AN ENHANCEMENT OF AUTHENTICATION AND ENERGY EFFICIENT CLUSTERING PROTOCOL FOR WIRELESS SENSOR NETWORK

AHMED ABDULHADI JASIM

FACULTY OF COMPUTER SCIENCE AND INFORMATION TECHNOLOGY
UNIVERSITY OF MALAYA
KUALA LUMPUR

2021

AN ENHANCEMENT OF AUTHENTICATION AND ENERGY EFFICIENT CLUSTERING PROTOCOL FOR WIRELESS SENSOR NETWORK

AHMED ABDULHADI JASIM

THESIS SUBMITTED IN FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

FACULTY OF COMPUTER SCIENCE AND INFORMATION TECHNOLOGY
UNIVERSITY OF MALAYA
KUALA LUMPUR

2021

UNIVERSITY OF MALAYA

ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: Ahmed Abdulhadi Jasim

Matric No: WVA170004-17006809/1

Name of Degree: Doctor of Philosophy

Title of Thesis: An Enhancement of Authentication and Energy Efficient Clustering Protocol for Wireless Sensor Network

Field of Study: Security Wireless Sensor Network

I do solemnly and sincerely declare that:

(1) I am the sole author/writer of this Work;
(2) This Work is original;
(3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
(4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
(5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya ("UM"), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
(6) I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate's Signature Date:

Subscribed and solemnly declared before,

Witness's Signature Date:

Name:

Designation:

AN ENHANCEMENT OF AUTHENTICATION AND ENERGY EFFICIENT CLUSTERING PROTOCOL FOR WIRELESS SENSOR NETWORK

ABSTRACT

Clustering is one of the popular techniques for Wireless Sensor Networks (WSNs) topology management. In each cluster, a leader is selected referred to as cluster head while the rest of the nodes in a cluster are referred to as cluster members. Clustering has been proven to be the most efficient approach in WSN. However, to realize the full benefit of clustering proper authentication and energy efficiency are needed to provide control to the WSN resources and prolonged network lifetime. Two of the main issues in implementing WSN clustering protocol are security issues related to authentication and energy efficiency issues. Therefore, in the first part of this thesis, the focus is given on how to overcome authentication and energy issues due to the problem of the security key and key length sharing at the base station. This can be done by enhancing the authentication of the Media Access Control (MAC) address and by utilizing the distance information and timestamp to detect attacks and reduces energy consumption, using a protocol called Secure and Energy-Efficient Data Aggregation method in clustering based on access control model (SEEDA). The proposed SEEDA protocol enhances the MAC address by utilizing a secret key and random timestamp in the verification process. The base station nodes also utilize the distance and timestamp between nodes to avoid delay in the network. The performance of the proposed SEEDA protocol is compared with SDA, SDAT, SDALFA, EESSDA, SDAACA, and EESDA, which is a widely used clustering protocol in the area of authentication. The simulation results show that the proposed SEEDA protocol outperforms the existing scheme with a 98.84% malicious nodes detection rate, 3.04 joules for energy consumption, the maximum delay of 0.038 seconds, and the resilient time of 0.054, 0.075 seconds with 8%,16% of malicious nodes affecting the network. Apart from looking at the authentication and energy issues, this

thesis also focuses on another important energy efficiency issue, which is the hot spots problem. In the WSNs environment, the sensor nodes closer to the base station nodes will take on more forwarding tasks. This will result in a massive overhead of the sensor nodes, and these nodes will run out of power sooner than the others. It causes a breakdown of the nodes and a loss of communication between sensor nodes; this breakdown is called the hot spots problem. Therefore, in the second part of this thesis, the focus is given on how to reduce the hot spots problem and balance the energy consumption among nodes. This can be done by utilizing an unequal clustering in the WSNs is able to reduce the hot spots problem, using a protocol called Energy-Efficient Unequal Clustering protocol based on a Balanced energy method (EEUCB). The proposed EEUCB protocol utilizes an unequal clustering mechanism based on the competition radius, double cluster head selection that reduces the energy consumption of head nodes in the clusters are proposed, applied the sleep and awake mechanism, and the base station calculates the distance based on the closest and farthest nodes and divides this distance into four layers as opposed to depends only on the residual energy of sensor nodes and the distance from all the sensor nodes to the base station node for calculating the competition radius in the prior methods, and also to enhance the data transmission process between sensor nodes and cluster head nodes in the network. EEUCB is compared with UDCH, EEFUC, FLEACH, and LEACH by performing various simulations. The simulation results show that the proposed EEUCB clustering protocol has achieved 13.06%, 14.7%, 19.75%, and 57.75% lifetime improvements against LEACH, EEFUC, FLEACH, and UDCH, respectively.

Keywords: Wireless sensor network, Clustering protocol, Security, Authentication, Energy efficiency, Unequal clustering.

[PENINGKATAN PENGASLIAN DAN PROTOKOL PENGKLUSTERAN CEKAP TENAGA UNTUK RANGKAIAN SENSOR WAYARLES]

ABSTRAK

Penggugusan adalah salah satu teknik popular untuk pengurusan topologi Rangkaian Sensor Tanpa Wayar (WSN). Dalam setiap gugusan, pemimpin yang dipilih disebut sebagai ketua gugusan sementara sisa nod dalam gugusan disebut sebagai anggota gugusan. Penggugusan terbukti merupakan pendekatan yang paling cekap di WSN. Walau bagaimanapun, untuk merealisasikan manfaat penuh penggugusan pengesahan yang tepat dan tahap kecekapan tenaga yang diperlukan untuk mengawal sumber daya WSN dan jangka hayat rangkaian yang berpanjangan. Dua masalah utama dalam melaksanakan protokol penggugusan WSN adalah masalah keselamatan yang berkaitan dengan masalah pengesahan dan kecekapan tenaga. Pada bahagian pertama tesis ini, tumpuan diberikan kepada cara mengatasi masalah pengesahan dan tenaga yang disebabkan oleh masalah kunci keselamatan dan perkongsian panjang kunci di stesen pangkal. Ini dapat dilaksanakan dengan meningkatkan pengesahan alamat kawalan capaian media (MAC) dan dengan menggunakan maklumat jarak dan cap masa untuk mengesan serangan dan mengurangkan penggunaan tenaga, menggunakan protokol yang disebut kaedah Pengagregatan Data Selamat dan Cekap Tenaga (SEEDA) dalam penggugusan berdasarkan model kawalan capaian. Protokol SEEDA yang dicadangkan meningkatkan alamat MAC dengan menggunakan kunci rahsia dan cap masa rawak ketika proses pengesahan. Nod stesen pangkal juga menggunakan jarak dan cap masa antara nod untuk mengelakkan kelewatan dalam rangkaian. Prestasi protokol SEEDA yang dicadangkan dibandingkan dengan SDA, SDAT, SDALFA, EESSDA, SDAACA, dan EESDA, yang merupakan protokol penggugusan yang kerap digunakan untuk pengesahan. Hasil simulasi menunjukkan bahawa protokol SEEDA yang dicadangkan mengatasi skema sedia ada dengan kadar pengesanan nod hasad 98.84%, 3.04 joule untuk

penggunaan tenaga, kelewatan maksimum 0.038 saat, dan masa yang berdaya tahan 0.054, 0.075 saat dengan 8%, 16% nod hasad yang mempengaruhi rangkaian. Selain melihat isu pengesahan dan tenaga, tesis ini juga memfokuskan kepada satu lagi masalah kecekapan tenaga yang penting, iaitu masalah hot spot. Dalam persekitaran WSN, nod sensor yang lebih dekat dengan nod stesen pangkal akan mengambil lebih banyak tugas penghantaran. Ini akan menyebabkan overhed besar-besaran dari nod sensor, dan nod ini akan kehabisan kuasa lebih cepat berbanding yang lain. Ia menyebabkan kerosakan nod dan kehilangan komunikasi antara nod sensor; kerosakan ini dipanggil masalah hot spot. Oleh itu, pada bahagian kedua tesis ini, tumpuan diberikan kepada bagaimana mengurangkan masalah hot spot dan mengimbangkan penggunaan tenaga di antara nod. Ini dapat dilaksanakan dengan memanfaatkan penggugusan yang tidak seimbang di WSN yang mampu mengurangi masalah hot spot, dengan menggunakan protokol yang disebut protokol Penggugusan Tidak Seimbang Cekap Tenaga (Energy-Efficient Unequal Clustering) berdasarkan kaedah tenaga seimbang(EEUCB). Protokol EEUCB yang dicadangkan menggunakan mekanisme penggugusan yang tidak seimbang berdasarkan radius persaingan, pemilihan ketua gugusan berpasang yang mengurangkan penggunaan tenaga nod kepala dalam kelompok dicadangkan, menerapkan mekanisme tidur dan jaga, dan stesen pangkal mengira jarak berdasarkan nod terdekat dan nod paling jauh serta membahagikan jarak ini menjadi empat lapisan berbanding hanya bergantung pada sisa tenaga nod sensor dan jarak daripada semua nod sensor ke nod stesen pangkal untuk mengira radius persaingan dalam kaedah sebelumnya, dan juga untuk meningkatkan proses penghantaran data antara nod sensor dan gugusan ketua nod di rangkaian. EEUCB dibandingkan dengan UDCH, EEFUC, FLEACH, dan LEACH dengan melaksanakan pelbagai simulasi. Hasil simulasi menunjukkan bahawa protokol penggugusan EEUCB yang dicadangkan masing-masing mencapai peningkatan sepanjang hayat 13.06%, 14.7%, 19.75%, dan 57.75% terhadap LEACH, EEFUC, FLEACH, dan UDCH.

Kata kunci: Rangkaian sensor tanpa wayar, Protokol pengelompokan, Keselamatan, Pengesahan, Kecekapan tenaga, Penggabungan yang tidak sama.

ACKNOWLEDGMENTS

First, my sincere appreciation and gratitude goes to **ALLAH** for giving me the strength, motivation, and inspiration to fulfill my PhD journey. I would like to extend my sincere appreciation to my kind supervisor, Associate Prof. Dr. **Mohd Yamani Idna Bin Idris**, who has inspired me and guided me both academically and morally with an open mind during my study. His academic and moral advice was essential to the success of my three years of PhD Journey. It was a privilege to have worked under your professional supervision. My gratitude also goes to my second supervisor, Dr. **Saaidal Razalli Azzuhri**, for serving as part of my supervisory committee and providing me with a guide and ideas on various aspects of my work. My appreciation also extends to the Faculty of Computer Science and Information Technology staff for their assistance during my study years at the University of Malaya.

I would like to express my deep appreciation and gratitude to my great parents: My loved father **Abdulhadi Jasim**, and my beloved mother **Rafeeah Al-Hashemi**, who there are stood beside me and supported me with their tenderness and prayers for me throughout my studies. I would like also to extend my sincere appreciation and gratitude to my dear brothers and sisters. I am especially indebted and grateful to my best brother Specialist Medical Radiologist Dr. **Oday Abdulhadi Jasim**, for supporting me financially and morally.

Finally, to my darling wife, **Noor Riyadh Issa**, and my beautiful daughters, **Mirna Ahmed** and **Lana Ahmed**: my sincere deepest appreciation and gratitude, those who would encourage me when the times got rough and supported me and sacrificed everything to achieve my noble goal. My heartfelt thanks.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVIATIONS

A_p	:	Approval
AMRP	:	Average of the Minimum Reachability Power
B-IBCA	:	Balanced-Imbalanced Cluster Algorithm
BS	:	Base station
B_z	:	Block size
CDAP	:	Concealed Data Aggregation Using Privacy Homomorphism
Cer	:	Certificate of the sensor node
CH	:	Cluster Head
CM	:	Cluster Member
2CH	:	Secondary Cluster Head
DAT	:	Data Aggregation Tree
DCFDAS	:	Distributed Collision-Free Data Aggregation Scheme
DEEC	:	Distributed Energy-Efficient Clustering Scheme for Heterogeneous
DoS	:	Denial-of-service
D_P	:	Data packets
D_t	:	Delay time
d_L	:	Distance length
$d_{i,BS}$	:	Distance from cluster node to the base station
d_{max}	:	Maximum distance
d_{min}	:	Minimum distance
d_0	:	Is the threshold distance
EDUC	:	Energy-Driven Unequal Clustering
EECHS	:	Energy-Efficient Cluster Head Selection Scheme
EEFUC	:	Energy-Efficient Fuzzy Logic for Unequal Clustering

EESDA	:	Two Secure and Energy-Efficient Data Aggregation Schemes
EESSDA	:	Energy-Efficient Secure Highly Accurate and Scalable Scheme for Data Aggregation
EEUCB	:	Energy-Efficient Unequal Clustering Protocol Based on a Balanced Energy Method
ESMR	:	Energy-Aware and Secure Multi-Hop Routing
E_{RX}	:	Is the energy consumption of the receiver
E_{TX}	:	Is the energy consumption of the transfer
$E_{avg(i)}$	:	The average energy of neighbor nodes
E_{rem}	:	Residual energy
FBUC	:	Fuzzy Based Unequal Clustering
FCFS	:	Fastest Collision-Free Scheduling
FLEACH	:	Impact of The Secondary Cluster Aggregation Based on Location
G	:	Aggregation Nodes
HE	:	Homomorphic Encryption
HEED	:	Hybrid Energy-Efficient Distributed clustering
IoT	:	Internet of Things
k	:	Number of clusters in the network
LAN	:	Local Area Network
LEACH	:	Low-Energy Adaptive Clustering Hierarchy
M	:	Monitor Nodes
MAC	:	Medium Access Control
MAI	:	Malicious Nodes Identification
MCFL	:	Multi-Clustering Algorithm Based on Fuzzy Logic
MEMS	:	Micro-Electro-Mechanical Systems
N	:	Number of nodes in the network

NN	:	Number of a neighbor node
PEGASIS	:	Power-efficient gathering in sensor information systems
R	:	Relay Nodes
SDA	:	Synopsis Diffusion Approach
SDAACA	:	Secure Data Aggregation Using Access Control Protocol
SDALFA	:	Secure Data Aggregation Based on Iterative Filtering Scheme
SDAT	:	Secure Data Aggregation Homomorphic Encryption Technique
SEEDA	:	Secure and Energy-Efficient data aggregation Method
TEDMA	:	Time-Division Multiple Access
TTDFP	:	A two-tier distributed fuzzy logic-based protocol
T_p	:	Throughput
UCR	:	Unequal Clustering Based Routing
UDCH	:	Energy-Efficient Unequal Double Clustering Routing Protocol
WLAN	:	Wireless Local Area Network
WSNs	:	Wireless Sensor Networks
W_t	:	Computation time

CHAPTER 1: INTRODUCTION

1.1 Introduction

In its simplest form, a Wireless Sensor Network (WSN) can be defined as a network composed of small units known as sensor nodes, whose placements follow a spatial distribution. These nodes communicate with one another wirelessly to gather information about the region under observation. The gathered data are relayed through multiple hops to the sink node. The data are then locally used or further transmitted to other networks by the sink, such as the internet (through a gateway). Clustering protocols in WSNs is an effective way to save limited resources. The main objective of the clustering protocols is to maximize the network lifetime by occasionally adopting a balanced energy consumption approach and distributing the load among nodes and to minimize the overhead to reduce wasteful energy consumption – e.g., bandwidth and battery energy – of sensor nodes.

On the other hand, the data aggregation method has become an important element in clustering protocol. The data aggregation is essential for the network's lifetime because the size of data generated by the sensor network can be massive owing to the large number of nodes in the network. Therefore, there can be elimination of the redundant packets via data aggregation if these packets carry the information about the same event as compared with other packets that are transferred to the base station (Sanli, Ozdemir, & Cam, 2004). As mentioned above, the clustering protocols in WSNs have many advantages to maximize the network lifetime, but at the same time it also faces many challenges and constraints that must be investigated in depth before widespread commercial deployment is expected. Two of the main concerns encountered in implementing clustering protocol in WSNs are authentication and energy efficiency issues. Authentication is an important research issue in WSNs because adversaries can easily attack the sensor nodes deployed in the sensitive and open environment, and the seepage of aggregated data causes damage

to the networks. These data cannot be recovered in a short period of time. The energy efficiency issue on another hand is very important in WSNs due to the battery-powered sensor nodes having limited energy and complicated battery changing procedures. It is challenging to implement authentication while preserving the energy consumption in the network. The prior techniques proposed secure data aggregation in clustering to address the authentication and energy issues related to the security problems of WSNs. While most of them addressed security with the authentication issue in clustering protocols, there are still some who addressed the security without considering the authentication issue such as (Sofiene Ben Othman, Trad, Alzaid, & Youssef, 2013). These techniques suffer from authentication and have several limitations, such as sharing the security key and the key length with a base station node, and not much attention is given to enhancing the authentication of the Medium Access Control (MAC) address (Sofiene Ben Othman et al., 2013; Soufiene Ben Othman, Trad, Youssef, & Alzaid, 2013; Razaque & Rizvi, 2017; Rezvani, Ignjatovic, Bertino, & Jha, 2015; S. Roy, Conti, Setia, & Jajodia, 2014; Wang, Qin, & Liu, 2013). Hence, these limitations motivated us to propose and address the authentication and energy issues in the first part of this thesis, where the focus is given on how to overcome authentication and energy issues due to the security issues of the WSNs. A protocol called SEEDA is proposed in this thesis to overcome these limitations.

In addition, apart from looking at the authentication and energy issues in WSNs as mentioned earlier, this thesis also focuses on another important energy efficiency issue in clustering protocol, which is the hot spots problem that has been a major problem that can degrade the performance of WSNs. Prior approaches proposed unequal clustering and double cluster head techniques to reduce the hot spots problem and to preserve the energy consumption in the network. However, these approaches only concentrate on utilizing residual energy and the distance of sensor nodes to the base station, but not much attention is given to enhance the data transmission process. This would lead to an imbalance of

energy distribution among nodes in the network (Amodu & Mahmood, 2018; W. R. Heinzelman, Chandrakasan, & Balakrishnan, 2000; Phoemphon, So-In, Aimtongkham, & Nguyen, 2020; F. Zhu & Wei, 2019). Therefore, to overcome these limitations, in the second part of this thesis, the focus is given on how to reduce the hot spots problem and balance the energy consumption among nodes in the network. A protocol called EEUCB is proposed in this thesis to overcome these limitations.

The typical topology setup for clustering protocol in WSNs is shown in Figure 1.1, together with the malicious nodes that might appear to disrupt the network.

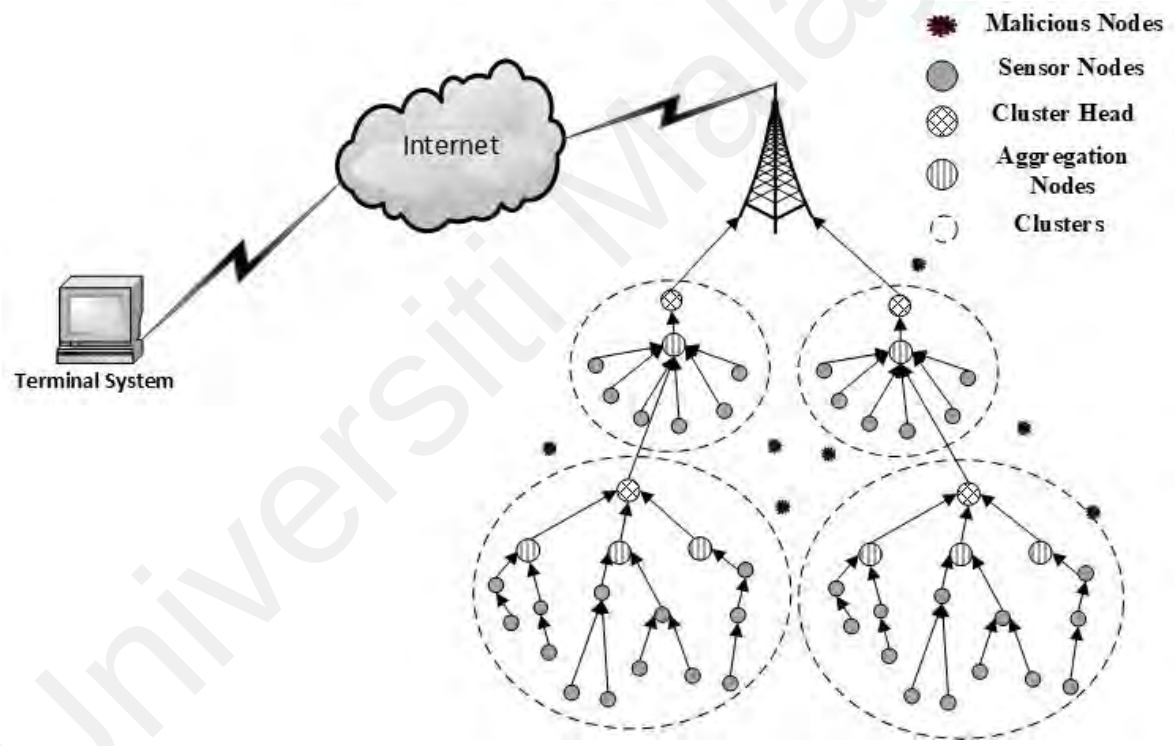


Figure 1. 1: WSNs Clustering Protocols Topology with Potential Disruptive Malicious Nodes

1.2 Research Motivation

The networks are going to carry sensitive information from sensor nodes as well as to be utilized in controlling numerous applications. Therefore, the security of WSN is a critical field in data transmission. For example, if the attacker can disturb or monitor the WSNs much more effortlessly than he may want to attack a similar wired system, there

will be little interest in the use of WSN in security-touchy applications. Therefore, the clustering protocols must work in combination with data communication security protocols, as any collision between these protocols would create loopholes in network security. To overcome this issue, the authentication and authorization process among nodes in the network is required to secure the original data from attacks and adversaries.

In addition, energy efficiency is also a problem that needs to be addressed because the battery-powered sensor node has limited energy and a complicated battery-changing procedure. The WSNs are usually used to monitor harsh and inaccessible environments, which reduces the use of infrastructure-based networks that may need constant human monitoring and interventions. Due to challenging circumstances and random sensor node deployment, however, replacing or recharging ‘dead’ sensor nodes' batteries are difficult. Therefore, these challenges will affect the quality, performance, and lifetime of WSNs. Furthermore, the location of sensor nodes in the network also affects energy efficiency. For example, when the sensor nodes are located at a distance from the base station, the sensor nodes will require more energy to receive data and forward it to the base station or server. Hence, this will decrease the lifetime of the network. In addition, sensor nodes that are closer to the base station nodes will take on more forwarding tasks. For these reasons, the hot spots problem is one of the major challenges of energy efficiency in WSNs. It should be borne in mind when considering solutions to this problem of energy consumption that data transmission of wireless communication consumes more energy than data processing.

1.3 Problem Statement

Previous researchers have proposed security schemes and techniques in clustering protocols for WSNs, particularly on authentication and energy efficiency issues; however, they lack an in-depth review of clustering protocols. WSNs are vulnerable to various

attacks because of their distributed wireless nature, which results in delays and loss of data in the network. Therefore, the security network is required to secure clustering protocols in the network.

Generally, aggregation nodes in clusters merge the data collected from their child nodes and forward the secure aggregated data to the base station node. Assuming the adversary nodes may be familiar with most of the security techniques in the WSN, they can reach the nodes by utilizing a wireless communication channel. The adversary may also exploit the process in an ad hoc network due to the unavailability of public-key cryptography techniques in a typical WSN. Therefore, a secure data aggregation in clustering protocol is necessary to have secure access control for successful data aggregation nodes. The data aggregation process with access control may improve the quality of service and reduce energy consumption. Hence, data aggregation requires secure access control to preserve data authenticity and integrity. Data aggregation should be obtained with high accuracy and low communication cost without compromising data privacy.

Authentication and authorization processes are possible mechanisms for detecting and preventing malicious attacks from accessing the network by checking the secure authentication of the new nodes. Authentication is the process of verifying the legitimacy of new nodes that join the network. This process is performed at the base station and aims to prevent the adversary nodes from joining the network and acting as original nodes to collect data from the network. Whereas authorization is the process that allows only authorized users to read and transmit the data. The implementation of both authentication and authorization processes in the network is essential because if malicious nodes manage to join the network, the authorization process can prevent these nodes from accessing the network's data.

Secure data aggregation methods in clustering protocol problems have been investigated in numerous studies from the literature. These problems range from sparse and small networks to large and dense ones with varying network applications and topologies (Sofiene Ben Othman et al., 2013; Soufiene Ben Othman et al., 2013; Razaque & Rizvi, 2017; Rezvani et al., 2015; S. Roy et al., 2014; Wang et al., 2013). The main focus of these works is to achieve network integrity and security via secure data aggregation nodes and prevent malicious attacks from accessing the network. However, the previous research that focuses on these issues has several limitations, such as sharing the security key and the key length with a base station node, and not much attention is given to enhancing the authentication of the Medium Access Control (MAC) address. This leads to the secure data aggregation of nodes in clustering protocol being exposed to malicious activities and unable to prevent attacks from accessing the network.

Based on the discussion, it can be seen that the implementation of authentication and authorization, together with the preservation of the network energy, poses a significant challenge in WSN. Therefore, it will be addressed in this thesis.

In addition, energy efficiency issues are very important to WSNs due to the battery-powered sensor nodes having limited energy and complicated battery changing procedures. In the original methods, the sensor nodes' initial idea was to collect the data and channel it directly to the base station node. When the process occurs, it leads to long-distance communication and results in higher energy consumption. Traffic and collisions will occur when sensors are transmitting data within the same period of time; this scenario will also result in data re-transmission and higher energy consumption. To avoid this issue and to increase network lifetime, a reconsideration of clustering is required. In a clustering mechanism, the sensor nodes are divided into several clusters. Each cluster contains a central cluster known as the cluster head and several nodes of the cluster called cluster

members. The cluster members receive data from an environment and send it to the cluster head. Once the cluster head receives the data, it will aggregate them in order to avoid data redundancy; the data will then be sent over to the base station via single-hop or multi-hop, as shown in Figure 1.2. One of the many advantages of the clustering protocol is balanced energy consumption, scalability, and an improved network lifetime. However, the clustering protocols have some problems. For example, during data transmission to the sink, sometimes there might be an increase in energy consumption in WSNs (Amodu & Mahmood, 2018) because the farthest distance CH will consume more energy to send data to BS. Multi-hop communication is typically used to save energy (Phoemphon et al., 2020). Furthermore, implementing security in WSNs can be challenging due to limited energy available as the energy is highly consumed during data transmission (Hsu, Leung, & Su, 2008; H. Hu, Chen, Ku, Su, & Chen, 2009). To extend the network lifetime and allocate the energy for implementing the security, the amount of transmission overhead should be reduced (Jariwala, Patel, Patel, & Jinwala, 2014). Therefore, efficient energy management of data aggregation must be considered in designing a secure network to protect it from attacks and prolong the network lifetime.

Aside from considering the authentication and energy issues, this thesis also focuses on another important energy efficiency issue, which is the hot spots problem. In this multi-hop communication process, the CH, which is closer to the BS, will do more forwarding tasks. This results in massive CH overhead, and these CHs run out of power sooner than the others. This leads to a breakdown of the cluster and a loss of communication between CHs. This breakdown is known as the hot spots problem. Furthermore, since the CH is responsible for the aggregating and transmission process, when it receives data from the CM, the CH will aggregate data and forward them to the BS. As the CH consumes more energy than CM, it leads to an unbalanced energy distribution in the overall network. If the CHs are not in close proximity, the furthest one from the BS node will dissipate more

energy and increase the overhead. To avoid this situation and save energy, this thesis proposes an unequal clustering protocol to reduce the hot spots problem, balance energy consumption throughout the network, and double cluster head selection that reduces the energy consumption of head nodes in the clusters are proposed, which is the second part of this thesis.

Unequal clustering and double cluster head methods have been proposed in numerous studies in the literature to balance the energy consumption among sensor nodes in the network (Amodu & Mahmood, 2018; W. R. Heinzelman et al., 2000; Phoemphon et al., 2020; F. Zhu & Wei, 2019). The main focus of these works is to solve the hot spots problem and to reduce the energy consumption of head nodes in the clusters. However, these methods only concentrate on utilizing residual energy and the distance of sensor nodes to the base station, and not much attention is given to enhancing the data transmission process. This would lead to an imbalance of energy distribution among nodes in the network. Thus, a clustering protocol is necessary for maximum energy conservation, feasible for the distance between nodes, distributing nodes, and data transmission among sensor nodes.

Hence, this study highlights two of the main issues in implementing WSNs clustering protocols, which are authentication and energy efficiency issues related to security problems and the hot spots problem which is related to energy efficiency issues as well. The main goal of this study is to prevent the adversary from accessing the network by checking the fake aggregated data and preserving the energy consumption of secure data aggregation, and at the same time, to balance energy consumption by reducing the hot spots problem. This goal will lead to an increased detection rate of malicious nodes and prolong the network lifetime. Therefore, the authentication and energy efficiency issues are addressed in this study.

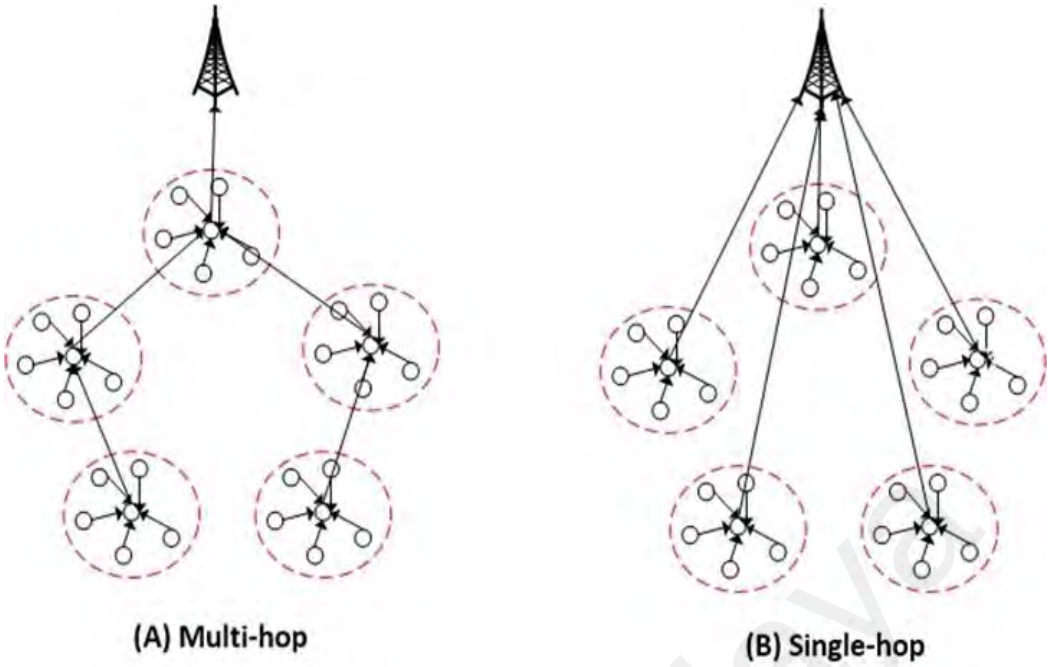


Figure 1. 2: The Structure of Clustering Protocol

1.4 Research Questions

This study was conducted to answer the following research questions:

Q1. What are the state-of-the-art clustering protocols in mitigating security and energy problems?

Q2. How to enhance an authentication to secure data aggregation and enable efficient energy usage of nodes in the cluster based on MAC address to prevent unauthorized access?

Q3. Is it possible to utilize minimum and maximum distance to balance the data transmission process for data aggregation?

Q4. Is it possible to utilize a double cluster head with a sleep-awake mechanism to reduce the energy consumption of head nodes in the clusters?

Q5. What are the evaluation metrics of clustering protocols for WSNs?

1.5 Research Aim and Objectives

The main purpose of this thesis is to develop a secure and energy-efficient clustering protocol for a wireless sensor network to increase the detection rate of the malicious node, reduce energy consumption in the network, and prolong the network lifetime. To accomplish this, the study aims to achieve the following specific objectives:

(i) To study existing, state-of-the-art clustering protocols for WSNs.

(ii) To enhance authentication for secure data aggregation and enabling efficient energy usage of nodes in the cluster based on MAC address.

(iii) To propose a clustering protocol with a double cluster head technique based on a balanced energy data transmission process for clustering that is able to reduce energy consumption and prolong network lifetime in WSNs.

(iv) To evaluate the proposed clustering protocols with different simulation scenarios and evaluation metrics.

According to the research aim and objectives mentioned above, Table 1.1 shows a summary of the problem statement, research questions, and objectives of data aggregation methods in clustering protocols for WSNs.

Table 1. 1: The Summary of Research Methodology

Problem Statement	Research Objectives	Research Questions
Lack of in-depth review of state-of-the-art clustering protocols, especially on security and energy issues.	To study existing, state-of-the-art clustering protocols for WSNs.	What are the state-of-the-art clustering protocols in mitigating security and energy problems?
Most of the existing security techniques in clustering protocols have several limitations, such as sharing of the security key and the key length with a base station node and not much attention is given to enhancing the	To enhance authentication for secure data aggregation and enabling efficient energy usage of nodes in the cluster based on MAC address.	How to enhance an authentication to secure data aggregation and enable efficient energy usage of nodes in the cluster based on MAC address to prevent unauthorized access?

authentication of the MAC address.		
Most of the existing clustering protocols for WSNs utilize residual energy and distance of sensor nodes to the base station, but not much attention is given to enhancing the data transmission process. This would lead to an imbalance of energy distribution among nodes in the network.	To propose a clustering protocol with a double cluster head technique based on a balanced energy data transmission process for clustering that is able to reduce energy consumption and prolong network lifetime in WSN.	Is it possible to utilize minimum and maximum distance to balance the data transmission process for data aggregation?
		Is it possible to utilize a double cluster head with a sleep-awake mechanism to reduce the energy consumption of head nodes in the clusters?
There is a need to identify a suitable evaluation technique to prove the performance of the proposed protocols.	To evaluate the proposed clustering protocols with different simulation scenarios and evaluation metrics.	What are the evaluation metrics of clustering protocols for WSNs?

1.6 Research Contributions

The novelty of this study is introducing new methods to address the security and energy issues in clustering protocols for WSNs, to increase the detection rate of malicious nodes, reduce energy consumption, and prolong the network lifetime. The following points summarizes the contributions of this work:

(i) To enhance the authentication of MAC address for secure data aggregation methods in the cluster, a method in clustering protocol called the Secure and Energy-Efficient Data Aggregation method in clustering based on an access control model (SEEDA) is proposed. The main considerations of the SEEDA protocol are the secure data aggregation of the sensor nodes and to preserve the energy consumption among sensor nodes in the cluster, which aims to enhance the authentication of MAC address by generating a random value and random timestamp with a secret key. Furthermore, the proposed SEEDA protocol detects and prevents malicious attacks such as Sybil and Sinkhole from joining and accessing the network. The base station nodes also utilize the

distance and timestamp between nodes to avoid delay in the network. The advantages of the SEEDA protocol are that it is able to increase the malicious node detection rate and reduce energy consumption based on an access control model by reducing redundant data transmission and communication overhead.

(ii) To reduce the hot spots problem, an unequal clustering protocol is also proposed in this study which is called the Energy-Efficient Unequal Clustering Protocol based on a balanced energy method (EEUCB). The EEUCB protocol aims to optimize energy usage in clustering based WSNs by adopting unequal clustering protocol to avoid the hot spots problem and to avoid long-distance data transmission among nodes. The EEUCB proposes to distribute the nodes depending on the divided network layers by calculating the farthest and closest node to the base station. Furthermore, the EEUCB protocol improves the secondary cluster head (2CH) selection to reduce the load and overhead on the primary cluster head by calculating the highest residual energy.

(iii) The sleep and awake mechanism among sensor nodes is proposed based on the distance from sensor nodes to CH and the energy level of sensor nodes to preserve the energy consumption and prolong the network lifetime.

(iv) A new cluster head rotation strategy and layers implementation scheme is proposed to balance the energy consumption between cluster members, cluster heads, and the base station nodes in the network. The strategy is based on the average distance threshold, average energy threshold, the layer implementation algorithm, and residual energy of clusters to construct the path to the base station node.

(v) A simulation model for clustering protocol in WSNs, which includes the conventional SDA, SDAT, SDALFA, EESSDA, SDAACA, EESDA, UDCH, EEFUC, FLEACH, LEACH techniques and schemes, and the proposed SEEDA and EEUCB

schemes in clustering protocols is performed. The performances of these clustering schemes are investigated, and the results are analyzed under various scenarios.

1.7 Thesis Organization

The structure of this thesis is organized as follows. **Chapter 1** gives a generalized background of the research study, presents the motivation for carrying out this work, and discusses the problem statement, followed by a definition of the research objectives. Noteworthy contributions of the research are summarized toward the end of the chapter.

Chapter 2 presents a background of sensor network technology and its evolution as well as briefly describes the unique features, challenges, and requirements of WSNs. Next, the data aggregation technique with its advantages and disadvantages is also presented. In addition, the types of attacks at different layers are also defined and classified. Furthermore, the authentication and energy efficiency issues related to the security problems are presented and classified with the existing techniques and schemes for secure data aggregation methods in clustering protocols. On the other hand, the energy efficiency in clustering protocols and the types of clustering algorithms are presented and described. The main concepts and representative techniques of energy issues in clustering protocols are discussed. Furthermore, a comparison between the different secure data aggregation and energy efficiency issues in clustering protocols is presented with their main advantages and limitations highlighted. Finally, the chapter briefly discusses the data correlation techniques and schemes in WSNs, and their relationship to this study is comprehensively reviewed.

Chapter 3 discusses the methodology used in this thesis and briefly discusses the proposed methodologies and protocols. The design, verification, and implementation of the proposed (SEEDA and EEUCB) protocols are presented. Finally, the research

methodology flowchart shows the sequence of the research phases and presented information on the connections between every phase's structural components.

Chapter 4 presents the proposed structure and functionalities of the SEEDA protocol. Moreover, a detailed description of the proposed SEEDA protocol is provided. The performance evaluation of SEEDA and the modeling and formulation followed in this thesis are introduced in this chapter. Finally, the simulation results of the proposed scheme performance evaluation are illustrated, discussed, and compared to those of existing schemes.

Chapter 5 describes the proposed EEUCB protocol along with its design and evaluation. The major operations involved in the EEUCB protocol are discussed in detail. Finally, the simulation results of the proposed scheme performance evaluation are illustrated, discussed, and compared to those of existing schemes.

Chapter 6 summarizes the main contributions of this work along with future research.

CHAPTER 2: LITERATURE REVIEW

2.1 Introduction

This chapter first presents the technological background of Wireless Sensor Networks (WSNs) and an overview of the paradigm, challenges, and requirements of WSNs in section 2.2. The concepts of the data aggregation techniques are crucial because the authentication and data aggregation functions are being executed inside the aggregation nodes in the cluster. For this reason, an introduction of the data aggregation techniques in clustering protocols is provided in section 2.3, including its types, essential components, advantages, and its challenges for WSNs. Section 2.4. presents the security requirements in clustering protocol for WSNs, while the authentication and energy efficiency issues related to security problems for secure data aggregation techniques in clustering protocol will be present in section 2.5, in a summary. In addition, section 2.6 presents another important energy efficiency issue in clustering protocols, which is the hot spots problem, with a summarization of the techniques and approaches of its. Finally, the discussion of this chapter will be presented in section 2.7.

2.2 Overview of Sensor Network Technology

During the last few years, sensor-enabled smart devices have utilized (WSNs), which are extremely important components in a smart city's infrastructure and the Internet of Things (IoT). Integration of the information world of the IoT and the physical world is made possible through WSNs clustering protocol (Čolaković & Hadžialić, 2018; Madakam, Lake, Lake, & Lake, 2015; D.-G. Zhang, 2012). Due to its low-cost implementation, WSNs are employed in various applications such as wildfire tracking, healthcare, disaster management, smart grid, military surveillance, homeland security, and monitoring (H. Li, Li, Qu, & Stojmenovic, 2014; Razaque & Rizvi, 2017). In such smart environments, WSN is considered a key technology in providing various IoT applications and services to users (Rawat, Singh, Chaouchi, & Bonnin, 2014).

In Australia, at least thirty-three people have been killed, and more than eleven million hectares of parks, forests, and bush burned in January 2020. This incident occurred because of the temperature rise and months of severe drought ("Australia fires," 2020). In addition, in Minnesota, the unexpected collapse of a highway bridge into the fast-flowing Mississippi River caused the deaths of nine people on August 2, 2007. According to the Minnesota National Transportation Safety Board, this incident occurred because of excessive bridge load combined with bad weather (Baranauckas, 2007). If sufficient information is available about the temperature, weather, load conditions, and areas where smoke plumes gather, we can take the necessary measures in time to reduce the damage.

A WSN consists of a huge number of static or mobile sensor nodes that form the wireless network with the usage of self-organization and multi-hop methods. Its purpose is to collaborate detection, processing, and transmit the object monitoring information in areas in which the network coverage of WSN comprises many nodes and sub-nodes. These nodes collect the information from their surrounding environment and send it to the base station or server. Many applications have been used in WSNs, such as wildfire tracking, healthcare, disaster management, smart grid, military surveillance, homeland security, and monitoring (Abdollahzadeh & Navimipour, 2016; Castillo-Effer, Quintela, Moreno, Jordan, & Westhoff, 2004; Jasim et al., 2019; Rawat et al., 2014; G. Sharma, Bala, & Verma, 2012). Three elements constitute sensor networks: the sensor node, the user node, and the sink node. The sensor node is the essence of the whole network; these sensor nodes' responsibility is to data processing, storing data, and transmitting. Each node consists of battery power, memory storage, and processor modules that collectively help in sensing (Kocakulak & Butun, 2017; Shafiq, Ashraf, Ullah, & Tahira, 2020; S. Zhang & Zhang, 2012). The sensor node can sense lots of environmental information, including pressure, mechanical pressure strength, temperature and humidity, vehicle movement, airflow speed, and other different characteristics. (H. Li et al., 2014). Micro-Electro-

Mechanical Systems (MEMS) have developed smart sensor nodes. These sensors are small, with limited processing and computation resources. The benefit of these smart sensor nodes is that they are inexpensive compared with traditional sensors (Barsocchi et al., 2020; Kocakulak & Butun, 2017). Sensor nodes execute three primary tasks: (i) physical quantity sampling for specific surrounding conditions, (ii) processing and storing sensed data, and (iii) transferring sensed data from the data collection point to the sink node or the base station (BS). The radios utilize to communicate between the sensor nodes and the BS to be exchanged with applications for fulfilling the desired tasks. Moreover, the communication between the sensor nodes and the BS allows for sharing information via different networks, such as LAN, WLAN, WPAN, and the Internet, with other computers.

Compared to conventional networks, WSNs offer several benefits, including low cost, easy deployment, flexibility, accurateness, and scalability. These advantages enabled the diverse usage of WSNs in various applications. However, the unique features of WSNs cause technical issues while processing data, conducting communication, and managing sensors. These issues pose serious challenges related to energy consumption, network control and detection, bandwidth utilization, and data exchange (Dehkordi et al., 2020; A. Kumar, Dadheech, & Chaudhary, 2020; Mahdi, Abdul Wahab, Idna Idris, et al., 2016). These networks also face many challenges and constraints that must be investigated in-depth before widespread commercial deployment can be expected. The most significant constraints that can affect the design of WSNs are security, energy efficiency, scalability, and data aggregation of nodes in cluster-based of WSNs. To analyze these challenges in-depth, the next section highlights the challenges and requirements of WSNs, which include security, energy consumption, node deployment, fault tolerance, implementation cost, and data aggregation.

2.2.1 Challenges and Requirements of WSNs

The previous section introduced the overview of WSNs and the advantages of the WSNs. This section introduces the challenges and requirements of WSNs, such as security, energy consumption, node deployment, fault tolerance, scalability, implementation cost, coverage, and data aggregation. WSNs have great potential, but this potential has not reached been reached because of different challenges in designing and deploying them. Researchers are, therefore, interested in the challenges of WSNs. The challenges of WSNs can be described as follow:

1. **Security**: Security is one of the important issues in WSNs. Security operates on many applications such as monitoring battlefields, time-critical applications, structural monitoring, and surveillance applications (Bala, Bhatia, Kumawat, & Jaglan, 2018). Sensor nodes in the network are faced with security threats because WSNs are vulnerable to various attacks due to their distributed wireless nature, which results in delays and loss of data in the network (Alsaedi, Hashim, Sali, & Rokhani, 2017). WSNs have high data sensitivity, thereby allowing the adversary to discreetly intercept information from the nodes (P. Kumar, Gurtov, Iinatti, Sain, & Ha, 2016). For example, the adversary can intercept the transmitted packets by disconnecting the link between the source and destination nodes, generating a fake node with a similar identity to the authentic node, or changing the transmission path. Therefore, network security in WSN is crucial to preserve the integrity of the network (Dabhade & Alvi, 2021; Grover & Sharma, 2016; Shim, 2015). All the sensor nodes in the network must be safe from unauthorized access to data in WSNs. There are some essential security requirements in order to protect against unauthorized access, such as data authentication, data confidentiality, data integrity, availability, and redundancy. These requirements will be described

in greater detail in section 2.4. Security also matters in the design of the hardware of sensor nodes in real life. The communication from one sensor node to another sensor node is very costly compared to instruction computations (Hari & Singh, 2016). The constraints, or we should we say the obstacles, can be divided into the following categories, these are:

A. **Limited Resource:** The security in WSNs requires resources that are restricted for utilization, consisting of vitality to control the sensor, code space, and information memory. Currently, these resources are exceptionally restricted in a modest remote sensor node, which is generally non-rechargeable.

   (a) **Limited Memory and Storage Space**: A sensor node is a small electronic device with a memory and a central processing unit. The messages in WSN are small compared to the alternate systems. There is no understanding of segmentation in the various uses of WSN (Amutha, Sharma, & Nagar, 2020; Ould Amara, Beghdad, & Oussalah, 2013).

   (b) **Restriction Power Energy:** There are three stages of consumption in the sensor node: communication between sensor nodes, microprocessor computing, and sensor transducer. Therefore, it the practical to recharge or change thousands of sensors.

B. **Unreliable Communication:** The center of monitoring, usually located in the monitoring region, monitors the communications in the network. The data and information are monitored and carried by a third-party service such as a 3G/4G network and satellite telecommunication;

these data are measured by network throughput. However, the range of sensor node communication is limited (Hari & Singh, 2016; Y.-C. Lin & Cheung, 2020).

(a) **Unreliable Transfer:** In WSNs, many sensor nodes are connected to the network. The packets sometimes disconnect from the network; this leads to the packets being damaged or dropped due to channel errors, so the result is missing or lost data packets.

(b) **Conflicts:** If the packet is found in the middle of a transfer for a high-density sensor network, there will be a conflict, and the connection will be re-established. The transfer will fail, and the sensor network's security will be weak.

(c) **Latency:** It is difficult to coordinate amongst sensor nodes inside the network. The exit node in the network which is already covered by the broadcast message changes the state and is called the "bridge." The bridge will start to broadcast the message, and the new active sensor nodes will obtain the message. Synchronization problems might be significant to sensor network security (R. Kumar, Tripathi, & Agrawal, 2020).

2. **Energy Consumption:** Energy consumption occurs during data sensing, processing, and transmission. However, among these activities, data transmission is often the costliest action as far as energy utilization is concerned. It will also affect the quality, performance, and lifetime of WSNs. Due to these factors, controlling energy consumption is one of the major problems with WSNs. An important fact to note when attempting to reduce this problem is that data transmission of wireless communication consumes more energy compared to data processing (Haseeb, Ud Din, Almogren, & Islam,

2020; Rashid & Rehmani, 2016; Sarkar & Murugan, 2019; F. Zhu & Wei, 2019). This issue will be further detailed in sections 2.5 and 2.6.

3. **Node Deployment:** Deployment means the implementation of sensor nodes in a real-world scenario. It is also important to consider deployment in WSNs. Two node deployment techniques are utilized in WSNs; firstly, static deployment is decided by nodes' location according to the optimization technique. The location of nodes in this technique is static; it will not change place during the implementation. Secondly, dynamic deployment is where the nodes are distributed in the network randomly based on the application and demographic location of the area (Amutha et al., 2020; Sumathi & Velusamy, 2020).
4. **Fault Tolerance:** In the lifetime of WSNs, many nodes in WSNs become blocked or fail due to less power or physical damage, or environmental interface. The low power of nodes must not affect the overall task of the sensor network; this can be achieved and this problem avoided by rerouting the data with other sensor nodes which have more energy for transmitting data to the base station (Mohapatra & Rath, 2020; Panda & Khilar, 2015; A. Sharma & Sharma, 2016).
5. **Scalability:** The variety of sensor nodes deployed in the network or sensing location is probably in the order of hundreds or thousands or more. Any routing protocol must be capable of work with a massive number of sensor nodes and be sufficient to respond to any activities that occur in the network (Shukla & Tripathi, 2020).
6. **Coverage:** A coverage area is also an important design parameter in WSNs. Because each node in the network for WSNs has a particular view of the

environment, specific the sensor's view of the environment is limited in accuracy and range; a limited physical area can only cover the environment.

7. **Implementation Cost:** WSNs are made up of a large number of sensor nodes. WSNs might be required software or a hardware tool or storage capacity, which generates the implementation cost of each node in the network (Battat, Seba, & Kheddouci, 2014).
8. **Data Aggregation:** Data aggregation is also one of the most important techniques in WSNs; it can reduce the number of transmission packets and reduce redundant data. Due to the data aggregation process occur inside the aggregation nodes and the MAC address authentication generate by aggregation nodes, so, the next section will discuss data aggregation in the cluster in more detail and describe the advantages and disadvantages of this technique.

2.3 Data Aggregation and Challenges in WSNs

In a typical WSN, a large number of sensor nodes collect the information, data, and specific application from the environment and transfer it to the base station or the server where it is analyzed, processed, and used by the application. The general approach of data processing is to process the data collected by sensor nodes and forwarding it's to the base station (Choudhari & Rote, 2021; Kaur & Munjal, 2020; Xiaowu Liu et al., 2019; Rana & Dudhgoankar, 2017). The distributed data processing in a network usually is referred to as data aggregation. The data aggregation technique is to merge the network data if these data belong to the same phenomenon. The main goal of the data aggregation technique is to jettison redundant data transmission and prolong the network lifetime by reducing the resource consumption of sensor nodes such as battery power and bandwidth. As the network increases its lifespan, data collection techniques may degrade the quality of service critical to WSN, such as security, latency, data accuracy, and fault tolerance.

Therefore, the design of efficient data aggregation of nodes in clustering-based protocols is a challenging task because the designer protocol should be a trade-off between security, latency, fault-tolerance, energy efficiency, and data accuracy. To achieve the trade-off of these challenges, the data aggregation techniques are closely related to how packets are routed through the network. Therefore, the architecture of the sensor network plays an essential role in the performance of different aggregation models (Maraiya, Kant, & Gupta, 2011; Ozdemir & Xiao, 2009). Figure 2.1 shows two models, the first is the data collection model, and the second is the non-data aggregation model. Sensor nodes 1, 2, 3, 4, 5, 6 are ordinary nodes that collect data from the environment and report it to the next level of the nodes. The sensor node (7) performs the data collection process before transferring the data to the base station performs the sensing and gathering at the same time. In the aggregation model, after the sensor nodes collect the data from the environment, the data travels through the network, and a single packet is transmitted to the base station or sink. Sensor nodes also travel within the network in the second model, but all sensor nodes transmit data packets to the base station. Thus, we can see that the non-data aggregation model increases the transmission number of data packets and consumes the energy of the sensor nodes in the network.

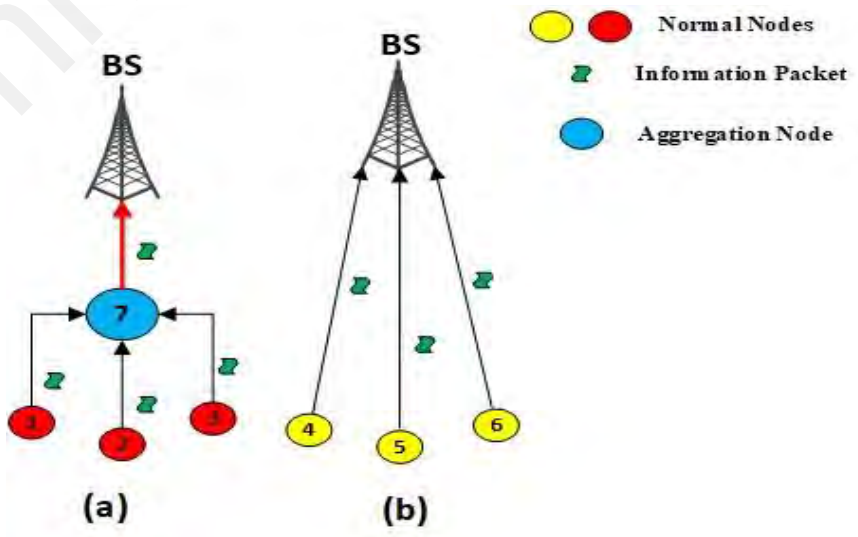


Figure 2. 1: (a) Data Aggregation (b) Non-Data Aggregation Model

2.3.1 Advantages of Data Aggregation in WSNs

The data aggregation in WSNs has several advantages that were defined by (Dehkordi et al., 2020; N. R. Roy & Chandra, 2020). The advantages of data aggregation as described as follow:

i. Data aggregation can help improve information accuracy and robustness obtained from the entire network.

ii. When collecting the data from the sensor nodes, certain redundancy will occur in the data. Therefore, the redundant data in the network needs to be reduced.

iii. Reduce the traffic load and preserve the energy of the sensor.

iv. Reduce the size and amount of data transmission.

v. The sensor nodes are able to integrate multiple aggregations.

2.3.2 Disadvantages of Data Aggregation in WSNs

The data aggregation process also has a disadvantage; the disadvantage of data aggregation was defined by (Mishra & Thakkar, 2012; N. R. Roy & Chandra, 2020).

i. The data aggregation collects the data from the sensor nodes and sends it to the base station or sink. However, if these sensors contain a compromised node exposed to a malicious attacker, the base station cannot guarantee the authenticity of the collected data received.

ii. Several copies of data results have been aggregated and then sent to the base station simultaneously; it will increase the energy consumption at these nodes.

iii. It does not apply to all measurement environments.

In the summary of this section, the data aggregation technique was presented along with its requirements in WSNs, also its advantages and disadvantages. The next section covers the security requirements in clustering protocol and the types of attacks in WSNs that will

be described. Furthermore, the different layers-based attacks in clustering techniques will also be presented.

2.4 Security Requirements in Clustering Protocol for WSNs

A WSN is a special type of network. It has some commonalities that it shares with a typical computer network and displays many of the unique characteristics. At WSNs, security services must protect data and information communicated over the network from node misbehavior and attacks. The sensor network should fulfill some requirements for providing successful security communication. The general requirements of security for WSNs are confidentiality, authentication, access control, integrity, and availability (Garg, Saini, & Gupta, 2020; Grover, Sharma, & Shikha, 2014; Grover & Sharma, 2016; Pardesi & Grover, 2015; V. P. Singh, Jain, & Singhai, 2010). The other requirements of security WSNs are called secondary requirements such as self-organization, data freshness, and time synchronization (Anwar, Bakhtiari, Zainal, Abdullah, & Qureshi, 2014). The essential security requirements in WSNs are described below:

(1) Data Confidentiality: The security mechanism should ensure that no one except the intended recipient can understand the message in the network when accessing it. The requirements of the confidentiality issues in WSNs should address the following: (a) the sensor nodes should not allow the neighbors to access and read the message except if they are authorized to do, (b) the distribution of keys mechanism must be quite strong, (c) the keys and public information such as sensor identity should also be encrypted to protect them from attacks (Ghosal & DasBit, 2015; R. Kumar et al., 2020; Moorthy, Bangera, Amrin, Avinash, & NS, 2020). Furthermore, the routing information should also be confidential because,

in some cases, malicious nodes can use this information to reduce network performance. Therefore, if keeping sensitive data secret, the data is encrypted with a secret key that only intended recipients have. Typically, a key-based mechanism for secure data aggregation is utilized to achieve data confidentiality. A key-based mechanism's responsibility is to guarantee data confidentiality by generating a secret key to protect information from adversaries (Abidin, Vadi, & Rana, 2021; Guo & Chen, 2011).

(2) Availability: Availability ensures the survivability of network services against Denial of Service (DoS) attacks. A DoS attack can be launched at any wireless sensor network layer and can permanently disable the victim nodes. Consequently, in DOS attacks, excessive communication or computation can drain the battery charge of the sensor node. The consequences of availability loss may be catastrophic. For example, if some sensor nodes' availability cannot be provided in a battlefield surveillance application, this might lead to an enemy invasion. WSNs are deployed with high node redundancy to tolerate such availability losses. Since data aggregators collect the data of several sensor nodes and send the aggregated data to the base station, data aggregators' availability is more important than normal sensor nodes. Thus, in WSNs, intruders launch DoS attacks to prevent data aggregators from performing their tasks so that some part of the network losses its availability (Bade & Garba, 2019; Bhushan & Sahoo, 2018).

(3) Data Integrity: Data integrity means that the sensor nodes in the network should ensure that the data is trustworthy or not corrupted, and the data should not be changed during the communication process. When

malicious nodes join the network, the data will be corrupted to prevent the network from normal operation. In fact, due to unreliable communication channels, the data might change without the presence of an attacker. Therefore, the message authentication code is used to prevent data integrity, and the data aggregation results will be changed. So, it is not possible to provide only data integrity for wireless communication, it should also be provided with data freshness to prevent a replay attack because the malicious nodes are able to listen to the transmitted data and replay them later. (Abood, Wang, Mahdi, Hamdi, & Abdullah, 2021; Ghormare & Sahare, 2015; A. V. Singh & Chattopadhyaya, 2015).

(4) Data Freshness: Data freshness means that when the data packets reach the destination, it ensures that the adversaries are not able to replay these packets (Ghosal, Halder, & DasBit, 2012; Qazi et al., 2021). Although some of the data freshness has the same meaning as data integrity, the difference between them can be found in some specific scenarios (Ozdemir & Xiao, 2009). For example, suppose the sensor nodes transmitted data packets to all of their neighbors by a single hop. In that case, the adversaries may send malicious nodes to perform a replay attack by replaying the data packets to corrupt the communication between the sender and receiver. The malicious nodes do not change the content of data packets, which does not break the law of integrity. Hence, the base station may receive duplicate packets, which leads to incorrect data aggregation results (Xiaowu Liu et al., 2019).

(5) Self-Organization: Each node in the network should be self-organizing because, in WSNs, no fixed infrastructure exists. Therefore, each node in the network is independent and has unique properties allowing them to

adapt to different situations. Hence, the self-organizing feature is a great challenge for security (Anwar, Bakhtiari, Zainal, Abdullah, & Qureshi, 2015; Grover & Sharma, 2016).

(6) Time synchronization: Many applications in WSNs require time synchronization. The security mechanism also requires synchronization among a group of sensor networks. The main objective of time synchronization is to equalize the local time for all nodes in a network. AS the WSNs have limited computation, the traditional time algorithms are not practical for synchronizing the network. The time synchronization method is defined by (Ahmad, Shiwei, Qi, Meixi, & Ling, 2016; Al-Shaikhi & Masoud, 2017; X. Sun et al., 2020).

(7) Authentication: Authentication is essential in WSNs because if the adversary joins the network, it can modify data packets and change the original data by injecting fabricated packets. Therefore, authentication, one of the major objectives, is proposed and introduced in this thesis. The authentication technique has been shown to resolve some issues in clustering approaches for WSNs. This issue of authentication will be further described and discussed in section 2.5 and chapter 4.

(8) Access Control: The access control technique is the main problem to consider when addressing the security data aggregation technique for WSNs. (Hari & Singh, 2016; Iqbal & Mir, 2020). It allows valid or legitimate users to access the data and denies invalid users access to the network. Since there is no standard infrastructure in WSNs, access control has solutions that are different from traditional solutions. The access control cannot provide a solution for every application because it (the

access control) has some challenges, as described by (Butun & Sankar, 2011; S.-K. Yang, Shiue, Su, Liu, & Liu, 2020).

2.4.1 Types of Attacks in Clustering Protocol for WSNs

In this sub-section, we introduce the type of attacks in clustering protocol for WSNs. The attacks are very important issues in secure data aggregation of nodes in the clustering-based protocols for WSNs because the most common attacks against WSNs occur in the information when transmitting the data between nodes. In the transmission process, an attacker can steal or modify the information by deploying some malicious nodes in the network that have a similar ID to normal nodes (Grover & Sharma, 2016). The type of attacks can be classified into two main categories: active and passive attacks. A passive attack means the attack can obtain the original data in the network without a break in the communication among nodes (Bade & Garba, 2019; Garg et al., 2020; Önen & Molva, 2007; Pawar & Agarwal, 2017; Rodhe & Rohner, 2008). While the active attack is clearly seen as an attack, it can attack the network by sending malicious nodes, thereby causing interrupted communication (F. Sun et al., 2014; Wu, Dreef, Sun, & Xiao, 2007; Xie, Yan, Yao, & Atiquzzaman, 2018; Y. Yang, Wang, Zhu, & Cao, 2008). The main types of attacks in the WSN clustering protocol are listed and analyzed as follows:

A. Denial of Service (DOS) Attacks

Denial of service (DOS) attacks is one of the most common attacks in WSNs. The main of DOS attacks is to disable any part of a WSNs from proper functioning or correctly on time in the network. The attackers use the previous types of attacks to prevent legitimate sensor nodes from using network resources (Gavric & Simic, 2018; Osanaiye, Alfa, & Hancke, 2018). On the other hand, in DOS attacks, some malicious nodes may refuse data transmission to high-level nodes during the sensor nodes' data transmission process. When designing a network without a security mechanism, two situations may

arise. First, high-level sensor nodes will maintain the reception requirements from the base station. If the lower-level nodes refuse to send any data to the higher-level nodes, they will consume more energy in the network. Second, high-level nodes will apply a "time-out" mechanism. Therefore, if it does not receive data within a certain period of time, it will only transmit data that has already been received. In this case, some information will be ignored. (Dharini, Balakrishnan, & Renold, 2015; KanagaSuba Raja & Pushpa, 2020; Pathan, Lee, & Hong, 2006). The DOS was considered by (Önen & Molva, 2007). In this technique, if the high-level sensor nodes do not receive data information from the child node at a certain time, the child node will be marked as a malicious node. Moreover, all the nodes will share a secret key between them and will ignore the shared binary key.

B. Sybil Attacks

The attacker disguises himself as a normal node or valid sensor node in the network. It can play more than one role during the data aggregation process to attack the network and steal the data without being detected (Aftab et al., 2015). Sybil attacks occur when a malicious node claims to have multiple identities either by creating new identities or impersonating the existing identities. For example, a malicious node may impersonate the identities of the neighboring nodes. This malicious node will repeat automatically and make several copies of itself to disrupt the network operation. Furthermore, the Sybil attacks can affect the network's data aggregation by claiming a fake ID. The malicious node can also steal an identity to enable them to join the network (Alsaedi et al., 2017; Raja & Beno, 2017; Santhi & Sowmiya, 2017). Figure 2.2 shows a Sybil attack in the network.

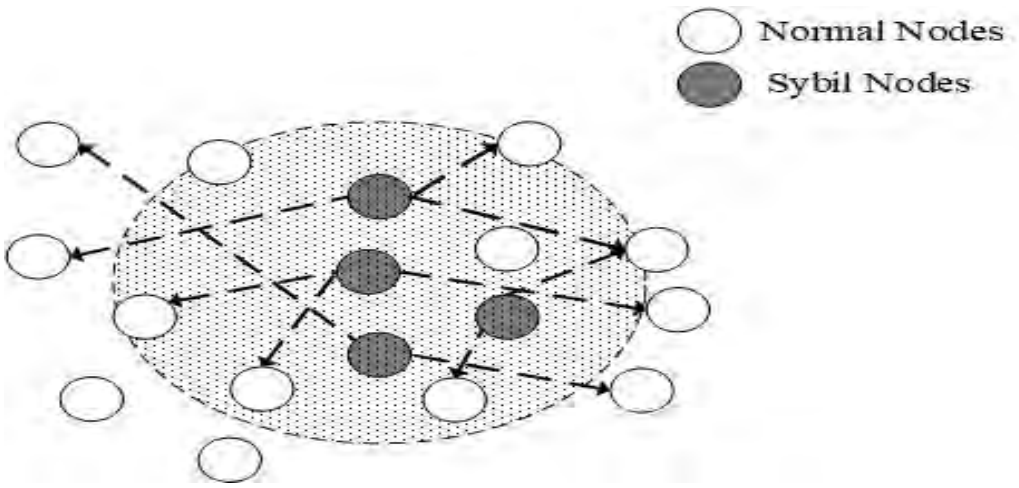


Figure 2. 2: Malicious node with Multi Identities

C. Stealth Attacks

The stealth attack is one of the most common attacks caused by the security problem in aggregating data across a network. This attack aims to inject fake data during the data aggregation process and will change the results and decisions of the base station. Most networks collect the data from the sensor nodes and make critical decisions based on specific rules. When the sensor nodes are injected from a stealthy attack, the decision may be reversed totally and change the original data in the network. (Guo & Chen, 2011).

D. Sinkhole Attacks

The sinkhole attacks affect the network layer by using a path or bandwidth among the nodes. The attack attracts the nearby distributed nodes. The adversary fakes the neighbor nodes. The sinkhole attacks are able to drop the data packets or forward them to another attack or tamper with aggregated data (Santhi & Sowmiya, 2017; Shafiei, Khonsari, Derakhshi, & Mousavi, 2014). In addition, the adversary's goal is to attract nearly all the traffic from a particular area through a malicious node. As shown in Figure 2.3, the sinkhole attacks typically work by making a compromised node look especially attractive to surrounding nodes. The malicious node has more power than the other nodes in the

network and connects to the base station node through a single hop. It is claimed and displayed to have the shortest possible path to the sink in order to attract more network traffic. Most routing algorithms choose the shortest path for data transmission.

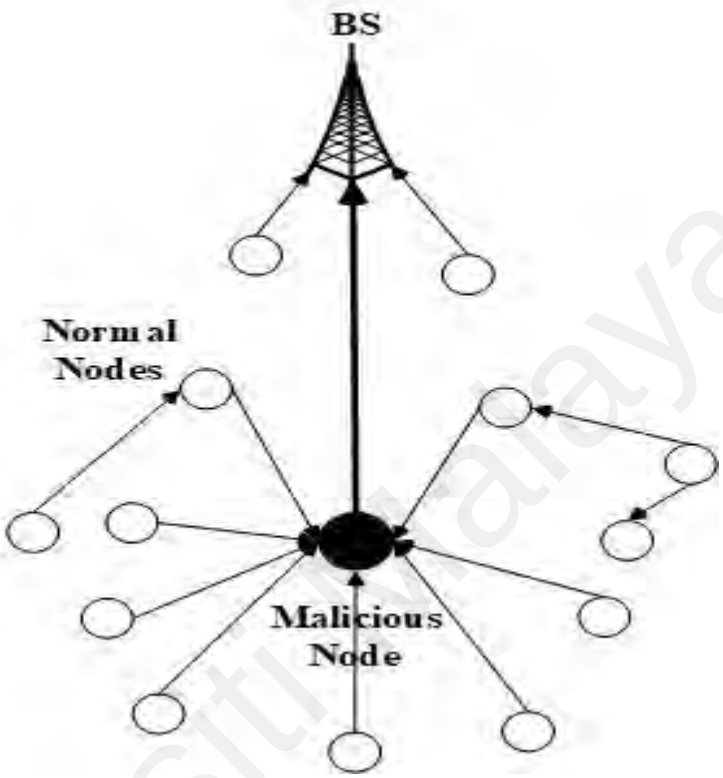


Figure 2. 3: Sinkhole Attacks

E. Replay Attacks

A replay attack is that the attacker repeatedly sends past sensor information, affecting the sensor node's freshness, and the base station cannot get the latest data from every node in the network. To prevent replay attacks, a specific time mark must be attached with each packet being sent to the following nodes. In (Kwon, Yu, Lee, Son, & Park, 2021; Y. Yang et al., 2008) used a random number to prevent replay attacks.

F. Wormhole Attacks

In wormhole attacks, the attacker can record the data packets at one location in the network, transfer to another location, and retransmit them into the network (Amish &

Vaghela, 2016; Salehi, Razzaque, Naraei, & Farrokhtala, 2013). Figure 2.4 shows that node A and node B are preserving the wormhole link in the network and are the two malicious nodes. There is a link between both malicious nodes known as a wormhole link. Node A sends a message to node (i). Node (i) sends a message to node (j) via a wormhole link, which forwards it to node B.

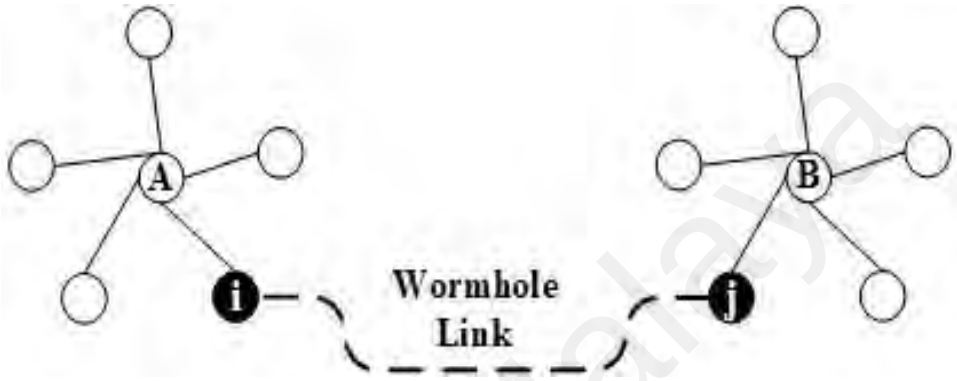


Figure 2. 4: Wormhole Attacks

G. HELLO Flood Attack

The attacker sends HELLO packets from one node to another node with high energy in routing protocols (Gill & Sachdeva, 2018; R. Kumar et al., 2020). The attacker uses HELLO packets to coax sensing nodes into WSNs. In this case, the opponent sends an attacker with a high wireless transmission range. Processing power sends HEELO packets to some isolated sensor nodes in a large area within the network. Hence, sensor nodes are affected such that the opponent is a neighbor. The normal node tries to pass through the attacker because it knows it is its neighbor, and finally, the normal node is spoofed by the attacker. Figure 2.5 shows that the attacker node broadcasts HELLO packets with a higher transmit packet from the base station node. The next sub-sequence from the sub-section discusses the attacks at different layers in clustering-based protocols for WSNs, such as application layers, transport layer, network layer, data link layer, and physical layer.

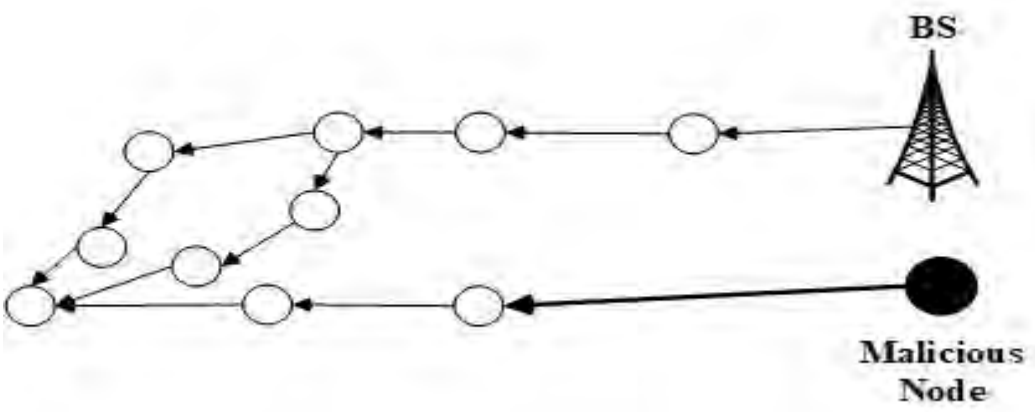


Figure 2. 5: HELLO Flood Attack

2.4.1.1 Different Layers Based Attacks in Clustering Protocol for WSNs

This sub-section introduces the different types of layers that can be attacked in WSNs, such as application layer, transport layer, network layer, data link layer, and physical layer. The different types of attacks in each layer are summarized in Table 2.1. Although the attacks exist in each layer, the network layer has the highest number of attacks. For this reason, the previous section described the attacks in the network layer. The five layers in WSNs clustering protocol can be described as follow:

- **Application Layer:** In the clustering-based protocols for WSN, the sensor nodes are deployed in remote environments, which leads adversaries to expose the sensor nodes and generate large traffic to the base station. Therefore, the application layer's responsibility is to provide the data requested for individual sensor nodes distributed remotely with the end-users (Gopika & Panjanathan, 2020; Raymond & Midkiff, 2008).
- **Transport Layer:** The transport layer is responsible for the end-to-end managing connection between sensor nodes to send and receive packets and data encryption in the network. This layer utilizes a simple technique to reduce the communication overhead; therefore, the transport layer is often vulnerable

to attacks. The two primary attacks on this layer are flooding and desynchronization (Kaushal & Sahni, 2015).

- **Network Layer:** The responsibility of the network layer is to forward data packets to the next hop via data link and transport layers. This layer also provides effective routing data between two nodes, nodes to cluster head nodes, or nodes to the base station in the network (Gunduz, Arslan, & Demirci, 2015).
- **Data Link Layer:** The fourth layer is the data link layer in WSNs. This layer is responsible for multiplexing data streams, data frame detection, medium access control, and error control. The data link layer is also responsible for point-to-point or point-to-multipoint connections in the network (Islam, Fahmin, Hossain, & Atiquzzaman, 2020).
- **Physical Layer:** The fifth layer is the physical layer in the WSNs is responsible for carrying out functions like frequency selection, signal detection, carrier frequency generation, and data encryption. The physical layer is vulnerable to various attacks due to the broadcast nature of wireless communication and the functions delicate in WSNs. Furthermore, the sensor nodes in WSNs are often deployed in hostile or insecure environments so that the attacker can have physical access to the network (Osanaiye et al., 2018).

Table 2. 1: Summary of Attacks at Different Layers in WSNs

Types of Layers	Types of attacks
Application Layer	Path-based, Overwhelming sensors, and Deluge attacks.
Transport Layer	Flooding and Desynchronization attacks.
Network Layer	Black holes, Hello Flood, Sinkholes, Sybil, Information, and Selective forwarding, Wormhole attacks.
Data Link Layer	Jamming, Collision, Exhaustion attacks.
Physical Layer	Collision and Tempering attacks.

As mentioned above, the previous section presented the security requirements and the types of attacks in clustering protocol for WSNs, where the examples of how the attacks affect the networks are also presented. Furthermore, the types of layers in the network and the effect of attacks on it were presented. The next section describes the authentication and energy efficiency issues related to security problems for secure data aggregation techniques in clustering protocol. The methods of encryption and decryption based on cryptography techniques along with their advantages and disadvantages will also be introduce in the next section. Finally, the contributions and limitations of security data aggregation techniques based on cryptography techniques can also be found in the next section.

2.5 Authentication and Energy Efficiency Issues for Secure Data Aggregation Techniques in Clustering Protocol

This section describes the authentication and energy efficiency issues related to security problems for securing data aggregation techniques in the clustering protocol of WSNs. Reliable security is essential in sensor networks as the distributed nature of sensor nodes makes them vulnerable to various attacks. Sensor nodes are mainly powered by batteries, and frequent replacement of batteries for a large number of nodes is impractical. Therefore, algorithms or security techniques should be highly efficient in terms of energy consumption. Another limitation of the security WSNs is preserving and delivering quality data to another wireless device without any interference from an adversary. Therefore, this section will present the authentication and energy efficiency together due to their overlapping in security issues of WSNs. To increase the detection rate of malicious nodes, the energy consumption of nodes in the network should be reduce, and vice versa.

Authentication is essential in WSNs because if the adversary joins the network, it can modify data packets and change the original data by injecting fabricated packets. Therefore, the receiver needs to have the mechanism to check and verify that received packets have indeed come from the actual sensor node (Grover & Sharma, 2016). If there is any communication between two nodes or more, then the authentication technique can be carried out through a Media Access Control (MAC) address (Cui, Shao, Zhong, Xu, & Liu, 2018; Yadav & Mishra, 2020). An authentication technique is a significant issue in clustering approaches. Therefore, authentication, one of the major objectives, is proposed and introduced in this thesis. The authentication technique has been shown to resolve some issues in security clustering protocols for WSNs.

In addition, the energy efficiency issue is also very important for WSNs due to the battery-powered sensor nodes having limited energy and complicated battery changing procedures. It is challenging to implement authentication while preserving the energy consumption in the network. Furthermore, security implementation in WSNs is crucial, in order to preserve the integrity of the network. However, implementing security in WSNs can be challenging due to limited energy available as the energy is highly consumed during data transmission. To extend the network lifetime and allocate the energy for implementing the security, the amount of transmission overhead should be reduced. Therefore, efficient energy management of data aggregation must be considered in designing a secure network to protect it from attacks and prolong the network lifetime.

In previous techniques, secure data aggregation and approaches were discussed to address the authentication and energy efficiency issues in security networks with their advantages and limitations at the end of this section. Furthermore, the summarization of their techniques is presented in Table 2.4.

Besides, the cryptography technique on the other hand is also important in security networks because this technique is able to protect the data by using encryption and decryption methods. This brings to the next subsection that introduces the cryptographic techniques, the types, and how they affect the network.

2.5.1 Cryptography Techniques

Cryptography is the basic technique for encryption and decryption of the data to protect the original data when being transferred or stored over the network. The encryption function is where data are transformed from plaintext into ciphertext, while the decryption function is where the data are transformed from ciphertext into plaintext. The secure data aggregation approaches based on cryptographic techniques are further classified based on which cryptographic technique is used. Two types of encryption and decryption methods exist, namely symmetric and asymmetric key encryption in cryptographic techniques. The former uses only one public key for data encoding and decoding, whereas the latter uses two different keys: one for encoding and the other for decoding. There other techniques for secure data aggregation are called non-cryptographic techniques. Non-cryptographic techniques can achieve security data without employing encryption and decryption functions (Beg, Al-Kharobi, & Al-Nasser, 2019; Nithya, 2020; Rani & Kaur, 2017). The symmetric and asymmetric key cryptography can be described as follow:

2.5.1.1 Symmetric Key Cryptography

The symmetric key cryptography utilizes one key for both encoding and decoding functions in WSNs, as shown in Figure 2.6. The symmetric key does not consume more energy, computation overhead, and memory; hence, symmetric-key cryptography is often preferred to asymmetric key cryptography. There are two types of symmetric key cryptography, namely keystream and block cipher. The keystream cipher mostly utilizes

one-bit of plaintext data in order to produce a one-bit ciphertext. This one-bit plaintext is obtained by XOR plaintext bits with a random sequence of bits called keystream. In contrast, the block cipher encrypts the information by breaking the plaintext into blocks and encrypting them to produce ciphertext in each block. In WSNs, the large sequence encrypting of plaintext consumes more energy and multiply block size; therefore, the keystream is better than block cipher (Gandara, Wang, & Utama, 2018; Ilayaraja, Shankar, & Devika, 2017; X. Zhang, Heys, & Li, 2010).

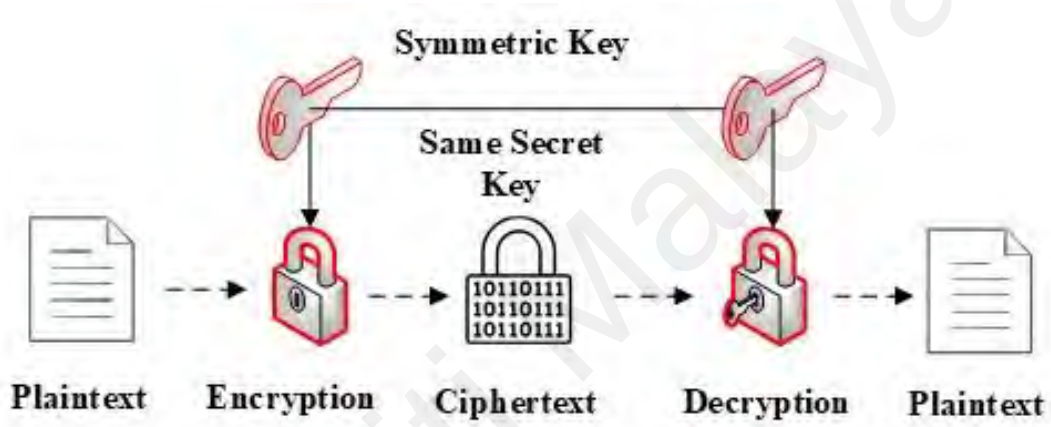


Figure 2. 6: Symmetric Key Cryptography

Several researchers have proposed symmetric key cryptography methods such as (Burhanuddin et al., 2018; Çam, Özdemir, Nair, Muthuavinashiappan, & Sanli, 2006; L. Hu & Evans, 2003; Huang, Shieh, & Tygar, 2010; M. Kumar, Verma, & Lata, 2015; Wu et al., 2007). The techniques proposed symmetric key cryptography with unencrypted data aggregation. The aggregated data was not at the immediate next hop, but it is forwarded over the first hop and aggregated in the second hop. The data integrity, authentication, and data confidentiality in these techniques were not addressed, which leads the network to be vulnerable to attacks. In addition, they are increasing the redundancy of data aggregation. To overcome these problems, a message authentication code (MAC) was introduced with symmetric encryption for secure data aggregation (Kurmi, Verma, & Soni, 2017; X. Li, Chen, Li, & Wang, 2015). These approaches

successfully secure the authentication of all the nodes in the network because the attacker cannot guess the MAC address of the author's message when the data is sent through the network. Hybrid cryptography for secure data aggregation was proposed by (Prakash & Rajput, 2018). This method is proposed to achieve data integrity and confidentiality for secure data aggregation. Furthermore, this method developed a hybrid algorithm for data encryption and decryption functions and obtained data integrity and confidentiality. However, data authentication was not addressed. On the other hand, symmetric-key cryptography is utilized to encrypt and reduce energy consumption in WSN (X. Zhang et al., 2010). This method utilizes both keystream cipher and block cipher to calculate the energy cost.

2.5.1.2 Asymmetric Key Cryptography

Asymmetric key cryptography utilizes a pair of keys: a public key and a private key. The public key is announced to the public, whereas the receiver keeps the private key. The sender uses the receiver's public key for encryption, and the receiver uses his private key for decryption, as shown in Figure 2.7. Even though the public key was broadcasted to every node, the hacker is not able to retrieve the private key with only the help of the public key (Bisht & Singh, 2015; SenthilKumar & Senthilkumaran, 2016; P. Singh & Chauhan, 2017). Many researchers have proposed asymmetric key cryptography methods such as (Boudia, Senouci, & Feham, 2015; Mykletun, Girao, & Westhoff, 2006; Ozdemir, 2007; Rafik & Mohammed, 2013). The following techniques are proposed Privacy Homomorphism (PH) based on homomorphic encryption techniques in order to obtain end-to-end data confidentiality and data aggregation. During network deployment, the aggregation node shares keys in pairs with its neighbor nodes, and each neighbor node can encrypt the data and send it to the aggregation node.

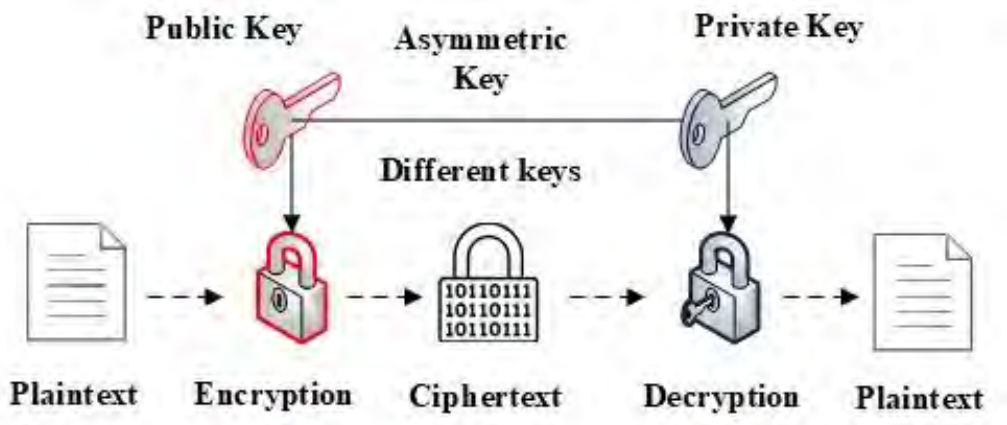


Figure 2. 7: Asymmetric Key Cryptography

The aggregator node decrypts all the data, aggregates the data, and encrypts data using a privacy homomorphic encryption algorithm. If the node does not have its own private key by the base station, the advantage is that these approaches will not decrypt data. An efficient and secure recoverable data aggregation scheme is proposed by (Zhong, Shao, Cui, & Xu, 2018). This technique focused on Homomorphic Encryption (HE) to solve the limited aggregation function and unauthorized aggregation. The advantage of this technique is that the base station can recover the original sensing data. Thus, the aggregation functions are not tied or restricted by the aggregation functions, and false data is filtered. However, this authentication process was not sufficient to cover all the nodes in the network.

As mentioned earlier, secure data aggregation based on cryptography techniques is a common issue in WSNs. Therefore, in this thesis, we need to compare the advantages and disadvantages of cryptography techniques, such as symmetric and asymmetric key cryptography, used by various researchers in WSNs. Table 2.2 compares symmetric and asymmetric cryptography. While Table 2.3 presents a summary of the techniques and approaches based on cryptography techniques in WSNs. The secure data aggregation techniques and schemes in the clustering protocol for WSNs will be further discussed in the following sub-sections.

Table 2. 2: Comparison Between Symmetric and Asymmetric Cryptography

Types of Cryptography	Advantages	Disadvantages
Symmetric	Utilize a single key for both encryption and decryption operations.	An individual communication link needs a secret key.
	Symmetric encryption is faster speed and efficient in implementation.	Due to the dynamic structure and the self-organization of nodes, key management is difficult.
	Smaller key size.	
	The communication resource minimum consumed.	
Asymmetric	Able to resolve the key distribution issue.	Longer keys require.
	Utilizes different keys, one key for encrypting data and the second one for decrypting data.	Require high power and bandwidth.
	It is often used for securely exchanging secret keys.	Low efficiency with small networks.

Table 2. 3: Summary of Techniques Based on Cryptography in WSNs.

REF	Description	Contributions	Limitations
(L. Hu & Evans, 2003)	Design the technique to reduce energy consumption and secure the data in the network.	The base station can directly send a broadcast message to all sensor nodes. In addition, the network is spread out enough, so there are many hops among the sensor nodes and the base station.	Does not provide data confidentiality.
(Çam et al., 2006)	This is proposed in order to avoid redundant data transmission between sensor nodes and cluster head nodes.	Introduced the sleep and active mode to decrease the number of active nodes in the network. In addition, using a symmetric key in order to secure data	The authentication technique is not addressed. Also, data privacy is not maintained in this scheme.

		between sensor nodes and cluster head nodes.	
(Ozdemir, 2007)	In CDAP, proposed in order to achieve end-to-end data confidentiality and data aggregation.	Improved the data aggregation, bandwidth, energy consumption.	High computational overhead in this scheme. Authentication and integrity are not addressed.
(Albath & Madria, 2009)	Utilize digital signature algorithm to achieve integrity.	Integrity for data information and aggregation process are provided.	The computational overhead in this scheme was high.
(Y. Zhang, Zhu, & Feng, 2009)	Introduced the mobile agent to exchange the private key among the sensor nodes in the network.	Introduced medium access control MAC address, and authentication to secure the message to all nodes in the network.	The communication overhead and energy consumption were high in this scheme.
(X. Zhang et al., 2010)	Proposed symmetric cryptography algorithm to calculate the energy cost in WSNs.	Reduced energy consumption.	The authentication and integrity techniques were not addressed.
(Huang et al., 2010)	In this technique, it was proposed to achieve security and privacy during the transmission of data in the network.	Reduced the communication overhead and used random keys to encrypt data.	Confidentiality and integrity are not addressed in this scheme.
(Ozdemir & Xiao, 2011)	Integrity protecting hierarchical concealed data aggregation is proposed to encrypt data in WSNs.	Achieved data integrity and confidentiality.	The MAC address in this technique was not strong.
(Parmar & Jinwala, 2014)	Proposed security data aggregation to achieve security and privacy in WSNs.	Provided end-to-end privacy data aggregation.	Energy consumption in this method was not addressed.
(Shim & Park, 2014)	In shim, the proposed secure data aggregation based on cryptographic primitives in heterogeneous clustered to reduce the total length of ciphertexts and to achieve end-to-end confidentiality.	Addressed hop by hop authentication.	In this method, the energy consumption and latency increased.

M. Kumar, Verma, & Lata, 2015	Proposed homomorphic encryption in order to secure data aggregation based on mobile agents.	Reduced energy consumption.	The data aggregation lacks redundancy. Also, integrity and confidentiality were not addressed.
X. Li, Chen, Li, & Wang, 2015)	Introduced fully homomorphic encryption in order to detect false data aggregation.	Reduced energy consumption.	Not much attention is given to enhancing the MAC address.
(X. Li et al., 2015)	Proposed fully homomorphic encryption to preserve the confidentiality data aggregation.	Provided false data detection and secure data aggregation.	Addressed confidentiality and integrity, but authentication was not addressed.
(Karthikeyan, Velumani, Kumar, & Inabathini, 2015)	Proposed secure hierarchical data aggregation algorithm to achieve security and reduce energy consumption.	Reduced computational overhead on the sensor nodes.	The types of attacks in this scheme were not considered.
(Zhong et al., 2018)	Focused on homomorphic encryption (HE) to solve the limited aggregation function and unauthorized aggregation.	The sink node can retrieve the original sensing data and the false data filtered.	The authentication does not cover all the nodes in the network.
Prakash & Rajput, 2018)	Proposed hybrid cryptography for secure data aggregation.	Addressed Data integrity and data confidentiality.	Data authentication was not introduced.
(Okay & Ozdemir, 2018)	Proposed technique in order to achieve end-to-end confidentiality based on smart grid data aggregation.	Reduced the amount of data stored in cloud servers.	Not has given much attention to preventing the attacks to join the network.
(Arora & Hussain, 2018)	The proposed technique in order to provide security is based on sharing a key among nodes.	Utilizes the same key for encryption and decryption methods.	Data confidentiality, integrity, privacy was not addressed.
(Tripathy, Pradhan, Tripathy, & Nayak, 2019)	Proposed a hybrid cryptosystem for security based on a public-key algorithm.	Takes less memory.	Data authentication was not introduced.

2.5.2 Secure Data Aggregation Homomorphic Encryption Technique (SDAT)

In (Soufiene Ben Othman et al., 2013), the authors proposed that SDAT uses the homomorphic encryption algorithm and message authentication codes (MAC) in order to achieve data confidentiality, integrity, and authentication for secure data aggregation in WSNs. This technique utilized MAC to generate a secret key to protect the message in

the network. The advantages of this technique have achieved integrity and authentication among nodes. However, the detection of a malicious attack was not considered. It focused only on encryption messages in the network, thereby making the network highly vulnerable, which will lead to an increase in energy consumption.

2.5.3 Synopsis Diffusion Approach (SDA)

In (S. Roy et al., 2014), the authors proposed the SDA technique that used the synopsis diffusion approach in order to secure data aggregation. The base station node calculates the aggregates such as count or sum despite the falsified sub-aggregate attack. The algorithm of this protocol consists of two phases. In the first phase, the base station estimates the count of aggregate data based on minimum authentication received from sensor nodes. In the second phase, the base station demands more authentication from the sub-nodes, while this sub-node is determined by the estimate of the first phase. Finally, the base station calculates the true value by filtering out the false, malicious nodes from the aggregates. The advantages of this algorithm reduced the communication overhead. However, the base station received the authentication from sub-nodes only in the network; thereby, the adversary can join the network by compromising the sensor nodes' authentication. Furthermore, trust between the nodes and the server may not be available.

2.5.4 Energy-Efficient Secure Highly Accurate and Scalable Scheme for Data Aggregation (EESSDA)

In (Wang et al., 2013), the authors proposed the EESSDA technique which was used to establish a secure channel between sensor nodes and their neighbors. Two sensor nodes share a common random number for transmitting data without encryption and decryption processes in the network. In this technique, the child node decomposes the data into slices and sends it to the aggregation node—the aggregation node receiving data and sends one message to BS to reduce traffic control. The contribution of this protocol provided data

confidentiality by decomposing the data into slices before the transmission process, moderate energy efficiency because the protocol was not using cryptographic techniques, and accuracy of the aggregation because data packets have less chance of collision. However, the EESSDA protocol did not address authentication among nodes, and the network is vulnerable to attacks because it does not provide security during the data aggregation process, which leads to reduced security in the network and increases communication overheads.

2.5.5 Two Secure and Energy-Efficient Data Aggregation Schemes (EESDA)

In (Sofiene Ben Othman et al., 2013), the authors proposed the EESDA technique to detect malicious nodes in the network. This technique aims to ensure that the base station does not accept any forged data aggregation. The two schemes are proposed to provide secure message integrity with the trade-offs between security and computation cost. The first scheme is a concrete homomorphic MAC proposed to achieve integrity of data aggregation. The second scheme proposed additive digital signatures and homomorphic encryption to achieve confidentiality and integrity in the network. However, authentication between nodes was not provided; hence, the security of all nodes in the network was not ensured and allowed attacks to join the network. In addition, this protocol increased energy consumption and delays in the network.

2.5.6 Secure Data Aggregation Based on Iterative Filtering Scheme (SDALFA)

In (Rezvani, Ignjatovic, Bertino, & Jha, 2014), the authors proposed SDALFA, which used an iterative filtering algorithm. This scheme aims to protect against sophisticated collision attacks and improves the iterative filtering algorithm by initially approximating the trustworthiness of sensor nodes, making the algorithms not only robust but also converge more accurately and faster. The benefits of this protocol reduced the number of iterations required to approach a stationary point within the prescribed range.

2.5.7 Energy-Aware and Secure Multi-Hop Routing (ESMR)

In (Haseeb et al., 2019), the authors proposed the ESMR technique to use a secret sharing scheme to multi-hop data security against malicious nodes and increase energy efficiency. The network in this technique is segmented into inner and outer zones based on the location of nodes. In each zone, the cluster head node aggregate data and secures the data using a secret sharing scheme and is forwarded to the base station node. However, the authentication and data confidentiality techniques were not considered, and the security network limited the only focus on securing the cluster head node. However, all the nodes were not secured in the network.

2.5.8 Secure Data Aggregation with Malicious Nodes Identification (MAI)

In (H. Li et al., 2014), the authors proposed MAI, a secure data aggregation scheme to detect the malicious nodes with a stable node communication overhead. To verify the malicious aggregator identification, this scheme performs aggregation recalculation. The child nodes verify the identification of the data aggregation before sending it to the parent nodes. If the child nodes are identified to be malicious, the aggregation stops, and the parent nodes will not receive the data from the child nodes. This scheme helps to avoid unnecessary data transmission and preserves energy consumption. In addition, the aggregation results are signed with the private key of aggregation, so the results cannot be changed by anyone. However, this scheme generates delays when the child nodes encrypt the private key between them; it also has high energy consumption.

2.5.9 Distributed Collision-Free Data Aggregation Scheme (DCFDAS)

In (Qin, Zhang, Ma, Ji, & Feng, 2018), the authors proposed the DCFDAS scheme in WSNs to aggregate the data in the network without conflicts to preserve the limited energy and to reduce the network delay at the same time. The DCFDAS consists of two parts: The Data Aggregation Tree (DAT) and the Fastest Collision-Free Scheduling

(FCFS). DAT is used to minimize data aggregation by controlling the number of nodes in the network. FCFS is used to reduce the working period of the node required for data aggregation and preserve the energy of nodes. The advantages of this scheme are that by changing the working states of nodes, energy consumption is reduced, and only a limited number of nodes are used to avoid conflict re-transmission. However, in this scheme, the security for nodes in the network was not considered, which will lead to the network being vulnerable to attacks.

2.5.10 Secure Data Aggregation Using Access Control Scheme (SDAACA)

In (Razaque & Rizvi, 2017), the authors proposed the SDAACA scheme to detect Sybil and Sinkhole attacks in WSNs and utilized access control to check the falsely calculated aggregated data. The SDAACA scheme consists of two algorithms: secure data fragmentation and node joining authorization to secure data aggregation in the network. First, secure data fragmentation is responsible for fragmenting the data into small pieces before sending it to the aggregation nodes and hiding the data from the adversary. Second, node joining authorization is responsible for verifying a new node's authorized process before joining the network. The advantage of this scheme is that data are fragmented into small pieces, so if the malicious nodes can access the network, they will not be able to access the original data. However, the MAC authentication of this scheme was not secured and shared the keys among nodes through the network, leading the malicious nodes to steal the keys and data in the network. Figure 2.8 shows the classification of security in WSNs. Also, the summarized secure data aggregation techniques and approaches are presented in Table 2.4.

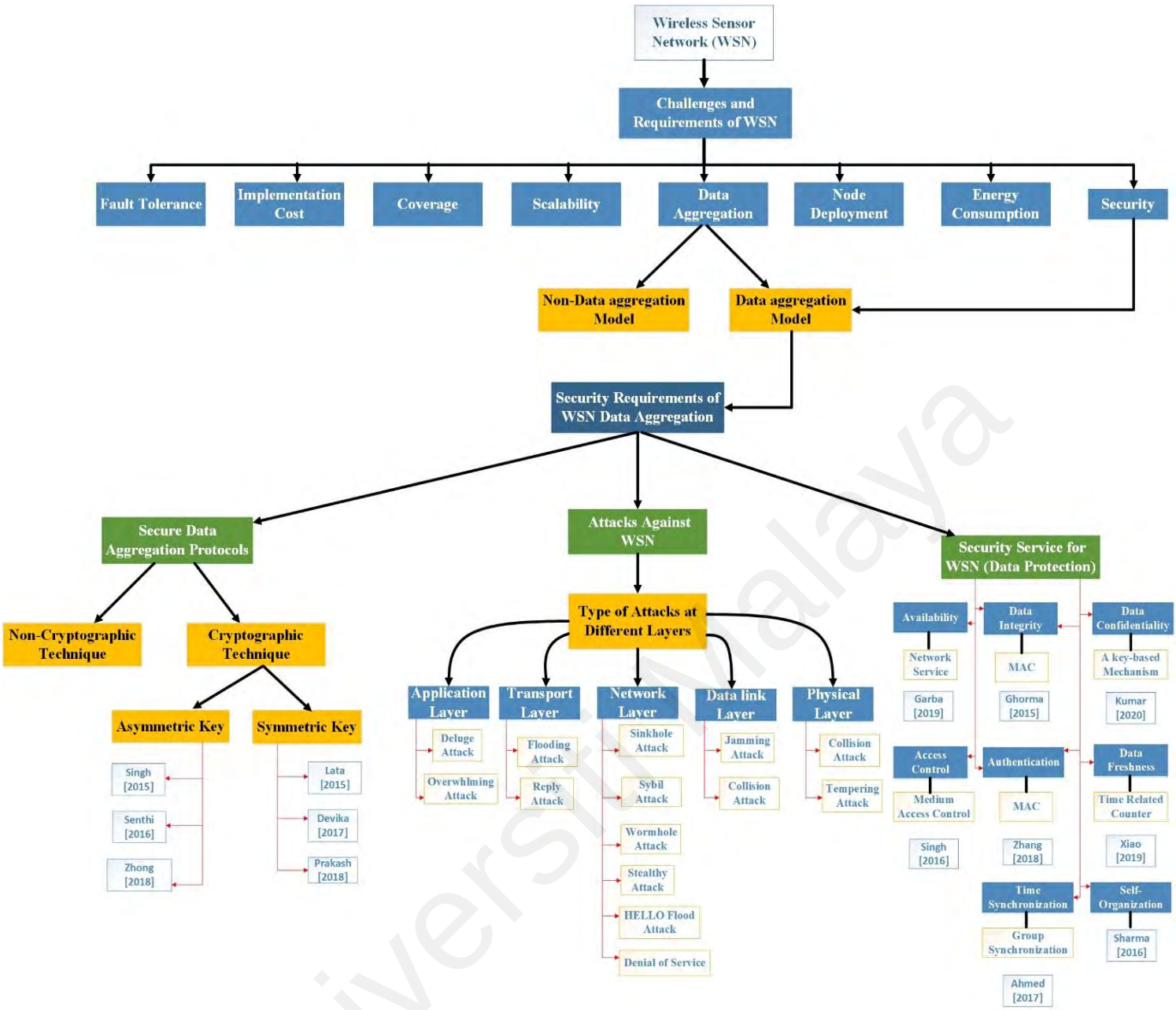


Figure 2. 8: The Classification of Security in WSNs

Table 2. 4: Summary of Secure Data Aggregation Techniques Based Clustering for WSNs

REF	Description	Contributions	Limitations
(Soufiene Ben Othman et al., 2013)	Proposed technique in order to achieve data confidentiality and integrity, and authentication process for secure data aggregation.	This technique has achieved integrity and authentication among nodes.	The detection of the malicious attack was not considered. It focused only on the encryption message in the network.

(Wang et al., 2013)	EESSDA, proposed in order to establish a secure channel between sensor nodes and their neighbors.	Provided data confidentiality by decomposing the data into slices before the transmission process, moderate energy efficiency, and accurate aggregation because data packets have less chance to collide.	Authentication among nodes was not addressed and did not provide security during the data aggregation process.
(Sofiene Ben Othman et al., 2013)	In EESDA, proposed to detect the malicious nodes in the network.	This technique provided secure message integrity with trade-offs between security and computation cost.	Authentication between nodes was not provided; there was increased energy consumption and delayed transmission in the network.
(Rezvani et al., 2014)	SDALFA this technique proposed in order to protect against sophisticated collision attacks.	The number of iterations required to approach a stationary point within the prescribed range was reduced.	The limitation of this method is that it only provides security for the sink nodes and does not support cryptographic methods.
(S. Roy et al., 2014)	SDA proposed a synopsis diffusion approach in order to secure data aggregation.	Reduced the communication overhead.	The base station received the authentication from sub-nodes only in the network.
(H. Li et al., 2014)	Proposed technique in order to detect the malicious nodes with a stable each node communication overhead.	This technique helped to avoid unnecessary data transmission.	The technique generates delays when the child nodes encrypt the private key between them and increased energy consumption.
(Mohan & Dayananda, 2016)	In EECSSDA, the proposed technique to increase network lifetime.	Reduced energy consumption and provide data confidentiality.	The detection of the malicious attack was not considered, the focus was only on encryption messages in the network.
(Prathima, Prakash, Venugopal, Iyengar, & Patnaik, 2016)	In SDAMQ, proposed secure data aggregation for multiple queries.	Provided authentication technique between nodes.	Increased energy consumption.
(Razaque & Rizvi, 2017)	SDAACA, proposed in order to detect Sybil and	This technique fragmented data into small pieces.	MAC, authentication of this technique was not secured, and keys were

	Sinkhole attacks in WSN and utilized access control to check the false calculated aggregated data.		shared among nodes through the network, which will lead to malicious nodes being able to steal the keys and data from the network.
(Qin et al., 2018)	In DCFDAS, proposed in order to aggregate the data in the network without conflicts to preserve the limit on energy and to reduce the network delay at the same time.	Changed the working states of nodes to preserve energy consumption and the number of nodes was limited used to avoid conflict retransmission.	The security of nodes in the network was not considered.
(Zhong et al., 2018)	Proposed technique to reduce energy consumption by filtering false data in the network.	Provided the end-to-end data confidentiality and integrity service.	Authentication among nodes was not addressed.
(P. Zhang, Wang, Guo, Wu, & Min, 2018)	MODA used multi-functional secure data aggregation in order to reduce the communication cost.	Provided end-to-end security.	Not much attention is given to address authentication and integrity among nodes in the network.
(Haseeb et al., 2019)	In ESMR, it used a secret sharing scheme to multi-hop data security against malicious nodes and increase energy efficiency.	The network is segmented into inner and outer zones based on the location of nodes.	The authentication and data confidentiality techniques were not considered. In addition, there was only a secured cluster head node, while all the nodes were not secured in the network.

2.5.11 Discussion of Authentication and Energy Efficiency Issues Related to Security Problems in Clustering Protocol for WSNs

In this discussion, the authentication and energy efficiency issues related to security problems for securing data aggregation techniques in clustering protocol were presented. Moreover, the cryptographic techniques, the types, and how they affect the network were also described. Different approaches and techniques were presented in this section. However, in all these studies, the authentication and energy efficiency remain challenging due to the sharing of the security key and the key length with a base station node, and not much attention is given to enhancing the authentication of the Medium Access Control (MAC) address. Also, the distance information among nodes was not calculated. The SDA, SDAT, SDALFA, EESSDA, SDAACA, and EESDA techniques were chosen to measure the reliability of our proposed SEEDA protocol. The SDA technique was chosen to check the base station for false aggregated data. The SDAT technique was also chosen for its encryption method, and the SDALFA technique was chosen due to the network's robustness against attacks. The EESSDA technique was chosen as it can provide data confidentiality and distance information, whereas the SDAACA technique was chosen for its ability to fragment data into pieces and its common energy consumption radio model with our protocol. Finally, EESDA technique was chosen for it is able to detect malicious nodes in the network.

In the previous section, a review of authentication and its energy efficiency issues has been conducted. Though energy issues have been covered, the review is only limited to energy consumption in the authentication process. In WSNs, many other factors affect energy issues. One of the important issues that affect energy efficiency, is the hot spots problem that will be presented in the next section. A review of this issue and the previous methods used to reduce it in the clustering protocols will be conducted. In addition, the review also includes the cover in the next section the types of clustering for WSNs with

different techniques and approaches and the equal and unequal clustering techniques to preserve the energy consumption of nodes in the network.

2.6 Energy-Efficient in Clustering Protocols for WSNs

Energy efficiency issues are a significant problem of the operation of WSNs since a battery-powered sensor node has limited energy and a complicated battery changing procedure; these affect the quality, performance, and lifetime of the WSNs clustering protocol. Furthermore, in this multi-hop communication process, the CH, which is closer to the BS, will do more forwarding tasks. This results in massive CH overhead, and these CHs run out of power sooner than the others. This leads to a breakdown of the cluster and a loss of communication between CHs. This breakdown is known as the hot spots problem. Furthermore, as the CH is responsible for the aggregating and transmission process, when the CH receives data from the CM, the CH will aggregate data and forward them to the BS. As the CHs consume more energy than CM, this also leads to unbalanced energy distribution in the overall network. If the CHs are not nearby, the furthest one from the BS node will dissipate more energy and increase the overhead.

Many researchers have proposed techniques and approaches to reduce hot spots problem, reduce energy consumption and prolong the network lifetime based on clustering algorithms will be described in subsection 2.6.3. This section describes the techniques and approaches proposed based on clustering algorithms such as static, dynamic, equal, unequal, and double clustering approaches and techniques. It also verifies how these clustering methods help to increase the energy efficiency and lifetime of the network.

2.6.1 Cluster-Based Protocols

Similar to tree-based protocols, cluster-based protocols are also used in hierarchically organized networks (Xuxun Liu, 2012; Naeimi, Ghafghazi, Chow, & Ishii, 2012). The cluster-based protocols are extensively used in hierarchical data aggregation. They can

efficiently manage data, reduce communication costs, enable traffic control, and improve energy efficiency and network stability (Gielow, Jakllari, Nogueira, & Santos, 2015). Node clustering is performed by virtually dividing the network into small sets of nodes or clusters. Each cluster consists of sensor nodes inside each other, and the nodes are grouped according to primary considerations. A cluster node can be designated either as the Cluster Head (CH) or as the Cluster Member (CM). First, data is collected from member nodes by CH. The data is then collected and sent to the upstream node (Abbasi & Younis, 2007; Shagari et al., 2020; X. Zhu, Shen, & Yum, 2009). Cluster-based protocols have the following advantages (Bouabdallah, Rivero-Angeles, & Sericola, 2009; O. Younis, Krunz, & Ramasubramanian, 2006):

(i) Enhance the bandwidth which is utilized to reduce the energy consumption (i.e., minimization of collisions caused by channel contention).

(ii) Minimize the overhead to reduce wasteful energy consumption.

(iii) Maximize the network lifetime by occasionally adopting a balanced energy consumption approach and distributing load among nodes.

(iv) Prevent long-distance transmission among nodes in order to boost resource utilization and lower energy consumption.

In a clustered network, the cost is classified as intra-cluster or inter-cluster cost. The communication cost from the nodes within a cluster to the CH is considered intra-cluster cost, whereas the one from the CH to the BS is the inter-cluster cost (S. P. Singh & Sharma, 2015). Scalability and effectiveness enhancements of a cluster are validated using the clustering cost. The drawbacks can be determined by qualitative or quantitative analysis of the cost of the clustering structure (Liao, Qi, & Li, 2013; Pukhrambam, Bhattacharjee, & Das, 2017; M. Younis, Youssef, & Arisha, 2003).

(i) When several mobile nodes are involved in the network's random topological change, a drastic increase occurs in the exchange of information due to clustering.

(ii) In some clustering mechanisms, the complete reconstruction of the entire network structure is involved if the remaining energy of the CH is depleted.

(iii) The number of rounds involved in forming a cluster is determined by the computation round metric.

Generally, cluster-based protocols operate in three stages: cluster formation, CH selection or election, and data transmission. Clustering algorithms can be either static or dynamic (Mahdi, Abdul Wahab, Idris, et al., 2016). The static and dynamic clustering techniques can be described as follow:

- **Static Clustering:** In static clustering, the clusters are formed prior to network operation and based on the network parameters, such as the residual energy in the nodes as in the study of (Wendi B Heinzelman, Chandrakasan, & Balakrishnan, 2002) or the physical distance as in the Voronoi diagram-based method by (W.-P. Chen, Hou, & Sha, 2004). Moreover, the updating and reestablishment of clusters do not occur adaptively. LEACH (W. R. Heinzelman et al., 2000) and HEED (O. Younis & Fahmy, 2004) are two classical models of static clustering. They differ in the method for selecting the CH. LEACH assumes that the energy levels of all the nodes are equal during the selection, whereas HEED considers the energy variation in the nodes to optimize the network lifetime.
- **Dynamic Clustering:** A dynamic cluster architecture is reactively formed close to the event sensing nodes. Once the event is located, a specific sensor node is selected as CH (ideally the node with the maximum energy or adjacent to the event), while the other event sensing nodes are designated as member nodes (Jain, Saini, & Bhooshan, 2014). The main benefit of this approach is that only

participating nodes are involved in data aggregation. Therefore, dynamic clustering conserves the energy of the idle nodes (Villas et al., 2013).

This thesis reviews some distributed clustering algorithms that can be divided into Equal and Unequal clustering algorithms. Some popular Equal clustering algorithms are LEACH (Wendi B Heinzelman et al., 2002), HEED (O. Younis & Fahmy, 2004), PEGASIS (Lindsey & Raghavendra, 2002), EECHS (Ren & Yao, 2020). These algorithms work well in homogeneous clustered networks of equal size. In unequal clustering algorithms, the network is partitioned into clusters of different sizes. Clusters near the base station are smaller than clusters that are far from the base station (Ali, 2015). These algorithms use the location of the base station as well as the residual energy as the CH selection parameter. Using unequal clusters decreases the intra-cluster work of the sensor nodes, which are closer to the base station or have a lower battery level. Some recent unequal clustering algorithms are proposed by researchers. (Ali, 2015; Simon et al., 2004; Yick, Mukherjee, & Ghosal, 2005, 2008; S. Zhang & Zhang, 2012). Uneven clustering can lead to more uniform power dissipation among cluster head nodes, effectively increasing network lifespan and solving the problem of hot spots.

In the following subsections, we introduce equal clustering and unequal clustering algorithms and protocols based on their cluster size.

2.6.2 Equal Clustering Protocol

On a large scale, WSN's energy efficiency and extended network lifetime were the main issues. Clustering the network has made data aggregation and communication between the node and the BS more efficient, thereby saving node power and extending the lifetime of the network. In this section, we discuss some distributed equal clustering algorithms as follows:

2.6.2.1 Low-Energy Adaptive Clustering Hierarchy (LEACH)

Low-energy adaptive clustering hierarchy (LEACH) is one of the pioneering cluster-based routing protocols in WSNs (Wendi B Heinzelman et al., 2002). Cluster structures are used for data aggregation, and the aggregation points are selected as the CHs. LEACH rotates CHs to achieve the fair and equal dissipation of energy among all the network nodes for communicating with the BS.

LEACH operation is divided into numerous rounds, each of which is divided into two phases: the setup and the steady-state phases. In the setup phase, the clusters are organized, whereas, in the steady-state phase, the data are delivered to the BS. The decision to become a CH in the ongoing round is made by every node in the setup phase. This choice is dependent on the proposed CH percentage for the network and the number of times the node has served as a CH up to that point. The decision involves selecting a number randomly between 0 and 1. If the selected number is less than the threshold calculated by Equation 2.1, then the node designates itself as the CH for the round in progress.

$$T(n) = \begin{cases} \dfrac{P}{1 - P\left(r \bmod \dfrac{1}{P}\right)}, & if\ n \in G \\ 0, & otherwise \end{cases} \qquad (\mathbf{2.1})$$

Where P denotes the desired percentage of CHs, r denotes the ongoing round, and G denotes the set of nodes that have not served as CHs in the previous $1/P$ rounds. An advertisement message is broadcast by the node on its successful election as the CH node. Depending on the received signal strength of the advertisement, a membership message is sent by the other nodes to the CH after they have decided to join it. For the even distribution of the energy load among the sensor nodes, rotation is involved in the CH election in every round by initiating a new advertisement phase depending on the

calculated result of Equation 2.1. The data are sensed and then transmitted to the CH in the steady-state phase. The data received by the CH from its respective cluster nodes are aggregated before being sent to the BS.

The advantages of LEACH include the following (Lai, Fan, & Lin, 2012; Xuxun Liu, 2012; Zungeru, Ang, & Seng, 2012): (i) LEACH reduces the amount of data to be transmitted to the BS by data aggregation. (ii) The load is shared fairly among the nodes as a CH cannot be reelected. (iii) LEACH does not require any control information from the BS or global knowledge of the network to operate; thus, LEACH is a completely distributed routing protocol. (iv) Unnecessary collisions are avoided by the use of a TDMA schedule. (v) Energy dissipation can be avoided by the cluster members by opening or closing communication interfaces following their allocated time slots.

Although LEACH is the simplest hierarchical technique and can reduce energy consumption in a WSN, it poses many problems, including the following (Abbasi & Younis, 2007; J. Chen, 2011; Xuxun Liu, 2012; Xuxun Liu & Shi, 2012): (i) the residual energy of the node is not considered in the CH selection process. As a result, the nodes with low initial energy can be selected as the CHs, and premature death, coverage, and energy hole problems can be consequently yielded. (ii) A significant amount of energy is wasted in constructing the clusters as the clusters are reformed in each phase. (iii) The CHs can be densely or sparsely deployed in different areas, as LEACH performs the CH selection in terms of probabilities. (iv) The CHs located far from the BS die earlier given that CHs transmit aggregated data to the BS directly. (v) LEACH is inappropriate for large-scale networks because single-hop transmission is adopted in inter-cluster and intra-cluster communications; thus, LEACH is not a scalable routing protocol.

Several modified versions of the original LEACH technique have been proposed in the literature, including TL-LEACH (Loscri, Morabito, & Marano, 2005), E-LEACH (Xiangning & Yulin, 2007), M-LEACH (Xiaoyan, 2006), V-LEACH (Yassein,

Khamayseh, & Mardini, 2009), LEACH-FL (Ran, Zhang, & Gong, 2010), W-LEACH (Abdulsalam & Kamel, 2010), and T-LEACH (Hong, Kook, Lee, Kwon, & Yi, 2009).

2.6.2.2 Hybrid Energy-Efficient Distributed Clustering (HEED)

In (O. Younis & Fahmy, 2004), the authors proposed the HEED technique, which uses a hybrid method for cluster head selection for the homogeneous network. The overall objective of HEED is to form efficient clusters to increase the lifetime of the network. The cluster head is selected based on a mixture of the node residual energy of each node and a secondary parameter, which is subject to the proximity of the node to its neighbors or the degree of the node. HEED ends in O (1) iterations, which reduce the communication cost (S. Kumar, Prateek, Ahuja, & Bhushan, 2014). It equally scatters cluster head overall the network. The HEED protocol can be used in various applications on sensor networks, such as fault tolerance, extended network lifetime, scalability, and load balanced. The cost of a cluster head is defined as its Average of the Minimum Reachability Power (AMRP). AMRP is the average of the minimum power levels required by all nodes within the range of the cluster to reach the head of the cluster. AMRP provides an approximation of the cost of communication.

2.6.2.3 Power-Efficient Gathering in Sensor Information Systems (PEGASIS)

In (Lindsey & Raghavendra, 2002), the authors proposed the PEGASIS technique for optimal gathering of data in WSNs. The main idea of this technique is to form a greedy chain between the sensor nodes so that each node receives data and transmits it to a close neighbor. The chain proposed to minimize the total length of data when the nodes transfer data to the base station. The advantages of this technique minimize the distance among nodes, nodes take turns in transmitting the fused data to the base station to balance energy consumption in the network, the number of transmissions and receiving's is limited

among all sensor nodes. However, the load on the cluster head was high, which leads to reducing the network lifetime.

2.6.2.4 Energy-Efficient Cluster Head Selection Scheme (EECHS)

In (Ren & Yao, 2020), the authors proposed the EECHS scheme in order to reduce energy consumption during the selection of cluster heads in the network. The cluster head can preserve more energy for data transmission, and the nodes used to store and monitor the real-time information of reminding energy for all sensor nodes. However, in this scheme, the cluster head is randomly and alternately selected among the network nodes based on probability and does not balance energy consumption between nodes in the network.

As mentioned above, the previous subsection presented the clustering protocols with types of its to reviewing the energy consumption of nodes in the network. The next subsection describes another important energy efficiency issue, which is the hot spots problem. The techniques and approaches that are proposed to reduce it and the types found in the clustering protocol will also be introduced in the next subsection.

2.6.3 Unequal Clustering Protocol

There is a hot spots problem with equal clustering, which results in unbalanced power consumption in equally formed clusters. Clusters far from the base station are dead than nearby clusters due to the higher communication cost. To overcome this problem, unequal clustering has been proposed. The discussion of some unequal clustering algorithms and protocols can be described as follows:

2.6.3.1 Distributed Energy-Efficient Clustering Scheme for Heterogeneous (DEEC)

In (Qing, Zhu, & Wang, 2006), the authors proposed the DEEC scheme to achieve energy efficiency and increases the scalability and lifetime of the network. The selection of the

cluster head is based on the probability of the ratio between the residual energy of each node and the average energy of the network. The node with high residual and initial energy will have more chances to become a cluster head than the node with low energy. The DEEC scheme assumed that all sensor nodes in the network are equipped with different amounts of energy, which is a source of heterogeneity. It could be the result of reactivating sensor networks in order to extend the lifetime of the network. The advantages of this scheme can prolong the lifetime of the network, especially the stability of a heterogeneous-aware clustering algorithm. However, this scheme does not address the closest distance cluster node to the base station, leading to a problem with hot spots when there is more than one cluster transmitting data to the base station.

2.6.3.2 Energy-Efficient Unequal Clustering Mechanism (EEUC)

In (C. Li, Ye, Chen, & Wu, 2005), the authors are proposed the EEUC technique in order to look at periodical data aggregation in WSN. This technique aims to segment the nodes into unequal clusters; the nodes closest to the base station are smaller than those furthest from the base station. This technique has been proposed to conserve energy for forwarding data between clusters. They also proposed an energy-aware multi-hop routing protocol for inter-cluster communication. In addition, the EEUC technique utilizes localized computation for the selection of cluster heads. The advantage of this technique is that it divided the network into unequal size clustering. This leads to reducing the distance between nodes and base station and avoids the problem of hot spots. However, this technique utilizes one cluster head for aggregation and data transmission to the base station, which leads to increased load on the main cluster head.

2.6.3.3 A fuzzy Energy-Aware Unequal Clustering Algorithm (EAUCF)

In (Bagci & Yazici, 2013), the authors are proposed the EAUCF technique to resolve the problem with hot spots and to handle the uncertainties in estimating the cluster head

radius. This technique utilizes a probabilistic model for the selection of the cluster head and utilizes the distance to the base station and residual energy to make wise decisions in the network. The advantages of this technique are that they addressed the node close to the base station, which will consume more energy for receiving and transmitting to the base station. However, the balance of energy consumption among nodes was not considered.

2.6.3.4 Fuzzy Based Unequal Clustering (FBUC)

In (Logambigai & Kannan, 2016), the authors are proposed the FBUC technique in order to enhance the fuzzy energy-aware unequal clustering algorithm (EAUCF). The fuzzy system algorithm is used to determine the radius of nodes with residual energy inputs and a number of neighbor nodes. The head of the cluster is elected in the network based on the energy level, and the CMs join the CHs based on a fuzzy system with distance from the CH, and the CH numbers are fuzzy system inputs. The advantage of this technique reduces the transmission delay. However, the overhead is increased by this method that leads to reducing network lifetime.

2.6.3.5 Energy Conserved Unequal Clusters with Fuzzy Logic (ECUCF)

In (Sundaran, Ganapathy, & Sudhakara, 2017), the authors proposed the ECUCF algorithm based on the distance of nodes from the base station in order to enhance the FBUC protocol. The EECUF algorithm is proposed to reduce energy consumption in the network. The network is divided into closest, middle, outside sectors from the base station. This algorithm utilizes an asleep-awake mechanism to preserve the energy of nodes. The selection of the cluster head is randomly based on probability between nodes, and the base station is randomized rotationally. However, this algorithm utilizes one cluster head for aggregation and data transmission to the base station.

2.6.3.6 Energy-Driven Unequal Clustering (EDUC)

In (Yu, Qi, & Wang, 2011), the authors proposed the EDUC approach for heterogeneous wireless sensor networks. The EDUC approach includes an energy-driven adaptive cluster head rotation method and a distributed unequal clustering algorithm. The unequal clustering algorithm used competition ranges to generate clusters of unequal size. The clusters farther away from the base station have smaller sizes than those closer to the base station. Thus, the cluster heads farther away from the base station can preserve some energy for the long-distance transmission, and an energy-driven cluster head rotation method is adopted to rotate the role of the cluster head and balance the energy consumption in the network. Each node acts as a cluster head no more than once during the whole network lifetime. However, this approach addressed balanced energy consumption but did not consider the distance length of nodes from the base station, which leads to reducing the network lifetime.

2.6.3.7 Unequal Clustering Based Routing (UCR)

In (G. Chen, Li, Ye, & Wu, 2009), the authors proposed the UCR approach to avoid the hot spot problem. The UCR approach consists of two algorithms; one is a greedy geographic and energy-aware routing protocol for inter-cluster communication; the second is an Energy-Efficient Unequal Clustering (EEUC) algorithm for topology management. The selection of the cluster head is based on the residual energy of neighboring nodes. The base station sends the beacon signal to all sensors to calculate the distance from each node based on the strength of the received signal. This helps to select the proper throughput for data transmission to the base station and to create unequal clustering. Each cluster head has a competition area and is used to create clusters of unequal size. After selecting the cluster head, the cluster head sends an advertisement message to the network. The node is connected as a cluster member to the cluster head with higher received signal strength, and the Voronoi area of the sensor node is also

established. For multi-hop inter-cluster routing, the relay nodes are selected based on the ratio of residual energy and energy costs of the relay paths. It achieves a maximum service life compared to HEED but is prone to errors and less robust due to the noise in the real environment.

2.6.3.8 Balanced-Imbalanced Cluster Algorithm (B-IBCA)

In (Sivakumar, 2020), the authors proposed the B-IBCA in order to reduce energy consumption and prolong the network lifetime based on the stabilized Boltzmann approach. The advantage of this approach, the selection of CH based on the distance and residual energy consumption in the network. However, this approach did not reduce the load on the cluster head node and was not applied the sleep-awake mechanism, which leads to increasing the energy consumption in the network.

2.6.3.9 A Two-Tier Distributed Fuzzy Logic-Based Protocol (TTDFP)

In (Sert, Alchihabi, & Yazici, 2018), the authors proposed the TTDFP protocol in order to increase the network lifetime of multi-hop WSNs based on a fuzzy logic protocol and address the aggregation problems. This protocol utilizes an unequal clustering protocol to solve the hot spots problem and to reduce the energy consumption of nodes. In the first tier distributed fuzzy based on the energy-based competition of provisional leaders, it was chosen by a probabilistic model. The selection of CHs was based on the maximum competition radius and threshold. While the second tier utilized fuzziness to enhance the routing. The selection of CHs is based on distance to BS, residual energy, and relative distance. The advantages of this protocol addressed the hot spots problem and utilized energy, the distance for the selection of CHs. However, this approach did not consider a double cluster head and the sleep-awake mechanism in the network. It would be an increased load on the main CH and consume more energy.

2.6.3.10 An Improved Energy-Aware Distributed Unequal Clustering (EADUC)

In (Gupta & Pandey, 2016), the authors proposed the EADUC approach in order to improve the selection of CHs and to solve the hot spots problem in WSNs. The selection of CHs was based on a number of nodes in the neighborhood. The advantage of this approach is that it utilizes relay matric and distance information for forwarding the data toward the BS. However, it does not utilize the sleep-awake mechanism of nodes, which leads to increases in energy consumption in the network.

2.6.4 Double Cluster Head Based Clustering Protocol

The double cluster head in clustering protocol is proposed aiming at the premature death of the cluster head due to speedy energy consumption and the unbalanced energy consumption of nodes. In addition, it proposed this scheme to reduce the overhead and energy consumption of the cluster head nodes and distribute operations between the cluster head nodes. In the sub-sections, we describe some of the protocols that proposed double clustering in the network.

2.6.4.1 Energy-Efficient Fuzzy Logic for Unequal Clustering (EEFUC)

In (Phoemphon et al., 2020), the authors proposed the EEFUC technique in order to reduce energy consumption in the network by multi-hop clustering using the fuzzy logic method. The advantage of this technique is that it has taken into consideration the distance among cluster members, cluster head with the base station, which leads to preserving energy consumption in the network. The clustering architecture is divided into four stages: CH selection, CH determination, computation radius, and selection of second CH. The selection of CH utilizes the fuzzy logic method, and they calculate the computation radius for unequal clustering distribution nodes. However, the balance of energy consumption was not considered; it will increase energy consumption and reduce network lifetime.

2.6.4.2 Energy-Efficient Unequal Double Clustering (UDCH)

In (F. Zhu & Wei, 2019), the authors proposed the UDCH technique in order to avoid the hot spots problem and reduce the energy consumption of the cluster head node. The UDCH technique used unequal clustering technology to solve the problem of hot spots. The cluster head closer to the base station takes on more forwarding tasks, so the cluster size should be smaller to reduce the overhead of the cluster head node. In addition, unequal clustering technology is adopted by calculating the competition radius by each node. This technique proposed two cluster heads to reduce energy consumption and overhead. The process selection of the main cluster head is based on calculating the delay time by each node, while the selection of the second cluster head is determined by the distance of sensor nodes to the main cluster head. The main cluster head is responsible for aggregating data and forwarding them to the base station, while the second cluster head is responsible for collecting the data from sensor nodes and transmitting them to the main cluster head. The advantage of the UDCH technique is that it addresses the hot spots problem in the network and distributes the responsibilities between cluster head nodes. However, in this scheme, the distance length between cluster members and the base station node was not considered, which lead to energy wastage across the network nodes while reducing the lifetime of the network. Moreover, in UDCH, not much attention is given to enhancing the data transmission process between sensor nodes and cluster head nodes in the network.

2.6.4.3 Impact of the Secondary Cluster Aggregation Based on Location (FLEACH)

In (Amodu & Mahmood, 2018), the authors proposed the FLEACH technique in order to prolong the network lifetime in WSN. The proposed technique employs the second cluster to aggregate the data and send them to the primary CH based on location threshold and energy threshold to reduce the energy consumption by reducing the overhead and increase network lifetime. The selection of the main cluster head is determined based on the

random selection of CH, which was based on probability, while the selection of the second cluster head was determined by the highest residual energy of nodes. However, this protocol does not address the hot spots problem. Furthermore, randomly selected CH, similar to the LEACH technique processing, leads to the loss of more energy for selection, and the nodes will die soon.

2.6.4.4 Multi-Clustering Algorithm Based on Fuzzy Logic (MCFL)

In (Mirzaie & Mazinani, 2018), the authors proposed the MCFL technique in order to reduce the number of transmitted messages between the nodes. This scheme proposed a fuzzy logic for selecting multi-clustering algorithm nodes that are clustered in different rounds using different algorithms without selecting any nodes as cluster heads in some rounds. The number of messages transmitted from each node to other nodes and to the base station has been reduced, saving more power on the network. The advantages of this method are that it increased the throughput by increasing the number of messages addressed to the base station. However, the hot spots problem was not addressed in this method.

Finally, as mentioned earlier, energy efficiency is a common issue in WSNs. Therefore, in this thesis, we compared and presented the limitations of clustering protocols and algorithms from the various researchers is in Table 2.5. In addition, the classifications of clustering techniques and approaches in WSNs are shown in Figure 2.9.

Table 2. 5: Summary of Techniques Based on Clustering Protocols for WSNs

REF	Description	Contributions	Limitations
(Wendi B Heinzelman et al., 2002)	In LEACH, the selection of CH is randomly based on probability between nodes, and the base station is randomized rotationally. The	Reduces communication between the nodes and base station to preserve energy in the network. Utilized data aggregation	LEACH does not address the hot spots problem. A randomly selected cluster head leads to energy loss, and the nodes are terminated sooner.

	responsibility of CH is to collect data from cluster members, perform data aggregation, and forward them to the base station directly.	technique in order to reduce the data redundancy transmission.	
(Eschenauer & Gligor, 2002)	The selection CH was based on remained energy, and the location address of nodes in the network was proposed.	This technique successfully balanced the network energy burden and dramatically improved energy efficiency.	The distance among nodes and the sleep and awake mode was not considered; this leads to increases in energy consumption in the network.
(Lindsey & Raghavendra, 2002)	Proposed technique to minimize the total length of data when the nodes transfer data to the base station.	Minimize the distance among nodes, and the number of transmissions and receiving is limited among all sensor nodes.	The load on the cluster head is high, which leads to reducing the network lifetime.
(O. Younis & Fahmy, 2004)	HEED, the selection of CH was based on the remained energy of nodes and the communication cost.	Multi-hop routing, also inter-cluster and intra-cluster transmission was used.	The overhead of the HEED technique was high.
(C. Li et al., 2005)	In EEUC, the selection of CH was based on the localized computation, also utilized the unequal clustering technique on nodes in the network.	Decreased the distance between nodes and the base station, whereby divided the network into unequal clustering in size.	One CH was selected to collect and simultaneously transmit data to the base station.
(Qing et al., 2006)	In DEEC, the selection of CH was based on a ratio of residual energy of each node and the average energy level of the network.	The advantage of this technique was that it improved energy efficiency in the network.	It does not address the close distance between cluster nodes and the base station; it leads to the problem of hot spots when more than one cluster transmitting data to BS.

(G. Chen et al., 2009)	In UCR, proposed in order to avoid hot spot problem.	Increased network lifetime.	Increased overhead in the network.
(Sen, 2010)	LEATCH, offers a two-level hierarchical clustering approach to guarantee communication between cluster nodes and the base station in the network.	The advantage of this technique was reduced delay and energy consumption.	The balanced energy consumption technique between cluster nodes was not addressed.
(Yu et al., 2011)	In EDUC, proposed for heterogeneous wireless sensor networks.	Addressed balance energy consumption	Did not consider the distance length of nodes from the base station, which leads to reducing the network lifetime.
(H. Lin, Wang, & Kong, 2015)	Fan-Shaped Clustering proposed a method in order to increase network lifetime.	Addressed hot spots problem.	It does not utilize the sleep-awake mechanism of nodes.
(Abo-Zahhad, Ahmed, Sabor, & Sasaki, 2015)	In MSIEEP, proposed in order to alleviate the energy holes.	Addressed hot spots problem.	Did not consider the distance length of nodes from the base station, which leads to reducing the network lifetime.
(Logambigai & Kannan, 2016)	Determined the radius of nodes was based on a fuzzy system, and the selection of CH was based on the energy level of nodes.	It reduced the transmission delay in the network.	Increased overhead in the network.
(Gupta & Pandey, 2016),	In EADUC, a technique was proposed in order to improve the selection of CHs and to solve the hot spots problem in WSNs.	Utilizes relay matric and distance information for forwarding the data toward the BS.	It does not utilize the sleep-awake mechanism of nodes, which leads to increases in energy

			consumption in the network.
(Mittal, Singh, & Sohi, 2017)	In SEECP, the technique is proposed to balance load among nodes.	Utilize residual energy for selection CHs.	The hot spots problem was not addressed.
(Han, Yang, Wang, & You, 2017)	Distributed energy-efficient clustering DEC proposed technique in order to reduce energy consumption.	Proposed a double cluster head.	The hot spots problem was not addressed.
(Bozorgi, Rostami, Hosseinabadi, & Balas, 2017)	Energy harvesting (EH-WSN) proposed a technique to increase the stability of the network	Balanced energy consumption.	The hot spots problem was not addressed.
(Sert et al., 2018)	TTDFP was proposed to address the aggregation problem and to increase network lifetime.	Addressed the hot spots problem and utilized energy, distance for the selection of CHs.	Did not consider a double cluster head and the sleep-awake mechanism in the network. It would be an increased load on the main CH and consume more energy.
(Amodu & Mahmood, 2018)	In FLEACH, the random selection of CH was based on probability. Selection of multi-level CH in order to reduce the load on primary CH.	The secondary cluster head was determined based on the highest residual energy of the nodes.	The hot spots problem was not addressed.

(Mirzaie & Mazinani, 2018)	Utilized fuzzy logic in order to select CH, clustering nodes in different rounds uses different clustering algorithms.	Increased throughput by increasing the number of messages addressed to the base station.	The hot spots problem was not addressed in this method.
(F. Zhu & Wei, 2019)	In UDCH, utilized double CH in order to reduce energy consumption and prolong the network lifetime.	They addressed the hot spots problem in the network.	The distance threshold among nodes was not calculated, which leads to the reduction of network lifetime and increase energy consumption.
(Sivakumar, 2020)	In B-IBCA, the selection of CH was based on the distance to the base station and the residual energy in the network.	This technique addressed the hot spots problem in the network.	This method has not reduced the load on the cluster head node and was not applied sleep-awake mechanism, which leads to increasing the energy consumption in the network.
(Ren & Yao, 2020)	In EECHS, the CH is randomly and alternately selected among the network nodes based on probability.	Reduced the delay transmission of data in the network.	Unbalanced energy consumption in the network.
(Phoemphon et al., 2020)	In EEFUC, utilized a fuzzy logic method in order to reduce energy consumption and multi-hop clustering in the network	The selection of multi-hop clustering was based on higher residual energy.	The balanced energy consumption among nodes was not addressed, which leads to an increase in communication overhead in the network.

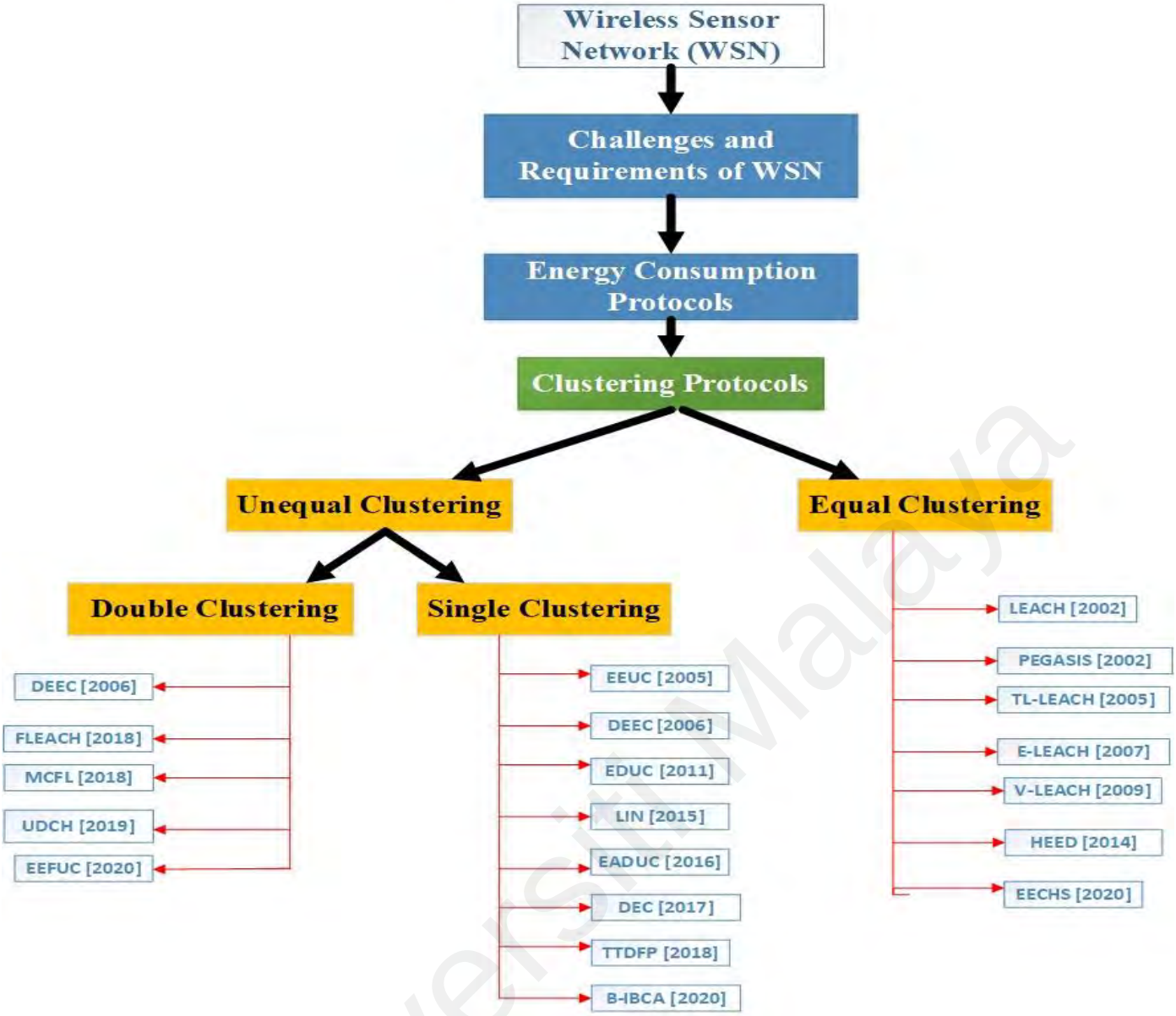


Figure 2. 9: Classification of Clustering Protocols and Types of Its in WSNs

2.6.5 Discussion of Equal and Unequal Clustering Protocols for WSNs

In the discussion of this section, one of the important issues that affect energy efficiency, which is the hot spots problem was presented. A review of this issue and the previous methods to reduce it in the clustering protocols were also presented. In addition, the types of clustering for WSNs with different techniques and approaches and the equal and unequal clustering techniques to preserve the energy consumption of nodes in the network were covered. Several techniques and approaches that were designed to reduce energy consumption and prolong network lifetime in WSNs were also reviewed. Different

approaches and schemes were employed by the clustering protocol, not limited to the selection of CHs, double cluster heads, CH rotation, and redundant data transmission. However, in all these studies, balancing energy consumption and network lifetime remains challenging due to the poor balance in energy consumption, the distance among nodes, and the load transmission on the CH node in the network. The novelty of this research consists of our ability to enhance the balanced energy consumption and to reduce the load transmission on the CH node in the network. The LEACH, FLEACH, EEFUC, and UDCH techniques were chosen to measure the reliability of our proposed EEUCB protocol. The LEACH technique was chosen for to its common energy consumption radio model with our protocol. Whereas the FLEACH technique was chosen due to its common selection of the 2CH with our protocol to reduce the load on the primary CH. The EEFUC on the other hand, was chosen for the selection of multi-hop clustering. Lastly, the UDCH was chosen in consideration of the hot spots problem and due to its common selection of the primary CH with our protocol in WSNs.

2.7 Chapter Discussion

This chapter provided background on sensor network technology and its evolution and also briefly described the unique features, challenges, and requirements of WSNs. Moreover, data aggregation was defined. Clustering protocols (i.e., security protocols, data aggregation functions, data aggregation scheduling algorithms, and data representation algorithms) in WSNs were also reviewed in detail. Every technique adopts an energy-saving mechanism to prolong the network lifetime. For this reason, using clustering protocols is vital in order to decrease the number of data transmissions and energy consumption. The secure data aggregation techniques in the clustering protocol for WSNs were reviewed and classified. In addition, the types of attacks at different layers are also defined and classified in the network.

In addition, the clustering protocols and the types of clustering algorithms were described. The main concepts and representative techniques of each type of clustering and security protocols were discussed. Furthermore, a comparison between the different security and clustering techniques was presented, with their main advantages and limitations highlighted. Finally, existing data aggregation techniques and their connection to the present study were comprehensively reviewed. These techniques and approaches were categorized as security based on cryptography, authentication techniques, and unequal clustering protocols in WSNs. The secure data aggregation and energy efficiency techniques in clustering protocols discussed are summarized in Tables 2.3, 2.4, and 2.5, respectively.

The next Chapter 3 describes the methodology and procedures adopted to achieve the objectives of the current study.

CHAPTER 3: RESEARCH METHODOLOGY

3.1 Introduction

This chapter presents the general methodology used in designing and implementing of the proposed protocols to achieve the outlined objectives, as illustrated in Figure 3.1. The first stage of this chapter reviews the previous works that focus on security problems related to authentication and energy issues for secure data aggregation technique in clustering protocol. In addition, aside from looking at the authentication and energy issues related to security problems, this stage also reviews another important energy efficiency issue in clustering protocol, which is the hot spots problem. Therefore, section 3.2 reviews the literature of previous works to address these issues and how to overcome them.

In the second stage of this chapter, the focus is given on how to overcome authentication and energy issues due to the security key and key length sharing problem at the base station. This can be done by enhancing the authentication of the Media Access Control (MAC) address and by utilizing the distance information and timestamp to detect attacks and reduces energy consumption, using a protocol called Secure and Energy-Efficient Data Aggregation method in clustering based on access control model (SEEDA) protocol. Then, section 3.3 presents the requirements, design, implementation, and verification of (SEEDA) protocol.

In addition, the third stage of this chapter will present on how to reduce the hot spots problem and balance the energy consumption among nodes. This can be done by utilizing an unequal clustering in WSNs that can be done through a protocol called Energy-Efficient Unequal Clustering protocol based on a Balanced energy method (EEUCB). Then, the requirements, design, implementation, verification of (EEUCB) protocol will be presented in section 3.4.

Finally, the performance evaluation metrics of this research were defined in the fourth stage. Each method will be further explained in Chapters 4 and Chapter 5.

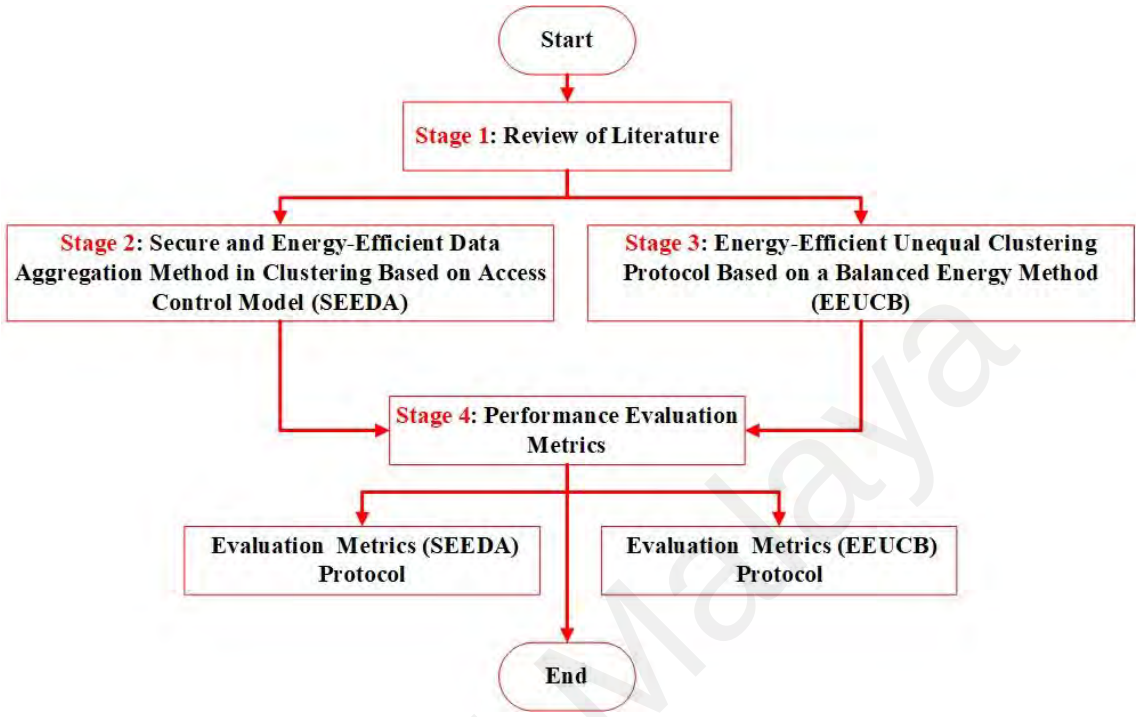


Figure 3. 1: Research Methodology

3.2 Stage 1: Review of Literature

This stage reviews relevant and credible state-of-the-art studies on key issues of security and energy in the clustering protocols of WSNs to investigate support provided by the clustering protocol to increase the detection rate of malicious nodes, network lifetime, and reduce energy consumption in the network. In WSNs, the malicious nodes cause the failure nodes to perform their tasks because WSNs are vulnerable to various attacks because of their distributed wireless nature, resulting in delays and loss of data in the network. As presented earlier in chapter 2, most of the techniques proposed in the literature attempt to solve several problems affecting security in WSNs by authentication, authorization, and preserving the original data in the network. However, many drawbacks are confronted by existing security strategies. Hence, efficient performance is not observed.

In chapter 2, different security clustering techniques are discussed and classified with the particular assertion on their security structure. According to the security classification in chapter 2, the benchmark techniques are selected from the hierarchical structure categories to investigate and evaluate the performance of proposed schemes. The main objective of the benchmark studies (SDA, SDAT, SDALFA, EESSDA, SDAACA, and EESDA) is to target the issues of security to facilitate authentication in WSNs. These works' main focus is to achieve network integrity and security via clustering protocols and to prevent malicious attacks from accessing the network. In addition, energy efficiency is also a problem that needs to be addressed because the battery-powered sensor node has limited energy and a complicated battery-changing procedure. The WSNs are usually used to monitor harsh and inaccessible environments, which restrain the use of infrastructure-based networks that may need constant human monitoring and interventions. Due to challenging circumstances and random sensor node deployment, however, replacing or recharging “dead” sensor nodes' batteries is difficult. Therefore, these challenges will affect the quality, performance, and lifetime of WSNs.

In addition, algorithms or security techniques should be highly efficient in terms of energy consumption. Therefore, this section will present the authentication and energy efficiency together due to their overlapping in security issues of WSNs. Hence, to increases the detection rate of malicious nodes, should be reducing the energy consumption of nodes in the network and on the contrary.

Apart from looking at the authentication and energy issues, this thesis also focuses on another important energy efficiency issue, which is the hot spots problem. This issue means the sensor nodes closer to the base station nodes will take on more forwarding tasks. This will result in a massive overhead of the sensor nodes, and these nodes will run out of power sooner than the others. It causes a breakdown of the nodes and a loss of

communication between sensor nodes; this breakdown is called the hot spots problem. As presented earlier in chapter 2, most of the techniques proposed in the literature attempted to solve several problems affecting energy in the clustering protocol of WSNs by unequal clustering and balanced energy consumption among nodes in the network. However, many drawbacks are confronted by existing equal and unequal clustering strategies, which hence leads to unbalanced energy distribution in the overall network.

In chapter 2, different unequal clustering and double cluster head approaches and techniques are discussed and classified with the particular assertion on their clustering route structure. According to the clustering classification in chapter 2, the benchmark protocols are selected from the hierarchical structure categories to investigate and evaluate the performance of proposed schemes. The main objective of the benchmark studies (LEACH, FLECH, EEFUC, and UDCH) is to target the issues of balanced energy consumption in WSNs. The main focus of these works is to solve the hot spots problem and to reduce the energy consumption of head nodes in the clusters.

3.3 Stage 2: Secure and Energy-Efficient Data Aggregation Method in Clustering Based on Access Control Model (SEEDA)

The second stage proposes a Secure and Energy-Efficient Data Aggregation Method in clustering Based on Access Control Model (SEEDA) Protocol. The proposed secure clustering protocol aims to increase the malicious node detection rate, providing authentication to prevent attacks from access to the network, and reduce energy consumption. At this stage, the methodology used is divided into four sub-stages: design, implementation, and validation. At this stage, the proposed SEEDA protocol used is divided into four sub-stages: requirements, design, implementation, and validation:

3.3.1 Requirement of Proposed SEEDA Protocol

The first sub-stage to establish a secure clustering scheme is to assess previous methods, analyze them all in order to identify a gap, and so define the system requirements. A comprehensive literature review is focused on discovering the problems of authentication schemes. This literature review, presented in chapter 2, discusses the advantages and drawbacks of different existing secure clustering protocols (Objective 1). In this chapter, the review of the literature section also provides the problem statements of different existing methods. The proposed protocol was simulated using the network simulator 3.25 running on the Ubuntu operating system, 16.4 LTS version. Several scenarios of attacks were performed to evaluate the ability of the proposed protocol to detect malicious nodes in the network.

3.3.2 Design of Proposed SEEDA Protocol

This second sub-stage oversaw the development of a secure and energy-efficient data aggregation method (SEEDA) using an access control model to address security and energy issues for WSNs. This protocol was conducted by modifying a few functionalities of SDA, SDAT, EESDA, SDALFA, EESSDA, and SDAACA techniques and schemes. Our SEEDA protocol follows the same scenario as presented in (Razaque & Rizvi, 2017), which considers the oil-refinery monitoring process using WSNs. The proposed SEEDA protocol aims to enhance the authentication between the nodes in the network and detect and prevent the malicious node from joining and accessing the network. For authentication, the SEEDA protocol improves medium access control (MAC) address by generating a random timestamp and random value with a secret key to verify the fake aggregated data when the base station received the packets, which allows the detection and prevention of the attacks when a new node attempts to join the network. Furthermore, the base station will perform checks on the fake aggregated data before sending them to the server by comparing the nodes' information with the broadcasted query message from

the cluster head node when a new node joins the network. The cluster head nodes have the cluster member nodes' information, such as node identity, message information, and time stamp.

In addition, the base station utilizes distance information to detect attacks. This proposed scheme reduces energy consumption by reducing the redundancy of the transmitted data. The SEEDA protocol also focuses on protecting the aggregated data from attacks such as Sybil and Sinkhole. These attacks attempt to engage all the nodes in the network, which also provides a platform for other forms of attacks, such as tricking and alleviating routing information. This scenario will increase traffic generation in the network, send fake routing data to nodes, and increase the redundancy of data transmission in the network.

The SEEDA protocol consists of three main algorithms: data fragmentation, secure node authentication, and fully homomorphic encryption algorithms based on the access control model. Details, descriptions of the data fragmentation, secure node authentication, and the fully homomorphic encryption algorithms are provided in chapter 4.

The data fragmentation algorithm breaks the data into smaller pieces before the data are transmitted to the next-hop nodes to hide them from being attacked. The SEEDA protocol utilizes a fragmentation algorithm to keep the original data from the attacker. For example, if the attackers can access the network, they will be able to read and transmit data to other malicious nodes, or attackers will just drop the original data. Therefore, to avoid these issues and to prevent the attacker from accessing the original data, data is fragmented into blocks. The details of this algorithm are given in subsection 4.2.1.

Meanwhile, the secure node authentication algorithm checks if any node is leaving or joining the network to prevent the data between nodes from being tampered with or interrupted. The secure node authentication algorithm utilizes an access control model

that has the ability to distribute the operation between nodes. The secure node authentication algorithm is helpful for authentication between nodes; for example, the sensor nodes can send data between them directly. If the attacks are carried out and act as valid nodes in the network, attackers will be able to steal the data and cause transmission delay or causing network interruption. The details of this algorithm are given in sub-section 4.2.3.1.

The fully homomorphic encryption algorithm which can protect end-to-end data confidentiality will be applied in this protocol. This ability allows more operations to be implemented without increasing the communication overhead. Thus, the proposed protocol can maintain or reduce the energy consumption in the network while implementing the secure node authentication algorithm. The details of this algorithm are given in sub-section 4.2.2.1.

This work proposes a secure and energy-efficient data aggregation protocol to aggregate the data and make the network highly secure from attacks by checking the aggregated data before transmitting it to the server. To support the energy consumption and to prolong the network lifetime, the SEEDA protocol employs cluster network topology involving static and mobile sensor nodes. The hierarchical cluster is built using six types of sensor nodes, namely child node, monitor node, relay node, aggregation node, cluster head, and base station node, as shown in Figure 3.2. The structure of each node is described in section 4.2. The process of each sensor node in the network can be defined as follows:

- **Child Nodes:** The responsibility of the child node is sensing and sending the data to the monitor nodes.
- **Monitor Nodes (M):** Fragment the data into smaller pieces using a data fragmentation algorithm. This is to protect and prevent the attackers from stealing

the data. Then, the monitor node will send the fragmented data to the next neighbor nodes.

- **Relay Nodes (R):** The relay node's responsibility is to group the fragmented data and send the information to the aggregation nodes.
- **Aggregation Nodes (G):** Aggregation nodes, as the name suggests, will aggregate the data using aggregation functions. The authentication is also performed on the new nodes attempting to join the network to verify their legitimacy.
- **Cluster Head Nodes (CH):** The cluster head nodes receive and send aggregated data to the base station node without decrypting the aggregated data. Other than that, the cluster head node sends a broadcast query message to all cluster member nodes to verify their identity and node's information. The broadcast query message includes address identification of cluster head node, cluster member nodes, and the time as well as data information of the member nodes. The query message with all the information is then sent to the base station node to store the information.
- **Base Station Nodes (BS):** The base station nodes receive the aggregated data from the cluster head nodes. The base station node analyzes the aggregated data and checks for fake aggregated data before sending them to the server by checking the authentication process such as a random value, random timestamp, and secret key. Furthermore, the base station node utilizes the distance and timestamp between nodes and checks them with cluster head node information when the new

nodes join the network. The base station is assumed to have substantial energy and memory compared with other nodes.

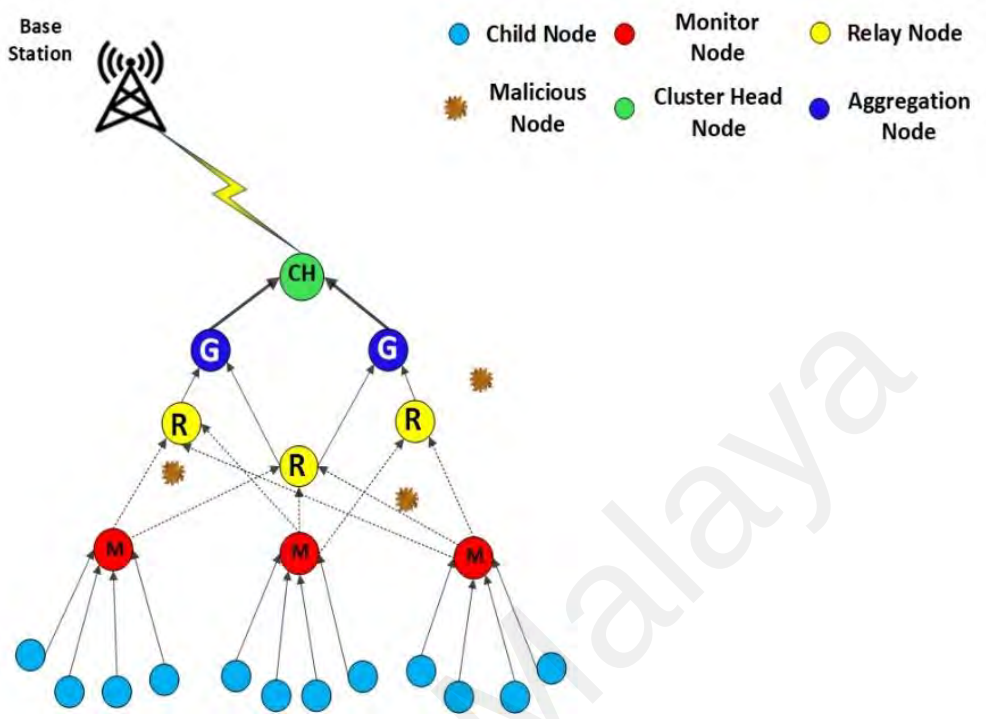


Figure 3. 2: The Network System Model for Secure Data Aggregation Method

3.3.3 Implementation of Proposed SEEDA Protocol

At this sub-stage of our work, the proposed protocol was simulated using the network simulator 3.25 running on the Ubuntu operating system 16.4 LTS version in a computing environment with an Intel(R) Core (TM) i7-4770 processor, 3.40 GHz CPU, and 8 GB RAM running on Microsoft Windows 10 Pro 64-bit operating system. Network Simulator NS-3 was used because to develop a free simulation environment suitable for networking research. In addition, NS-3 is committed to building a solid simulation core that is well documented, easy to use, and debug. Several scenarios of attacks were performed to evaluate the ability of the proposed protocol to detect malicious nodes. Between 8% and 30% of Sybil and Sinkhole attacks in the network were considered during the evaluation. A total of 400 sensor nodes such as child nodes, monitor nodes, relay nodes, aggregation nodes, cluster head nodes, and base station nodes were deployed in the area of 400×400

m² and 1000×1000 m². The quality of the SEEDA protocol results is described in section 4.6, Figures from 4.4 to 4.7.

3.3.4 Verification and Validation of Proposed SEEDA Protocol

In this sub-stage, verification confirms that the proposed SEEDA protocol is correctly converted from pseudo code to functional application. In this section, we verify and validate the implementation of the proposed secure clustering protocol by performing extensive simulation tests. Initially, we verify the functionalities of MAC authentication for secure data aggregation nodes in the cluster, base station verifies the fake aggregated data and the timestamp of all nodes, the distance between nodes to choose the best path, and detected attacks to ensure that it will provide the same expected results using a simulation approach. Different evaluation metrics were used to assess detection rate, energy consumption and accuracy, end-to-end delay, and resilient time in the network. Such an assessment can provide evidence that the proposed scheme satisfies design requirements and provides high malicious node detection rates, preventing and detecting malicious attacks, and reducing energy consumption. Chapter 4 discusses further details regarding implementation, verification, validation, and performance analysis of the SEEDA protocol.

In the previous section, SEEDA protocol was proposed, which is using an access control model to address security and energy issues for WSNs. The proposed SEEDA protocol aims to enhance the authentication between the nodes in the network and detect and prevent the malicious node from joining and accessing the network. The next section will focus on another important energy efficiency issue, which is the hot spots problem were Energy-Efficient Unequal Clustering protocol based on a Balanced energy method (EEUCB) is proposed. The proposed unequal clustering protocol aims to balance energy consumption, increase network lifetime, and solve the hot spots problem.

3.4 Stage 3: Energy-Efficient Unequal Clustering Protocol Based on a Balanced Energy Method (EEUCB)

The third stage proposes an Energy-Efficient Unequal Clustering protocol based on a Balanced energy method (EEUCB). The proposed unequal clustering protocol aims to balance energy consumption, increase network lifetime, and solve the hot spots problem. At this stage, the proposed EEUCB protocol used is divided into four sub-stages: requirements, design, implementation, and validation:

3.4.1 Requirement of Proposed EEUCB Protocol

This sub-stage introduces the requirements of the proposed EEUCB protocol to assess previous techniques and schemes to identify and analyze the problem statement. A comprehensive literature review is focused on discovering the problems of unequal and double clustering protocol. This literature review, presented in chapter 2, discusses the advantages and drawbacks of different existing unequal and double clustering methods (Objective 1). In this chapter, the review of the literature section provides the problem statements of different existing methods. The proposed protocol was simulated using MATLAB 2019b. IEEE 802.15.4/ZigBee presented the network topologies of this research. Several scenarios of unequal clustering were performed to investigate the energy consumption efficiency and network lifetime extension of the proposed protocol.

3.4.2 Design of Proposed EEUCB Protocol

This second sub-stage oversaw an energy-efficient routing protocol with an unequal clustering scheme based on a balanced energy method (EEUCB). This protocol was conducted by modifying a few functionalities of the LEACH, FLECH, EEFUC, UDCH protocols. It is focused on improving energy consumption, network lifetime, and solving the hot spots problem by adopting unequal clustering technology. The size of the cluster depends on the distance between the cluster nodes to the BS. The CH that is closer to the

BS nodes takes more energy to receive data and forward them to the BS. To resolve this issue, the EEUCB protocol is proposed to reduce the size of the CH, thus reducing its overhead. Meanwhile, EEUCB considers the energy of the CH as another metric in cluster size decisions, in addition to the distance. The CHs with more residual energy will form more massive clusters. This can eventually help us to avoid the node dying early when more than CHs share the data relay task in these areas. Besides, the sleep and awake mechanism are utilized in our protocol based on distance range from sensor nodes to CH as well as to the energy level, so it will preserve energy consumption and improve to prolonged lifetime network.

In addition, we propose a double cluster head to reduce the overhead and energy consumption of the CH node. In this method, each cluster has two CHs. The primary CH is responsible for aggregating and forwarding data to the BS node if its distance is greater than the distance threshold, and the energy consumption is less than the energy threshold. The 2CH is responsible for receiving and aggregating data within each cluster and sending them to the primary CH if the distance of the 2CH is less than the distance threshold and the energy is greater than the energy threshold. EEUCB proposes a different election mechanism for both the CH and radius for each node depends on the residual energy of sensor nodes and the distance from the cluster nodes to the BS; however, it does not consider the fact that the minimum distance is the closest distance of a node from the BS, and that the maximum distance is the farthest distance of the node from the BS, which leads to an increase in energy consumption. Furthermore, UDCH proposes a double cluster head in order to reduce the load on the primary CH. The selection process of the primary CH is determined by computing the delay time of each node. The second CH is determined based on the distance from the sensor nodes to the primary CH. However, in UDCH, not much attention is given to enhancing the data transmission process between 2CH. Calculating the delay-time for each CH node improves the election mechanism of

the primary CH. As for the 2CH election mechanism, we consider the highest residual energy and the distance between nodes and the BS. The minimum distance between non-CHs and CHs can reduce the delay, improve energy consumption, and reduce transmission time.

To balance the energy consumption among CMs, CHs, and the BS, a clustering rotation strategy based on the average energy threshold, average distance threshold, and performance of layering by the BS node is proposed; this can increase network lifetime and the efficiency of energy consumption. Two techniques are proposed for transmission: (i) intra-cluster transmission for head clusters to share the data between them; and (ii) the inter clustering transmission proposed when the CH is placed at a great distance from the BS. These two techniques will reduce the overhead and avoid delays in the network. The network system architecture design based on the EEUCB protocol is shown in Figure 3.3.

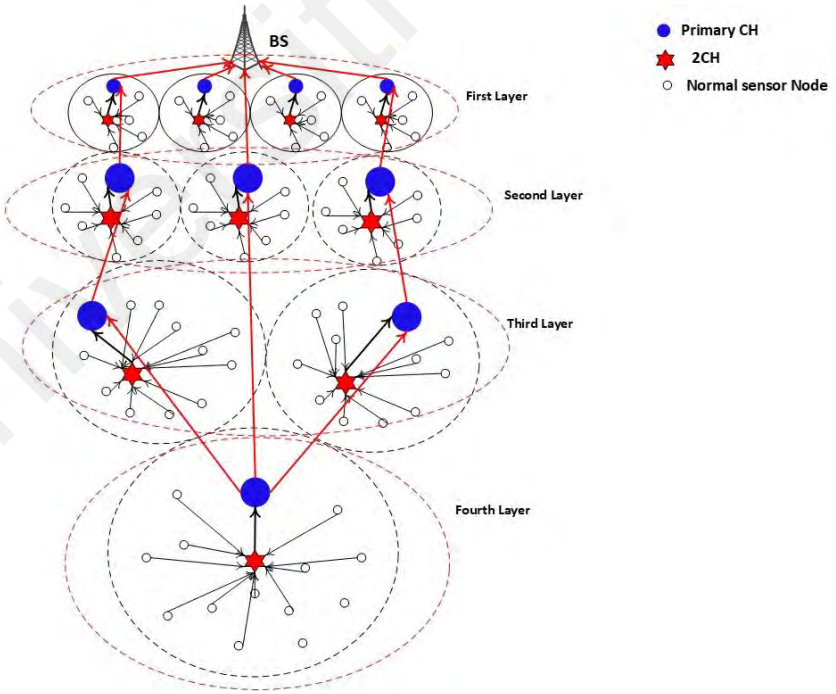


Figure 3. 3: The Network System Design

The EEUCB protocol contains four phases: Processing phase, Initialization phase, Cluster setup phase, and Transmission phase. To better understand these phases and processes in the EEUCB protocol, Figure 3.4 shows the flowchart of the proposed EEUCB protocol.

1. Processing Phase: The processing phase function estimates the distance between cluster nodes and neighbor nodes and checks the number of neighbor nodes.

2. Initialization Phase: The initialization phase calculates the radius of clustering to generate unequal clusters in the network and the election of the primary cluster head by calculating the delay time of nodes.

3. Cluster Setup Phase: The clustering setup phase will elect the nodes to become CH; the election of CH depends on the delay time processing. This section contains three sub-sections: primary CH selection and sleep awake mechanism, cluster formation, and secondary CH selection.

4. Transmission Phase: The process of data transmission between CHs and cluster members through the network. These phases will be further described in chapter 5.

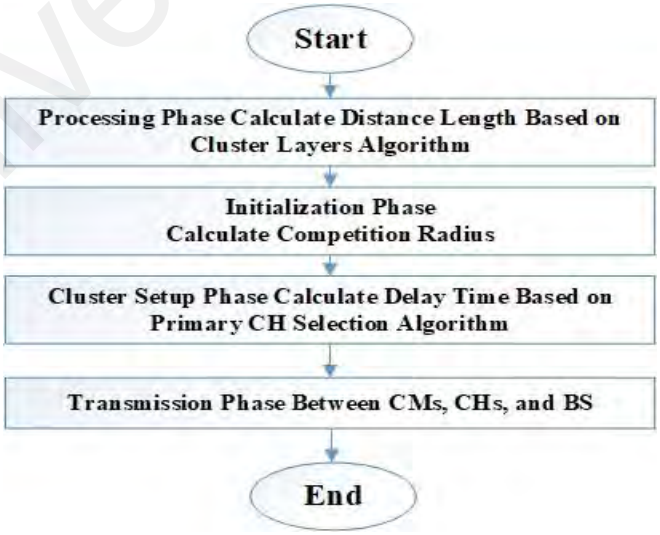


Figure 3. 4: Flowchart of EEUCB Protocol

3.4.3 Implementation of Proposed EEUCB Protocol

At this sub-stage, the proposed protocol was simulated using MATLAB R2019b in a computing environment with an Intel(R) Core (TM) i7-4770 processor, 3.40 GHz CPU, and 8 GB RAM running on Microsoft Windows 10 Pro 64-bit operating system. The network topologies of this research were presented by IEEE 802.15.4/ZigBee. MATLAB was used because the author can efficiently use this tool. However, this tool requires writing many small functions. MATLAB is a high-level programming language that provides an interactive environment for algorithm development, data analysis, and plotting capabilities. There were four different scenarios, such as a different number of nodes and area sizes. All the examined scenarios show similar results regardless of node numbers and network area sizes. A total of 1000 sensor nodes were deployed in the area of 200×200 m², 300×300 m², 400×400 m², and 1000×1000 m². The quality of the results of the EEUCB protocol is described in section 5.6, Figures from 5.8 to Figure 5.12.

3.4.4 Verification and Validation of Proposed EEUCB Protocol

In this sub-stage, verification confirms that the proposed EEUCB protocol is correctly converted from pseudo code to functional application. This section verifies and validates the implementation of the proposed unequal and double clustering method by performing extensive simulation tests. We verify the functionalities of delay time, sleep-awake mechanism, and network layer for balanced energy consumption, distance to the base station, cluster head, and representative nodes election to ensure that it will provide the same expected results using a simulation approach. Different evaluation metrics were used to assess network lifetime, average energy consumption, average residual energy, end-to-end delay, and throughput. Such an assessment can provide evidence that the proposed scheme satisfies design requirements and reduces energy consumption, the load on the primary cluster head, and prolongs the network lifetime. Chapter 5 discusses

further details regarding implementation, verification, validation, and performance analysis of EEUCB protocol.

3.5 Stage 4: Performance Evaluation Metrics

By carrying out intense simulations, the performance of the proposed schemes was evaluated. The following performance evaluation metrics were considered and used for benchmarking with previous schemes to assess the performance of the proposed schemes:

- **Detection Rate:** This metric is defined as checking the false data inserted by malicious nodes into aggregated data.
- **Resilience Time:** This metric is defined as the time offsets are identified when the malicious nodes are joining the network. The malicious time offsets will be excluded, while the rest of the time offsets are used to estimate the actual time offsets.
- **Network Lifetime:** The time during which the network is able to carry out its desired functions is given by network lifetime. It is the network time before the death of the first node, which is also termed as nodal lifetime. It can also be defined as the time till a specific node portion die.
- **Energy Consumption:** This metric gives information about the total energy consumed in sensing, processing, and transmitting the data packets that the source generates (measured in Joules per data processed). Average energy consumption is given by determining the ratio of energy consumed by all nodes to the total number of network nodes.
- **End-To-End Delay:** is defined as the time taken when the packets transfer from the sensor nodes to the base station node.
- **Throughput:** is the number of data packets successfully transmitted to the destination in a period of time.

The details and description of each metric are presented in sections 4.3 and 5.4.

3.6 Chapter Summary

This chapter provides an organized description of the adopted methodology. The first section started with a brief introduction. This was followed by a discussion about the stages of the methodology from the review of the literature and identification of the problem statement to the pre-analysis stage of the conventional routing schemes (SDA, SDAT, SDALFA, EESSDA, SDAACA, EESDA, LEACH, FLECH, EEFUC, and UDCH) then to the design and implementation of the proposed (SEEDA and EEUCB) protocols. The research methodology flowchart showed the sequence of the research stages and presented information on the connections between every stage's structural component. Finally, the performance evaluation metrics that were considered to study the performance of the proposed schemes were defined. The next chapter will present the proposed SEEDA protocol with its performance evaluation metric and experimental results.

CHAPTER 4: DEVELOPMENT OF SECURE AND ENERGY-EFFICIENT DATA AGGREGATION METHOD IN CLUSTERING BASED ON ACCESS CONTROL MODEL (SEEDA)

4.1 Introduction

In this chapter, a secure and energy-efficient data aggregation method in clustering based on an access control model is presented. This scheme is referred to as secure data aggregation in clustering for WSN (hereinafter referred to as SEEDA). SEEDA protocol aims to enhance authentication by generating a random value and random timestamp with a secret key. The base station node will verify the fake aggregated data when the packets are received using the generated key earlier. Furthermore, the attacks are detected and prevented by utilizing secure node authentication, data fragmentation algorithms, fully homomorphic encryption, and an access control model, as shown in Figure 4.1. The performance of the proposed protocol is compared with SDA, SDAT, SDALFA, EESSDA, SDAACA, and EESDA, which is a widely used protocol in the area of secure data aggregation.

In the previous chapter, the theoretical concepts and utilized system models were presented. In this chapter, the design architecture of SEEDA, its verification, and the simulation results of its performance evaluation are presented and discussed.

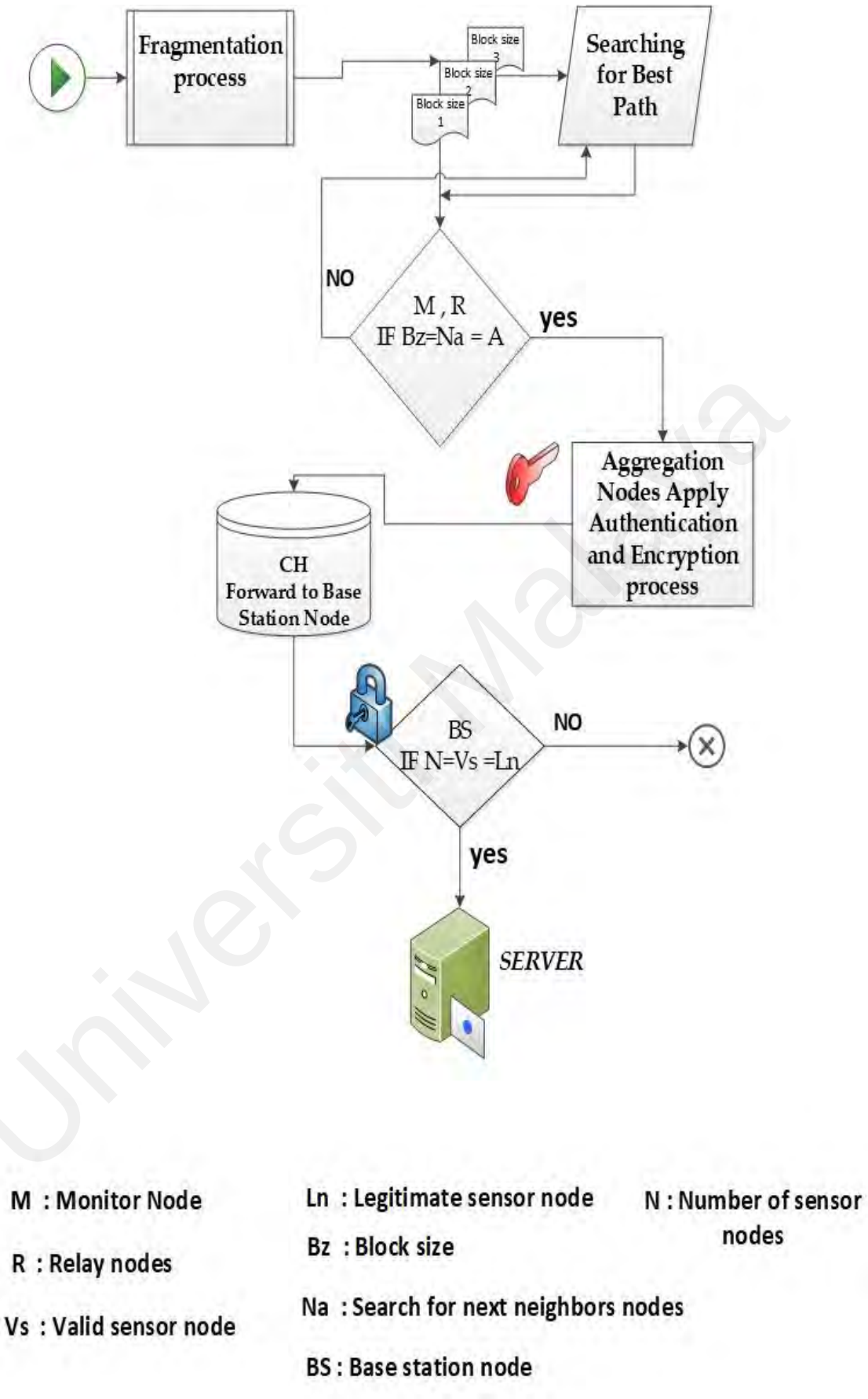


Figure 4. 1: The Access Control Model

4.2 Development of Proposed SEEDA Protocol

This section introduces the development of proposed secure and energy-efficient data aggregation in clustering to aggregate the data and make the network highly secure from attacks by checking the aggregated data before transmitting it to the server. The previous chapter defined the hierarchical cluster of SEEDA protocol that included the six types of sensor nodes, namely, child node, monitor node, relay node, aggregation node, cluster head, and base station node built in the network. The format of each type of node is described in **Algorithm 1**. Each node has its own operation and data transmission procedure. They are built to preserve and reduce the energy of nodes, lessen communication overhead, and preventing transmission redundancy in the network.

Algorithm 1. Formatting Type of Nodes
1. hierarchical clustering (child, M, R, A, CH, BS)
2. **For** each cluster member in access control **do**
3. child nodes start to send data to the next neighbor nodes
4. **If** M&R nodes receive data **then** fragment data into block size
5. while there are more than one cluster members
6. find the nodes closest distance with M nodes
7. **If** the C1 closest to C2 **then** select and called relay nodes (R)
8. C1== C2 and M == R nodes **Else**
9. **repeat** to step 6
10. **End if**
11. **End if**
12. **If** the $E \geq 11.2 \leq 14$ J **then** select aggregation nodes (A)

13. R== A nodes and A receive all data information
14. The A nodes calculate MAC
15. **If** the E ≥ 10 ≤ 16 J **then** select cluster head nodes (CH)
16. CH== BS
17. **End if**
18. **End if**
19. End for

In this work, the SEEDA protocol uses different types of nodes. To determine and understand the node's type, each node in the network is assumed to have different energy and bandwidth, depending on the deployment location of the node. The energy and bandwidth of different type of nodes are as follows; the relay nodes are set to 11 joules energy and 55 kbps bandwidth, the cluster head nodes are set to 15 joules energy and 80 kbps bandwidth, the aggregation nodes are set to 14 joules energy and 95 kbps bandwidth, and the monitor nodes are set to 7 joules and 45 kbps bandwidth. The following describes the function of each node.

(a) Child Nodes

Sense and send the data to the monitor nodes.

(b) Monitor Nodes (M) and Relay Nodes (R)

The monitor nodes and relay nodes fragment the data into smaller pieces using a data fragmentation algorithm as described in **Algorithm 2**. The purpose of data fragmentation is to protect and prevent attackers from stealing the data. Consequently, the monitor node

will send the fragmented data to the relay nodes, and the relay nodes will group the fragmented data and send the information data to the next neighbor nodes.

Algorithm 2. Process of Monitor Nodes & Relay Nodes
1. If M collect the data from child nodes **then**
2. The M fragment the data block $'\boldsymbol{D_b}'$, into block size $(\mathbf{D_b} = \mathbf{B_{z1}}, \mathbf{B_{z2}}, \mathbf{B_{z3}} \ldots \boldsymbol{B_{zn}})$
3. The monitor and relay nodes search for the best path next to neighbor nodes
4. If the $'\boldsymbol{B'_z}, =$ M, '$\boldsymbol{R}$'= $'\boldsymbol{N'_a}$, **then**
5. The monitor and relay nodes sending fragment data to the next neighbor nodes
6. End if
7. End if

(c) Aggregation Nodes (G) and Cluster Head Nodes (CH)

Aggregation nodes, as the name suggests, will aggregate the data using aggregation functions. Firstly, the authentication is performed on the new nodes attempting to join the network to verify their legitimacy as described in **Algorithm 3**. After that, the aggregation nodes will perform encryption processes and send the encrypted data to the cluster head nodes. The cluster head nodes receive and send aggregated data to the base station node without decrypting the aggregated data. Other than that, the cluster head node sends a broadcast query message to all cluster member nodes to verify their identity and node's information. The broadcast query message includes address identification of cluster head node, cluster member nodes, and the time as well as data information of the member

nodes. The query message with all the information is then sent to the base station node to store the information.

Algorithm 3. Process of Aggregation (G) & Cluster head Nodes (CH)
1. The CH sends a broadcast query message to all cluster members
2. The cluster members receive the query message and send the nodes information to cluster head nodes
3. If the CH = M_q **then**
4. CH send and store all information to BS
5. The aggregation nodes calculate MAC
6. The aggregation nodes encrypt the aggregated data before sending it to CH nodes
7. The aggregation nodes send the encryption data to CH
8. End if

(d) Base Station Nodes (BS)

The base station nodes will receive the aggregated data from the cluster head nodes. The base station node analyzes the aggregated data and checks for the fake aggregated data before sending them to the server by checking the authentication process such as a random value, random timestamp, and secret key, as described in **Algorithm 4**. Furthermore, the base station node utilizes the distance and timestamp between nodes and checks them with cluster head node information when the new nodes join the network. The base station is assumed to have substantial energy and memory compared to other nodes.

Algorithm 4. Process of Base Station Nodes
1. The CH forward aggregated data to BS
2. BS decrypt the aggregated data
3. BS investigate the data information from cluster members and check the broadcast message with cluster head nodes BS= Cer (N_{id}, D_P, T,D_C) A = (B_s,M_i,D_P, T,D_C,N_{id})
4. If N== V_s = L_n(N) **then**
5. BS approval N == V_s $\in S_n$ **Else**
6. $B_s \neq$ N and not approval N
7. End if

4.2.1 Data Fragmentation Algorithm

The fragmentation algorithm is used to hide and preserve the original data from being tampered with by the malicious nodes. The fragmented blocks information is shown in Figure 4.2., where it contains SEEDA protocol version, type of service, block size, fragmentation data, address of the source node (i.e., monitor node), address of the destination node (i.e., relay node), and the data packet size, which enable them to be reconstructed back.

The malicious node needs to group all the blocks generated from the data fragmentation algorithm in order to intercept the original message. Since each block produced by the fragmentation algorithm has a privacy protection component acquired from the cluster head node and aggregation nodes, they are not easily tampered with. The monitor node and relay node will then search for the best path based on the distance calculation and distribute the data to the relay node. Following the data distribution, the relay node will reconstruct the fragmented data and send them to the aggregation node.

The data fragmentation process is also described in **Algorithm 5**. The input and output parameters ($\mathbf{M}, \boldsymbol{D_b}, \mathbf{S}, \mathbf{U}$) of the algorithm is specified in 1–2. In Step 3, the monitor node 'M' collects the data block $'\boldsymbol{D_b}'$, from the child nodes and checks the number of block data. The monitor node fragments the data block into small size blocks $'\boldsymbol{B'_z}$, ($\mathbf{D_b} = \mathbf{B_{z1}}, \mathbf{B_{z2}}, \mathbf{B_{z3}} \ldots \boldsymbol{B_{zn}}$) in Step 4-5. In Steps 6, due to the random deployment of sensor nodes in the network, the monitor node searches for the next neighbor nodes $'\boldsymbol{N'_a}$, with the nearest distance to send the fragmented data to the relay nodes'$\boldsymbol{R}$.' This process decreases energy consumption. When the monitor node finds the neighbor nodes, the monitor node keeps sending the data until all the data are forwarded to the relay node $\boldsymbol{M} = \boldsymbol{N_a} \rightarrow \boldsymbol{B_z to}$ **R**, as described in Steps 7–8. In Step 9, all the fragmented data are received in relay nodes $\mathbf{B_z} = \mathbf{R}$, then the relay nodes send the fragmented data to the aggregation node'$\boldsymbol{D'_a}$, for the aggregation process S ^{v} U. After the aggregation and encryption process is completed, the aggregation node forwards the encrypted data to the cluster head node, 'CH,' as described in Step 10. Finally, in Steps 11–12, the base station receives the encrypted aggregation data from the cluster head node CH $\Vdash B_s$ using the algorithm defined in (X. Li et al., 2015).

Algorithm 5. **The Data Fragmentation**
M: monitor node; D_b: data block; B_z: block size; N_f: number of fragment block size; data; N_a : search for next neighbors hope; B_s: base station; CH: cluster head node; R: relay nodes; Da: data aggregation; S: set functions for aggregation; U: authentications.
1. Input (M,R, D_b,S,U)
2. Output (B_s,N_f)
If the M collect the data from child nodes; **then**
3. The M, R, fragment data D_b into small size blocks B_z
4. ($\mathrm{D_b} = \mathrm{B_{z1}}, \mathrm{B_{z2}}, \mathrm{B_{z3}} \ldots B_{zn}$)
5. The M search the N_a to send the B_z

6.	**If** the M find the N_a; and M ∈ N_a **then**
7.	The $M = N_a \rightarrow B_z to$ R
8.	$N_a \rightarrow B_z to$ R forward process, and
9.	N_a *repeats to steps* $6-8$ until
10.	$\boldsymbol{D}_a$ *received* D_b
11.	The D_a collect all data D_b, N_f
12.	The D_a gather the data fragment by applying a process S ᵛ U
13.	CH collect the aggregated data from D_a
14.	The CH forwards the encrypted data to base station nodes
15.	CH ‖ B_s
16.	**End if**
17. End if	

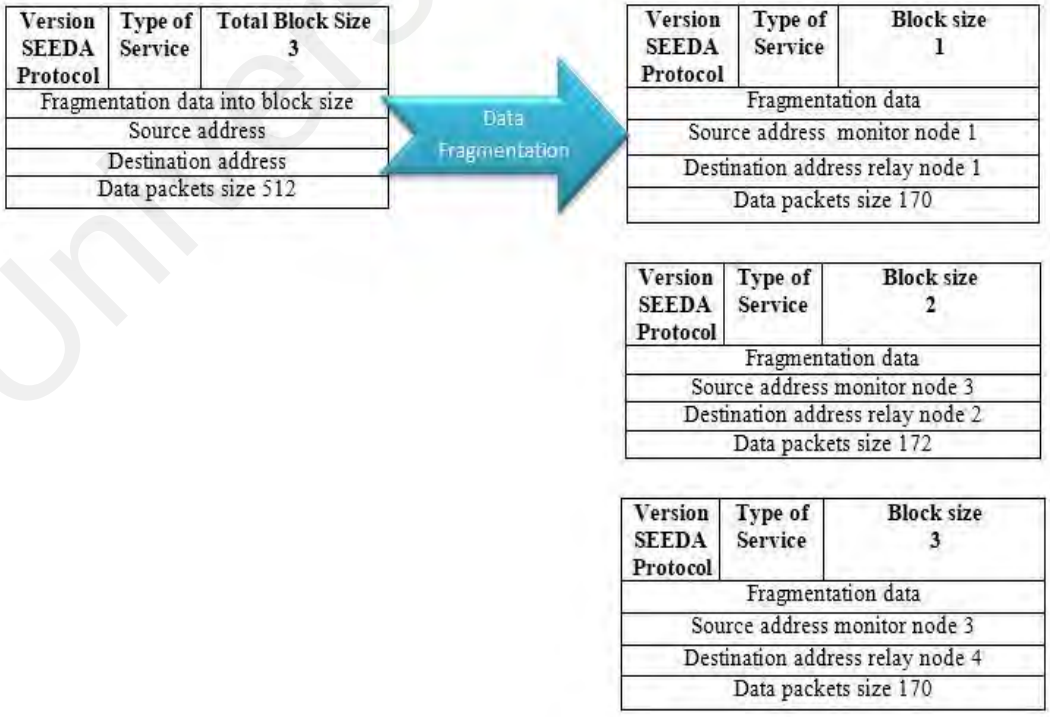


Figure 4. 2: The Data Fragmentation Process

4.2.2 Secret Key

In this section, discuss the secret key with the SEEDA protocol. The SEEDA protocol is using a fully homomorphic encryption algorithm that can protect end-to-end data confidentiality as defined in (X. Li et al., 2015) and perform encryption on the aggregated data before sending it to the base station nodes. This scheme just uses addition and multiplication over the integers and its concept is simple. The generation of the secret key with the SEEDA protocol follows the same process as presented in (X. Li et al., 2015). For the key generation phase, the BS generates the secret key and public key as follow:

$$\boldsymbol{K_e = S \quad and \quad P_k = (P_k, y)} \tag{4.1}$$

Where K_e is the secret key, S is secret key was replaced by the vector of the subset S, P_k is the public key and y is the set of the public key. Then base station informs the aggregation nodes of the public key for them to encrypt the data information which is sensed by the aggregation nodes.

SEEDA is employing the encryption process when the data information gets to the aggregation node because the monitor nodes and relay nodes are focusing on the fragmentation process to hide the original data from the adversary. By doing so, the encryption process from monitor nodes is not needed. This action will reduce the transmission delay. The aggregation nodes will aggregate the received data from relay nodes and check the authentication process before sending it to cluster head nodes. When the aggregation nodes completed the authentication process, the aggregation nodes will encrypt the aggregated data using an encryption algorithm to make the data highly secure and preventing attackers from stealing the data in the network. At the same time, the aggregation nodes send the encryption data to cluster head nodes, and the cluster head nodes will then forward it to the base station nodes without decrypting the aggregated data.

4.2.3 Access Control Model

There are three important modules involved in the access control model: authentication and authorization, medium access control and data integrity, and authentication and redundancy. The proposed secure node authentication algorithm detects and prevents malicious attacks from accessing the network by checking the new nodes' secure authentication. The medium access control generates and calculates the MAC and data aggregation functions. The authentication process describes the authentication procedure and checks the distance between nodes.

Our method aims to enhance the authentication between nodes and to decrease redundancy. This is because many researchers focused on the security and aggregation process without addressing the authentication and authorization issues between nodes. The details of the three important modules are described in the following subsections, which discuss the authentication and authorization process, MAC and data integrity detection, and authentication and redundancy.

4.2.3.1 Authentication and Authorization Process

Prior works on authentication lack of focus on authentication with authorization method. Authentication is the process of verifying the legitimacy of the new nodes that join the network. This process is performed at the base station. The aim of this process is to prevent the adversary nodes from joining the network and act as original nodes to collect data from the network. The authorization is the process that allows only authorized users to read and transmit the data. The implementation of both authentication and authorization processes in the network is important because if the malicious nodes have successfully joined the network, the authorization process can prevent these nodes from accessing the data in the network. Therefore, this work includes both secure node

authentication and authorization algorithms in the proposed SEEDA protocol, as described in **Algorithm 6**.

Steps 1-2 presents the input and output parameters of the algorithm. In Steps 3-7, the cluster head node sends the broadcast query message to the cluster sensor nodes. The sensor nodes that received the query message from the cluster head node computes its node identity and information message $M_q = (CH_{id} \| S_{n\,id} \| M_i \| T)$. In steps 8-12, the aggregation node informs the base station node to verify the aggregated data from the new sensor node by checking the node ID, message, timestamp, and distance (N_{id}, D_P, T,D_C).

The base station checks the query message in the cluster head node, as described in steps 13-14. The authorization process is performed if the distance and certificates for the new sensor node are similar to the original nodes in which the new node is considered valid and authorized. After completing the authorization process, the base station authorizes the new node to join the network and allows the data to be sent between nodes, as described in Steps 15-17. Conversely, if the new sensor node is malicious and unauthorized, the base station will reject the new node from joining the network. This process is described in Steps 18-19.

Algorithm 6. **Secure Node Authentication**
B_s: base station; N: number of the sensor node; CH: cluster head nodes; Cer: certificate of the sensor node; M_i: information message; M_q: query message; A: authorization; A_p: approval; S_n: sensor the network; D_P: data packets; D_C: distance between nodes; T: broadcast the time of nodes; V_s: valid sensor node; L_n: legitimate sensor node; L_i: illegitimate sensor node; En: entry the network
1. Input (M_i,N_{id}, N, T,D_C)
2. Output (A, Cer,)
3. **For** each member sensor nodes in access control model **do**

4. CH sends a query message to it all member sensor nodes (M_q)
5. after receiving (M_q) the sensor nodes compute the **6.** $M_q = (CH_{id} \| S_{n\,id} \| M_i \| T)$ **7.** CH $\rightarrow M_q$
8. **If** the new sensor node join the network
9. N$\rightarrow$En $\in S_n$ **then** CH inform B_s
10. B_s recall N
11. B_s investigate Cer (N_{id}, D_P, T,D_C) for the V_s **12.** set A for N; A = (B_s,M_i,D_P, T,D_C,N_{id})
13. B_s check broadcast A(M_q) with CH
14. **If** N== V_s = L_n(N) and N = = M_q **then**
15. B_s approval N == V_s $\in S_n$
16. **Else**
17. N $\neq$ V_s and N $\neq$ M_q L_i= V_s **then**
18. $B_s \neq$ N and not approval N
19. **End if**
20. **End if**
21. End for

4.2.3.2 Medium Access Control and Data Integrity

This section explains the access control model and the procedure of the base station to check the authentication process and to secure the data aggregation. The base station utilizes the distance and timestamp to examine the authenticity of all nodes in the network. We assume node N acts as a malicious node that creates fake sub-aggregate data for authentication. First, the malicious node N creates false data with a random value. Then,

node N sends the medium access control (MAC), which contains the random value and personal identity of node N to the base station for authentication. When the base station receives the MAC, the base station will verify the legitimacy of node N by checking the node identity, data packets, distance between nodes, and their timestamp. The N_{MAC} is calculated using equation 4.2, given as below:

$$\boldsymbol{N_{MAC} = \int_{k+1}^{n}\{k(R_v) + n_i} \qquad (4.2)$$

Where, k is a set of sensor nodes, R_v is the random value, $n_i$is the node ID. In Equation 4.3, the base station creates the random value and random timestamp to authorize the node. A random value, R_v is an arbitrary number that is used to avoid malicious attacks due to duplication. A random timestamp, R_t, is a timestamp encoded with a random number.

The malicious node needs to know the time it takes for a specific node to transfer data to the base station and their random number in order to masquerade an attack. The base station with a random value and random timestamp can be generated as follows:

$$\boldsymbol{B_{s(v,t)} = \sum_{j+1}^{j} R_t + \sum_{i+1}^{i}(R_v 1)^{gv} \times n(s)} \qquad (4.3)$$

Where, $B_{s(v,t)}$ are the base station with a random value and random timestamp, R_t is the random timestamp, gv is the random value generated by a malicious node, n(s) is the number of the sensor node, i, j are the set of a random value and random timestamp. The data aggregate can be computed as follows:

$$\boldsymbol{DA = \sum_{n=1}^{n}(B_s) * N_{MAC}} \qquad (4.4)$$

Where, DA is the data aggregation, n is the set of sensor nodes.

Lemma: The malicious sensor nodes unable to create MAC with fake data that is similar to the original data recorded at that base station.

Proof: Let's assume node N can create the random value with false data and send it to the base station for authentication

$$\boldsymbol{N = (n_{id}, R_v, b1 \ldots bn)} \qquad (4.5)$$

To improve the authentication and allow the base station to determine the fake data, we not only create the random value with aggregated data, but we also create a random timestamp and secret key. The following equation shows the medium access control with the aforementioned security measures.

$$\boldsymbol{N_{MAC} = (k_e + R_v + R_t + D_p + b1..bn)} \qquad (4.6)$$

$$\boldsymbol{N_{MAC} = N}$$

Where, the N_{MAC} is the medium access control, k_e *is the secret key*, D_p is the data packets, R_t is the number of random timestamps, R_v is the number of a random value, b1...bn is the number of bit data.

The proposed SEEDA protocol design the secret key by using fully homomorphic encryption to make the network highly secured. An encryption process uses aggregation and base station nodes to preserve energy. Our protocol distributes data to all nodes to enable the valid nodes to share the data packets between them in the network. The data packet format is as shown in Figure 4.3. This messaging security is very helpful in preventing attacks from accessing the network. The malicious nodes cannot create similar messages, such as the time and the secret key. Apart from that, the base station also holds the distance and ID between nodes from the cluster head nodes. For this reason, the malicious node will not be able to join the network and share its data.

Data Packets				
Fragment data	Random value	Random timestamp	Information data packets	Number of data packets
Distance between nodes	Secret key	MAC Authentication		Data packets size
Source address				
Best path				
Destination address				

Figure 4. 3: The Data Packets Format

4.2.3.3 Authentication and Redundancy

The authentication process makes it challenging for an attacker to join the network. This authentication method is expected to enhance the network's security since the network's design and key encryption only allow authorized users to transmit the data. This work assumes the aggregation node (G) sends the secure aggregated data to the cluster head nodes (CH). This operation can be described as:

$$\boldsymbol{p_e = \left\{ n_{id,} CH_{id,} N_{MAC_{k(n,CH),}} D_{p,} D_c \right\}} \qquad (4.7)$$

Where, p_e is the packet encryption, $MAC_{k(n,CH)}$ is the key to message authentication code for cluster member nodes and cluster head nodes, D_P is the data packets, D_c is the distance between nodes. The cluster head nodes receive the packet encryption from aggregation nodes, and then the cluster head node forwards the packet encryption to the neighbors of cluster head nodes or to the base station. The process of cluster head node (CH) forwarding the data to the base station nodes can be written as:

$$\boldsymbol{p_{e1} = \left\{ CH_{id,} NCH_{id,} N_{MAC(CH,NCH),} D_{p,} D_c \right\}} \qquad (4.8)$$

Where, NCH_{id} is the identity of the next hop cluster head node, $MAC_{k(CH,NCH)}$ is the group of encryption key transmission between cluster member nodes or cluster member nodes with the base station.

SEEDA protocol proposes these equations to enhance the authentication and integrity of the message encryption when the data are sent to the nodes to reach the base station node. Finally, substituting equation (4.7 -4.8) into equation (4.9): where, T_s is the total data encryption sent through the network.

$$\boldsymbol{T_S} = [\{\boldsymbol{N_{id,}CH_{id,}N_{MAC(n,CH),}D_PD_c}\} + \{\boldsymbol{CH_{id,}NCH_{id,}N_{MAC(CH,NCH),}D_{P,}D_c}\}] \quad (4.9)$$

The base station node calculates the distance between nodes to determine the best path to the next nodes for data transmission and to check for a node that joins the network. This operation helps to avoid the redundancy of data transmission in the network because the nodes will send the data in a short time and will not generate traffic control through the sending process. The distance between the nodes can be calculated as:

$$\mathbf{D_c = N \times \frac{T_r}{S} + C_p} \quad (4.10)$$

where, D_C is the distance between nodes, N is the number of nodes in the network, T_r is the transmission range, C_p is the ID number of the clustering node, **S** is the propagation speed of the signal. Due to the nodes' random deployment, equation (4.10) checks the distance between nodes that helps to choose the best path for data transmission to next-hop nodes.

Consequently, it also reduces the delay and data redundancy in the network.

4.3 Evaluation Metrics

The lemma and proof of the proposed method have been presented in the previous section. To further test the reliability of the proposed method, four evaluation metrics are considered. These metrics measure the network's security and performance, namely, the detection rate of the malicious nodes, energy consumption and accuracy, end-to-end

delay, and resilient time in the network. These metrics are explained in the following sub-section.

i. Detection Rate

The performance of the proposed SEEDA protocol is evaluated by simulating Sybil and sinkhole attacks. The malicious nodes are detected by checking the false data inserted into the aggregated data. The detection rate of malicious attacks can be written as equation 4.11 (Wazid, Das, Kumari, & Khan, 2016):

$$\boldsymbol{D_m = \frac{A_d}{A_d + F_d}} \qquad (4.11)$$

Where, D_m is the detection rate of fake aggregated data, A_d is the number of aggregated data and, F_d is the number of false aggregated data. The number of false aggregated data depends on how many malicious nodes in the network.

ii. Energy Consumption and Accuracy

The efficient management of energy consumption in the network is very important for secure data aggregation. One of our proposed protocol goals is to reduce or maintain the energy even when a malicious attack occurs in the network. It is also designed to prolong the network lifetime by reducing the communication overhead.

Let's assume the clusters sensor nodes P_{d1} and aggregation nodes P_{d2} send the data packets and messages between them. The equation can be written as:

$$\boldsymbol{C_1 = (1 - P_{d1}) * S_n * P_{d2}} \qquad (4.12)$$

$$\boldsymbol{C_2 = (1 - P_{d1}) * S_n * \sum_{S=0}^{Pd2} S * (D_a - 1)} \qquad (4.13)$$

where, The C_1 is the communication overhead for node 1, C_2 is the communication overhead for node 2, P_{d1}, P_{d2} are the data packets sent between the nodes, S_n is the number of sensor nodes, D_a is the aggregated data. On the other hand, the total communication overhead C_t, of the exchanged message can be written as:

$$\boldsymbol{C_t = C_1 + C_2} \qquad (4.14)$$

Where, C_t is the total communication overhead. The energy consumption can be evaluated and computed as:

$$\boldsymbol{E_C = C * V_S} \qquad (4.15)$$

Where, E_c is the energy consumption, C is the initial energy to send data, V_S is the average of send bit data per second. The wasted energy when node N transmits the packets to the next node N_1 can be calculated as:

$$\boldsymbol{W_E(N, N_1, D, P_Z) = (C1 + C2 * D(N, N_1) * V_S * P_Z} \quad (4.16)$$

Where, W_E is the wasted energy, $D(N, N_1)$ is the distance between (N, N_1) nodes, P_Z is the packet size, C1 is the cluster node 1, C2 is the cluster node 2. The energy when packets are received between nodes can be written as:

$$\mathbf{E}= \boldsymbol{(N, N_1, P_Z) = (V_S * P_Z)} \qquad (4.17)$$

iii. End-to-End Delay

The end-to-end delay is defined as the time difference between the time when the packet aggregation occurs and the time when the packet arrives at the aggregate queue. The end-to-end delay can be computed as:

$$\mathbf{D} = \frac{\sum_{i+1}^{P}(T_{rec\ i} - T_{send\ i})}{P} \qquad (4.18)$$

Where, D is the end-to-end delay, T_{rec} is the time when packets are received, T_{send} is the time when sending packets, and P is the total number of packets.

iv. Resilient Time in The Network

After collecting a set of time offsets from multiple nodes, the malicious time offsets from Sybil and sinkhole attacks are identified. These malicious time offsets will be excluded, and the rest of the time offsets are used to estimate the actual time offset. The resilience time shows the performance of the network after the network is compromised. The resilience time can be written as:

$$\boldsymbol{R_S = T_i - T_j} \qquad (4.19)$$

Where, R_S is the resilience time, T_i is the set of time offsets from nodes, T_j is the set of time offsets under malicious nodes.

4.4 Complexity Analyses of the Proposed SEEDA Protocol

In this section, suppose that they are N total number of nodes in the network, and C the constant of complexity. In our protocol, we have one loop in algorithm and iteration, so the computational complexity is O(N). The complexity is acceptable for a large network with a large number of sensor nodes in the network. In addition, the SEEDA protocol has found the optimal distance and best path to reduce energy consumption, and significant improvement over the many O(N) algorithms and protocols in the literature. The following describes the communication control message:

a. Calculate the cluster member's access control in the network, these are called C0.
b. The CH sends a broadcast message to all cluster members N.

c. If the new node joins the network, the base station will check the authentication in the network, these can be described as C1+C2+C3.

The overall control message can write as:

C0= N * (C1+C2+C3)

= C0+C1 C2 C3N

= O(N)

4.5 Experimental Setup

The proposed SEEDA protocol used an access control model to secure the data aggregation in clustering and to conserve energy efficiently. The SEEDA protocol design the model and simulation as in SDAACA. The proposed SEEDA protocol was simulated using the network simulator 3.25 running on Ubuntu operating system 16.4 LTS version. Several scenarios of attacks were performed to evaluate the ability of the proposed protocol to detect malicious nodes. 8% to 30% of Sybil and sinkhole attacks in the network were considered during the evaluation.

A total of 400 sensor nodes, such as child nodes, monitor nodes, relay nodes, aggregation nodes, cluster head nodes, and base station nodes, were deployed in the area of 400×400 m² and 1000×1000 m². The parameters used in the simulation are presented in Table 4.1. Various values of energies and bandwidths for the sensor nodes were tested, depending on the deployment location of the node. For example, the relay nodes are set to 11 joules energy and 55kbps bandwidth; the cluster head nodes are set to 15 joules energy and 80kbps bandwidth, the aggregation nodes are set to 14 joules energy and 95kbps bandwidth, the monitor nodes are set to 7 joules and 45kbps bandwidth. The

standard energy for the sensor network is 3.5 joule. The sensor nodes require the pause time in some situations. We set a 23 second pause time with a 22-minute simulation time.

Table 4. 1: Simulation Parameters

Parameters in SEEDA	Value
Transmission range	50 meters
Initial energy of relay node, monitor node, aggregation node, cluster head node	7,11,14,15 joules
Simulation time	22 minutes
Network size	400*400 m^2,1000*1000 m^2,
Number of sensor nodes	400
Bandwidth for relay, monitor, cluster head, aggregation nodes	45,55,80,95 kb/sec
Buffering capacity	50 packets at each node
Data packet size	512 bytes
Initial pause time	16 seconds
Power intensity	-14dbm to 13dbm

4.6 Simulation Results

The simulation results show that the performance of the proposed protocol successfully keeps the network highly secure. This is because the proposed protocol enhances the authentication by generating a random value and random timestamp with a secret key, which makes it difficult for the adversary to replicate. Thus, prevent an unauthorized node from joining the network. The sensor nodes are also protected by fragmenting the data into small pieces before transmitting it to the next-hop nodes. The base station node verifies the fake aggregated data and checks the certificate nodes.

Furthermore, the base station node utilizes the distance between nodes and timestamps to secure the network. Apart from security, the distance information is used to speed up time by choosing the optimal next-hop nodes.

The performance comparison between the proposed protocol and the other six protocols in securing data aggregation is presented in Figure 4.4 to Figure 4.7. Table 4.3 presents the different scenarios generated, such as a different number of nodes and different area sizes. All the examined scenarios show similar results when the nodes and area sizes of the network are smaller or larger.

In Table 4.2, it can be seen that the detection rate of the proposed SEEDA protocol is approximately 98.84%, with the presence of 24% malicious nodes in the network, as shown in Figure 4.4. This indicates that the proposed protocol is only slightly affected by the increment of the malicious nodes in the network. The detection rate of other prior protocols ranges between 88.02–96.6%, with 24% of malicious nodes in the network. The proposed protocol uses the authentication process in all nodes to validate the new nodes and only allows the authorized nodes to join the network.

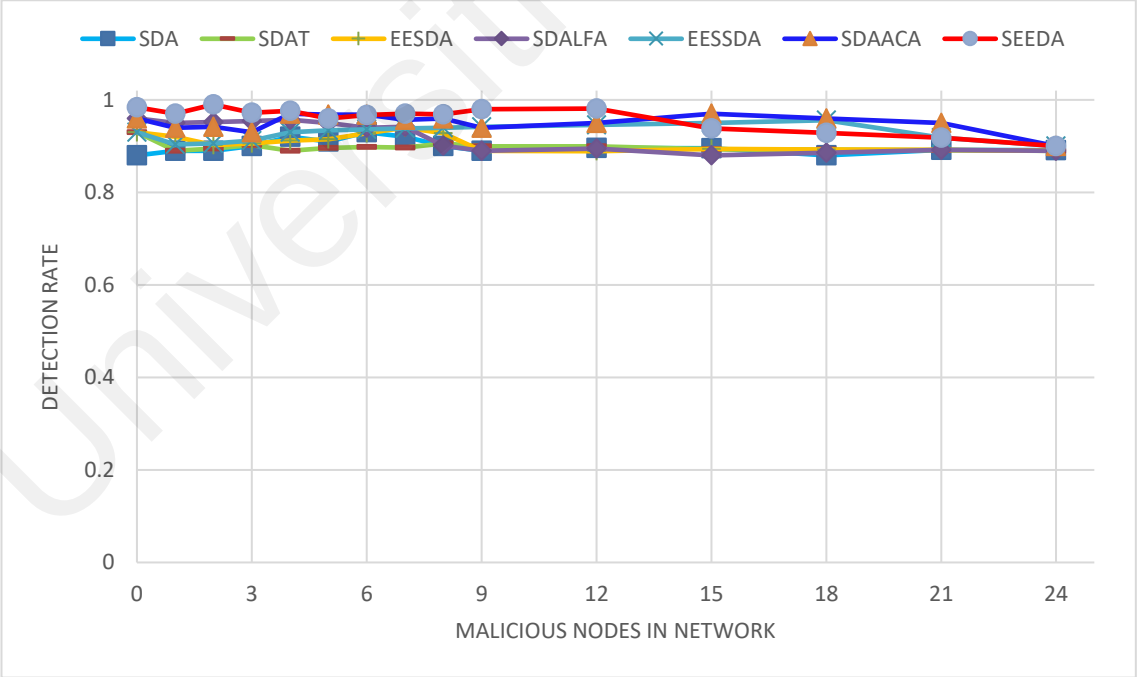


Figure 4. 4: The Detection Rate with Malicious Nodes From 0% to 24%.

Table 4. 2: Malicious Node Detection Rate % Comparison with Different Protocols

Malicious Node	SDA	SDAT	EESDA	SDALFA	EESSDA	SDAACA	SEEDA
0	0.88	0.93	0.93	0.96	0.93	0.96	0.984
1	0.89	0.89	0.92	0.95	0.904	0.94	0.97
2	0.89	0.895	0.902	0.952	0.906	0.942	0.99
3	0.9	0.902	0.906	0.954	0.912	0.93	0.972
4	0.92	0.89	0.912	0.957	0.93	0.97	0.976
5	0.91	0.896	0.915	0.95	0.934	0.968	0.96
6	0.93	0.898	0.928	0.94	0.936	0.968	0.967
7	0.92	0.897	0.94	0.942	0.938	0.957	0.97
8	0.9	0.906	0.928	0.901	0.94	0.96	0.969
9	0.89	0.8997	0.89	0.89	0.942	0.94	0.98
12	0.896	0.8995	0.889	0.895	0.946	0.95	0.981
15	0.8955	0.89355	0.893855	0.88	0.95	0.97	0.93856
18	0.88	0.8923	0.89283	0.886	0.956	0.96	0.928395
21	0.892	0.8912	0.89182	0.891821	0.91821	0.95	0.918292
24	0.891	0.8901	0.89	0.89001	0.9001	0.9	0.900198

Table 4. 3: The Average Outcome of Energy Consumption Comparison with Different Number of Nodes and Area

Total nodes	Deployed area	Malicious nodes %	Other protocols	SEEDA protocol
100	400×400 m²	10%	3.34-3.9 joules	2.45 joules
		20%	3.66-3.96 joules	2.87 joules
		30%	3.57-4.0 joules	3.32 joules
250	400×400 m²	10%	3.5-4.0 joules	2.51 joules
		20%	3.71-3.89 joules	2.90 joules
		30%	3.60-4.0 joules	3.4 joules
350	1000×1000 m²	10%	3.5-4.0 joules	2.48 joules
		20%	3.68-3.98 joules	2.89 joules
		30%	3.61-4.0 joules	3.36 joules
400	1000×1000 m²	10%	3.6-4.0 joules	2.48 joules
		20%	3.73-3.96 joules	2.86 joules
		30%	3.56-4.0 joules	3.4 joules

Energy consumption is an important factor for a secure data aggregation system. The data packets and messages that send between clusters sensor nodes and aggregation nodes consume energy and will continue to consume more energy with the adversary's

participation. This evaluation will investigate the effect of Sybil and sinkhole attacks on energy consumption. Figure 4.5 plots the energy consumption in joules as a function of the process concerning an increase in the malicious nodes with an additional increase of monitoring process time (days). From the figure, it can be observed that the proposed SEEDA protocol consume the least energy in all three different scenarios with 10% (Figure 4.5(a)), 20% (Figure 4.5(b)), and 30% (Figure 4.5(c)) malicious node with 2.51 joules, 2.90 joules, and 3.4 joules respectively in time (days). The other protocols consume about 3.5-4.0 joules, 3.71-3.89 joules, 3.60-4.0 joules with 10%, 20%, 30% malicious nodes in the network respectively in time (days). The reason for the efficient performance of our protocol is caused by reducing the communication overhead and reducing the delay among nodes in the network. The base station utilizes the distance information to detect attacks and to choose the best path next hop. Besides, we used data aggregation, so this mechanism helps us to reduce the energy consumption among nodes in the network. In addition, our protocol proposed a fully homomorphic encryption algorithm this method can maintain or reduce the energy consumption in the network while implementing the secure node authentication algorithm. Furthermore, the SEEDA protocol employs cluster network topology involving static and mobile sensor nodes. The hierarchical cluster is built using six types of sensor nodes, namely child node, monitor node, relay node, aggregation node, cluster head, and base station node to support the energy consumption and to prolong the network lifetime.

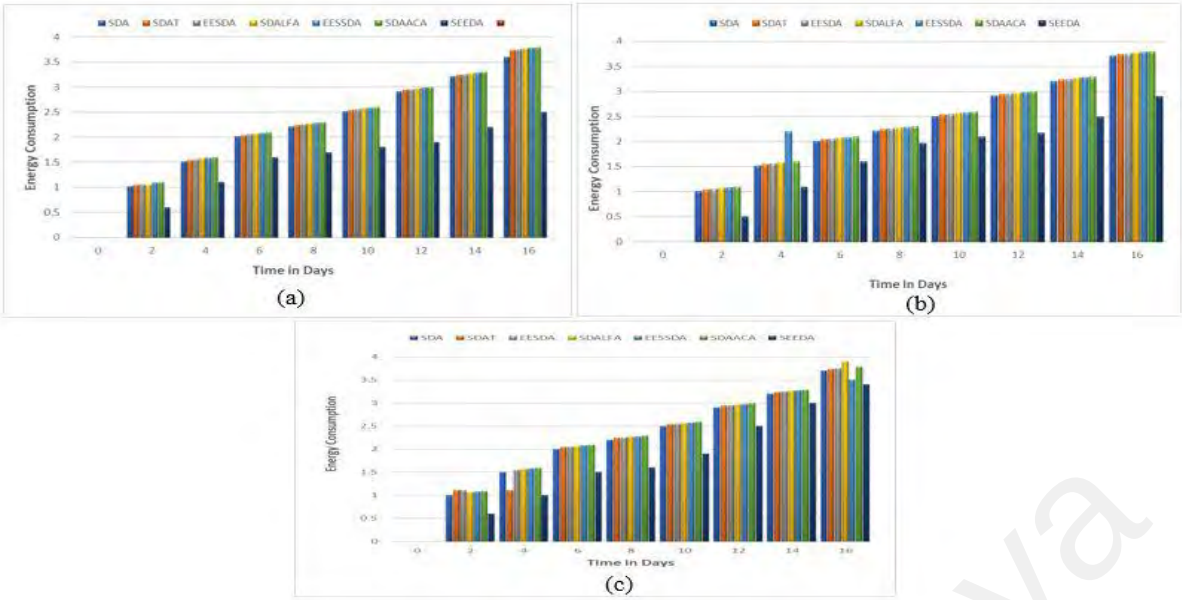


Figure 4. 5: The Energy Consumption with a) 10% Malicious Nodes, b) 20% Malicious Nodes, c) 30% Malicious Nodes

The end-to-end delay (in seconds) for all protocols is shown in Figure 4.6. The delay of the proposed protocol is minimal compared to other data aggregation in clustering protocols because the base station nodes utilize the distance and timestamp between the nodes to prevent the attacks from accessing the network. Consequently, it helps to reduce the delay and to avoid network traffic. The proposed protocol recorded a maximum delay of 0.038 seconds, whereas the prior protocols have higher latency between 0.059–0.08 that is considered very high for sensitive applications.

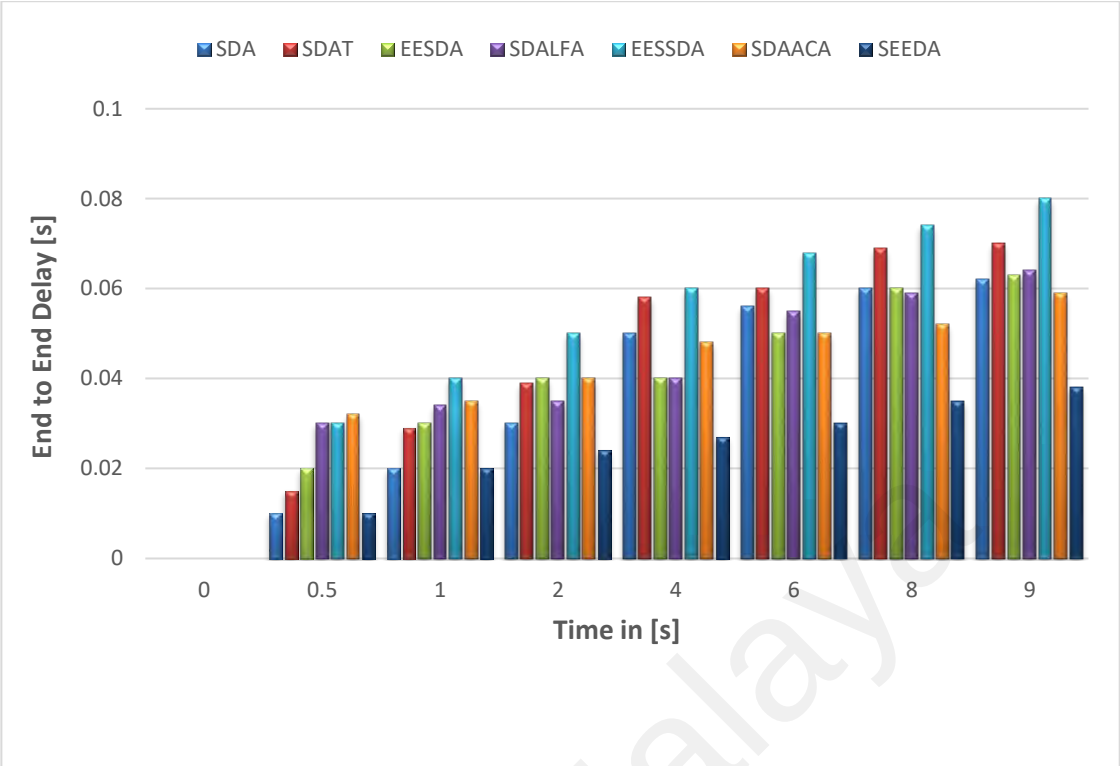
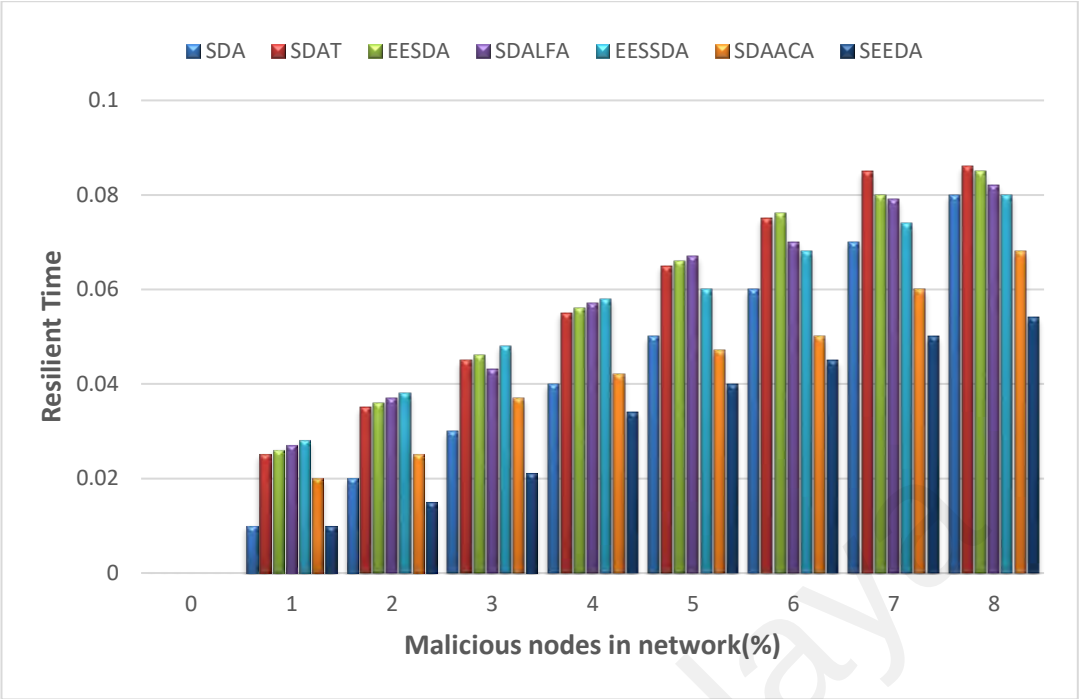
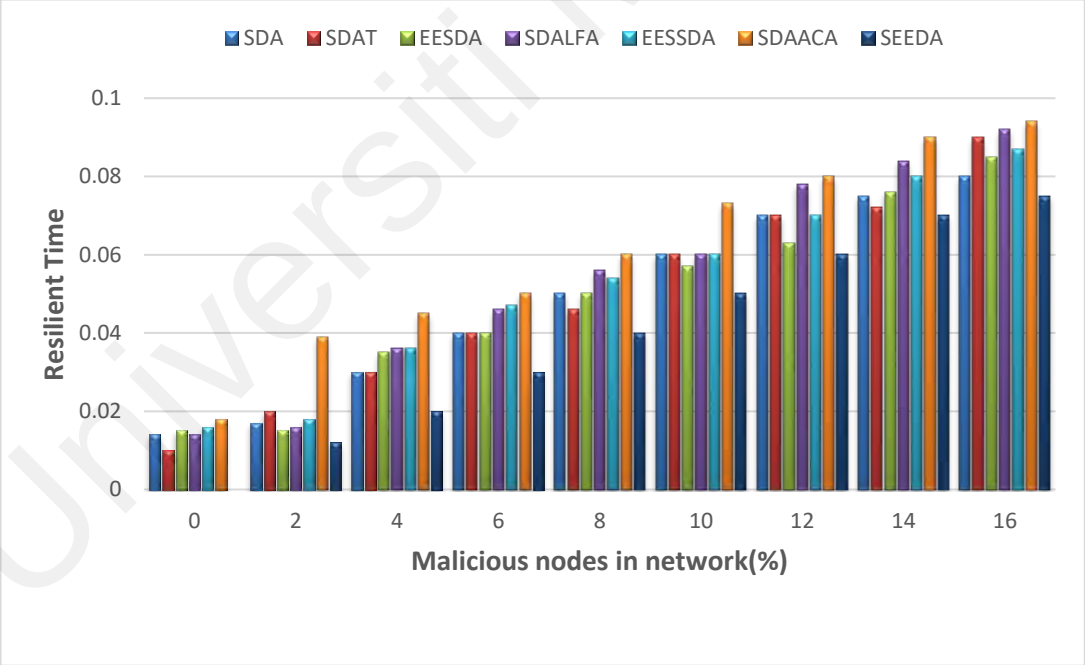


Figure 4. 6: End-to-End Delay of Our Proposed SEEDA Protocol

Figure 4.7(a) shows the resilient time when 0% to 8% malicious node affected the network, while Figure 4.7(b) shows the resilient time when 0% to 16% malicious node affected the network. Both simulations are conducted separately. Compared to other protocols, the proposed SEEDA protocol shows the best result with 0.054 seconds resilient time when affected by 8% malicious node (Figure 4.7(a)) and 0.075 seconds when affected by 16% malicious node (Figure 4.7(b)). The least resilient time shown by other protocols is at 0.068 seconds when affected by 8% malicious node (Figure 4.7(a)) and 0.094 seconds when affected by 16% malicious node (Figure 4.7(b)). The proposed SEEDA protocol outperforms the others because of the secure node authentication, and the base station node checks for false data aggregation to avoid attacks. Also, the SEEDA protocol can identify the integrity of the nodes. These operations have led to a low resilient time compared to other protocols.



(a)



(b)

Figure 4. 7: (a) Resilient Time when 0% to 8% Malicious Nodes Affected a Network, (b) Resilient Time when 0% to 16% Malicious Nodes Affected a Network.

4.7 Discussion

The proposed SEEDA protocol is an improved version of the SDAACA protocol. Both methods utilize a data fragmentation algorithm and a random value, but the proposed SEEDA protocol also includes query mechanism, distance information, and random timestamp. The data fragmentation algorithm is used to hide and preserve the original data from being tampered with by the malicious nodes. The malicious node needs to group all the blocks generated from the data fragmentation algorithm to reconstruct the original message. In the proposed SEEDA protocol, the cluster head node sends a broadcast query message. This message consists of data such as node ID and distance, cluster head node ID, and data packets used to protect the fragments and data. The fragmented data can be distributed to different or multiple node locations. If one block is compromised, the other block can be used as a backup. In a situation where data retransmission is needed, the fragmentation method still has an advantage since only a compromised block will need to be retransmitted. The data distribution is also able to increase the network lifetime due to the selection of high-energy nodes.

As the location of the legitimate node is known, the utilization of distance may differentiate them from the malicious node. If a node has a different estimated distance measurement as the legitimate node, it can be regarded as a malicious node. The distance can be calculated via transmission range and signal strength; therefore, it is not easily tampered with. Other than that, the utilization of distance information between nodes is used to search for the best path and to reduce transmission costs. Rather than transmitting data to further distance nodes, which requires more energy, distance information allows efficient task distribution among available nodes. This can reduce the delay in the network and also able to conserve energy.

A timestamp is a sequence of characters or encoded information identifying when the event occurred. The timestamp is used to calculate the time it takes to transmit data from nodes to the base station. This transmission time is recorded and is used as a signature to determine the legitimacy of a node. To strengthen the security, the timestamp is encoded with a random number to produce a random timestamp. A random timestamp is proposed so that the malicious node will not be able to duplicate the data packet even though they know the legitimate node's typical events occurrence.

The base station nodes check the distance and the timestamp of all the nodes broadcasted by the cluster head node. If they are different from the recorded value, they can be regarded as an adversary. The adversary can also be detected by comparing the transmission time and distance. Usually, faraway nodes would have higher transmission times compared to nearer nodes.

4.8 Chapter Summary

This chapter presented the structural design, functions, and simulation results of the proposed SEEDA protocol, secure and energy-efficient data aggregation in clustering for WSN using an access control model. The proposed protocol enhanced the authentication for MAC by generating a random value and random timestamp with a secret key. The base station node verifies the fake aggregated data before sending it to the server. Other than that, the proposed protocol detects and prevents attacks such as Sybil and sinkhole. Our protocol consisted of three algorithms: data fragmentation, secure node authentication, and fully homomorphic encryption algorithms. The data fragmentation algorithm was utilized to partition the data into small pieces before transmitting them to the next hop nodes to hide them from being attacked. The secure node authentication algorithm was utilized to check the node's authentication that is leaving or joining the network to prevent tampering or interrupting the data transmission between nodes. The

fully homomorphic encryption algorithm was utilized to encrypt the aggregated data before sending it to the base station nodes. The performance of the proposed protocol is compared with SDA, SDAT, SDALFA, EESSDA, SDAACA, and EESDA, which is a widely used protocol in the area of secure data aggregation. The simulation results show that the proposed SEEDA method outperformed these protocols in terms of the detection rate of the malicious nodes, energy consumption and accuracy, end-to-end delay, and resilient time in the network.

CHAPTER 5: DEVELOPMENT OF ENERGY-EFFICIENT WIRELESS SENSOR NETWORK WITH AN UNEQUAL CLUSTERING PROTOCOL BASED ON A BALANCED ENERGY METHOD (EEUCB)

5.1 Introduction

The hot spots problem is one of the most important challenges in WSNs. It is the primary reason for unbalance of energy consumption when multi-hop transmission between clusters with BS is performed. CHs closest to the BS will consume more energy than other CHs due to the receiving of more forwarding tasks.

In this chapter, an energy-efficient unequal clustering protocol based on a balanced energy method (EEUCB) is proposed. EEUCB protocol aims to optimize energy usage in clustering-based wireless sensor networks by adopting unequal clustering technology to achieve the following: (i) to avoid the hot spots problem; (ii) to address the distance among sensor nodes; and (iii) to apply the sleep-awake mechanism. Furthermore, to decrease the overhead and energy consumption of the CH node, a double cluster head technique has been proposed. In order to balance the distribution of energy consumption among CMs and the CH, we performed a clustering rotation strategy based on the average energy threshold, average distance threshold, and BS layering node. The performance of the proposed EEUCB protocol with UDCH, FLEACH, EEFUC, and LEACH are compared by performing various simulations.

In the previous chapter 3, the theoretical concepts and utilized system models were presented. In this chapter, the design architecture of EEUCB, its verification, and the simulation results of its performance evaluation are presented and discussed.

5.2 The Network and Energy Model of EEUCB Protocol

In this section, introduce the proposed energy-efficient with an unequal clustering protocol based on a balanced energy method (EEUCB) for WSN to reduce energy consumption and to prolong the network lifetime. Before explaining the details of our protocol, we will first describe the network model, and the energy model in the following sub-sections.

5.2.1 Network Model

In this sub-section, sensor nodes are randomly distributed in an environment of M*M square area, BS node has unlimited power, and outside of monitoring area, each node can be adjusted as a CM or a CH. All nodes and BS are stationary and can adjust the transmission power in order to estimate the distance and communication range; each sensor node has limited power and has a unique ID, and aware of each of the locations in the network. Each cluster has two CHs, called primary cluster head CH and second cluster head 2CH, the primary CH responsible for receiving data from nodes and send multi-hop to BS. While 2CH is responsible for receiving data from nodes and aggregated the data and sends it to primary CH.

5.2.2 Energy Model

To evaluate the performance of the proposed EEUCB scheme, we use the energy model, which is similarly described in the LEACH protocol (Wendi Beth Heinzelman, 2000). The communication between nodes consumes the vast majority of energy, so the energy consumption is neglected for sensing and processing in this work. In the process of communication, transmission and receiving consume more energy than monitoring. Therefore, we consider the energy of transmitting and the energy of receiving as energy consumption for communication. The energy consumed by transmitting l-bit data can be calculated as:

$$E_{TX} = \begin{cases} k \times E_{elec} + k \times efs \times d^2 & when\ d \leq d_0 \\ k \times E_{elec} + k \times emp \times d^4 & when\ d \geq d_0 \end{cases} \quad (5.1)$$

Where, E_{TX} is the energy consumption of transmitter, k is the length of transmission data, E_{elec} is the energy consumption for sending or receiving 1-bit data, efs is the data energy consumption of 1-bit in free space mode (d^2 power loss), emp is the data energy consumption of a 1-bit in multi-path attention mode (d^4 power loss), and d_0 is the threshold distance value, if the node distance is less than the distance threshold, it will send data via free space, while if the node distance is greater than the distance threshold, it will send data via multi-path to avoid the high energy consumption during the transmission or receiving data. The threshold distance value d_0 can be calculated as:

$$d_0 = \sqrt{\frac{efs}{emp}} \quad (5.2)$$

The energy consumed by transmitting k-bit data can be calculated as:

$$E_{RX}(i) = k.E_{elec} \quad (5.3)$$

Where, E_{RX} is the energy consumption of the receiver, and E_{elec} is the energy consumption of the receiver circuit or sender circuit for 1-bit data.

5.3 Development of Proposed EEUCB Protocol

In this section, introduce the development of the proposed EEUCB protocol. The EEUCB protocol contains four phases: Processing phase, initialization phase, cluster setup phase, and transmission phase. The processing phase is responsible for estimating the distance length with the base station node and calculate the average residual energy of neighbor nodes. The responsibility of the initialization phase is to calculate the radius of clustering to generate unequal clusters in the network. The cluster setup phase

responsible for calculating the delay time for the election of the primary cluster head node. Last, the transmission phase is responsible for transmission data between cluster members and base station node and balancing energy consumption between sensor nodes and cluster head nodes. These phases are explained in the following sections. To better understand the protocol, Figure 5.1 shows the EEUCB flowchart.

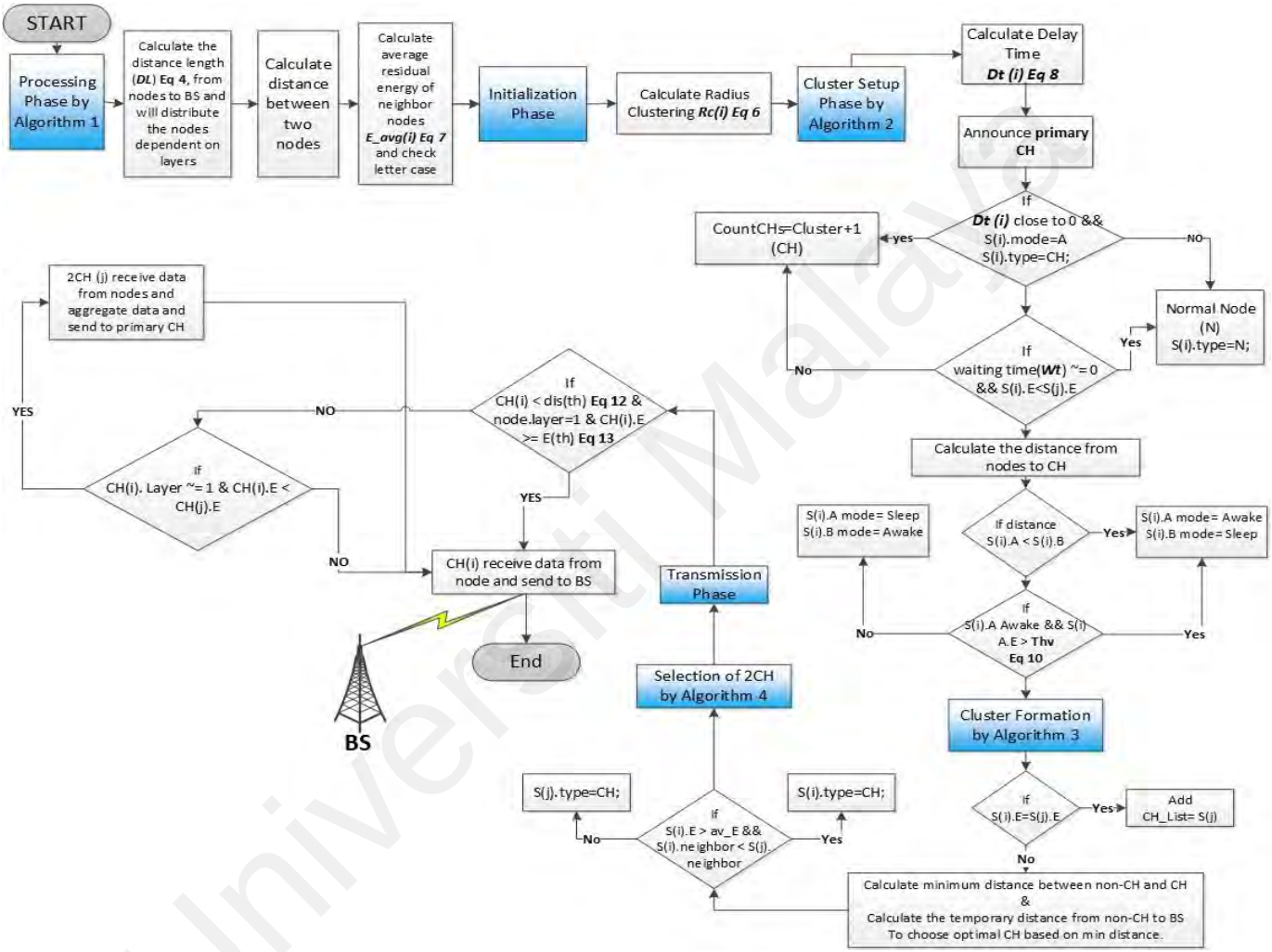


Figure 5. 1: Flowchart of EEUCB Protocol

5.3.1 Processing Phase

The processing phase function estimates the distance between cluster nodes and neighbor nodes and checks the number of neighbor nodes. First, after the sensor nodes are deployed in the network, The BS broadcast a message in the network, and then each sensor node calculates its distance to BS based on the signal strength indicator (RSSI) mechanism to identify the location of the sensor nodes (Bozorgi et al., 2017). Then the

sensor nodes send their location information to the BS. In this case, to be more realistic, our proposed method does not assume that all the sensor nodes can reach the BS through their own signal strength. Instead, we utilized intermediate neighbor nodes that forward information to other intermediate neighbor nodes until they reach the BS. The BS then calculates the distance difference among the sensor nodes from the BS and divided it into four layers empirically to transmit data from cluster head to BS. These four layers are used to identify the distance of each node from the BS and helps for multi-hop communication between clusters to reduce energy consumption. In addition, the place of cluster nodes is distributed in the network based on the layering mechanism. After dividing the sensor nodes into various layers according to their distance from BS, the BS notifies each node to which layer it belongs. The nodes are in the first layer (which is close to the BS), the data is transmitted to the BS in a single hop. On the other hand, nodes in the second, third, and fourth layers utilize a multi-hop mechanism. The first layer has the smallest cluster size (with more clusters), followed by the second, third, and finally the fourth with the biggest cluster size (with lesser cluster numbers). Unlike the prior equal clustering method, the relay tasks in the proposed method can be shared among cluster heads within the same layer. Nodes closest to the sink tend to drain their energy at a faster rate than other nodes as they have to perform more communications. Therefore, sharing relay tasks can avoid a single cluster head from the need to handle all incoming data from the higher layer (i.e., hot spots problem). Besides that, the sleep-awake mechanism is also utilized in our proposed protocol. The sleep-awake mechanism relies on two criteria, which are the distance from sensor nodes to CH and the energy level.

To make the BS determines clustering strategy, and performs layering, each node calculates the distance length of each layer. The distance length $\boldsymbol{d_L}$ can be calculated as:

$$\boldsymbol{d_L = (d_{max} - d_{min})/4} \qquad \textbf{(5.4)}$$

Where, $\mathbf{d_{max}}$ is the farthest distance of the node from the BS, $\mathbf{d_{min}}$ is the closest distance of a node from the BS. The clustering is performed independently by each of the layerings. The equation (5.4), improved to avoid the long-distance, so this will lead to the extension of the network's lifetime and improvement of network stability. In addition, we proposed an unequal clustering routing protocol; therefore, the location of the BS node will be outside the sensing area as described by (Amodu & Mahmood, 2018). If the location of the BS inside the sensing area, the network lifetime will be increasing because more nodes will share the high energy consumption around BS. Therefore, we located the BS outside the sensing area to verify the network lifetime of our protocol. On the other hand, if the location of the base station inside the sensing area and the sensor nodes opposite side the base station, the Pythagorean Theorem can be used to verify the difference between the farthest and closest distance of nodes from the BS. Therefore, in this case, the BS will distribute nodes in the network based on the competition radius $\boldsymbol{R_c(i)}$ such as in equation (5.7), if the $\boldsymbol{R_c(i)}$ is greater than the maximum competition radius $\boldsymbol{RL_{max}}$, the location of nodes will be either in the second, third, or fourth layer in the network. The Pythagorean Theorem can be written as:

$$(\boldsymbol{P})^2 = (\boldsymbol{x})^2 + (\boldsymbol{y})^2 \qquad \textbf{(5.5)}$$

Where, $\boldsymbol{P}$ is the hypotenuse, $\boldsymbol{x}$ is the right triangle, $\boldsymbol{y}$ is the width triangle.

The clustering layers algorithm is described in **Algorithm 1**. The input and output parameters (Node (i), d_{min}, d_{max}, BS, $\mathrm{d_L}$) of the algorithm is specified in 1–2. The sending and receiving of broadcast messages between nodes and BS are specified in steps 3–5. Step 6 calculates the distance length(d_L). Each node evaluates the layer with the BS in step 7. In steps 8-17, each node compares the layer depending on the maximum and minimum distance with BS.

Algorithm 1. The Clustering Layers of EEUCB
1. Input (Node(i), $\mathbf{d_{min}}$, d_{max}, BS)
2. Output (Node(i). layers)
3. BS broadcast message to all nodes
4. Nodes receive a message from BS
5. Nodes send a status message to BS
6. Calculate the distance length according to Equation (**5.4**)
7. **For** all of the nodes, BS **do**
8. **If** the distance node (i) to BS < d_{min} + d_L **then**
9. Node(i). layer = L1;
10. **Else**
11. **If** the distance node (i) to BS > d_{min} + d_L && distance node (i) to BS < d_{min} + (L2×d_L) **then**
12. Node(i). layer = L2;
13. **Else**
14. **If** the distance node (i) to BS > d_{min} +(L2×d_L) && distance node (i) to BS < d_{min} +(L3×d_L) **then**
15. Node(i). layer = L3;
16. **Else**
17. Node(i). layer = L4;
18. **End**
19. **End**

A. Complexity Analysis of Algorithm 1

We assume C to be the constant of complexity, and Node (i) is the name of sensor nodes. In the clustering layers algorithm, we have one loop and iteration, so the time complexity of this algorithm is ***O(N)***. The following describes the time complexity of algorithm 1:

1. Iteration to find the distance length from sensor nodes to BS, which is represented as ***n***.

2. We have three condition statements to distribute the sensor nodes into different layers. The total complexity of these statements is ***C1+C2+C3***.

The overall time complexity of algorithm 1:

$T(n) = n * (C1+C2+C3)$

$=C1\ C2\ C3n$

$= O(N)$

5.3.2 Initialization Phase

If the sensor nodes are closer to the BS and are forwarding more data to it, the node will be heavier and consume more energy. To balance out the energy in the whole network, the initialization phase proposes to generate unequal clustering in the network and elect the main CH by calculating the delay time of the nodes. An unequal algorithm could balance the energy among sensor nodes. Hence, the size of cluster nodes should be smaller than others if the cluster nodes are closer to the BS. The initialization phase is divided into three, namely unequal clustering generation, competition radius calculation, and delay time.

5.3.2.1 Unequal Clustering Generation and Calculating Competition Radius

In this sub-phase, the generation of unequal clusters is based on the competition radius. The competition radius is responsible for generating unequal clustering for each node and determining the size of the cluster with a BS node. Existing unequal clustering, such as in UDCH, utilizes the residual energy of sensor nodes and the distance from all the sensor nodes to the BS node. However, this method does not consider the length of the distance between CMs and the BS node, which leads to energy wastage across the network nodes and a reduced network lifetime. Apart from residual energy and the distance from all sensor nodes to the BS as in the prior method, our proposed EEUCB method also considers the minimum distance of the closest node from the BS and the maximum distance of the furthest node from the BS. These two criteria were considered to avoid the

long-distance; this will extend network lifetime and improve network stability. Furthermore, our protocol utilizes the maximum capacity of node energy to improve the existing method. The calculation of the competition radius (R_c) of the prior method (UDCH) is shown below:

$$\boldsymbol{R_c(i) = a\frac{E_{rem}(i,r)}{E_{init}} + b\frac{d_{i,BS}}{D_{max}}\, RL_{max}} \qquad \textbf{(5.6)}$$

While the calculation of competition radius (R_c) of our proposed EEUCB protocol is shown below:

$$\boldsymbol{R_c(i) = 1 - a\left(\frac{D_{max}-d_{i,BS}}{d_{max}-d_{min}}\right) - b\left(1 - \frac{E_{rem}(i,r)}{E_{max}}\right) RL_{max}} \quad \textbf{(5.7)}$$

Where, $\boldsymbol{R_c(i)}$ is the radius of node i, $\boldsymbol{D_{max}}$ is the maximum distance from nodes to BS, $\boldsymbol{d_{i,BS}}$ is the distance from node i to BS, $\boldsymbol{E_{rem}(i,r)}$ is the residual energy of node i at round r, $\boldsymbol{E_{init}}$ is the initial energy of node i, $\boldsymbol{E_{max}}$ is the maximum capacity of node energy, $\boldsymbol{RL_{max}}$ is the maximum competition radius for becoming CH, ***(a)*** and ***(b)*** is the weighted factor between [0,1] to adjust the scope of $\boldsymbol{R_c(i)}$ and determines the impact of the energy and distance on the competition radius. When (a) and (b) increase, the range of competition radius value decreases; conversely, when (a) and (b) decreases, the range of competition radius value increases, also include the effect of competition radius value on the energy.

5.3.2.2 Delay Time

After calculating the competition radius $\boldsymbol{R_c(i)}$, the sensor nodes then calculate the delay time $\boldsymbol{D_t(i)}$ to announce being a CH; the delay time is computed in equation (5.9). The calculation of the said process depends on the residual energy of neighbor nodes $\boldsymbol{E_{rem(j,r)}}$, the number of neighbor nodes $\boldsymbol{\vdots\, NN(i,r)\, \vdots}$, and the average residual energy of

neighbor nodes $\boldsymbol{E_{avg(i)}}$ such as in equation (5.8). Therefore, the sensor nodes will send a broadcast packet to the neighbors. The content of the packet includes the node ID, residual energy, location of the node, packet size, and the type of packet (packet type_1) because different types of broadcast packets are sent in the network to calculate the average energy of the neighbor nodes. Each node records the information of neighbor nodes when they receive the packet, and the nodes are identified by the neighbors; the nodes will then calculate the average energy of neighbor nodes as defined in (Bozorgi & Bidgoli, 2019). The average energy of neighbor nodes can be calculated as:

$$\boldsymbol{E_{avg(i)} = \frac{\sum j \epsilon NN(i,r) E_{rem(j,r)}}{\max(\vdots NN(i,r) \vdots, \sigma)}} \qquad \textbf{(5.8)}$$

Where, $E_{avg(i)}$ is the average energy of neighboring node i, $NN(i,r)$ is the set of neighbor nodes, $\vdots NN(i,r) \vdots$ is the number of neighbors, $\boldsymbol{j}$ is the neighbor node of node i, σ is a tiny number which have no effect on the Eq (5.9) if the result becoming zero. When many nodes have the same number of neighbors, σ plays a role to get a different number of neighbors for each node. The delay time of each node can be calculated as:

The delay time of each node can be calculated as:

$$\boldsymbol{D_t(i) = \left(1 - \frac{E_{rem(i)}}{E_{avg(i)}}\right) * W_t + R_v} \qquad \textbf{(5.9)}$$

Where, $D_t(i)$ is the delay time on node i, $E_{rem(i)}$is the residual energy of node i, W_t is the competition time of primary CH, R_v is the random value. The random value can play a role in reducing communication conflicts when the nodes have the same residual energy. The format of the (packet type_1) is shown in figure 5.2.

Packet size	**Node ID**	**Location**	**Residual Energy**

Figure 5. 2: Format of Data Packet Type_1

5.3.3 Cluster Setup Phase

In this section, the clustering setup phase will elect the nodes to become CH; the election of CH depends on the delay time processing. This section contains three sub-sections: primary CH selection and sleep awake mechanism, cluster formation, and secondary CH selection.

5.3.3.1 Primary CH Selection and Sleep Awake Mechanism

The Primary CH selection method for the proposed EEUCB is similar to the UDCH (F. Zhu & Wei, 2019). However, in EEUCB, further improvement is made by taking into account the Sleep-Awake mechanism. The method starts with the calculation of the delay time, as in equation (5.9). When the delay time closes in on zero, the primary CH will be elected. If a node is calculated to have a shorter delay time, it will have a higher chance to become a CH and send a broadcast packet. The content of the packet includes the packet size, node ID, location of the node, residual energy, and the type of packet (packet type_2). The format of the packet type_2 is shown in Figure 5.3.

Packet size	Node ID	Location	Residual Energy

Figure 5. 3: Format of Data Packet Type_2

Each node should wait until the delay time process ends. If a node [i] received packet type_2 from nodes before the delay time process ends, the node [i] will become normal nodes "N."

In addition, if the other nodes broadcast the packet type_2 and the candidate CH receive the packet before the end of the competition time $\boldsymbol{W_t}$, it will compare the residual energy of node S_i, and node S_j. If the node S_i is greater than node S_j, it will become primary CH; otherwise, it becomes a normal node "N." Furthermore, the candidate CH will broadcast messages to sensor nodes. The sensor nodes will verify the messages from CH. If the CH

has limited residual energy and might die sooner, the nodes will inform neighbor nodes in the network to change the path and send data to another CH.

Since nodes with a shorter distance to the CH consume less energy, we propose a round-robin sleep-awake rotation method to select nodes to transmit data to the CH based on two stages. This method is illustrated in Figure 5.4. Let it be assumed that node A and node B are 3m and 5m away from the CH, respectively. In the first stage (Fig 5.4a), the round-robin-based selection depends only on the node distance to the CH. Since node A has the shortest distance to the CH, it is the first selected (awake) to transmit data to the CH (Round_1). The non-selected nodes will remain asleep. In the next round (Round_2), node B—which has the second shortest distance to the CH—is then selected to be awake and sends data to the CH. The process will continue until all nodes have transmitted their data to the CH. In the second stage (Figure 5.4b), the selection is not based on distance alone; the energy level becomes the additional criterion that is included in the rule.

Apart from that, the utilization of distance, R, in the second stage, is modified based on equation (5.10). Starting from Round_ 3, the distance, R, is used to calculate the variable Z as follows:

$$\boldsymbol{Z = R \times n} \qquad \textbf{(5.10)}$$

Where, $\boldsymbol{Z}$ is the name of each node, $\boldsymbol{R}$ is the radius, and $\boldsymbol{n}$ is the variable increment if $\boldsymbol{Z}$ has the lowest value.

From the example in Fig 5.4 (i.e., Round_3), node A is selected to be awake to transmit its data to the CH since it has the lowest Z value (i.e., ZA=RAx1=3x1=3). Node B (and others, if any), on the other hand, will remain asleep. Following the awake selection, the selected node A will then update its Z value (i.e., ZA=RAx2=3x2=6) in Round_4, while

the Z value for the other nodes (which were not selected) remain as before. In round_4, the lowest Z value (i.e., Node B with ZB=RBx1=5x1=5) is selected to be awake once again. The selected node will update its Z value in the 5th Round; in this example, it can be seen that ZB is updated to ZB=10 (i.e., ZB=RBx2=5x2=10). The process continues in the next round.

Besides utilizing the Z value, the sleep-awake mechanism in the second stage also depends on the energy level according to the threshold value, as shown in equation (5.11). Assuming that the $\boldsymbol{Th_v}$ is equal to 0.02j, the current awake node—as long as it remains awake and continues to transmit data to the CH—should have residual energy greater than the threshold value. From the example in Fig 5.4 (i.e., Round_6), node A will remain awake and transmit its data to the CH since it has the lowest Z value and the fact that its residual energy is greater than the threshold value. Node B, on the other hand, will remain asleep.

In addition, if the two nodes have the same Z value in the cluster, the node with the higher residual energy is chosen. From the example in Fig 5.4 (i.e., Round_9), node A and B have the same Z value, whereas node A will have higher residual energy than node B. Therefore, node A will remain awake and continue to transmit data to the CH, while node B remains asleep.

The threshold value of the node in the cluster is calculated as below:

$$\boldsymbol{Th_v = \frac{T_E}{N}} \qquad \textbf{(5.11)}$$

Where, $\boldsymbol{T_E}$ is the total energy of sensor nodes in the cluster, $\boldsymbol{N}$ is the total number of sensor nodes.

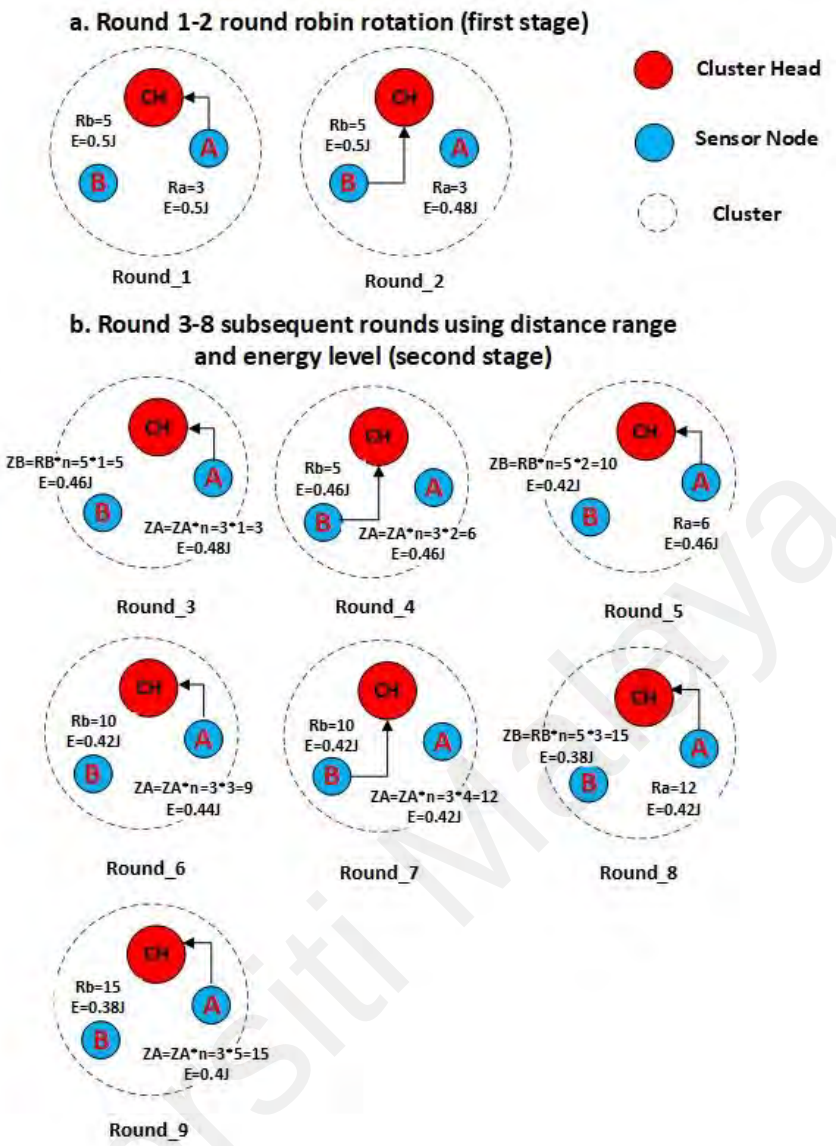


Figure 5. 4: Sleep-Awake Rotation

After the sleep and awake mechanism, the primary CH will broadcast a message within the maximum competition radius $\boldsymbol{RL_{max}}$ at the end of competition time $\boldsymbol{W_t}$. It consists of a packet, which includes node ID, residual energy, location of the node, and packet size, distance to BS, and the type of this packet (packet type_3). The format of the packet type_3 is shown in Figure 5.5.

Packet size	Node ID	Location	Distance to BS	Residual Energy

Figure 5. 5: Format of Data Packet Type_3

The primary CH algorithm is described in **Algorithm 2**. The input and output parameters (S_i, W_t, A, B, Th_v, RL_{max}, *a, b,* $R_v, R_c(i)$, $D_t(i)$, CH) of the algorithm is specified in 1–2. In steps 3-6, calculate the clustering radius $\boldsymbol{R_c(i)}$ and broadcasting to all neighbor nodes. The number of neighbor nodes $\boldsymbol{NN(i,r)}$ and the average energy of neighbor nodes $\boldsymbol{E_{avg(i)}}$ are calculated in steps 7-8. In step 9, is calculated the delay time for each node $\boldsymbol{D_t(i)}$. In steps, 10-20, the processing of nodes to select primary CH is conducted as explained in the above section. The residual energy between nodes in steps 21-30. The sleep and awake mechanism in steps 31-44. Finally, the primary CH broadcast type_3 packet within the competition time $\boldsymbol{W_t}$ in steps 45-48.

Algorithm 2. Primary CH Selection and Sleep Awake Mechanism
1. Input (S_i, W_t, A, B, Th_v, RL_{max}, *a, b,* R_v)
2. Output (primary CH and mode)
3. **For** each node S_i **do**
4. Calculate $\boldsymbol{R_c(i)}$ according to Equation (5.6)
5. Broadcast packet type_1
6. The nodes identified their neighbors
7. Calculate the number of neighbor nodes $\boldsymbol{NN(i,r)}$
8. Calculate the average energy of neighbor nodes $\boldsymbol{E_{avg(i)}}$
9. Each node calculates the delay time $\boldsymbol{D_t(i)}$ according to Equation (5.9)
10. S_i. Type =" N"
11. **If** S_i. $\boldsymbol{D_t(i)}$ = close to 0
12. CountCH= countCH +1
13. S_i. Type =" CH"
14. Broadcast packet type_2
15. **End**
16. **While** S_i. $\boldsymbol{D_t(i)} \neq$ close to 0
17. **If** S_i. $\boldsymbol{D_t} > S_j$. $\boldsymbol{D_t}$
18. S_i. Type =" N"
19. **End**
20. **End**
21. **While** S_i. $\boldsymbol{W_t} \neq 0$
22. **If** S_i. $\boldsymbol{D_t} < S_j$. $\boldsymbol{D_t}$
23. **If** S_i. Type =" CH"
24. **If** S_i. E < S_j. E
25. S_i. Type =" N"
26. **Else** S_i. Type =" CH"
27. **Endif**
28. **Endif**

29. **Endif**
30. **Endwhile**
31. **If** S_i. Type =" CH"
32. **While** all nodes send data once to CH
33. Calculate the distance from nodes to CH
34. **If** $S_i(A)$. distance to CH $<$ $S_i(B)$. distance to CH
35. $S_i(A)$. mode =" Awake"
36. $S_i(B)$. mode =" Sleep"
37. **Endif**
38. **If** $S_i(A)$ $Awake$ && $S_i(A).E > \boldsymbol{Th_v}$
39. $S_i(A)$. mode =" Awake"
40. $S_i(B)$. mode =" Sleep"
41. **Else** $S_i(A)$. mode =" Sleep" **&&** $S_i(B)$. mode =" Awake"
42. **Endif**
43. **Endwhile**
44. **Endif**
45. **If** S_i. Type =" CH"
46. Broadcast packet type 3 to BS
47. **Endif**
48.**Endfor**

B. Complexity Analysis of Algorithm 2

In primary selection CH and sleep-awake mechanism algorithm, the time complexity is ***O(N^2).*** The following describes the time complexity of algorithm 2:

1. Calculate the competition radius of each node in the network, which is represented as ***n***.

2. Calculate the delay time; if the delay time close to zero, the node will become primary CH. The total complexity of these statements is ***(C1+C2).***

3. If the delay time does not equal zero, the node will become a normal node, which can be computed as (***n*C3).***

4. Check the waiting time, and the residual energy of sensor nodes in the network can be computed as ***n*(C4+C5+C6).***

5. Check the distance from sensor nodes to CH, which can be computed as ***n*(C7+C8+C9).***

6. Broadcast message to BS. The complexity of this statement is ***C10***.

The overall time complexity of algorithm 2:

= n * (C1+C2) + (n*C3) + n*(C4+C5+C6) + n*(C7+C8+C9) + C10

= n* (n + n + n) + C10

= n* (3n)

= O(N^2) ⇒ O(N2).

5.3.3.2 Clustering Formation

The cluster formation function selects the optimal non-CH to become a member of the CH. After the primary CH selection in this sub-section, the CH will broadcast packet type_3 and the node S_i as it awaits receipt of the packet. If the node S_i receives the message from node S_j, it will add and store it to the list of candidate nodes for the CH (CH_list) and change it into a non-CH state. On the other hand, the node S_i possibly receives more messages from different CHs. In this case, it will choose the optimal nodes as its CH by calculating the distance from non-CHs to the CH. The optimal selection of the CH is dependent on the minimum distance, high residual energy, and a smaller number of neighbor nodes. In addition, the maximum number of nodes in each cluster will affect the performance of the network. So, to balance the distribution of nodes among the cluster heads and to achieve balanced energy consumption, in this paper, we define the maximum number of each cluster as calculated by $= \frac{N}{T_{CH}}$, where the N is the total number of sensor nodes, T_{CH} is the total number of CH

in the network. At the time when sensor nodes join the cluster, the cluster head will compare the number of nodes with the t value; if the value is less than the t value, it will accept the node; otherwise, it will reject the request.

Algorithm 3, the input and output parameters (S_i , S_j , CH_list) of the algorithm is specified in 1–2. Explains the procedure of cluster formation that chooses the optimal CHs in steps 3-11.

Algorithm 3. Cluster Formation
1. Input (S_i, S_j)
2. Output (CH_list)
3. **For** each node S_i **do**
4. **If** S_i. Type =”N” && S_i. E>0 && S_i. Type =” Awake”
5. **If** $S_i.Head = S_j.Head$
6. compute the minimum distance from non-CH to CHs
7. **If** S_i. E > av_energy && S_i. Neighbor < S_j. Neighbor
8. S_i. Type =” CH”
9. broadcast packet type_3
10. CH_list store S_i
11. **Elseif** S_j. Type =” CH”
12. **Endif**
13. **Endif**
14. **End for**

C. Complexity Analysis of Algorithm 3

In the cluster formation algorithm, we have one loop and iteration, so the time complexity of this algorithm is ***O(N)***. The following describes the time complexity of algorithm 3.

1. Iteration to the selection of non-CH to become a member of the CH, which is represented as ***n***.

2. Checking the status of each node in the cluster can be computed as ***(C1 * C2 * C3)***.

The overall time complexity of algorithm 3:

= n * (C1 * C2 * C3).

= C1 C2 C3n

= O(N).

5.3.3.3 Secondary Cluster Head Selection

To reduce the overhead, processing load, and to save energy by minimizing energy consumption of primary CH, we propose to elect secondary cluster head 2CH. The primary cluster head CH is responsible for the aggregation of data and their transmission to the BS node if the distance from primary CH is less than the distance threshold and the residual energy is greater than the energy threshold. Meanwhile, the 2CH is responsible for receiving and aggregating data within each cluster and send the aggregated data to the primary CH if the distance from the primary CH is greater than the distance threshold and the residual energy is less than the energy threshold. The MAC algorithm between cluster members and CHs based on Time-Division Multiple Access (TDMA). Whenever if the primary CH aggregates data from nodes, the primary CH creates the TDMA schedule for other normal nodes and broadcast the schedule to all the cluster members. Otherwise, if the 2CH aggregates data from nodes, it will create the TDMA schedule for the normal nodes within the cluster. In this case, each node will send data to the 2CH based on the schedule.

In **Algorithm 4**, we assume that CH is represented by 'j'; and 'f' represented the sensor nodes distributed in the cluster. The input and output parameters of the algorithm are specified in 1-2. In steps 3-6, we define the cluster-ID, as 'j' and the energy of non-CH, as 'f' in per round in the network. To select the 2CH, our EEUCB protocol proposes that the sensor nodes with the highest residual energy max_E will become 2CH. If the N(f)

belongs to the cluster N(j) and the N(f) equal to N(j), each cluster will compute the energy temporarily and find the maximum residual energy in each cluster as described in steps 7-15. Steps 16-19 will check the residual energy of N(f). If it is high, it will be selected as 2CH; otherwise, it will become a normal cluster node.

Algorithm 4. Selection of Secondary Cluster Head 2CH
1. Input (total number of nodes (n), the total number of CH (n_{CH}))
2. Output (2CH)
3. **for** each non-ch to f **do**
4. Cluster ID j = n (r+1, f)
5. Node 'f ' = E(r+1, f) % energy per round
6. **End**
7. **if** non-ch $-1 \geq 1$
8. **for** each cluster head j **do**
9. $max_E = 0$
10. **if** N(f) ϵ j
11. **if** N(f)= non-ch && N(f).E > 0
12. energy_temp= Max (max_E, E(r+1, f))
13. **if** energy temp > max_E
14. max_E = energy_temp
15. **End**
16. **if** N(f).E == max_E
17. I=f %% 2CH ID
18. **End**
19. **Endif**
20. **Endif**
21. **End for**
22. **Endif**

D. Complexity Analysis of Algorithm 4

In the selection of the secondary cluster head 2CH algorithm, we have two separate loops, so the time complexity of this algorithm is ***O(N)***. The following describes the time complexity of algorithm 4:

1. Iteration to Selection of secondary cluster head 2CH, which is represented as ***n***.

2. Checking the status of the non-cluster head, and iteration for each cluster head can be computed as ***n*(C1*C2*C3*C4).***

3. Checking the highest residual energy of sensor nodes in the cluster. The complexity of this statement is ***C5***.

The overall time complexity of algorithm 4:

$= n + n * (C1 * C2 * C3 * C4) + C5$

$= n + n + C5$

$= 2n \Rightarrow O(N).$

5.3.4 Transmission Phase

The transmission phase is the process of data transmission between CHs and CMs through the network. The transmission phase consists of a CH rotation strategy and a layering implementation. The CH rotation strategy and the layered implementation scheme will be further described in the next sub-section. The process of the transmission is utilized after the selection of the primary CH, which sends a broadcast schedule to all CMs by creating a TDMA schedule. The CMs will send data to the primary CH, which aggregates data. Thereafter, aggregation operations will be sent to the BS.

5.3.4.1 CH Rotation Strategy and Layers Implementation

The CH rotation strategy and layering implementation scheme are proposed in our EEUCB protocol to balance the energy consumption between CMs, CHs, and the BS nodes. We proposed this CH rotation strategy as the unbalanced energy consumption between sensor nodes and CHs during data transmission in the network will affect the network lifetime and increase energy consumption throughout the network. The rotation

strategy functions to balance energy consumption between sensor nodes and CHs so that each node in the network has a chance to become a CH.

A layering implementation is proposed in our EEUCB protocol in order to extend network lifetime and reduce energy consumption. The function of layering implementation is to estimate the location of the primary CH in the network that was estimated based on the distance to the BS. The layering implementation is performed, as shown in the processing phase section (5.3.1).

Concerning the transmission phase of prior methods as in FLEACH and UDCH, the FLEACH utilizes average distance while UDCH utilizes the average energy threshold between CMs and CHs and constructs the path to the BS through the network. On the other hand, the proposed EEUCB utilizes both the average distance and energy threshold. The EEUCB also makes use of layering implementation to reduce energy consumption and prolong the network lifetime.

In our proposed EEUCB protocol, the CH rotation strategy includes two sub-phases, namely the Intra and inter clustering transmission. In the first sub-phase, the CH performs the Intra-cluster transmission. The CH rotation strategy is utilized between CMs and CHs. If the distance from the primary CH is less than the distance threshold such as in equation (5.13), and the residual energy is greater than the energy threshold such as in equation (5.14), then the nodes will send data directly via single hop to the primary CH; then, the primary CH receives the data and applies aggregate functions. These nodes help the CH to consume less energy in receiving data. However, if the distance from the primary CH is greater than the distance threshold, and the residual energy is less than the energy threshold, the system will use multi-hop routing to transfer data to the CH, which consumes more energy. As such, the nodes will send data to the 2CH, which will aggregate data to be sent to the primary CH. Finally, the primary CH receives the

aggregated data from the 2CH and sends it to the BS. In addition, if the distance is greater than the distance threshold, and residual energy is greater than the energy threshold, the sensor nodes will send data to the 2CH. Conversely, if the distance is less than the distance threshold, and the residual energy is less than the energy threshold, the sensor nodes will send data directly to primary CH. As the transmission process depends on distance, the transmission of data from CMs to CHs will not consume more energy in the network.

The second sub-phase is constructing a path to BS via inter-cluster layering. In this sub-phase, the MAC algorithm between the CHs and BS of our EEUCB protocol uses the carrier-sense multiple access (CSMA) for transmission data to BS. For the data transmission in this sub-phase, the sensor nodes may change the state into sleep mode. Therefore, the energy status and the number of node members will change in the cluster. Therefore, if the CH [i] is not located in the first layer (not close to BS), and the residual energy of CH [i] is less than the residual energy of CH [j], the CH [i] will send aggregated data to CH [j], and the CH [j] will transmit the data to BS; otherwise, the CH [i] transmit aggregated data to BS directly. The intra and inter clustering transmission based on the EEUCB protocol is shown in Figure 5.6. The flowchart of the main CH rotation strategy is shown in Figure 5.7.

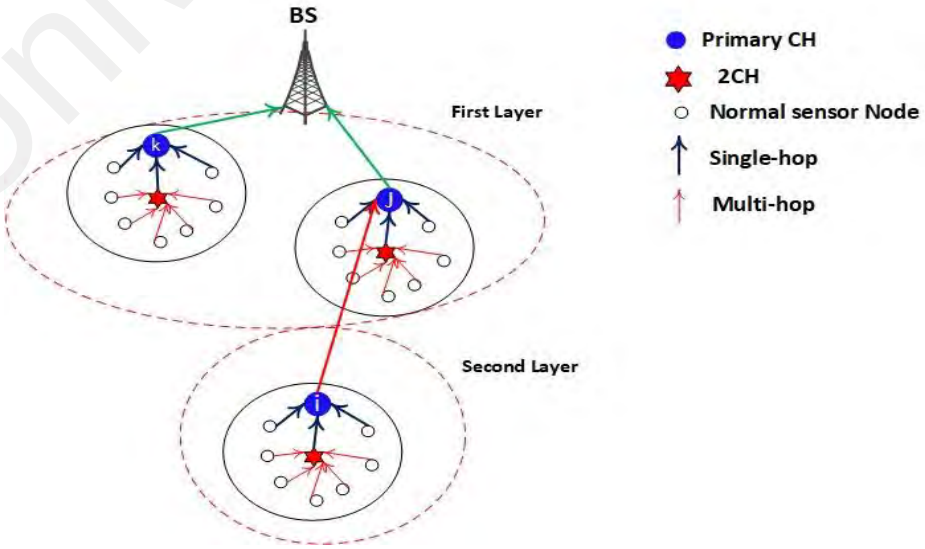


Figure 5. 6: The Intra and Inter Clustering Transmission

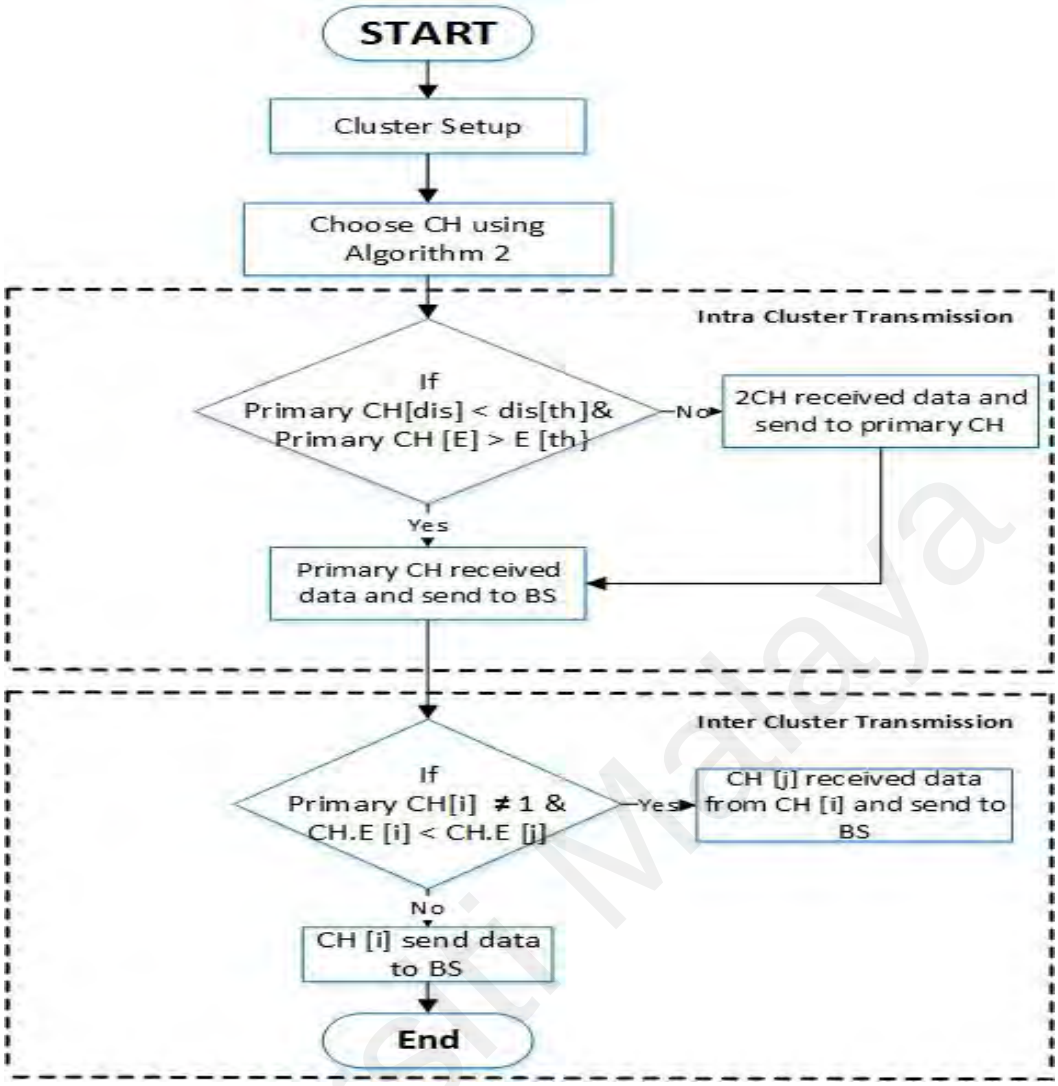


Figure 5. 7: The Flowchart of Main CH Rotation Strategy

The distance threshold was considered because of the sensor node's different distributions within the network area. So, if the primary CH distance is greater than the distance threshold, it means that the CH distance is farthest from sensor nodes and will lead to more energy-consuming for data aggregation and transmission process. Therefore, the sensor nodes will send to the 2CH, and the 2CH will then forward it to the primary CH. The distance threshold depends on the average distance. The average distance means the distance of all sensor nodes to BS. The average distance can be written as:

$$avg_D = \frac{1}{N}\sum_{i=1}^{N} D_{itoBS} \quad (5.12)$$

Where, avg_D is the average distance, N is the set of sensor nodes, D_{itoBS} is the distance of each node to the BS. The equation of the distance threshold can be calculated as:

$$\boldsymbol{D_{th} = \partial \times avg_D} \qquad \textbf{(5.13)}$$

Where, D_{th} is the distance threshold and ∂ is the weight factor to determine the impact of the distance between the CMs and CHs. When the ∂ increases, the distance range of CHs is greater than the distance threshold. Hence the 2CH will consume more energy for the receiving and transmission process, and data will not be aggregated.

The energy threshold was also considered in our protocol in order to extend the network lifetime and to reduce energy consumption. When the energy of the primary CH is lesser than the energy threshold, the CH node will die early and disconnect the communication between nodes. To avoid this problem, we elect the 2CH to help the primary CH to be "alive" by receiving the data from nodes and execute aggregated functions, and after that send it to the primary CH. The computation of the energy threshold can be written as follow:

$$\boldsymbol{E_{th} = \beta \times E_{CH}(i)} \qquad \textbf{(5.14)}$$

Where, E_{th} is the energy threshold, β is the weight factor to determine the impact of the energy level on the energy threshold, and E_{CH} is the initial energy of CH after each cluster selection.

5.4 Evaluation Metrics

The proposed method has been presented in the previous section. To further test the reliability of the proposed method, four evaluation metrics are considered; these metrics measure the network lifetime, average energy consumption, average residual energy, and throughput. These metrics are explained in the following sub-sections.

5.4.1 Network Lifetime

The network lifetime is the time when the first sensor node in the network runs out of energy and dies. The measure network lifetime can be written as:

$$N_{lifetime} = \sum_{r=1}^{r\ max} \sum_{i=1}^{N} (Node^{(i)}.E \leq 0, (dead = dead + 1; \quad \textbf{(5.15)}$$

$$If\ (dead == 1, first\_dead = r)))$$

Where, $N_{lifetime}$ is the network lifetime, r is the number of rounds, N is the total number of sensor nodes, and $dead = dead$ is the number of dead nodes around.

5.4.2 Average Energy Consumption

Efficient energy consumption is significant in cluster network technology. One goal of our proposed protocol is to reduce and preserve the energy consumption in the network. The average energy consumption can be calculated as:

$$\boldsymbol{E_{avg} = \sum_{i}^{N} E_{ini}(i) / N} \qquad \textbf{(5.16)}$$

Where, E_{avg} is the average energy consumption, $E_{ini}(i)$ is the initial energy of sensor node (i) and *N* is the total number of sensor nodes.

5.4.3 Average Residual Energy

The residual energy is energy within the system that is not being used, but when released, it can execute the work. The residual energy is calculated when each sensor node is transmitting and receiving the packets among them in the network. The average residual energy can be computed as:

$$\boldsymbol{E_{rem} = \sum_{i}^{N} E_{rem}(i)/N} \qquad \textbf{(5.17)}$$

Where, E_{rem} is the average residual energy of nodes and $total\ E_{rem}(i)$ is the remained energy of sensor node (i).

5.4.4 End-To-End Delay

The end-to-end delay is defined as the time is taken when the packets transfer from the sensor nodes to the base station node. The end-to-end delay can be computed as:

$$\boldsymbol{D = \sum_{i+1}^{P(i)} \frac{P_{t(i)} - P_{r(i)}}{Total\ packets}} \qquad \textbf{(5.18)}$$

Where, D is the end-to-end delay, $\boldsymbol{P_{t(i)}}$ is the time when sending packets, $\boldsymbol{P_{r(i)}}$ is the time when the packets are received.

5.4.5 Throughput

Throughput is the number of data packets successfully transmitted to the destination in a period of time. The formula of throughput can be written as:

$$\boldsymbol{T_p = \frac{T_s}{T_r} sec} \qquad \textbf{(5.19)}$$

Where, T_p is the throughput, T_s is the total number of packets send to BS and T_r is the total number of packets received at BS.

5.5 Time and Space Complexity Analysis of EEUCB Protocol

In this section, we calculate the time and space complexity of our EEUCB protocol. EEUCB protocol includes four algorithms to reduce energy consumption and prolong the network lifetime for WSN. Let's calculate the time complexity first; after that, we calculate the space complexity of our EEUCB protocol. We have calculated the time complexity of each algorithm in this paper, so we will combine all the time complexity

of algorithms to calculate the overall time complexity of the proposed EEUCB protocol. The following describes the time complexity of our EEUCB protocol:

1. The time complexity of algorithm 1 is ***O(N).***

2. The time complexity of Algorithm 2 is ***O(N2).***

3. The time complexity of Algorithm 3 is ***O(N).***

4. The time complexity of Algorithm 4 is ***O(N).***

The overall time complexity of EEUCB protocol can be compute as follow:

T(n)= O(N) + O(N2) + O(N) + O(N)

= 3N + N^2

= N^2

= O(N^2) ⟹O(N2).

According to the mentioned above, we have calculated the time complexity of our EEUCB protocol. We need now to calculate the space complexity to check the total amount of memory used by our EEUCB protocol. The calculation of the space complexity process is the same as the process of time complexity. Therefore, we will also combine all the space complexity of algorithms to calculate the overall space complexity of the proposed EEUCB protocol. The following describes the space complexity of our EEUCB protocol:

1. The space complexity of algorithm 1 is ***O(N).***

2. The space complexity of Algorithm 2 is ***O (1).***

3. The space complexity of Algorithm 3 is ***O(N).***

4. The space complexity of Algorithm 4 is ***O(N).***

The overall space complexity of the EEUCB protocol can be computed as follow:

$$S(n) = O(N) + O(1) + O(N) + O(N)$$

$$= 3N + 1$$

$$= O(N)$$

5.6 Simulation Results

The objective of our simulation is to compare the performance of EEUCB with other protocols using a MATLAB 2019b. The network topologies of this paper were presented by IEEE 802.15.4/ZigBee because this protocol supported clustering technology and carrier-sense multiple access (CSMA). On top of that, we want to investigate the energy consumption efficiency and network lifetime extension of our proposed protocol. Table 5.1 presents the three different scenarios, such as a different number of nodes and different area sizes. All the examined scenarios show similar results when regardless of nodes numbers and network area sizes. The parameters used in the simulation are presented in Table 5.2.

Table 5. 1: Parameters of Simulation

Parameters in EEUCB	Value
Sensing area	$200\times200m^2$ (Scenario_1)
	$300\times300m^2$ (Scenario_2)
	$400\times400m^2$ (Scenario_3)
	$1000\times1000m^2$ (Scenario_4)

Number of nodes	100,300,400, 1000
The initial energy of sensor nodes	0.5 joules
Data packet size	4000 bits
Control message size	200 bits
Maximum communication radius RL_{max}	70 m
Waiting time, W_t	5 second
Transmission energy, efs	10 pJ/bit/m^2
Transmission energy (long-distance, emp)	0.0013 pJ/bit/m^4
Electronic circuit energy, E_{elec}	50 pJ/bit
Aggregation energy	5 pJ/bit

Table 5. 2: Scenarios for The Proposed EEUCB Protocol

Scenario	Number of sensor nodes	Network space
Scenario_1	100	200×200 m^2
Scenario_2	300	300×300 m^2
Scenario_3	400	400×400 m^2
Scenario_4	1000	1000×1000 m^2

The simulation results show the performance of the proposed protocol successfully prolongs the network lifetime and reducing energy consumption. The aim is to balance the energy consumption among cluster members and cluster heads. The clustering rotation strategy is based on the average energy threshold, average distance threshold and performs layering by base station node. This can ensure more efficient energy consumption and network lifetime increment. On the other hand, we estimate a double cluster head to reduce the overhead considerably and energy consumption of the cluster head node. In this method, each cluster has two CH. The primary cluster head CH is

responsible for aggregating data and forwarding it to the BS node if the distance of primary CH is greater than the distance threshold and the energy is less than the energy threshold. Meanwhile, the secondary cluster head 2CH is responsible for receiving and aggregating data within each cluster and send the aggregated data to primary CH if the distance 2CH is less than the distance threshold and the energy of the 2CH is greater than the energy threshold.

The performance comparison between the proposed protocol and the other three cluster protocols is presented in Figure 5.8 to Figure 5.12.

Figure 5.8 plots the network lifetime, including the first node die (FND), half node die (HND), and the last node dies (LND) between the proposed EEUCB, LEACH, FLEACH, EEFUC, UDCH protocols in the first, second, third, and fourth scenarios is performed. From the figure, it can be observed that the proposed EEUCB protocol has increased network lifetime in all four different scenarios. In the first one, we deployed 100 sensor nodes in a $200m^2 \times 200m^2$ as shown in (Fig 8(a)), in (Fig 8(b)) we implemented the second scenario, the area of network $300m^2 \times 300m^2$ with 300 sensor nodes, and the third scenario, as shown in (Fig 8(c)) contains $400m^2 \times 400m^2$ the area with 400 sensor nodes. Lastly, in (Fig 8(d)), we implemented the fourth scenario in order to check the scalability of the protocol, the area of network $1000m^2 \times 1000m^2$ with 1000 sensor nodes. In addition, several different scenarios are generated to show the performance and to check the ability of the proposed protocol to preserve energy consumption and to prolong the network lifetime.

Fig 8(a) shows the LEACH protocol FND is at 600 rounds, and HND at 1150 rounds. In the FLEACH, EEFUC, and UDCH protocols, the FND is 1280, 1290, 1301 rounds, and HND at 2250, 2388, and 2390 respectively. In our EEUCB protocol, the FND is at 1420 rounds, and HND at 2600 rounds. Fig 8(b) shows that the FND of LEACH, FLEACH,

EEFUC, and UDCH protocols appear at 670, 1320, 1395, and 1421 rounds respectively, and the HND appeared at 1176, 2320, 2400, and 2485, while in the EEUCB FND is at 1620 rounds, and HND at 2800. In Fig 8(c), the LEACH, FLEACH, EEFUC, and UDCH protocols, the FND is at 695, 1322, 1402, and 1430 rounds, and HND is at 1200, 2350, 2420, and 2500, and LND at 3000, 3320, 3600, and 3679. The FND of our EEUCB protocol stands at 1645 rounds, while HND and LND at 2810 and 3800 rounds respectively. In Fig 8(d), the LEACH, FLEACH, EEFUC, and UDCH protocols, the FND is at 750, 1400, 1490, and 1525 rounds, and HND is at 1300, 2398, 2500, and 2580 rounds, and LND at 3200, 3620, 3765, and 3800 rounds. The FND of our EEUCB protocol stands at 1700 rounds, while the HND and LND at 2925 and 3920 rounds respectively.

The results show that our EEUCB protocol outperforms other protocols. The reason is that the proposed EEUCB protocol balance the energy consumption among the nodes in the network. This is achieved by an unequal clustering mechanism based on different competition radius calculations. The calculation of the competition radius for each node depends on a few factors: the closest and farthest distance of the nodes from the BS, the distance from all the sensor nodes to the BS node, the residual energy of each node at each round, and the maximum capacity of the node energy compared to UDCH and EEFUC protocol, there are only calculating the distance form all nodes to BS, whereas the LEACH has not shown good performance in terms of the network lifetime because it was not proposed an unequal clustering mechanism. The equal clustering mechanism was proposed instead. In addition, our EEUCB, UDCH, and FLEACH protocols utilize a double cluster head node in order to reduce the load on primary CH. FLEACH has not shown good performance in terms of the network lifetime due to the selection of primary CH is randomly and alternately selected among the network nodes based on probability, whereas the selection of primary CH of EEUCB and UDCH is based on the minimum computing delay time of each node. However, EEUCB has shown an improved network

lifetime because it utilizes the Sleep-Awake mechanism based on the distance from sensor nodes to CH, and the energy level of sensor nodes. The selection of 2CH of UDCH is based on the distance from the sensor nodes to the primary CH, so it has not shown an improved network lifetime, while our EEUCB protocol has shown an improved network lifetime because the selection of 2CH is based on calculating the highest residual energy of nodes. Hence, avoid the selection of 2CH with low residual energy. Both LEACH and EEFUC protocols have not shown an improved network lifetime because they only utilize one CH for aggregation and forward data transmission at the same time to the base station.

In addition, the transmission round also helps to reduce the energy consumption of nodes by reducing the communication overhead and prolong network lifetime in the network. As an example, if the location of the CHs is far from BS, and the residual energy of the CHs is low, the CH may not be able to transmit data to BS. This will lead to increases in energy consumption and packet loss during the data transmission in the network, and the CH might die sooner. The transmission round of EEUCB uses the average distance threshold, average energy threshold between CMs and CHs. Use the layer implementation and residual energy for the construct of a path to BS. Therefore, it has shown good performance in terms of the network lifetime compared to UDCH, FLEACH, LEACH, and EEFUC. The FLEACH protocol has not more shown an improved network lifetime because utilizing distance threshold only for transmission rounds between CMs and CHs and for constructing a path to BS in the network. The transmission round between CMs and CHs, and for construct a path to BS of EEFUC protocol utilizes residual energy of sensor node and distance from CH to BS. The UDCH utilize between CMs and CHs uses the average energy threshold and uses the average energy to construct a path to BS in the network.

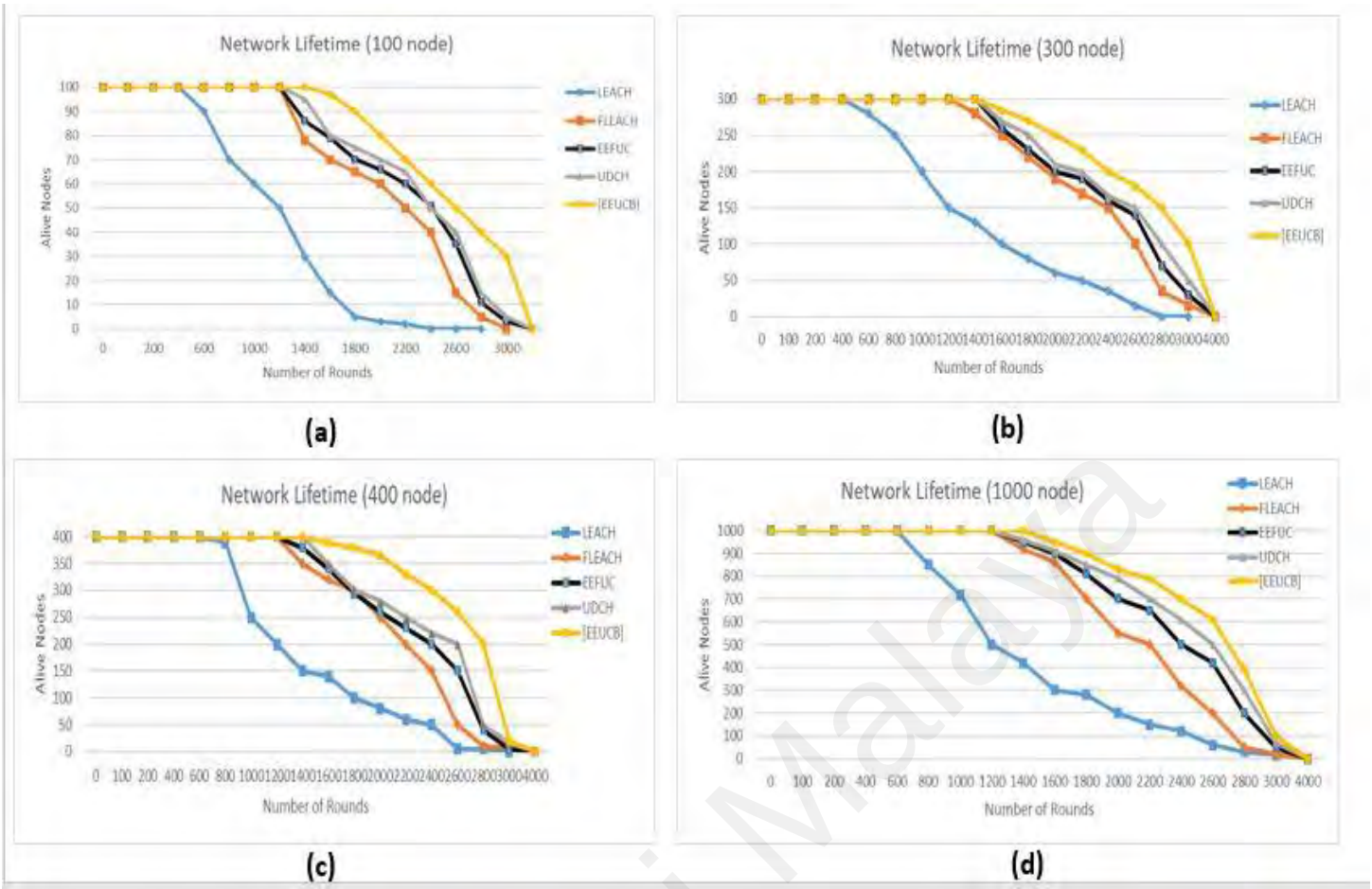


Figure 5. 8: The network lifetime with a) 100 number of nodes, b) 300 number of nodes, c) 400 number of nodes, d) 1000 number of nodes

In addition, we tested the statistical significance of the number of alive nodes (NOA) by using a paired T.TEST to draw a statistical inference as defined by (Mondal, Dutta, Ghosh, & Biswas, 2016; Mondal, Ghosh, & Biswas, 2016). A large sample consisting of pair of (NOA) in the proposed EEUCB protocol with other protocols such as LEACH, or FLEACH, or EEFUC, or UDCH are taken over different rounds behaves like a normal co-related variable. Table 5.3 shows the results of paired T.TEST for our proposed EEUCB protocol with other protocols. Our testing hypothesis has four cases that can be described as follows:

Null Hypothesis H_0: $(NOA_{EEUCB} = NOA_{LEACH})$. Alternative Hypothesis H_1: $(NOA_{EEUCB} > NOA_{LEACH})$.

Null Hypothesis H_0: $(NOA_{EEUCB} = NOA_{FLEACH})$. Alternative Hypothesis H_1: $(NOA_{EEUCB} > NOA_{FLEACH})$.

Null Hypothesis H_0: $(NOA_{EEUCB} = NOA_{EEFUC})$. Alternative Hypothesis H_1: $(NOA_{EEUCB} > NOA_{EEFUC})$.

Null Hypothesis H_0: $(NOA_{EEUCB} = NOA_{UDCH})$. Alternative Hypothesis H_1: $(NOA_{EEUCB} > NOA_{UDCH})$.

The test statistic T with n-1 degrees of freedom can be computed as:

$$\mathbf{T= D_{avg} \div (S_d \div \sqrt{(n-1)})} \quad \textbf{(5.20)}$$

Where $\boldsymbol{D_{avg}}$ and $\boldsymbol{S_d}$ denote the mean and standard deviation of the difference of NOA in two equal-sized correlated large samples of size n. The 95% confidence limits for $\boldsymbol{D_{avg}}$.

$$\boldsymbol{D_{avg} \pm T_{0.05} \times (S_d \div \sqrt{(n-1)})} \quad \textbf{(5.21)}$$

Where $\boldsymbol{T_{0.05}}$ is the 5% point of t-distribution with n − 1 degree of freedom.

Let p indicate the probability of the calculated value for our statistical t-test with n-1 degrees of freedom to obey the null hypothesis. A value of $p < 0.05$ indicates that the null hypothesis is rejected at a 5% significance level and the alternative one be accepted at a 95% confidence level. Our results were obtained by t-test of our proposed EEUCB protocol with LEACH, FLEACH, EEFUC, and UDCH protocols. In all the cases $p < 0.05$, so the null hypothesis is rejected at a 5 % significance level, and the alternative hypothesis is accepted at a 95% confidence level. Also, the lower and upper limits for the 95% confidence interval for $\boldsymbol{D_{avg}}$ is shown in Table 5.3. Therefore, it can be observed that the proposed EEUCB protocol outperforms LEACH, FLEACH, EEFUC, and UDCH.

Table 5. 3: T-Test NOA Results

[EEUCB]	T-test	Significance of Null hypothesis	Confidence interval 95%	
			Lower	Upper
LEACH	47.21	< 5%	35.37	38.63
FLEACH	21.19	< 5%	9.6	17.24
EEFUC	17.69	< 5%	5.33	13.66
UDCH	13.46	< 5%	3.64	10.58

Energy consumption is an essential factor in clustering protocols. Figure 5.9 plots the average energy consumption in joules as a function of the process number of rounds. The average energy consumption of our EEUCB protocol is less than the other four protocols, namely LEACH, FLEACH, EEFUC, and UDCH. The reason for this is because we improved the selection of CHs algorithms, the competition radius calculation, and transmission data operations, which leads to the distribution of CHs being more sensible and the reduction in energy consumption of the sensor nodes. By contrast, the UDCH calculated the computation radius and the distance from nodes to the BS but did not calculate the maximum capacity of the node energy in leading to an increase in the energy consumption of nodes. The LEACH has larger results than other protocols due to the cluster distributed randomly in the network and did not utilize a double cluster head. Therefore, it increases the load on the main CH, while the FLEACH utilizes a double cluster head but did not address the hot spots problem and randomly selected primary CH, which leads to the consumption of more energy and unbalanced energy consumption for receiving and transmitting data to the BS. The EEFUC addressed the hot spots problem but did not utilize a double cluster head, therefore, increasing the load on the main CH in the network.

In addition, we tested the statistical significance of the energy consumption for a single round using a pair-wise t-test to draw a statistical inference as defined as shown in Table 5.4. The calculation of the t-test of the energy consumption process is the same as the t-

test of (NOA), as in Equations (5.20) and (5.21). In all the cases $p < 0.05$, so the null hypothesis is rejected at a 5 % significance level, and the alternative hypothesis is accepted at a 95% confidence level. Furthermore, we checked the impact of the network capacity of energy consumption per round by testing our protocol with different scenarios are shown in Table 5.5.

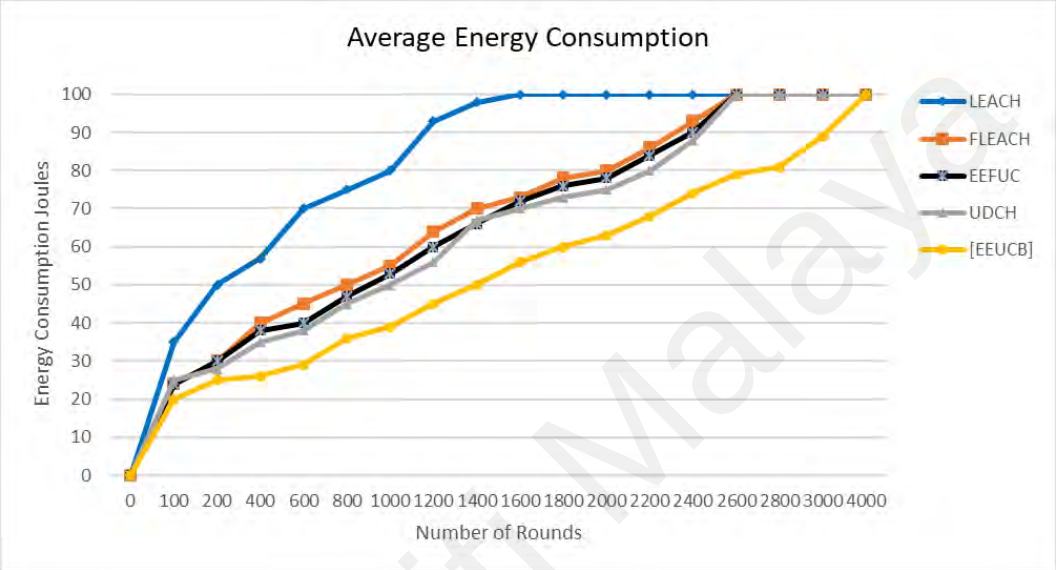


Figure 5. 9: The Average Energy Consumption

Table 5. 4: Results of T-Test of Energy Consumption for a Single Round.

[EEUCB]	*t*-test	Significance of the Null Hypothesis	Confidence Interval 95%	
			Lower	Upper
LEACH	12.68	< 5%	12.66	12.71
FLEACH	7.96	< 5%	7.91	8.033
EEFUC	4.13	< 5%	4.11	4.14
UDCH	2.37	< 5%	1.99	2.38

Table 5. 5: The Energy Consumption with Different Scenarios

Network capacity	LEACH	FLEACH	EEFUC	UDCH	[EEUCB]
Scenario_1	0.0495	0.0397	0.0298	0.0294	0.014
Scenario_2	0.0582	0.0412	0.0284	0.0226	0.0196
Scenario_3	0.0532	0.0434	0.0356	0.0297	0.0135
Scenario_4	0.0576	0.0422	0.0284	0.0221	0.0179

The average residual energy is calculated from the remaining energy after the packet's transmission process between sensor nodes. Figure 5.10 shows the energy balance evaluation of EEUCB. To get an accurate evaluation, we run a simulation ten times, and average results are computed. Figure 5.10 shows the average residual energy consumption comparison of our proposed EEUCB protocol with four protocols: LEACH, FLEACH, EEFUC, and UDCH. From the result, the average residual energy of the four protocols is less than the proposed EEUCB protocol. The average residual energy starts to drop below the half initial energy from 1000 rounds in LEACH and FLEACH, and in EEFUC and UDCH from 1200 rounds. The residual energy of the proposed EEUCB protocol, on the other hand, only starts to drop at 1400 rounds. Table 8 presents the standard deviation of residual energy against a different number of rounds. Results show that our EEUCB protocol outperforms other protocols and that EEUCB has a better energy balance compared to other protocols. The EEUCB protocol reduces energy consumption by reducing the communication overhead and utilizes the sleep-awake mechanism based on the distance from sensor nodes to the CH and the energy level of the sensor nodes. The selection of CH methods used in the proposed EEUCB protocol enhances load balancing based on minimum delay time with the sleep-awake mechanism and the highest residual energy methods. In contrast, in UDCH, the selection of the CHs process focuses on the delay time for the primary CH and the distance from the nodes to the CH. This could lead to the selection of CH nodes with low residual energy that may die early.

The selection of CH in LEACH FLEACH, and EEFUC are randomly selected based on probability, which leads to consuming more energy for the selection and transmission process. In addition, the transmission process is very important in clustering protocols due to more dissipated energy in the transmission process. Therefore, EEUCB has good results because we enhance the transmission process based on the energy and distance

thresholds between the CMs and CHs and the layering implementation methods between the CHs and the BS node. For example, if the distance from the primary CH is less than the distance threshold and the residual energy is greater than the energy threshold, the nodes will send data directly via single hop to the primary CH; the primary CH then receives the data and applies aggregate functions. These nodes help the CH to consume less energy in receiving data. However, if the distance from the primary CH is greater than the distance threshold, and the residual energy is less than the energy threshold, the system will use multi-hop routing to transfer data to the CH, which consumes more energy. On this basis, the nodes will send data to the 2CH, which will aggregate data to be sent to the primary CH, whereas the transmission between CHs and BS node beads on the layering process. If the CH [i] is not located in the first layer (not close to the BS), and the residual energy of the CH [i] is less than the residual energy of CH [j], the CH [i] will send aggregated data to the CH [j], and the CH [j] will transmit the data to BS; otherwise, the CH [i] transmit aggregated data to the BS directly. Therefore, these methods helped us to preserve energy consumption in the network. The transmission process between the CMs, CHs, and BS was based on the distance threshold in FLEACH. The distance threshold was perhaps insufficient to preserve the residual energy of nodes, which leads to a loss of residual energy for nodes. In EEFUC, the transmission was based on residual energy; however, the CH consumed more energy. Hence, it did not save much residual energy in sensor nodes, whereas UDCH uses only the average energy threshold for the transmission process.

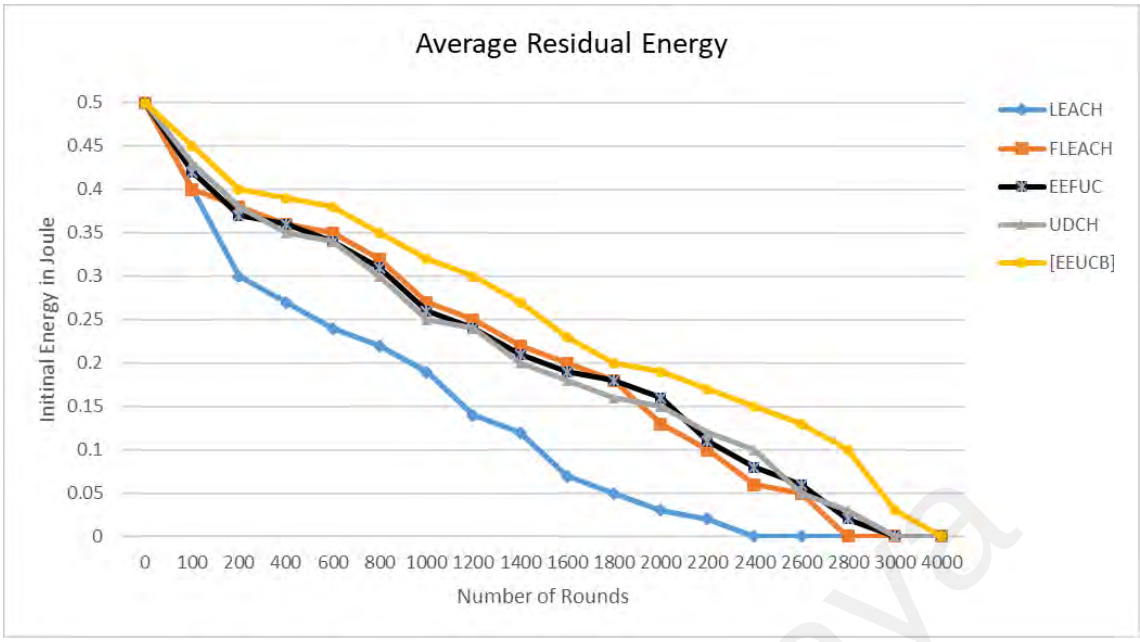


Figure 5. 10: The Average Residual Energy

Table 5. 6: Standard Deviation Residual Energy

No. Rounds	LEACH	FLEACH	EEFUC	UDCH	[EEUCB]
1000	0.413502	0.273344	0.252321	0.229902	0.143445
2000	0.426451	0.282043	0.255009	0.232022	0.146405
3000	0.427665	0.286485	0.261123	0.238551	0.149894
4000	0.431013	0.288559	0.264544	0.242256	0.153445

Figure 5.11 shows the end-to-end delay results (in seconds). The end-to-end delay can be defined as the packets transfer from sensor nodes to the sink node. The maximum delay of our proposed EEUCB protocol was recorded at 0.04 seconds, whereas the prior protocols have higher latency between 0.05–0.07. The lower delay time obtained by the proposed EEUCB is due to the inclusion of distance information between the sensor node to the base station by calculating the minimum and maximum distance and also using a double cluster head in each cluster. The distance information helps in terms of the layers' implementation method to estimate the distance between cluster nodes to the BS and dividing into four layers based on the furthest and closest distance to BS. The FLEACH and UDCH used a double cluster head without calculating the minimum and maximum

distance. Therefore, the location of CHs may be furthest from the BS, generating a delay and increasing energy consumption, whereas the LEACH and EEFUC do not propose a double cluster head nor calculate the minimum and maximum distance in the network.

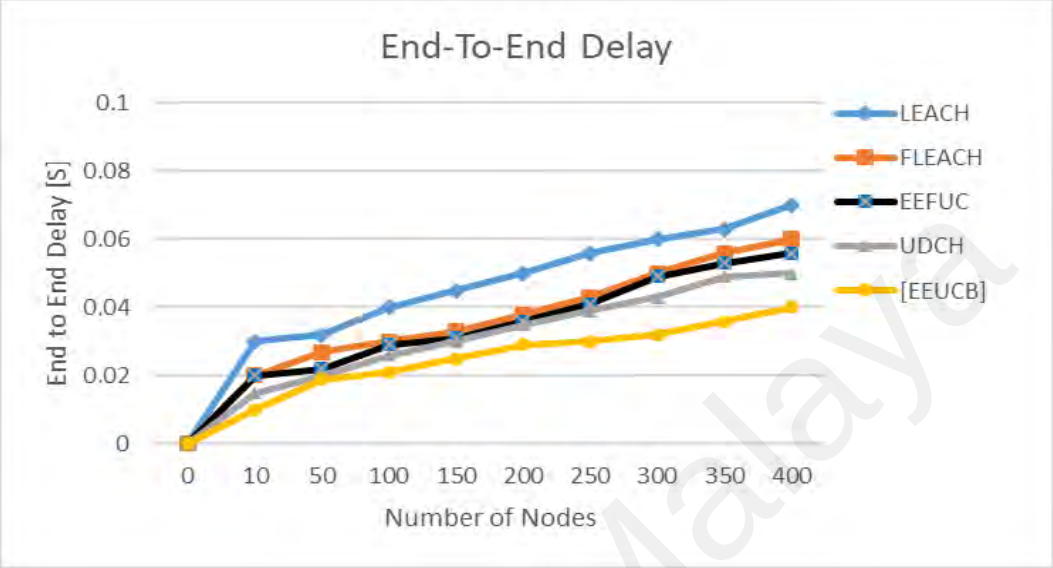


Figure 5. 11: The End-to-End Delay

Figure 5.12 shows the throughput results, which can be defined as the number of data packets received at the BS in a period of time. We have tested the throughput with the 1000 number of nodes. Results show that the throughput of our proposed EEUCB protocol performed better than the other four protocols. In LEACH, the received data packets are about 350,000, FLEACH about 520,000, EEFUC about 550,000, and UDCH around 610,000, while the EEUCB received data packets at about 700,000. Table 5.7 shows the results of a paired t-test for our proposed EEUCB protocol with other protocols. The calculation of the t-test of the throughput process is the same as the t-test of (NOA), as in Equations (5.20) and (5.21). A value of $p < 0.05$ indicates that the null hypothesis is rejected at a 5% significance level, and the alternative is accepted at a 95% confidence level. In every case where $p < 0.05$ the null hypothesis is rejected at a 5% significance level and the alternative hypothesis is accepted at a 95% confidence level. The proposed EEUCB managed to obtain the highest throughput because it used the sleep-awake

mechanism among nodes in the network. Therefore, the redundant data transmission among nodes is limited, restricted contrary to UDCH and EEFUC. Moreover, the EEUCB used the layering implementation method between CHs and BS. The LEACH has the lowest throughput because it did not consider the unequal clustering and sleep-awake mechanism. The proposed EEUCB protocol has shown an increase in throughput to the BS over other protocols.

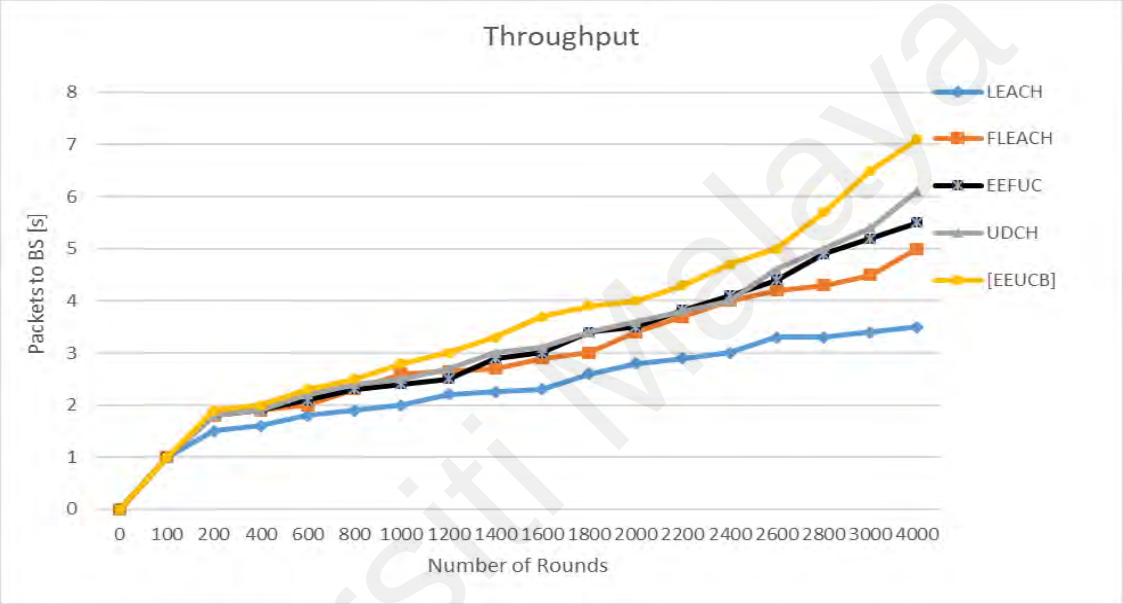


Figure 5. 12: The Throughput

Table 5. 7: T.Test Results for Throughput

[EEUCB]	T-test	Significance of Null hypothesis	Confidence interval 95%	
			Lower	Upper
LEACH	33.24	< 5%	24.46	37.80
FLEACH	18.55	< 5%	14.30	7.60
EEFUC	14.42	< 5%	10.74	12.90
UDCH	8.66	< 5%	2.55	5.16

5.7 Discussion

The proposed EEUCB protocol is an improved version of the UDCH protocol. Both methods utilize the delay time method to select the CH in the network. However, further

improvements are made in EEUCB by considering the Sleep-Awake mechanism. In addition, the proposed EEUCB protocol takes into account the highest residual energy for the selection of the 2CH and the distance between two nodes to check the node ID, location, neighbor nodes, and to check the number of neighbor nodes. In addition, the distance of cluster nodes is distributed in the network by layering calculations. In our protocol, the BS calculates the distance based on the closest and farthest nodes and divides this distance into four layers. The advantage of this method is that the nodes in the first layer can send the data to the BS through a single hop. If the nodes are in the second, third, or fourth layer, it will send the data to the CH through a multi-hop. This method helps to reduce and maintain energy consumption in a network. Besides that, the sleep-awake mode is utilized to maintain the energy of the node and to prolong the lifetime of the network. This is due to the fact that the active nodes furthest from the CH will reduce energy efficiency and will die early.

On the other hand, the selection of non-CHs is dependent on the minimum distance between non-CHs and the CH because the closest distance between the CM and the CH will not dissipate more energy. At the same time, the node can send data during long rounds rather than transmitting with a longer distance, which requires more energy and time. In addition, our protocol utilizes the energy and distance threshold to balance energy consumption within nodes; this allows the two CHs to be selected to reduce the overhead on the primary CH and to enable them to distribute the operations between them. The layering implementation and residual energy construct the path to the BS. These methods are proposed for the transmission of data from the CH to the BS node in the scenario where the primary CH is situated far from the BS node and is not located in the first layer. This will increase energy consumption and overhead. In order to resolve this, the primary CH will send the data to the other CH that is closest to the BS node. The comparison between our method and prior methods is listed in Table 5.6.

Table 5. 8: Comparison Between Our Method and Prior Methods

FLEACH	***EEFUC***	***UDCH***	***EEUCB (Proposed protocol)***
The placement of the sensor does not depend on the network layer.	The placement of the sensor does not depend on the network layer.	The placement of the sensor does not depend on the network layer.	The placement of the sensor nodes is based on the network layer as in Algorithm 1.
Does not propose an unequal clustering mechanism. The equal clustering mechanism was proposed instead.	Propose unequal clustering mechanism based on competition radius: The calculation of the competition radius for each node depends on: • The residual energy of sensor nodes. • The distance $\boldsymbol{d_{i,BS}}$ from all the sensor nodes to the base station node.	Propose unequal clustering mechanism is based on the competition radius $\boldsymbol{R_c(i) suchasinequation(}$ The calculation of the competition radius for each node depends on: • The residual energy of sensor nodes $\boldsymbol{E_{rem}(i)}$. • The distance $\boldsymbol{d_{i,BS}}$ from all the sensor nodes to the base station node.	Propose unequal clustering mechanism is based on the competition radius $\boldsymbol{R_c(i)\ such\ as\ in\ equation}$ The calculation of the competition radius for each node depends on: • The residual energy of each node at each round $\boldsymbol{E_{rem}(i,r)}$. • The distance $\boldsymbol{d_{i,BS}}$ from all the sensor nodes to the base station node. • The minimum distance $\mathbf{d_{min}}$ of the closest node from the base station. • The maximum distance $\mathbf{d_{max}}$ of the furthest node from the base station. • The maximum capacity of node energy $\boldsymbol{E_{max}}$.
FLEACH utilizes a double cluster head node in order to reduce the load on primary CH. The selection of CHs is as follows: •During the cluster formation phase, a CH is randomly and alternately selected among the network nodes based on probability.	EEFUC utilizes one CH for aggregation and forwards data transmission at the same time to the base station.	UDCH utilizes a double cluster head node in order to reduce the load on primary CH. The selection of CHs is as follows: • The selection of primary CH is based on the minimum computing delay time $\boldsymbol{D_t(i)}$ of each node.	Our EEUCB utilizes a double cluster head node in order to reduce the load on primary CH. The selection of CHs is as follows: • The Primary CH selection method for the proposed EEUCB is similar to the method in UDCH. However, in EEUCB, further improvement is done by considering the Sleep Awake mechanism based on the distance from sensor nodes to CH, and the energy level of sensor nodes.

• The selection of 2CH is based on calculating the highest residual energy of nodes.		• The selection of 2CH is based on the distance from the sensor nodes to the primary CH.	• The selection of 2CH is based on calculating the highest residual energy of nodes.
Transmission round between CMs and CHs use distance threshold $\boldsymbol{D_{th}}$ and use the distance threshold to construct a path to BS in the network.	Transmission round between CMs and CHs, and for construct a path to BS in the network use residual energy of sensor node and distance from CH to BS.	The transmission round between CMs and CHs use the average energy threshold $\boldsymbol{E_{th}}$ and use the average energy to construct a path to BS in the network.	Transmission round between CMs and CHs use average distance threshold $\boldsymbol{D_{th}}$, average energy threshold $\boldsymbol{E_{th}}$. While uses the layer implementations and residual energy to construct a path to BS in the network.

5.8 Chapter Summary

This chapter discussed the structural components and modules of the proposed EEUCB protocol. Its salient features of route computation based on the developed EEUCB protocol and capability to reduce energy consumption and to prolong the network lifetime. In the processing phase of EEUCB protocol, The BS calculates the distance difference of farthest and closest nodes from the BS and divide it into four (4) layers empirically to estimate the distance between cluster nodes and neighbor nodes and also the number of neighbor nodes. The initialization phase process generates unequal clustering in the network by calculating the radius of clustering to balanced energy consumption and electing the primary CH by calculating the nodes' delay time. In the cluster setup phase, the node has become primary CH depends on the delay in processing and further improvement by taking into account the sleep and awake mechanism. In addition, this phase also includes the selection of 2CH based on the highest residual energy to reduce the load of the primary CH in the network. Lastly, the transmission phase is the process of data transmission between CMs and CHs through the network based on energy threshold, distance threshold, and the use of the layering implementation to construct the path to the BS. These phases increase the efficiency of energy consumption and increase the network lifetime.

Four main scenarios related to network energy efficiency and scalability were considered in evaluating the effects of different parameters on the performance of the proposed EEUCB protocol. The evaluation results of EEUCB were compared with those of UDCH, EEFUC, FLEACH, and LEACH by using several performance metrics related to energy efficiency and network lifetime. The proposed EEUCB protocol exhibited the best performance results among the compared routing schemes. Such superiority is attributed to the capability of EEUCB to reduce energy consumption and prolong the network lifetime.

Consequently, EEUCB enables the system to achieve high reliability in different WSN scenarios. In the next chapter, the conclusions of this thesis, along with recommendations for future work, are presented.

CHAPTER 6: CONCLUSION AND RECOMMENDATIONS

6.1 Conclusions

WSNs have grown to be an attractive and interesting field of research in both the industry and academia, mainly due to the decentralization and dynamism of WSNs. However, WSNs are still vulnerable to various attacks because of their distributed wireless nature. Hence, the results in delays, loss of data in the network, and increase the energy consumption of nodes. Therefore, network security and energy consumption are very important in WSNs to secure the data from attacks. Based on the mentioned above, two of the main issues that were addressed in implementing the clustering protocol of WSNs are authentication and energy efficiency issues. The first part of this thesis focused on how to overcome authentication and energy issues due to the security issues of WSNs. Numerous security techniques and approaches have been successfully proposed. However, there are other challenging problems associated with these security techniques, such as the secure data aggregation in clustering proposed without addressing the authentication issues. It is challenging to implement authentication while preserving the energy consumption in the network. The well-known SDA, SDAT, SDALFA, EESSDA, SDAACA, and EESDA are proposed techniques for secure data aggregation in clustering that are expected to be effective solutions to the aforementioned authentication issues among sensor nodes in the network. However, these techniques provide security for the sink nodes but do not support the security for all nodes, they share the same security key and the key length with a base station node and pay not much attention in enhancing the authentication of the Medium Access Control (MAC) address. Therefore, SEEDA is proposed in this thesis to address these limitations. This protocol aims to make full utilization of the advantages associated with secure data aggregation in clustering to increase the malicious nodes' detection rate. Furthermore, managing efficient energy consumption and redundancy of data during the transmission in the network can be

challenging when sending a packet between nodes. In addition, the battery-powered sensor node has limited energy and a complicated battery changing procedure; these affect the quality, performance, and lifetime of WSNs. Therefore, algorithms or security techniques should be highly efficient in terms of energy consumption. Therefore, the first part of this thesis was focused on authentication and energy efficiency together due to their overlapping in security issues of WSNs. Hence, to increase the detection rate of malicious nodes, the energy consumption of nodes in the network should be reduced and vice versa.

Aside from considering the authentication and energy issues, this thesis also focused on another important energy efficiency issue, which is the hot spots problem. This issue started because the sensor nodes closer to the base station nodes will take on more forwarding tasks. This will result in a massive overhead of the sensor nodes, and these nodes will run out of power sooner than the others. It causes a breakdown of the nodes and a loss of communication between sensor nodes; this breakdown is called the hot spots problem. Many techniques and approaches were proposed to reduce the hot spots problem, energy consumption, and prolong the network lifetime. However, there are other challenging problems associated with these techniques, such as the energy-efficient technique, which is proposed without much attention given to address the problem of hot spots. Most of the existing clustering techniques, such as LEACH, FLEACH, EEFUC, and UDCH utilize residual energy and distance of sensor nodes to the base station, but not much attention is given to enhance the data transmission process between cluster members, cluster heads, and base station. This would lead to the imbalance of energy distribution among nodes in the network. Therefore, EEUCB is proposed in this thesis to address these limitations. The proposed EEUCB protocol also considered minimum and maximum distance from the sensor node to the base station to calculate the distance difference among the sensor nodes from the base station and empirically divided it into

four layers to transmit data from the cluster head to the base station, rather than only the pure distance. This leads to avoiding the long-distance, which leads to an extension of the network's lifetime and improvement of network stability. However, the closest and farthest distance of the nodes from the base station was not considered in the previous techniques, which led to energy wastage across the network nodes and a reduced network lifetime. Moreover, the previous techniques do not use the sleep-awake mechanism. Therefore, EEUCB is proposed to address these limitations.

Accordingly, this chapter concludes the thesis by reflecting on the research objectives set in the first chapter. In addition, it offers some possible and worthy recommendations for future research directions to extend the work presented in this thesis.

6.2 Achievement of the Research Objectives

The security problem related to authentication and the energy efficiency issues, and other important energy efficiency issues such as the hot spots problem and the distance among nodes are all tackled in this research. Four objectives were defined in Section 1.5 to achieve the main goal of this thesis. The following discussions demonstrate how each one of the objectives is met in the research study:

Objective 1: *To study existing, state-of-the-art clustering protocols for WSNs:*

This objective has been achieved by investigating the related literature of secure data aggregation in clustering protocols and unequal clustering techniques and approaches. Several representatives for secure data aggregation and unequal clustering techniques were analyzed, and their key features, merits, and demerits were highlighted. Based on the review of the existing data aggregation in clustering protocols and the results obtained from their performance evaluation, the issues related to authentication, detection rate, and balanced energy consumption has become clearer. The contribution of this objective has

been achieved by a taxonomy of secure data aggregation techniques and unequal clustering protocols with challenges. The structure of each technique and approach was summarized and discussed in Chapter 2.

Objective 2: *To enhance authentication for secure data aggregation and enabling efficient energy usage of nodes in the cluster based on the MAC address:*

The second objective of this research work focuses on providing a step-by-step procedure for presenting secure data aggregation in clustering protocols for WSNs. This scheme was developed and named as secure and energy-efficient data aggregation in WSNs using an access control model (SEEDA). Instead of sharing the security key that can be exposed as a security threat, the proposed SEEDA protocol improves the MAC address by utilizing a secret key and random timestamp in the verification process. The random timestamp is proposed to avoid duplication of the data packet by malicious nodes that exploits typical event occurrences. The base station node verifies the fake aggregated data before sending it to the server. Other than that, the base station nodes check the distance and the timestamp of all the nodes broadcasted by the cluster head node. If they are different from the recorded value, they can be regarded as an adversary. The adversary can also be detected by comparing the transmission time and distance. Usually, faraway nodes would have a higher transmission time compared to the nearer nodes. These operations reduce redundant data transmission and energy consumption, and also help prolong the network lifetime. Our protocol consists of three algorithms: data fragmentation, secure node authentication, and fully homomorphic encryption algorithms. The data fragmentation algorithm partitions the data into small pieces before transmitting them to the next hop nodes to hide them from being attacked. The secure node authentication algorithm checks the authentication of the node that is leaving or joining the network to prevent any tampering or interrupting of the data transmission between nodes. The fully homomorphic

encryption algorithm encrypts the aggregated data before sending it to the base station nodes. Additionally, the proposed protocol reduces energy consumption using an access control model by reducing the communication overhead. The core and auxiliary functionalities of the proposed SEEDA protocol are discussed in detail in Chapter 4.

Objective 3: *To propose a clustering protocol with a double cluster head technique based on a balanced energy data transmission process for clustering that is able to reduce energy consumption and prolong network lifetime in WSNs:*

The third objective has been achieved by investigating the energy efficiency and solutions to the hot spots problem by proposing a protocol called an Energy-Efficient Unequal Clustering protocol based on a Balanced energy method (EEUCB). Unlike the UDCH technique, the proposed EEUCB protocol utilizes an unequal clustering mechanism based on the competition radius. The calculation of the competition radius for each node depends on a few factors: (i) the closest and farthest distance of the nodes from the BS; (ii) the residual energy of each node at each round; and (iii) the maximum capacity of the node energy. The utilization of the distance between nodes and the BS reduces and maintains energy consumption in a network.

Other than unequal clustering, the proposed EEUCB protocol also considers double cluster head implementation. Instead of utilizing the distance from the sensor nodes to primary CHs for selecting 2CHs, the EEUCB improves the selection of 2CHs by calculating the highest residual energy of nodes. This would reduce the overhead on the primary CH and enable them to distribute the operations between them.

Besides that, to balance the energy consumption among CMs and the CH, a clustering rotation strategy based on a few factors is proposed, namely the average energy threshold, average distance threshold, and the use of the layering implementation to construct the

path to the BS; this increases the efficiency of energy consumption and increases the network lifetime. A comprehensive discussion of the EEUCB protocol design architecture is provided in Chapter 5.

Objective 4: *To evaluate the proposed clustering protocols with different simulation scenarios and evaluation metrics:*

The final objective is met through several extensive simulation experiments conducted to test and evaluate the proposed schemes under different scenarios. The efficiency of the SEEDA protocol in terms of the authentication function and energy consumption was evaluated in Section 4.3. The results obtained from the simulation conducted in this study demonstrated that the proposed scheme has superiority over other conventional schemes in terms of the malicious activity detection rate, energy consumption, end-to-end delay, and resilience time. The simulation results show that the proposed SEEDA method outperforms SDA, SDAT, SDALFA, EESSDA, SDAACA, and EESDA with 98.84% malicious nodes detection rate, 3.04 joules for energy consumption, the maximum delay of 0.038 seconds, and the resilient time of 0.054 to 0.075 seconds when 8% to 16% of malicious nodes are affecting the network.

Correspondingly, with the integration of authentication, and energy efficiency for secure data aggregation in the network, the developed EEUCB protocol has likewise succeeded in solving the hot spots problem, avoiding the nodes with long-distance, and balanced energy consumption among nodes, thus preserving the energy and network lifetime in large-scale WSNs scenarios.

The experiment results showed that the EEUCB protocol outperforms LEACH, FLEACH, EEFUC, and UDCH protocols in terms of network lifetime. The EEUCB has achieved 57.75%, 19.75%, 14.7%, and 13.06% against LEACH, FLEACH, EEFUC, and

UDCH, respectively. A detailed discussion of the performance evaluation and comparison of the EEUCB with LEACH, FLEACH, EEFUC, and UDCH protocols is provided in Section 5.6.

6.3 Recommendations for Future Work

Two techniques of clustering protocol centered on authentication, balanced energy consumption, and prolong network lifetime for WSNs were proposed in this research. The simulation result shows that our proposed protocols outperform other previous techniques in terms of authentication and energy efficiency. However, there are still some other unsolved issues needs to be addressed in the secure data aggregation techniques and unequal clustering protocols in WSNs, which may be considered for further research. These unsolved issues include the followings:

(i) We enhanced the authentication and prevented Sybil and Sinkhole attacks in secure data aggregation protocol. Therefore, the proposed protocol's improvement will focus on preventing more attacks and solving more challenges, especially when involving mobile nodes.

(ii) The time complexity for the cluster heads' election were high, whereas the space complexity was constant in unequal clustering protocol. Therefore, new techniques may be required to reduce the time complexity for the election of cluster heads.

(iii) Several parameters were used in the proposed schemes. Further investigation to find the optimum values of these parameters using optimization techniques with the aim to enhance the detection rate of malicious nodes and prolong network lifetime based on multiple objectives.

REFERENCES

Abbasi, A. A., & Younis, M. (2007). A survey on clustering algorithms for wireless sensor networks. *Computer Communications, 30*(14), 2826-2841.

Abdollahzadeh, S., & Navimipour, N. J. (2016). Deployment strategies in the wireless sensor network: A comprehensive review. *Computer communications, 91*, 1-16.

Abdulsalam, H. M., & Kamel, L. K. (2010). *W-LEACH: Weighted Low Energy Adaptive Clustering Hierarchy aggregation algorithm for data streams in wireless sensor networks.* Paper presented at the 2010 IEEE International Conference on Data Mining Workshops.

Abidin, S., Vadi, V. R., & Rana, A. (2021). On Confidentiality, Integrity, Authenticity, and Freshness (CIAF) in WSN. In *Advances in Computer, Communication and Computational Sciences* (pp. 87-97): Springer.

Abo-Zahhad, M., Ahmed, S. M., Sabor, N., & Sasaki, S. (2015). Mobile sink-based adaptive immune energy-efficient clustering protocol for improving the lifetime and stability period of wireless sensor networks. *IEEE Sensors Journal, 15*(8), 4576-4586.

Abood, M. S., Wang, H., Mahdi, H. F., Hamdi, M. M., & Abdullah, A. S. (2021). *Review on secure data aggregation in Wireless Sensor Networks.* Paper presented at the IOP Conference Series: Materials Science and Engineering.

Aftab, M. U., Ashraf, O., Irfan, M., Majid, M., Nisar, A., & Habib, M. A. (2015). A review study of wireless sensor networks and its security. *Communications and Network, 7*(04), 172.

Ahmad, B., Shiwei, M., Qi, F., Meixi, W., & Ling, R. (2016). An Accurate Global Time Synchronization Method in Wireless Sensor Networks. In *Theory, Methodology, Tools and Applications for Modeling and Simulation of Complex Systems* (pp. 17-24): Springer.

Al-Shaikhi, A., & Masoud, A. (2017). Efficient, single hop time synchronization protocol for randomly connected WSNs. *IEEE Wireless Communications Letters, 6*(2), 170-173.

Albath, J., & Madria, S. (2009). *Secure hierarchical data aggregation in wireless sensor networks.* Paper presented at the 2009 IEEE Wireless Communications and Networking Conference.

Ali, A. W. (2015). *Energy efficieny in routing protocol and data collection approaches for WSN: A survey.* Paper presented at the International Conference on Computing, Communication & Automation.

Alsaedi, N., Hashim, F., Sali, A., & Rokhani, F. Z. (2017). Detecting sybil attacks in clustered wireless sensor networks based on energy trust system (ETS). *Computer communications, 110*, 75-82.

Amish, P., & Vaghela, V. (2016). Detection and prevention of wormhole attack in wireless sensor network using AOMDV protocol. *Procedia Computer Science, 79*, 700-707.

Amodu, O. A., & Mahmood, R. A. R. (2018). Impact of the energy-based and location-based LEACH secondary cluster aggregation on WSN lifetime. *Wireless Networks, 24*(5), 1379-1402.

Amutha, J., Sharma, S., & Nagar, J. (2020). WSN strategies based on sensors, deployment, sensing models, coverage and energy efficiency: Review, approaches and open issues. *Wireless Personal Communications, 111*(2), 1089-1115.

Anwar, R. W., Bakhtiari, M., Zainal, A., Abdullah, A. H., & Qureshi, K. N. (2014). Security issues and attacks in wireless sensor network. *World Applied Sciences Journal, 30*(10), 1224-1227.

Anwar, R. W., Bakhtiari, M., Zainal, A., Abdullah, A. H., & Qureshi, K. N. (2015). *Enhanced trust aware routing against wormhole attacks in wireless sensor networks.* Paper presented at the 2015 International Conference on Smart Sensors and Application (ICSSA).

Arora, S., & Hussain, M. (2018). *Secure session key sharing using symmetric key cryptography.* Paper presented at the 2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI).

Australia fires. (2020, January 31). *BBC*. Retrieved from https://www.bbc.com/news/world-australia-50951043

Bade, A. M., & Garba, A. A. (2019). A REVIEW ON SECURITY ISSUES IN WIRELESS SENSOR NETWORKS.

Bagci, H., & Yazici, A. (2013). An energy aware fuzzy approach to unequal clustering in wireless sensor networks. *Applied Soft Computing, 13*(4), 1741-1749.

Bala, T., Bhatia, V., Kumawat, S., & Jaglan, V. (2018). A survey: Issues and challenges in wireless sensor network. *Int. J. Eng. Technol, 7*(2-4), 53.

Baranauckas, B. C. (Producer). (2007, August 2). Bridge Collapse in Minneapolis Kills at Least 7. *The New York Times*. Retrieved from https://www.nytimes.com/2007/08/02/us/02bridge.html

Barsocchi, P., Bartoli, G., Betti, M., Girardi, M., Mammolito, S., Pellegrini, D., & Zini, G. (2020). Wireless sensor networks for continuous structural health monitoring of historic masonry towers. *International Journal of Architectural Heritage*, 1-23.

Battat, N., Seba, H., & Kheddouci, H. (2014). Monitoring in mobile ad hoc networks: A survey. *Computer networks, 69*, 82-100.

Beg, A., Al-Kharobi, T., & Al-Nasser, A. (2019). *Performance Evaluation and Review of Lightweight Cryptography in an Internet-of-Things Environment.* Paper presented

at the 2019 2nd International Conference on Computer Applications & Information Security (ICCAIS).

Bhushan, B., & Sahoo, G. (2018). Recent advances in attacks, technical challenges, vulnerabilities and their countermeasures in wireless sensor networks. *Wireless Personal Communications, 98*(2), 2037-2077.

Bisht, N., & Singh, S. (2015). A comparative study of some symmetric and asymmetric key cryptography algorithms. *International journal of innovative research in science, engineering and technology, 4*(3), 1028-1031.

Bouabdallah, N., Rivero-Angeles, M. E., & Sericola, B. (2009). Continuous monitoring using event-driven reporting for cluster-based wireless sensor networks. *IEEE transactions on vehicular technology, 58*(7), 3460-3479.

Boudia, O. R. M., Senouci, S. M., & Feham, M. (2015). A novel secure aggregation scheme for wireless sensor networks using stateful public key cryptography. *Ad Hoc Networks, 32*, 98-113.

Bozorgi, S. M., & Bidgoli, A. M. (2019). HEEC: A hybrid unequal energy efficient clustering for wireless sensor networks. *Wireless Networks, 25*(8), 4751-4772.

Bozorgi, S. M., Rostami, A. S., Hosseinabadi, A. A. R., & Balas, V. E. (2017). A new clustering protocol for energy harvesting-wireless sensor networks. *Computers & Electrical Engineering, 64*, 233-247.

Burhanuddin, M., Mohammed, A. A.-J., Ismail, R., Hameed, M. E., Kareem, A. N., & Basiron, H. (2018). A review on security challenges and features in wireless sensor networks: IoT perspective. *Journal of Telecommunication, Electronic and Computer Engineering (JTEC), 10*(1-7), 17-21.

Butun, I., & Sankar, R. (2011). *A brief survey of access control in wireless sensor networks.* Paper presented at the 2011 IEEE Consumer Communications and Networking Conference (CCNC).

Çam, H., Özdemir, S., Nair, P., Muthuavinashiappan, D., & Sanli, H. O. (2006). Energy-efficient secure pattern based data aggregation for wireless sensor networks. *Computer communications, 29*(4), 446-455.

Castillo-Effer, M., Quintela, D. H., Moreno, W., Jordan, R., & Westhoff, W. (2004). *Wireless sensor networks for flash-flood alerting.* Paper presented at the Proceedings of the Fifth IEEE International Caracas Conference on Devices, Circuits and Systems, 2004.

Chen, G., Li, C., Ye, M., & Wu, J. (2009). An unequal cluster-based routing protocol in wireless sensor networks. *Wireless Networks, 15*(2), 193-207.

Chen, J. (2011). *Improvement of LEACH routing algorithm based on use of balanced energy in wireless sensor networks.* Paper presented at the International Conference on Intelligent Computing.

Chen, W.-P., Hou, J. C., & Sha, L. (2004). Dynamic clustering for acoustic target tracking in wireless sensor networks. *IEEE Transactions on Mobile Computing, 3*(3), 258-271.

Choudhari, M. R., & Rote, U. (2021). *Data Aggregation Approaches in WSNs.* Paper presented at the 2021 International Conference on Computer Communication and Informatics (ICCCI).

Čolaković, A., & Hadžialić, M. (2018). Internet of Things (IoT): A review of enabling technologies, challenges, and open research issues. *Computer networks, 144*, 17-39.

Cui, J., Shao, L., Zhong, H., Xu, Y., & Liu, L. (2018). Data aggregation with end-to-end confidentiality and integrity for large-scale wireless sensor networks. *Peer-to-Peer Networking and Applications, 11*(5), 1022-1037.

Dabhade, V. D., & Alvi, A. (2021). Review of wireless sensor network security schemes. In *Intelligent Computing and Networking* (pp. 41-51): Springer.

Dehkordi, S. A., Farajzadeh, K., Rezazadeh, J., Farahbakhsh, R., Sandrasegaran, K., & Dehkordi, M. A. (2020). A survey on data aggregation techniques in IoT sensor networks. *Wireless Networks, 26*(2), 1243-1263.

Dharini, N., Balakrishnan, R., & Renold, A. P. (2015). *Distributed detection of flooding and gray hole attacks in Wireless Sensor Network.* Paper presented at the 2015 International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM).

Eschenauer, L., & Gligor, V. D. (2002). *A key-management scheme for distributed sensor networks.* Paper presented at the Proceedings of the 9th ACM conference on Computer and communications security.

Gandara, R. B., Wang, G., & Utama, D. N. (2018). *Hybrid Cryptography on Wireless Sensor Network: A Systematic Literature Review.* Paper presented at the 2018 International Conference on Information Management and Technology (ICIMTech).

Garg, K. D., Saini, V., & Gupta, J. (2020). WSN Protocols, Research challenges in WSN, Integrated areas of sensor networks, security attacks in WSN. *European Journal of Molecular & Clinical Medicine, 7*(3), 5145-5153.

Gavric, Z., & Simic, D. (2018). Overview of DOS attacks on wireless sensor networks and experimental results for simulation of interference attacks. *Ingeniería e Investigación, 38*(1), 130-138.

Ghormare, S., & Sahare, V. (2015). *Implementation of data confidentiality for providing high security in Wireless Sensor Network.* Paper presented at the 2015 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS).

Ghosal, A., & DasBit, S. (2015). A lightweight security scheme for query processing in clustered wireless sensor networks. *Computers & Electrical Engineering, 41*, 240-255.

Ghosal, A., Halder, S., & DasBit, S. (2012). A dynamic TDMA based scheme for securing query processing in WSN. *Wireless Networks, 18*(2), 165-184.

Gielow, F., Jakllari, G., Nogueira, M., & Santos, A. (2015). Data similarity aware dynamic node clustering in wireless sensor networks. *Ad hoc networks, 24*, 29-45.

Gill, R. K., & Sachdeva, M. (2018). Detection of hello flood attack on LEACH in wireless sensor networks. In *Next-Generation Networks* (pp. 377-387): Springer.

Gopika, D., & Panjanathan, R. (2020). Energy efficient routing protocols for WSN based IoT applications: A review. *Materials Today: Proceedings*.

Grover, J., Sharma, M., & Shikha. (2014). Reliable SPIN in Wireless Sensor Network. *Proceedings of 3rd International Conference on Reliability, Infocom Technologies and Optimization*, 1-6.

Grover, J., & Sharma, S. (2016). *Security issues in wireless sensor network—a review.* Paper presented at the 2016 5th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO).

Gunduz, S., Arslan, B., & Demirci, M. (2015). *A review of machine learning solutions to denial-of-services attacks in wireless sensor networks.* Paper presented at the 2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA).

Guo, J., & Chen, X. (2011). *Survey on secure data aggregation for wireless sensor networks.* Paper presented at the Proceedings of 2011 IEEE International Conference on Service Operations, Logistics and Informatics.

Gupta, V., & Pandey, R. (2016). An improved energy aware distributed unequal clustering protocol for heterogeneous wireless sensor networks. *Engineering Science and Technology, an International Journal, 19*(2), 1050-1058.

Han, R., Yang, W., Wang, Y., & You, K. (2017). DCE: A distributed energy-efficient clustering protocol for wireless sensor network based on double-phase cluster-head election. *Sensors, 17*(5), 998.

Hari, P. B., & Singh, S. N. (2016). *Security issues in Wireless Sensor Networks: Current research and challenges.* Paper presented at the 2016 International Conference on Advances in Computing, Communication, & Automation (ICACCA)(Spring).

Haseeb, K., Islam, N., Almogren, A., Din, I. U., Almajed, H. N., & Guizani, N. (2019). Secret Sharing-Based Energy-Aware and Multi-Hop Routing Protocol for IoT Based WSNs. *IEEE Access, 7*, 79980-79988.

Haseeb, K., Ud Din, I., Almogren, A., & Islam, N. (2020). An energy efficient and secure IoT-based WSN framework: An application to smart agriculture. *Sensors, 20*(7), 2081.

Heinzelman, W. B. (2000). *Application-specific protocol architectures for wireless networks.* Massachusetts Institute of Technology,

Heinzelman, W. B., Chandrakasan, A. P., & Balakrishnan, H. (2002). An application-specific protocol architecture for wireless microsensor networks. *IEEE Transactions on wireless communications, 1*(4), 660-670.

Heinzelman, W. R., Chandrakasan, A., & Balakrishnan, H. (2000). *Energy-efficient communication protocol for wireless microsensor networks.* Paper presented at the Proceedings of the 33rd annual Hawaii international conference on system sciences.

Hong, J., Kook, J., Lee, S., Kwon, D., & Yi, S. (2009). T-LEACH: The method of threshold-based cluster head replacement for wireless sensor networks. *Information Systems Frontiers, 11*(5), 513-521.

Hsu, K., Leung, M.-K., & Su, B. (2008). *Security analysis on defenses against sybil attacks in wireless sensor networks.* Paper presented at the IEEE J.

Hu, H., Chen, Y., Ku, W.-S., Su, Z., & Chen, C.-H. J. (2009). Weighted trust evaluation-based malicious node detection for wireless sensor networks. *International Journal of Information and Computer Security, 3*(2), 132-149.

Hu, L., & Evans, D. (2003). *Secure aggregation for wireless networks.* Paper presented at the 2003 Symposium on Applications and the Internet Workshops, 2003. Proceedings.

Huang, S.-I., Shieh, S., & Tygar, J. (2010). Secure encrypted-data aggregation for wireless sensor networks. *Wireless Networks, 16*(4), 915-927.

Ilayaraja, M., Shankar, K., & Devika, G. (2017). A modified symmetric key cryptography method for secure data transmission. *International Journal of Pure and Applied Mathematics, 116*(10), 301-308.

Iqbal, U., & Mir, A. H. (2020). Secure and practical access control mechanism for WSN with node privacy. *Journal of King Saud University-Computer and Information Sciences*.

Islam, M. N. U., Fahmin, A., Hossain, M. S., & Atiquzzaman, M. (2020). Denial-of-Service Attacks on Wireless Sensor Network and Defense Techniques. *Wireless Personal Communications*, 1-29.

Jain, T. K., Saini, D. S., & Bhooshan, S. V. (2014). Cluster head selection in a homogeneous wireless sensor network ensuring full connectivity with minimum isolated nodes. *Journal of Sensors, 2014*.

Jariwala, V., Patel, H., Patel, P., & Jinwala, D. C. (2014). Integrity and privacy preserving secure data aggregation in wireless sensor networks. *International Journal of Distributed Systems and Technologies (IJDST), 5*(3), 77-99.

Jasim, A. A., Idris, M. Y. I. B., Azzuhri, S. R. B., Issa, N. R., Noor, N. B. M., Kakarla, J., & Amiri, I. S. (2019). Secure and Energy-Efficient Data Aggregation Method Based on an Access Control Model. *IEEE Access, 7*, 164327-164343.

KanagaSuba Raja, S., & Pushpa, S. X. (2020). A Review on detection mechanisms used in Wireless Sensor Network for DoS attacks.

Karthikeyan, B., Velumani, M., Kumar, R., & Inabathini, S. R. (2015). *Analysis of data aggregation in wireless sensor network.* Paper presented at the 2015 2nd International Conference on Electronics and Communication Systems (ICECS).

Kaur, M., & Munjal, A. (2020). Data aggregation algorithms for wireless sensor network: a review. *Ad Hoc Networks, 100*, 102083.

Kaushal, K., & Sahni, V. (2015). DoS attacks on different layers of WSN: A review. *International Journal of Computer Applications, 130*(17).

Kocakulak, M., & Butun, I. (2017). *An overview of Wireless Sensor Networks towards internet of things.* Paper presented at the 2017 IEEE 7th Annual Computing and Communication Workshop and Conference (CCWC).

Kumar, A., Dadheech, P., & Chaudhary, U. (2020). *Energy Conservation in WSN: A Review of Current Techniques.* Paper presented at the 2020 3rd International Conference on Emerging Technologies in Computer Engineering: Machine Learning and Internet of Things (ICETCE).

Kumar, M., Verma, S., & Lata, K. (2015). Secure data aggregation in wireless sensor networks using homomorphic encryption. *International Journal of Electronics, 102*(4), 690-702.

Kumar, P., Gurtov, A., Iinatti, J., Sain, M., & Ha, P. H. (2016). Access control protocol with node privacy in wireless sensor networks. *IEEE Sensors Journal, 16*(22), 8142-8150.

Kumar, R., Tripathi, S., & Agrawal, R. (2020). *A Review On Security in Wireless Sensor Network.* Paper presented at the 2020 International Conference on Emerging Smart Computing and Informatics (ESCI).

Kumar, S., Prateek, M., Ahuja, N. J., & Bhushan, B. (2014). MEECDA: multihop energy efficient clustering and data aggregation protocol for HWSN. *arXiv preprint arXiv:1408.3110*.

Kurmi, J., Verma, R. S., & Soni, S. (2017). An Approach for Data Aggregation Strategy in Wireless Sensor Network using MAC Authentication. *Advances in Computational Sciences and Technology, 10*(5), 1037-1047.

Kwon, D., Yu, S., Lee, J., Son, S., & Park, Y. (2021). WSN-SLAP: Secure and lightweight mutual authentication protocol for wireless sensor networks. *Sensors, 21*(3), 936.

Lai, W. K., Fan, C. S., & Lin, L. Y. (2012). Arranging cluster sizes and transmission ranges for wireless sensor networks. *Information Sciences, 183*(1), 117-131.

Li, C., Ye, M., Chen, G., & Wu, J. (2005). *An energy-efficient unequal clustering mechanism for wireless sensor networks.* Paper presented at the IEEE International Conference on Mobile Adhoc and Sensor Systems Conference, 2005.

Li, H., Li, K., Qu, W., & Stojmenovic, I. (2014). Secure and energy-efficient data aggregation with malicious aggregator identification in wireless sensor networks. *Future Generation Computer Systems, 37*, 108-116.

Li, X., Chen, D., Li, C., & Wang, L. (2015). Secure data aggregation with fully homomorphic encryption in large-scale wireless sensor networks. *Sensors, 15*(7), 15952-15973.

Liao, Y., Qi, H., & Li, W. (2013). Load-balanced clustering algorithm with distributed self-organization for wireless sensor networks. *IEEE Sensors Journal, 13*(5), 1498-1506.

Lin, H., Wang, L., & Kong, R. (2015). Energy efficient clustering protocol for large-scale sensor networks. *IEEE Sensors Journal, 15*(12), 7150-7160.

Lin, Y.-C., & Cheung, W.-F. (2020). Developing WSN/BIM-based environmental monitoring management system for parking garages in smart cities. *Journal of Management in Engineering, 36*(3), 04020012.

Lindsey, S., & Raghavendra, C. S. (2002). *PEGASIS: Power-efficient gathering in sensor information systems.* Paper presented at the Proceedings, IEEE aerospace conference.

Liu, X. (2012). A survey on clustering routing protocols in wireless sensor networks. *Sensors, 12*(8), 11113-11153.

Liu, X., & Shi, J. (2012). Clustering Routing Algorithms In Wireless Sensor Networks. *KSII Transactions on Internet and Information Systems (TIIS), 6*(7), 1735-1755.

Liu, X., Yu, J., Li, F., Lv, W., Wang, Y., & Cheng, X. (2019). Data aggregation in wireless sensor networks: from the perspective of security. *IEEE Internet of Things Journal*.

Logambigai, R., & Kannan, A. (2016). Fuzzy logic based unequal clustering for wireless sensor networks. *Wireless Networks, 22*(3), 945-957.

Loscri, V., Morabito, G., & Marano, S. (2005). *A two-levels hierarchy for low-energy adaptive clustering hierarchy (TL-LEACH).* Paper presented at the IEEE vehicular technology conference.

Madakam, S., Lake, V., Lake, V., & Lake, V. (2015). Internet of Things (IoT): A literature review. *Journal of Computer and Communications, 3*(05), 164.

Mahdi, O. A., Abdul Wahab, A. W., Idna Idris, M. Y., Abu znaid, A. M., Khan, S., Al-Mayouf, Y. R. B., & Guizani, N. (2016). A comparison study on node clustering techniques used in target tracking WSNs for efficient data aggregation. *Wireless Communications and Mobile Computing, 16*(16), 2663-2676.

Mahdi, O. A., Abdul Wahab, A. W., Idris, I., Yamani, M., Khan, S., Al-Mayouf, Y. R. B., & Guizani, N. (2016). A comparison study on node clustering techniques used in target tracking WSNs for efficient data aggregation. *Wireless Communications and Mobile Computing, 16*(16), 2663-2676.

Maraiya, K., Kant, K., & Gupta, N. (2011). Wireless sensor network: a review on data aggregation. *International Journal of Scientific & Engineering Research, 2*(4), 1-6.

Mirzaie, M., & Mazinani, S. M. (2018). MCFL: an energy efficient multi-clustering algorithm using fuzzy logic in wireless sensor network. *Wireless Networks, 24*(6), 2251-2266.

Mishra, S., & Thakkar, H. (2012). Features of WSN and Data Aggregation techniques in WSN: A Survey. *Int. J. Eng. Innov. Technol.(IJEIT), 1*(4), 264-273.

Mittal, N., Singh, U., & Sohi, B. S. (2017). A stable energy efficient clustering protocol for wireless sensor networks. *Wireless Networks, 23*(6), 1809-1821.

Mohan, B., & Dayananda, K. (2016). *Energy efficient clustering scheme with secure data aggregation for mobile Wireless Sensor Networks (EECSSDA).* Paper presented at the 2016 Online International Conference on Green Engineering and Technologies (IC-GET).

Mohapatra, H., & Rath, A. K. (2020). Survey on fault tolerance-based clustering evolution in WSN. *IET Networks, 9*(4), 145-155.

Mondal, S., Dutta, P., Ghosh, S., & Biswas, U. (2016). *Energy efficient rough fuzzy set based clustering and cluster head selection for WSN.* Paper presented at the 2016 2nd International Conference on Next Generation Computing Technologies (NGCT).

Mondal, S., Ghosh, S., & Biswas, U. (2016). *ACOHC: ant colony optimization based hierarchical clustering in wireless sensor network.* Paper presented at the 2016 international conference on emerging technological trends (ICETT).

Moorthy, R., Bangera, V., Amrin, Z., Avinash, N., & NS, K. R. (2020). *WSN in Defence Field: A Security Overview.* Paper presented at the 2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC).

Mykletun, E., Girao, J., & Westhoff, D. (2006). *Public key based cryptoschemes for data concealment in wireless sensor networks.* Paper presented at the 2006 IEEE International Conference on Communications.

Naeimi, S., Ghafghazi, H., Chow, C.-O., & Ishii, H. (2012). A survey on the taxonomy of cluster-based routing protocols for homogeneous wireless sensor networks. *Sensors, 12*(6), 7350-7409.

Nithya, B. (2020). Cluster Based Key Management Schemes in Wireless Sensor Networks: A Survey. *Procedia Computer Science, 171*, 2684-2693.

Okay, F. Y., & Ozdemir, S. (2018). *A secure data aggregation protocol for fog computing based smart grids.* Paper presented at the 2018 IEEE 12th International Conference on Compatibility, Power Electronics and Power Engineering (CPE-POWERENG 2018).

Önen, M., & Molva, R. (2007). *Secure data aggregation with multiple encryption.* Paper presented at the European Conference on Wireless Sensor Networks.

Osanaiye, O. A., Alfa, A. S., & Hancke, G. P. (2018). Denial of service defence for resource availability in wireless sensor networks. *IEEE Access, 6*, 6975-7004.

Othman, S. B., Trad, A., Alzaid, H., & Youssef, H. (2013). Secure and energy-efficient data aggregation for wireless sensor networks. *International Journal of Mobile Network Design and Innovation, 5*(1), 28-42.

Othman, S. B., Trad, A., Youssef, H., & Alzaid, H. (2013). *Secure data aggregation with MAC authentication in wireless sensor networks.* Paper presented at the 2013 12th IEEE International Conference on Trust, Security and Privacy in Computing and Communications.

Ould Amara, S., Beghdad, R., & Oussalah, M. (2013). Securing wireless sensor networks: A survey. *EDPACS, 47*(2), 6-29.

Ozdemir, S. (2007). *Concealed data aggregation in heterogeneous sensor networks using privacy homomorphism.* Paper presented at the IEEE international conference on pervasive services.

Ozdemir, S., & Xiao, Y. (2009). Secure data aggregation in wireless sensor networks: A comprehensive overview. *Computer networks, 53*(12), 2022-2037.

Ozdemir, S., & Xiao, Y. (2011). Integrity protecting hierarchical concealed data aggregation for wireless sensor networks. *Computer networks, 55*(8), 1735-1746.

Panda, M., & Khilar, P. M. (2015). Distributed Byzantine fault detection technique in wireless sensor networks based on hypothesis testing. *Computers & Electrical Engineering, 48*, 270-285.

Pardesi, P., & Grover, J. (2015). *Improved multiple sink placement strategy in wireless sensor networks.* Paper presented at the 2015 International Conference on Futuristic Trends on Computational Analysis and Knowledge Management (ABLAZE).

Parmar, K., & Jinwala, D. C. (2014). Symmetric-key based homomorphic primitives for end-to-end secure data aggregation in wireless sensor networks. *Journal of Information Security, 6*(01), 38.

Pathan, A.-S. K., Lee, H.-W., & Hong, C. S. (2006). *Security in wireless sensor networks: issues and challenges.* Paper presented at the 2006 8th International Conference Advanced Communication Technology.

Pawar, M., & Agarwal, J. (2017). A literature survey on security issues of WSN and different types of attacks in network. *Indian J. Comput. Sci. Eng, 8*(2), 80-83.

Phoemphon, S., So-In, C., Aimtongkham, P., & Nguyen, T. G. (2020). An energy-efficient fuzzy-based scheme for unequal multihop clustering in wireless sensor networks. *JOURNAL OF AMBIENT INTELLIGENCE AND HUMANIZED COMPUTING*.

Prakash, S., & Rajput, A. (2018). Hybrid cryptography for secure data communication in wireless sensor networks. In *Ambient Communications and Computer Systems* (pp. 589-599): Springer.

Prathima, E., Prakash, T. S., Venugopal, K., Iyengar, S., & Patnaik, L. (2016). SDAMQ: secure data aggregation for multiple queries in wireless sensor networks. *Procedia Computer Science, 89*, 283-292.

Pukhrambam, P., Bhattacharjee, S., & Das, H. S. (2017). *A Multi-level Weight Based Routing Algorithm for Prolonging Network Lifetime in Cluster Based Sensor Networks.* Paper presented at the Proceedings of the International Conference on Signal, Networks, Computing, and Systems.

Qazi, R., Qureshi, K. N., Bashir, F., Islam, N. U., Iqbal, S., & Arshad, A. (2021). Security protocol using elliptic curve cryptography algorithm for wireless sensor networks. *JOURNAL OF AMBIENT INTELLIGENCE AND HUMANIZED COMPUTING, 12*(1), 547-566.

Qin, D., Zhang, Y., Ma, J., Ji, P., & Feng, P. (2018). A distributed collision-free data aggregation scheme for wireless sensor network. *International Journal of Distributed Sensor Networks, 14*(8), 1550147718795847.

Qing, L., Zhu, Q., & Wang, M. (2006). Design of a distributed energy-efficient clustering algorithm for heterogeneous wireless sensor networks. *Computer communications, 29*(12), 2230-2237.

Rafik, M. B. O., & Mohammed, F. (2013). *Fast and secure implementation of ECC-based concealed data aggregation in WSN.* Paper presented at the Global Information Infrastructure Symposium-GIIS 2013.

Raja, K. N., & Beno, M. M. (2017). Secure data aggregation in wireless sensor network-Fujisaki Okamoto (FO) authentication scheme against sybil attack. *Journal of medical systems, 41*(7), 1-6.

Ran, G., Zhang, H., & Gong, S. (2010). Improving on LEACH protocol of wireless sensor networks using fuzzy logic. *Journal of Information & Computational Science, 7*(3), 767-775.

Rana, S., & Dudhgoankar, A. (2017). Review for Secure Data Aggregation In Wireless Sensor Networks.

Rani, S., & Kaur, H. (2017). Technical review on symmetric and asymmetric cryptography algorithms. *International Journal of Advanced Research in Computer Science, 8*(4).

Rashid, B., & Rehmani, M. H. (2016). Applications of wireless sensor networks for urban areas: A survey. *Journal of Network and Computer Applications, 60*, 192-219.

Rawat, P., Singh, K. D., Chaouchi, H., & Bonnin, J. M. (2014). Wireless sensor networks: a survey on recent developments and potential synergies. *The Journal of supercomputing, 68*(1), 1-48.

Raymond, D. R., & Midkiff, S. F. (2008). Denial-of-service in wireless sensor networks: Attacks and defenses. *IEEE Pervasive Computing, 7*(1), 74-81.

Razaque, A., & Rizvi, S. S. (2017). Secure data aggregation using access control and authentication for wireless sensor networks. *computers & security, 70*, 532-545.

Ren, Q., & Yao, G. (2020). An energy-efficient cluster head selection scheme for energy-harvesting wireless sensor networks. *Sensors, 20*(1), 187.

Rezvani, M., Ignjatovic, A., Bertino, E., & Jha, S. (2014). Secure data aggregation technique for wireless sensor networks in the presence of collusion attacks. *IEEE Transactions on Dependable and Secure Computing, 12*(1), 98-110.

Rezvani, M., Ignjatovic, A., Bertino, E., & Jha, S. (2015). Secure Data Aggregation Technique for Wireless Sensor Networks in the Presence of Collusion Attacks. *IEEE Transactions on Dependable and Secure Computing, 1*(12), 98-110.

Rodhe, I., & Rohner, C. (2008). *n-LDA: n-layers data aggregation in sensor networks.* Paper presented at the 2008 The 28th International Conference on Distributed Computing Systems Workshops.

Roy, N. R., & Chandra, P. (2020). *Analysis of data aggregation techniques in wsn.* Paper presented at the International Conference on Innovative Computing and Communications.

Roy, S., Conti, M., Setia, S., & Jajodia, S. (2014). Secure data aggregation in wireless sensor networks: Filtering out the attacker's impact. *IEEE Transactions on Information Forensics and Security, 9*(4), 681-694.

Salehi, S. A., Razzaque, M. A., Naraei, P., & Farrokhtala, A. (2013). *Security in wireless sensor networks: Issues and challanges.* Paper presented at the 2013 IEEE International Conference on Space Science and Communication (IconSpace).

Sanli, H. O., Ozdemir, S., & Cam, H. (2004). *SRDA: secure reference-based data aggregation protocol for wireless sensor networks.* Paper presented at the IEEE 60th Vehicular Technology Conference, 2004. VTC2004-Fall. 2004.

Santhi, G., & Sowmiya, R. (2017). A survey on various attacks and countermeasures in wireless sensor networks. *Int J Comput Appl, 159*(7), 0975-8887.

Sarkar, A., & Murugan, T. S. (2019). Cluster head selection for energy efficient and delay-less routing in wireless sensor network. *Wireless Networks, 25*(1), 303-320.

Sen, J. (2010). A survey on wireless sensor network security. *arXiv preprint arXiv:1011.1529*.

SenthilKumar, U., & Senthilkumaran, U. (2016). Review of asymmetric key cryptography in wireless sensor networks. *International Journal of Engineering and Technology, 8*(2), 859-862.

Sert, S. A., Alchihabi, A., & Yazici, A. (2018). A two-tier distributed fuzzy logic based protocol for efficient data aggregation in multihop wireless sensor networks. *IEEE Transactions on Fuzzy Systems, 26*(6), 3615-3629.

Shafiei, H., Khonsari, A., Derakhshi, H., & Mousavi, P. (2014). Detection and mitigation of sinkhole attacks in wireless sensor networks. *Journal of Computer and System Sciences, 80*(3), 644-653.

Shafiq, M., Ashraf, H., Ullah, A., & Tahira, S. (2020). Systematic Literature Review on Energy Efficient Routing Schemes in WSN–A Survey. *Mobile Networks and Applications*, 1-14.

Shagari, N. M., Idris, M. Y. I., Salleh, R. B., Ahmedy, I., Murtaza, G., & Shehadeh, H. A. (2020). Heterogeneous Energy and Traffic Aware Sleep-Awake Cluster-Based Routing Protocol for Wireless Sensor Network. *IEEE Access, 8*, 12232-12252.

Sharma, A., & Sharma, S. (2016). A Comparative Review on Reliability and Fault Tolerance Enhancement Protocols in Wireless Sensor Networks. *International Research Journal of Engineering and Technology (IRJET), 3*(1), 622-626.

Sharma, G., Bala, S., & Verma, A. K. (2012). Security frameworks for wireless sensor networks-review. *Procedia Technology, 6*, 978-987.

Shim, K.-A. (2015). A survey of public-key cryptographic primitives in wireless sensor networks. *IEEE Communications Surveys & Tutorials, 18*(1), 577-601.

Shim, K.-A., & Park, C.-M. (2014). A secure data aggregation scheme based on appropriate cryptographic primitives in heterogeneous wireless sensor networks. *IEEE transactions on parallel and distributed systems, 26*(8), 2128-2139.

Shukla, A., & Tripathi, S. (2020). A multi-tier based clustering framework for scalable and energy efficient WSN-assisted IoT network. *Wireless Networks*, 1-23.

Simon, G., Maróti, M., Lédeczi, Á., Balogh, G., Kusy, B., Nádas, A., . . . Frampton, K. (2004). *Sensor network-based countersniper system.* Paper presented at the Proceedings of the 2nd international conference on Embedded networked sensor systems.

Singh, A. V., & Chattopadhyaya, M. (2015). *Mitigation of DoS attacks by using multiple encryptions in MANETs.* Paper presented at the 2015 4th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO)(Trends and Future Directions).

Singh, P., & Chauhan, R. (2017). A Survey on Comparisons of Cryptographic Algorithms Using Certain Parameters in WSN. *International Journal of Electrical & Computer Engineering (2088-8708), 7*(4).

Singh, S. P., & Sharma, S. (2015). A survey on cluster based routing protocols in wireless sensor networks. *Procedia computer science, 45*, 687-695.

Singh, V. P., Jain, S., & Singhai, J. (2010). Hello flood attack and its countermeasures in wireless sensor networks. *International Journal of Computer Science Issues (IJCSI), 7*(3), 23.

Sivakumar, N. R. (2020). Stabilizing Energy Consumption in Unequal Clusters of Wireless Sensor Networks. *CMC-COMPUTERS MATERIALS & CONTINUA, 64*(1), 81-96.

Sumathi, J., & Velusamy, R. L. (2020). A review on distributed cluster based routing approaches in mobile wireless sensor networks. *JOURNAL OF AMBIENT INTELLIGENCE AND HUMANIZED COMPUTING*, 1-15.

Sun, F., Zhao, Z., Fang, Z., Du, L., Xu, Z., & Chen, D. (2014). A review of attacks and security protocols for wireless sensor networks. *Journal of Networks, 9*(5), 1103.

Sun, X., Su, Y., Huang, Y., Tan, J., Yi, J., Hu, T., & Zhu, L. (2020). Edge Computing-Based ERBS Time Synchronization Algorithm in WSNs. *Wireless Communications and Mobile Computing, 2020*.

Sundaran, K., Ganapathy, V., & Sudhakara, P. (2017). *Fuzzy logic based unequal clustering in wireless sensor network for minimizing energy consumption.* Paper presented at the 2017 2nd International Conference on Computing and Communications Technologies (ICCCT).

Tripathy, A., Pradhan, S. K., Tripathy, A. R., & Nayak, A. K. (2019). A New Hybrid Cryptography Technique in Wireless Sensor Network. *International Journal of Innovative Technology and Exploring Engineering (IJITEE), 8*(10), 121-131.

Villas, L. A., Boukerche, A., Guidoni, D. L., De Oliveira, H. A., De Araujo, R. B., & Loureiro, A. A. (2013). An energy-aware spatio-temporal correlation mechanism to perform efficient data collection in wireless sensor networks. *Computer communications, 36*(9), 1054-1066.

Wang, T., Qin, X., & Liu, L. (2013). An energy-efficient and scalable secure data aggregation for wireless sensor networks. *International Journal of Distributed Sensor Networks, 9*(12), 843485.

Wazid, M., Das, A. K., Kumari, S., & Khan, M. K. (2016). Design of sinkhole node detection mechanism for hierarchical wireless sensor networks. *Security and Communication Networks, 9*(17), 4596-4614.

Wu, K., Dreef, D., Sun, B., & Xiao, Y. (2007). Secure data aggregation without persistent cryptographic operations in wireless sensor networks. *Ad Hoc Networks, 5*(1), 100-111.

Xiangning, F., & Yulin, S. (2007). *Improvement on LEACH protocol of wireless sensor network.* Paper presented at the Sensor Technologies and Applications, 2007. SensorComm 2007. International Conference on.

Xiaoyan, M. (2006). *Study and design on cluster routing protocols of wireless sensor networks.* Ph. D. Dissertation. Zhejiang University, Hangzhou, China,

Xie, H., Yan, Z., Yao, Z., & Atiquzzaman, M. (2018). Data collection for security measurement in wireless sensor networks: a survey. *IEEE Internet of Things Journal, 6*(2), 2205-2224.

Yadav, R., & Mishra, R. (2020). An authenticated enrolment scheme of nodes using blockchain and prevention of collaborative blackhole attack in WSN.

Yang, S.-K., Shiue, Y.-M., Su, Z.-Y., Liu, I.-H., & Liu, C.-G. (2020). An authentication information exchange scheme in WSN for IoT applications. *IEEE Access, 8*, 9728-9738.

Yang, Y., Wang, X., Zhu, S., & Cao, G. (2008). SDAP: A secure hop-by-hop data aggregation protocol for sensor networks. *ACM Transactions on Information and System Security (TISSEC), 11*(4), 1-43.

Yassein, M. B., Khamayseh, Y., & Mardini, W. (2009). *Improvement on LEACH protocol of wireless sensor network (VLEACH.* Paper presented at the Int. J. Digit. Content Technol. Appl. 2009.

Yick, J., Mukherjee, B., & Ghosal, D. (2005). *Analysis of a prediction-based mobility adaptive tracking algorithm.* Paper presented at the 2nd International Conference on Broadband Networks, 2005.

Yick, J., Mukherjee, B., & Ghosal, D. (2008). Wireless sensor network survey. *Computer networks, 52*(12), 2292-2330.

Younis, M., Youssef, M., & Arisha, K. (2003). Energy-aware management for cluster-based sensor networks. *Computer Networks, 43*(5), 649-668.

Younis, O., & Fahmy, S. (2004). HEED: a hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks. *IEEE Transactions on mobile computing, 3*(4), 366-379.

Younis, O., Krunz, M., & Ramasubramanian, S. (2006). Node clustering in wireless sensor networks: recent developments and deployment challenges. *IEEE network, 20*(3), 20-25.

Yu, J., Qi, Y., & Wang, G. (2011). An energy-driven unequal clustering protocol for heterogeneous wireless sensor networks. *Journal of Control Theory and Applications, 9*(1), 133-139.

Zhang, D.-G. (2012). A new approach and system for attentive mobile learning based on seamless migration. *Applied Intelligence, 36*(1), 75-89.

Zhang, P., Wang, J., Guo, K., Wu, F., & Min, G. (2018). Multi-functional secure data aggregation schemes for WSNs. *Ad Hoc Networks, 69*, 86-99.

Zhang, S., & Zhang, H. (2012). *A review of wireless sensor networks and its applications.* Paper presented at the 2012 IEEE international conference on automation and logistics.

Zhang, X., Heys, H. M., & Li, C. (2010). *Energy efficiency of symmetric key cryptographic algorithms in wireless sensor networks.* Paper presented at the 2010 25th Biennial symposium on communications.

Zhang, Y., Zhu, L.-n., & Feng, L. (2009). Key Management and Authentication in Ad Hoc Network based on Mobile Agent. *J. Networks, 4*(6), 487-494.

Zhong, H., Shao, L., Cui, J., & Xu, Y. (2018). An efficient and secure recoverable data aggregation scheme for heterogeneous wireless sensor networks. *Journal of Parallel and Distributed Computing, 111*, 1-12.

Zhu, F., & Wei, J. (2019). An energy-efficient unequal clustering routing protocol for wireless sensor networks. *International Journal of Distributed Sensor Networks, 15*(9), 1550147719879384.

Zhu, X., Shen, L., & Yum, T.-S. P. (2009). Hausdorff clustering and minimum energy routing for wireless sensor networks. *IEEE transactions on vehicular technology, 58*(2), 990-997.

Zungeru, A. M., Ang, L.-M., & Seng, K. P. (2012). Classical and swarm intelligence based routing protocols for wireless sensor networks: A survey and comparison. *Journal of Network and Computer Applications, 35*(5), 1508-1536.