networks, vol. 28, no. 4, pp. 211–229, 2018.
[16] R. N. Enam and R. Qureshi, "*An adaptive data aggregation technique for dynamic cluster based wireless sensor networks*," 2014 23rd International Conference on Computer Communication and Networks (ICCCN), 2014.

[17] N. A. M. Alduais, J. Abdullah, and A. Jamil, "*Enhanced payload data reduction approach for Cluster Head (CH) nodes*," Telkomnika (Telecommunication Comput. Electron. Control., vol. 15, no. 3, pp. 1477–1484, 2017.
[24] Q. Ye, Y. Liu, and L. Zhang, "*An extended DISCUS scheme for distributed source coding in wireless sensor networks*," in 2006 International Conference on Wireless Communications, Networking and Mobile Computing, 2006, pp. 1–4.
[25] M. I. Husni, M. K. Hussein, N. A. M. Alduais, J. Abdullah and I. Marghescu, *"Performance of Various Algorithms to Reduce the Number of Transmitted Packets by Sensor Nodes in Wireless Sensor Network,"* 2019 11th International Conference on Electronics, Computers and Artificial Intelligence (ECAI), 2019, pp. 1-7, doi: 10.1109/ECAI46879.2019.9042081.
[27] J. Kammerl and E. Steinbach, *"Deadband-based Offline-Coding of Haptic Media,"* MM'08 - Proc. 2008 ACM Int. Conf. Multimedia, with co-located Symp. Work., pp. 549–557, 2008.
[28] B. Huang, W. Liu, T. Wang, X. Li, H. Song, and A. Liu, *"Deployment Optimization of Data Centers in Vehicular Networks," IEEE Access*, vol. 7, pp. 20644–20663, 2019.
[29] N. Chervyakov, M. Babenko, A. Tchernykh, N. Kucherov, V. Miranda-López, and J. M. Cortés-Mendoza, *"AR-RRNS: Configurable reliable distributed data storage systems for Internet of Things to ensure security," Futur. Gener. Comput. Syst.*, vol. 92, pp. 1080–1092, 2019.
[30] J. Azar, A. Makhoul, M. Barhamgi, and R. Couturier, *"An energy efficient IoT data compression approach for edge machine learning,"* Futur. Gener. Comput. Syst., vol. 96, pp. 168–175, 2019.
[31] G. Huo, C. Yu-cheng, L. Wang, Y. Zhao, and Y. Jin, *"Design of a High Accuracy and Real-Time Indoor Positioning System Based on Coding Point Identification and its FPGA Implementation,"* pp. 21–24, 2020.
[32] L. Yin, C. Liu, X. Lu, J. Chen, and C. Liu, *"Efficient compression algorithm with limited resource for continuous surveillance,"* KSII Trans. Internet Inf. Syst., vol. 10, no. 11, pp. 5476–5496, 2016.
[33] A. González-Vidal, F. Jiménez, and A. F. Gómez-Skarmeta, *"A methodology for energy multivariate time series forecasting in smart buildings based on feature selection,"* Energy Build., vol. 196, pp. 71–82, 2019.
[37] G. M. Dias, B. Bellalta, and S. Oechsner, *"The impact of dual prediction schemes on the reduction of the number of transmissions in sensor networks,"* Comput. Commun., vol. 112, pp. 58–72, 2017.
[54] N. A. M. Alduais, J. Abdullah, A. Jamil and L. Audah, *"An efficient data collection and dissemination for IOT based WSN,"* 2016 IEEE 7th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), 2016, pp. 1-6.
[57] S. K. Gupta, S. Vanjale, S. Rasal and M. Vanjale, *"Securing IoT Devices in Smart City Environments,"* 2020 International Conference on Emerging Smart Computing and Informatics (ESCI), 2020, pp. 119-123.
[58] L. Mesin, S. Aram and E. Pasero, *"A neural data-driven approach to increase Wireless Sensor Networks' lifetime,"* 2014 World Symposium on Computer Applications & Research (WSCAR), 2014, pp. 1-3.

[59] M. A. Rassam, A. Zainal, and M. A. Maarof, *"An adaptive and efficient dimension reduction model for multivariate wireless sensor networks applications,"* Applied Soft Computing, vol. 13, no. 4, pp. 1978–1996, 2013.

[60] L. Tan and M. Wu, *"Data Reduction in Wireless Sensor Networks: A Hierarchical LMS Prediction Approach,"* in IEEE Sensors Journal, vol. 16, no. 6, pp. 1708-1715, March15, 2016.

[61] C. Carvalho, D. G. Gomes, N. Agoulmine, and J. N. de Souza, *"Improving prediction accuracy for WSN data reduction by applying multivariate spatio-temporal correlation,"* Sensors, vol. 11, no. 11, pp. 10010–10037, 2011.

[70] M. K. Hussein, *"Data Reduction Algorithms for Wireless Sensor Networks Applications: Review,"* 2021 3rd International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA), 2021, pp. 1-7.

[71] M. K. Hussein, *"Impact of various data packet sizes on the performance of WSN-based clusters: Study,"* 2021 5th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT), 2021, pp. 424-428.

[72] M. K. Hussein, I. Marghescu, and N. A. M. Alduais, *"Performance of Data Reduction Algorithms for Wireless Sensor Network (WSN) using Different Real-Time Datasets: Analysis Study,"* Int. J. Adv. Comput. Sci. Appl., vol. 13, no. 1, pp. 649–661, 2022.