ENERGY EFFICIENT AND LOAD-BALANCED ROUTING SCHEMES FOR IN-NETWORK DATA AGGREGATION IN WIRELESS SENSOR NETWORKS

OMAR ADIL MAHDI

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

2017

University of Malaya

ENERGY EFFICIENT AND LOAD-BALANCED ROUTING SCHEMES FOR IN-NETWORK DATA AGGREGATION IN WIRELESS SENSOR NETWORKS

OMAR ADIL MAHDI

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

2017

University of Malaya

UNIVERSITY OF MALAYA

ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: Omar Adil Mahdi

Matric No: WHA130048

Name of Degree: Doctor of Philosophy

Title of Project Paper/Research Report/Dissertation/Thesis (“this Work”): Thesis

Field of Study: Communication & Networks (Computer Science)

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: / 9 / 2017

Subscribed and solemnly declared before,

Witness’s Signature Date: / 9 / 2017

Name:

Designation:

University of Malaya

ABSTRACT

Wireless sensor networks (WSNs) constitute a group of small autonomous units known as sensor nodes within an infrastructure-less and self-configuring wireless network. In such networks, the deployment of high-density sensor nodes creates a possibility for close nodes to produce redundant data that leads to unnecessary energy consumption. Therefore, efficient energy management strategies that reduce data redundancy have been proposed by prior researchers. This includes the well-known Data Routing for In-Network Aggregation (DRINA) and Efficient Data Collection Aware of Spatio-Temporal Correlation (EAST) that utilize in-network data aggregation to reduce data redundancy. The two protocols maximize the formation of overlapping paths to increase aggregation rate. Their methods, however, do not consider the actual nodes energy resources and the traffic load that could result in early multi-hop relay nodes failure. This affects the network ability to forward packet successfully, thus, lead to unstable network structure. In this thesis, energy-efficient and load-balanced routing schemes are proposed. The proposed schemes extend the work in DRINA and EAST protocols. The first scheme is the Weighted Data Aggregation Routing Scheme (WDARS) that aims to maintain an acceptable trade-off between energy efficiency and routes overlapping by considering metrics related to energy efficiency and load balancing for multi-hop relay nodes and routes overlapping point selection. The second scheme is the Weight Aware Spatial-Temporal Correlation Routing Scheme (WST-RS) which exploits the spatial and temporal correlation among the sensor observations in a fully distributed manner. This thesis also introduced a new Multi-Criteria Node Weight (MCNW) metric to weight the node’s status in terms of various metrics that contributes as a link cost function for link quality assessment and selecting the lightweight routes to the destination depending on the status of nodes. The results show that WDARS scheme outperforms DRINA and InFRA in terms of energy consumption by 11.65% from the

University of Malaya

initial energy as compared to 21.75% and 35.71% respectively. Similarly, the nodes in the proposed WST-RS scheme records the lowest energy consumption with 16.25% from the initial energy compared to the DRINA (35.79%), YEAST-CF (24.83%), and EAST (20.58%). The first node to die in WST-RS is much longer, nearly 3.06, 2.22, 1.82 times longer than DRINA, YEAST-CF, EAST. This research will contribute toward effectively alleviating the energy constraints associated with WSNs by extending the network lifetime and increasing the availability of services even in highly dense networks with heavy traffic load.

University of Malaya

ABSTRAK

Rangkaian Sensor Tanpa Wayar (WSNs) merupakan sekumpulan kecil unit autonomi yang dikenali sebagai nod sensor tanpa infrastruktur dan berkonfigurasi kendiri. Dalam rangkaian tersebut, penempatan nod sensor berkepadatan tinggi berkemungkinan menyebabkan nod yang berdekatan menghasilkan lebihan data dan menjurus kepada penggunaan tenaga secara berlebihan. Oleh itu, kecekapan strategi pengurusan tenaga yang mengurangkan lebihan data telah dicadangkan oleh penyelidik terdahulu.. Ini termasuk Data Routing for In-Network Aggregation (DRINA) dan Efficient Data Collection Aware of Spatio-Temporal Correlation (EAST) yang terkenal dengan menggunakan pengagregatan data dalam rangkaian bagi mengurangkan lebihan data. Kedua-dua protocol ini memaksimumkan pertindihan antara laluan untuk meningkatkan kadar pengagregatan. Bagaimanapun kaedah tersebut tidak mempertimbangkan sumber tenaga sebenar nod dan beban trafik yang menyebabkan kegagalan awal bagi multi-hop relay nod. Ini memberi kesan pada keupayaan rangkaian untuk menghantar paket dengan jayanya, oleh itu, menyebabkan struktur rangkaian yang tidak stabil. Dalam tesis ini, skim penghalaan yang mempunyai kecekapan tenaga dan keseimbangan beban dicadangkan. Skim yang dicadangkan melanjutkan kerja dalam DRINA dan EAST. Skim yang pertama ialah Weighted Data Aggregation Routing Scheme (WDARS) yang bertujuan mengekalkan keseimbangan yang memuaskan antara kecekapan tenaga dan pertindihan laluan dengan mempertimbangkan metrik yang berkaitan dengan kecekapan tenaga dan pengimbangan beban untuk multi-hop relay nod dan pemilihan titik pertindihan laluan. Skim yang kedua ialah Weight aware Spatial-Temporal Correlation Routing Scheme (WST-RS) yang mengeksploitasi korelasi ruang dan masa antara pemerhatian sensor secara pengagihan sepenuhnya. Tesis ini juga memperkenalkan metrik Multi-Criteria Node Weight (MCNW) yang baru untuk mengukur status nod dari pelbagai metrik yang menyumbang pada fungsi kos untuk penilaian kualiti pautan dan

pemilihan laluan yang paling ringan ke destinasi berdasarkan status nod. Keputusan menunjukkan bahawa skim WDARS mengatasi prestasi DRINA dan InFRA dari segi penggunaan tenaga dari tenaga awal sebanyak 11.65% berbanding 21.75% dan 35.71%. Begitu juga nod dalam skim WST-RS yang dicadangkan dengan mencatatkan penggunaan tenaga terendah iaitu sebanyak 16.25% dari tenaga awal berbanding DRINA (35.79%), YEAST-CF (24.83%), dan EAST (20.58%). Nod pertama yang mati dalam WST-RS mengambil masa lebih lama, hampir 3.06 kali lebih lama berbanding DRINA, 2.22 kali lebih lama dari YEAST-CF, dan 1.82 kali lebih lama dari EAST. Hasil kajian ini menyumbang pada keberkesanan mengurangkan kekangan tenaga pada WSNs dengan memanjangkan jangka hayat rangkaian dan meningkatkan ketersediaan perkhidmatan walaupun dalam rangkaian yang sangat padat dengan beban trafik yang berat.

University of Malaya

ACKNOWLEDGEMENTS

First, my sincere appreciation and gratitude goes to ALLAH for giving me the strength, motivation and inspiration to fulfill my PhD study. Secondly, I would like to extend my sincere appreciation to my kind supervisor, Dr. Ainuddin Wahid Bin Abdul Wahab, 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 PhD Journey. It was a privilege to have work under your professional supervision. My gratitude also goes to my co-supervisor Dr. Mohd Yamani Idna Bin Idris 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 entire staff of Faculty of Computer Science and Information Technology for their assistance during my study years in UM.

There are also some other institutions as well as people that should be thanked for their remarkable help and support they have provided me with during years of the PhD study. A special thanks to my employer, the Ministry of Higher Education & Scientific Research, University of Baghdad, and to my homeland, IRAQ. The financial supports by research grants RP036 (A, B, C)-15AET and UM.C/625/1/HIR/MoE/FCSIT/17 from the Ministry of Education Malaysia is also gratefully acknowledged. I say many thanks to the University of Malaya for giving me the chance to study and do research in a highly scientific environment. It is also my pleasure to thank all my fellow graduate students for their encouragement and collaboration, especially, Dr. Suleman Khan, Dr. Aminu Mustapha Bagiwa, and Roziana Ramli.

Innumerable appreciations to all sacrificed their love and care so as to enlighten my own life. My appreciation and gratitude also goes to all my family members and relatives my mother Hayfaa Ibrahim, my brothers, my sisters, my darling wife Yusor Rafid Bahar and my loving son Zainulabdeen Omar, their love, affection, patience, encouragement and continuing support throughout these years kept me motivated.

Finally, I dedicate this work to the memory of my late father Adil Mahdi Wayis (1955–2007) and my late sister Mayada Adil Mahdi (1982–2007).

University of Malaya

TABLE OF CONTENTS

University of Malaya

University of Malaya

University of Malaya

University of Malaya

University of Malaya

LIST OF FIGURES

University of Malaya

University of Malaya

LIST OF TABLES

University of Malaya

LIST OF SYMBOLS AND ABBREVIATIONS

ADA	:	Adaptive Data Aggregation
APTEEN	:	Adaptive Periodic Threshold-sensitive Energy Efficient sensor Network protocol
α-SPT	:	Weighted Shortest Path Tree
B_{ava}	:	Available buffer memory
BS	:	Base Station
B_{total}	:	Total buffer size
CAG	:	Cluster AGgregation
CCM	:	Cluster Configuration Message
CH	:	Cluster Head
CM	:	Cluster Member
CoGKDA	:	Combined Grey model and Kalman Filter Data Aggregation
CPU	:	Central Processing Unit
CSMA	:	Carrier Sense Multiple Access
CTS	:	Clear-To-Send
DAA + RW	:	Data-Aware Anycast and Randomized Waiting
DCT	:	Discrete Cosine Transform
DRINA	:	Data Routing for In-Network aggregation
EAST	:	Efficient data collection Aware of Spatio-Temporal correlation
ECBSN	:	Energy Chain Based Sensor Network
Ed	:	Euclidean distance
EDGA	:	Effective Data Gathering Algorithm
EECDA	:	Energy Efficient Clustering and Data Aggregation
EEHCA	:	Energy Efficient Hierarchical Clustering Algorithm

University of Malaya

E_{init}	:	Initial battery energy level
E-LEACH	:	Energy aware Low-Energy Adaptive Clustering Hierarchy
E_{res}	:	Residual energy
GBDAS	:	Grid-Based Data Aggregation Scheme
GIT	:	Greedy Incremental Tree
GPS	:	Global Positioning System
GSC	:	Gridiron Spatial Correlation
HCM	:	Hop Configuration Message
HEAP	:	Hybrid Energy-efficient Aggregation Protocol
HEED	:	Hybrid Energy Efficient Distributed
H_T	:	Hard Threshold
HtS	:	Hop to Sink
HtT	:	Hop to Tree
InFRA	:	Information-Fusion-based Role Assignment
IoT	:	Internet of Things
LAN	:	Local Area Network
LEACH	:	Low-Energy Adaptive Clustering Hierarchy
LEACH-FL	:	Low-Energy Adaptive Clustering Hierarchy Fuzzy Logic
MAC	:	Medium Access Control
MCNW	:	Multi-Criteria Node Weight
MEMS	:	Micro-Electro-Mechanical Systems
MHR	:	Multi-Hop Relay
M-LEACH	:	Multi-hop Low-Energy Adaptive Clustering Hierarchy
Node_Dens	:	Node density in the event zone
NW	:	Node Weight
NW_B_{ava}	:	Available buffer-based node weight

University of Malaya

NW_E_{res}	:	Residual energy-based node weight
NW_HtS	:	Hop to sink-based node weight
NW_HtT	:	Hop to tree-based node weight
NW_P_r	:	Position-based node weight
NWAUF	:	Normalized Weighted Additive Utility Function
PAs	:	Primary Aggregators
PEGASIS	:	Power-Efficient GAthering in Sensor Information Systems
P_r	:	Position of node
QoS	:	Quality of Service
R_c	:	Communication range
R_e	:	Event's maximum radius
REM	:	Route Establishment Message
RN	:	Representative Node
RPM	:	Revolutions Per Minute
R_s	:	Sensing range
RTS	:	Request-To-Send
SAs	:	Secondary Aggregators
SCR	:	Spatial Correlated Region
SCR_Center	:	Spatial correlated region center
SCR_D	:	Spatial correlated region dimension
SCR_OP	:	Spatially correlated region ordered pair
SCR_Quad	:	Spatial correlated region quadrant
SCR_Size	:	Spatial correlated region size
SFEB	:	Structure-Free and Energy-Balanced data aggregation protocol
SPT	:	Shortest Path Tree
S_T	:	Soft Threshold

University of Malaya

TAG	:	Tiny AGgregation
T_C	:	Count time
TDMA	:	Time Division Multiple Access
TEEN	:	Threshold sensitive Energy Efficient sensor Network protocol
$Th_{reshold\_Eres}$	:	Residual energy threshold
$Th_{reshold\_tt}$	:	Temporal tolerance threshold
$Th_{reshold\_w}$	:	Weight threshold
T-LEACH	:	Threshold-based Low-Energy Adaptive Clustering Hierarchy
TL-LEACH	:	Two-level Low-Energy Adaptive Clustering Hierarchy
UBC	:	UnBalanced Clustering
V-LEACH	:	Vice-CH Low-Energy Adaptive Clustering Hierarchy
WDARS	:	Weighted Data Aggregation Routing Scheme
WLAN	:	Wireless Local Area Network
W-LEACH	:	Weighted Low-Energy Adaptive Clustering Hierarchy
WPAN	:	Wireless Personal Area Network
WSNs	:	Wireless Sensor Networks
WST-RS	:	Weight aware Spatial-Temporal correlation Routing Scheme
YEAST-CF	:	dYnamic and scalablE tree Aware of Spatial correlaTion with variation of Closest First

University of Malaya

CHAPTER 1: INTRODUCTION

1.1 Introduction

During the last few years, sensor-enabled smart devices have utilized wireless sensor networks (WSNs), which are extremely important components in the infrastructure of a smart city and the Internet of Things (IoT). Integration of the information world of the IoT and the physical world is made possible through WSNs (Buttyán, Gessner, Hessler, & Langendörfer, 2010; Zhang, 2012). In such smart environments, WSNs are considered a key technology in providing various IoT applications and services to users. Furthermore, WSN applications are found in health care, surveillance missions, missile target tracking, and in dealing with natural catastrophes, such as storms, earthquake, floods, and widespread fire (Rawat, Singh, Chaouchi, & Bonnin, 2014).

WSN in its simplest form can be defined as a network composed of small units known as sensor nodes, whose placements follow a spatial distribution. These nodes wirelessly communicate with one another to gather information regarding 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). Figure 1.1 illustrates a typical WSN.

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. Nevertheless, these networks face many challenges and constraints that must be investigated in depth before a widespread commercial deployment can be expected. The most significant constraints that can affect the design of WSN are the resource constraints. Energy constraints are faced by sensor nodes because the frequency of data gathering depletes their energy. Energy consumption occurs during data sensing, processing, and transmission. However, among these activities, data transmission is often the costliest action as far as

University of Malaya

energy utilization is concerned. For this reason, energy consumption is primarily considered in the design of protocols and algorithms of WSNs (AbdelSalam & Olariu, 2009; Pantazis, Nikolidakis, & Vergados, 2013). The data-driven nature of WSNs likewise generates considerable information that requires routing via multiple hops to the sink/base station from the network (Abid, Nguyen, & Seba, 2015; Cheng, Liao, Tseng, & Hsu, 2012). Thus, routing protocols in such networks play a significant part in the data gathering and forwarding procedure. In-network data aggregation possibly helps in optimizing the routing process by utilizing the processing ability of intermediate nodes. Hence, the correlation among the sensor observations is either avoided or used efficiently by configuring the sensor nodes to smartly gather and report data with a minimum resource utilization (Efthymiou, Nikoletseas, & Rolim, 2006; Mahdi, Abdul Wahab, Idris, Abu Znaid, et al., 2016), which is a problem that needs to be addressed.

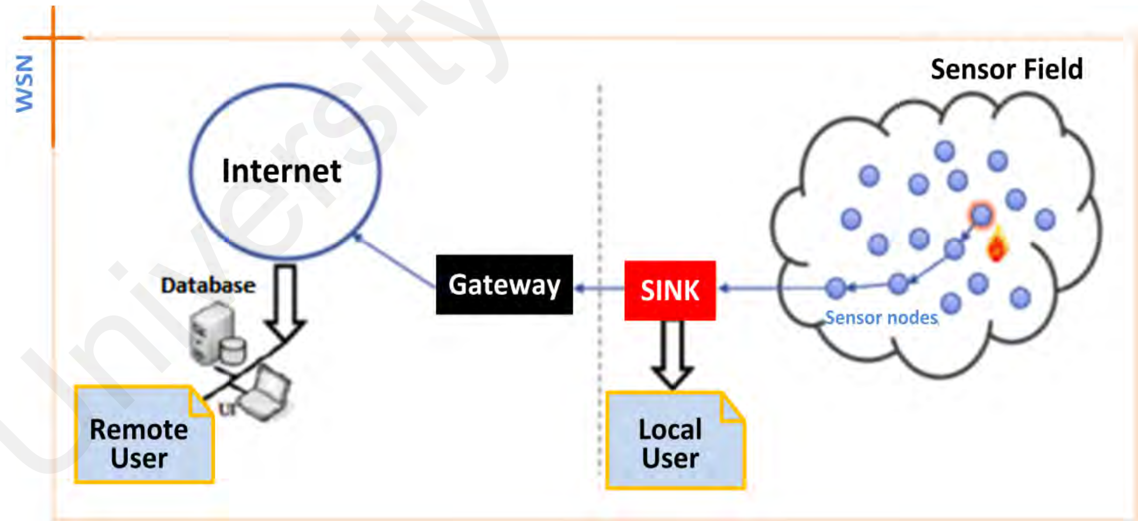


Figure 1.1: Wireless Sensor Network (WSN)

Source: (Rawat et al., 2014)

University of Malaya

1.2 Research Motivation

Energy and memory limitations are considerable constraints of sensor nodes because the limited energy supplied to network nodes causes WSNs to face crucial functional limitations. Therefore, the problem of limited energy resource on sensor nodes can only be addressed by using them efficiently. To address this problem, an energy efficient procedure is required to prolong network lifetime. The well-organized usage of the limited memory in sensors is also a problem that can be addressed by considering memory-consuming problems, such as load unbalancing and data duplication. Therefore, this research is motivated by the aforementioned issues that persist in the real-world network of sensors given the lack of resources. The following are the significant observations that motivate this research in WSN.

(i) Network energy: WSNs are normally used to monitor harsh and inaccessible environments, which restrain the use of infrastructure-based networks that may need constant human monitoring and interventions. However, due to tough circumstances and random sensor node deployment, replacing or recharging "dead" batteries of sensor nodes is difficult. Therefore, improving network energy efficiency and prolonging network lifetime are tough research challenges in WSNs.

(ii) In-network processing: Data transmission is probably the most energy-intensive operation performed by a sensor node. Figure 1.2 displays the number of TIMSP430F1611 (popular sensor processor) machine cycles equivalent in energy to the transmission of a single byte over the CC2420, CC1000, and MaxStream XTend radios. This figure indicates that any additional processing to reduce at least a single data bit might still be advantageous in terms of energy efficiency.

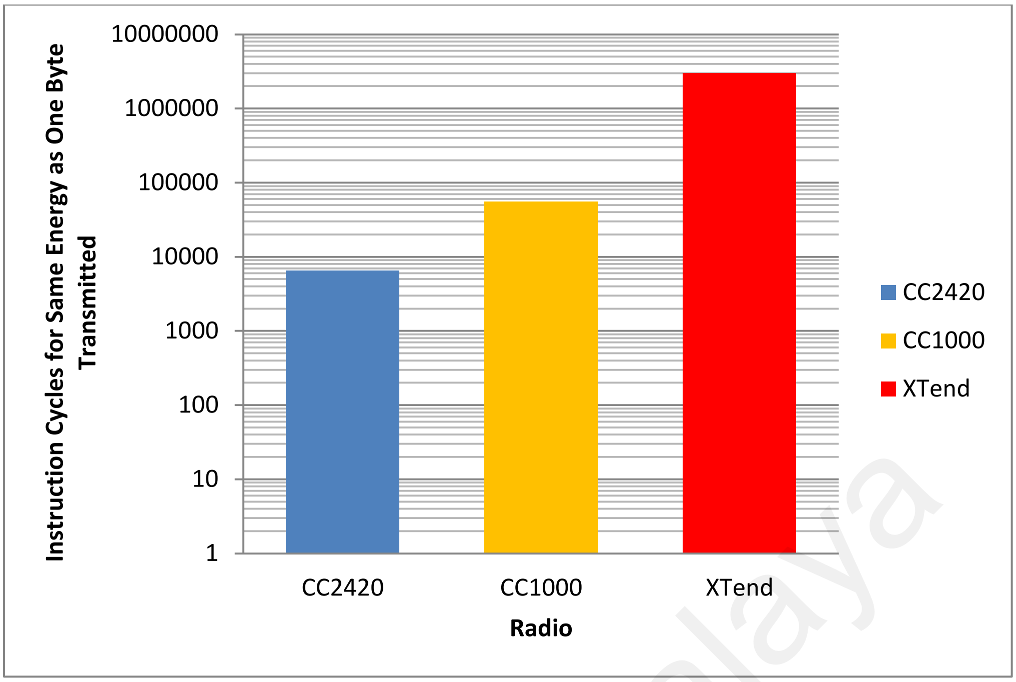


Figure 1.2: CPU Compute Cycles Versus Transmission Energy of One Byte Over Three Radios

Source: Redrawn from (Sadler & Martonosi, 2006)

In-network processing involves controlling data transmission by means of compression (Xu, Ansari, Khokhar, & Vasilakos, 2015) and/or aggregation (Draves, Padhye, & Zill, 2004) techniques. As an example, consider a cluster head that receives two packets containing temperature measurements from two dissimilar sources. A statistical aggregate such as AVERAGE or MAX or MIN of the two readings contained in the packets can be computed by the sensor node and sent in a single packet instead of forwarding them separately (Adya, Bahl, Padhye, Wolman, & Zhou, 2004).

(iii) **Sensor data correlation:** In most WSN applications, sensor nodes are normally overly deployed in the event areas to have a full sensing coverage. Thus, this densely populated sensor node distribution leads to high correlation data in the environment for geographically close sensors (spatial correlation) and is highly correlated for a period of time (temporal correlation). Moreover, data redundancy is discovered in the correlated sensor data. The transmission and processing of these redundant data cost extra network resource utilization. Hence, the manner in which the redundant data is processed and transmitted in these correlated sensor data poses an additional challenge.

1.3 Problem Statement

WSNs are growingly used in the process of sensing environmental data due to their aforementioned advantages. These networks collect environmental data using ad-hoc communications without any specific infrastructure or centralized control. The sensor nodes send the sensed data to the sink via wireless channels. The wireless modules only allow the nodes to communicate over limited and short radio ranges. Therefore, a multi-hop link is required to route network traffic from the event source region to the sinks (which may not be geographically close). However, transmitting raw data samples requires the establishment of a number of communication links that consume a significant amount of network resources.

In-network data aggregation technique is one possible mechanism to eliminate data redundancy and reduce the amount of data transmitted in the network. In-network data aggregation is dependent on the whereabouts and sensing time of data, implying that the convergence of raw data in space and time should be possible. Furthermore, an essential role is played by routing in data aggregation procedure, because routing helps make appropriate data forwarding decisions to achieve effective data aggregation, also known as in-network data aggregation routing protocol. In-network data aggregation routing protocol is a significant area in WSNs and has attracted interest from the research community over the years, because size and number of transmissions are reduced by in-network data aggregation, which lowers communication overhead and saves network resources.

A number of in-network data aggregation routing problems in WSNs 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, topologies, and homogeneities (Krishnamachari, Estrin, & Wicker, 2002; Nakamura, de Oliveira, Pontello, & Loureiro, 2006; Leandro Aparecido Villas et al., 2013; Wan, Zhang, & Chen, 2016). The main focus in these works is to maximize the formation of overlapping paths and increase aggregation rate. However, the excessive pursuit of routes overlapping regardless of the actual energy resources of nodes as well as the traffic load could result in prematurely dead backbone nodes. The reason for this is that no involvement occurs in some nodes in the routing process, whereas some nodes are heavily involved and congested, leading to an unstable distribution of load among nodes. Hence, network performance is degraded. Moreover, considering the assumption that the data from different event areas could be adequately aggregated, sometimes the data have to be routed over lengthier paths, which boost the total energy depletion. Based on the discussion, this exaggerated routes overlapping regardless of the actual nodes status presents a significant challenge for in-network data aggregation routing protocols that will be addressed in this thesis.

WSNs are data-driven and constantly monitor and collect data from the physical environment using the high concentration of sensor nodes scattering. Moreover, WSNs are deployed in distant and dangerous regions, where human access is risky and difficult. Thus, a backup system in the form of high dense scatter nodes is necessary to counter node failures. The highly dense scenarios in WSNs can produce data with high degree of correlation using geographically close nodes (spatial correlation). Moreover, successive sensor node readings during a certain time period have a significant temporal correlation due to the nature of the events in WSNs. This correlation creates significant data redundancy, which leads to unnecessary energy consumption in the processing and

University of Malaya

forwarding of the redundant sensed data to base station. Moreover, the propagation of readings from all sensors up to the network is impossible within definite time periods in WSNs scenarios with high successive events. To avoid this situation and to save energy, spatial and temporal data correlation approaches that reduce the number of data transmissions are utilized by the routing protocols. However, existing spatial, temporal, and spatio-temporal correlation solutions do not consider the status of instant nodes in terms of residual energy, available buffer memory, and nodes position within the event area during data gathering and forwarding for efficient representative and forwarding nodes selection. Moreover, the instant node density in the event area is ignored by existing solutions during the formation of correlated regions. Thus, a communication protocol is necessary to exploit the spatial and temporal correlations among the sensor observations in a fully distributed manner and for a maximum energy conservation, which is feasible for a large scale, highly dense, and event-driven WSN paradigm.

1.4 Research Questions

This study was conducted to answer the following research questions:

Q1. How does routing protocol affect the process of data aggregation?

Q2. Can a new Multi-Criteria Node Weight (MCNW) metric be useful in weighing the node status in various terms?

Q3. Is it possible to trade off multiple conflicting improvements based on the developed MCNW?

Q4. How is it possible to exploit the correlations in sensor data based on spatial and temporal convergences?

Q5. How can a multi-hop routing scheme be developed for in-network data aggregation to forward data in an energy efficient way with a desired load balance?

University of Malaya

Q6. Is the proposed in-network data aggregation routing schemes based on the developed MCNW superior to other existing data aggregation routing schemes in literature in terms of energy efficiency and load balancing?

1.5 Research Aim and Objectives

The main purpose of this thesis is to develop energy-efficient and load-balanced routing schemes for in-network data aggregation to prolong the network lifetime and enhance the data collection and aggregation processes. To accomplish this, the study aimed to achieve the following specific objectives:

(i) To investigate the existing in-network data aggregation and correlation awareness routing protocols.

(ii) To propose an innovative in-network data aggregation routing scheme for energy-efficient and load-balanced routing.

(iii) To propose a novel routing scheme for efficient data collection and aggregation that considers both spatial and temporal correlations to lower communication and data exchange.

(iv) To evaluate the developed schemes with different simulation scenarios and evaluation metrics.

1.6 Research Contributions

The novelty of this study is a new multi-constraint consideration routing scheme for in-network data aggregation in WSNs, which are dynamic, energy efficient, and load balanced. The following points summarize the contributions of this work:

(i) A new multi-hop routing scheme for in-network data aggregation called Weighted Data Aggregation Routing Scheme (WDARS) is proposed. The main considerations of WDARS are the energy efficiency and load balancing among sensor nodes, which aims to maintain an acceptable trade-off between energy efficiency and routes overlapping especially in high-traffic scenarios. WDARS combines the advantages of multi-hop concepts, which enhance data aggregation along the communication path, in the well-known Data Routing for In-Network Aggregation (DRINA) with the benefits of considering metrics related to energy efficiency and load balancing for multi-hop relay (MHR) nodes selection and routes overlapping point.

(ii) A new energy efficient, spatial and temporal awareness routing for in-network data aggregation scheme in WSNs referred to as Weight Aware Spatial-Temporal Correlation Routing Scheme (WST-RS) was also proposed in this study. The WST-RS takes into account the correlations in sensor data depending on the spatial convergence among the nodes and the temporal convergence between successive sensor readings when collecting and aggregating sensor data.

(iii) A new MCNW metric that weighs the node status in terms of various metrics is introduced. This new metric is a combination of metrics related to energy efficiency and load balancing in WSNs. The metric also contributes as a link cost function for link quality assessment and selects the more reliable routes to the destination depending on the status of nodes.

(iv) A new mechanism for forming spatial correlation among sensor nodes is developed. The new mechanism forms the spatial correlation by dividing an event area into spatial correlated regions and considering the density of nodes in the zone.

(v) A new mechanism that applies temporal suppression on consecutive sensor node readings by keeping extra information to utilize the temporal tolerance threshold

University of Malaya

is developed. The information kept depends on the roles in the sensor routing arrangement (i.e., representative node or cluster head).

(vi) A simulation model for routing protocol in WSN, which includes the conventional InFRA, DRINA, YEAST-CF, EAST, and the proposed schemes of WDARS and WST-RS, is performed. The performances of these routing schemes are investigated, and the result is 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 an overview of sensor network technology and its unique features. Next, it provides a brief review of in-network data aggregation in WSNs and explores the ongoing efforts of routing protocols on in-network data aggregation in WSNs as successors to structure, structure-free, and hybrid routing protocols. Then, several representative in-network data aggregation routing protocols are described, and their key features, merits, and demerits are highlighted. Finally, the chapter briefly discusses the data correlation protocols 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. A flowchart is presented to depict various research stages, their patterns, and their integration procedures. System models are introduced in the second part, including network model, data aggregation model, sleep scheduling and duty cycle model, queuing model, role assignment model, and energy consumption model. Finally, performance metrics are overviewed in this chapter.

Chapter 4 presents the proposed structure and functionalities of the WDARS scheme. Moreover, a detailed description of the proposed MCNW metric is provided. The performance evaluation of WDARS along with the modeling and formulation followed in this thesis is 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 provides a description of the proposed WST-RS scheme along with its design and evaluation. The major operations involved in WST-RS are discussed in detail. Finally, the proposed method is validated by presenting and analyzing simulation results.

Chapter 6 summarizes the main contributions of this work along with recommending potential areas for conducting future research.

University of Malaya

CHAPTER 2: LITERATURE REVIEW

2.1 Introduction

This chapter first presents the technological background of sensor networks and an overview of the paradigm, unique features, challenges, and requirements of wireless sensor networks (WSNs). The solutions to the challenges are likewise provided. Then, an introduction of in-network data aggregation is provided, including its types, basic components, and importance. Subsequently, the classification of the existing techniques and protocols for in-network data aggregation is expounded, with every protocol discussed separately by describing it as well as highlighting its key features, advantages, and drawbacks. Finally, the data correlation protocols in WSNs are reviewed and their significance to this study is comprehensively explained.

2.2 Overview of Sensor Network Technology

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 National Transportation Safety Board of Minnesota, this incident occurred because excessive bridge load, bad weather, and wear and tear (Borzi & Baranauckas, 2007). In California, temperature rise and choking caused the deaths of 19 firefighters in a fire incident on July 2, 2013 (Yadav & Yadav, 2016). These events and similar future incidents could be avoided if sufficient information about the weather, temperature, load conditions, and dense smoke areas is provided so that timely measures can be implemented. The development of micro-electro-mechanical systems (MEMS) has significantly contributed to the advancement of cost-effective and small wireless sensor nodes with diverse functions. Monitoring physical conditions, handling sensed information, and making appropriate decisions are possible with the help of these nodes. Border surveillance, healthcare provision, tracking operation, and disaster monitoring are some of the critical applications of WSNs (Hu, Wang, & Ji, 2013; Jaigirdar & Islam,

University of Malaya

2016; Lazarescu, 2017). A sensor network comprises small wireless sensor nodes with data-acquisition, battery, storage, and mote (processor/radio board) modules that collectively help in sensing. Various wireless sensor boards are available in the market for different applications, including magnetometers, accelerometers, barometric pressure sensors, wind speed sensors, acoustic sensors, solar radiation sensors, light sensors, humidity sensors, moisture sensors, GPS modules, magnetic RPM sensors, rainfall meters, temperature sensors, and seismic sensors (Wheeler, 2007). 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) (Akyildiz, Su, Sankarasubramaniam, & Cayirci, 2002). The radios are used for the communication between the sensor nodes and the BS so that data can be exchanged with applications for fulfilling the desired tasks. Moreover, the communication between the sensor nodes and the BS allows for the sharing of information via additional networks, like LAN, WLAN, WPAN, and the Internet, with other computers.

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, and bandwidth utilization and data exchange (Shen, Srisathapornphat, & Jaikaeo, 2001; Sohraby, Minoli, & Znati, 2007). Thus, the next section discusses the distinctive features, challenges, and requirements of WSNs.

2.3 Distinctive Features, Challenges, and Requirement of WSNs

Built-in smart processing capability, ad hoc nature, easy deployment, and flexibility are some of the advantages of WSNs over traditional networks. However, these distinctive features raise several challenges in the design of the hardware, communication protocols, and applications of WSNs. These challenges must be addressed in the development of WSN technologies to promote is extensive use in future applications. New protocols should be developed, or algorithms and communication protocols of traditional wireless ad hoc networks should be modified to achieve this goal (Karl & Willig, 2007; M. Younis, Senturk, Akkaya, Lee, & Senel, 2014). The significant challenges of WSNs and their possible solutions are listed in Table 2.1.

Table 2.1: Challenges and Solutions of WSNs

Challenges	Solutions
Large-scale deployment	• WSNs must employ architectures and protocols that address the scalability requirements. • The nodes must be small in size, cost effective, and energy efficient while demonstrating high intelligence. • The changes in the topology of the network due to node failures, high mobility, and node addition must be addressed through the self-organizing behavior of WSN.
Dynamic and extreme environmental conditions	• The WSN requires the exhibit of adaptive characteristics in its operation and protocols so that end-users can handle diverse conditions of wireless channels and fluctuation connectivity conditions.
Resource constraints	• The efficient usage of resources can solve the problems related to resource constraints. • Network lifetime can be prolonged by implementation of energy-efficient protocols such as energy-aware routing. • Efficient usage of memory is required by the sensors by considering storage issues that arise due to routing tables, security, and data replication.
No global identification (ID) for sensor nodes	• Large overhead is incurred in WSNs if global ID is used. A data-centric paradigm should be used by WSNs because of the restrictions of insufficient processing power and memory. This solution could help in focusing on the data produced by a sensor group instead of considering the sensed data of every individual node.
Unreliable wireless communication	• The reliability of WSNs should be ensured for proper operation. For applications that have strict reliability requirements, data delivery to the destination should be guaranteed.
Data redundancy	• Correlation and data aggregation mechanisms should be employed to eliminate data redundancy.
Prone to node failures	• WSNs should be robust to node failures by utilizing fault-tolerance techniques.

University of Malaya

The design of sensor nodes should be simple and cost-effective for widespread deployment. Limited energy resources also pose challenges to WSNs, which operate for long time periods. Figure 2.1 depicts the energy consumption percentages of different functions during the active period of a node, as drawn from literature. Packet communication consumes more than 50% of the energy of the sensor node (Halgamuge, Zukerman, Ramamohanarao, & Vu, 2009).

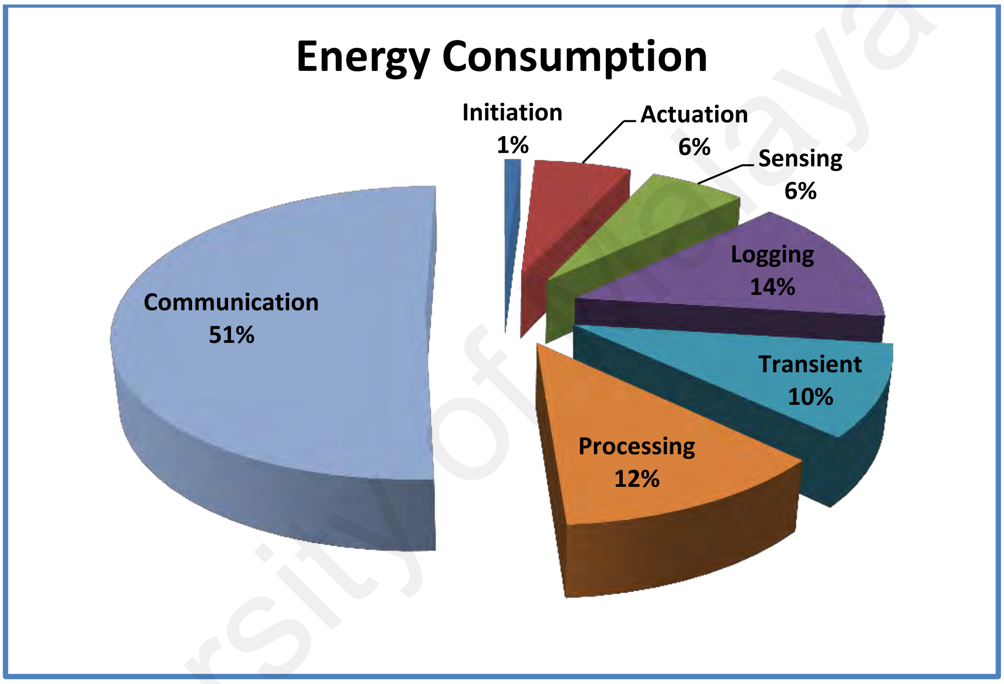


Figure 2.1: Energy Consumption by Numerous Activities of WSNs

Source: Modified from (Halgamuge et al., 2009)

The network lifetime can be prolonged and network energy can be saved by limiting data communication, as it is the most expensive function. Data communication can be minimized in WSNs by employing an important mechanism of in-network data aggregation (Patel & Kanawade, 2017).

2.4 Basics of In-Network Data Aggregation

Conventionally, sensor nodes collect data from a certain area and transmit it wirelessly to the sink node. The analysis, processing, and usage of data take place at the

University of Malaya

sink node. In certain scenarios, data sensed by various nodes and related to the same event can be aggregated and forwarded for joint processing either at the sink node or locally before being transmitted. This distributed data processing is handled by in-network data aggregation. In-network data aggregation is defined as a systematic distribution of data collection and routing via a multi-hop network; this procedure reduces resource usage by data processing at intermediate nodes and enhances the network lifetime (Fasolo, Rossi, Widmer, & Zorzi, 2007; Naeimi, Ghafghazi, Chow, & Ishii, 2012). In-network data aggregation requires four basic ingredients, namely, *routing protocols*, *data aggregation functions, data aggregation scheduling algorithms,* and *data representation algorithms* (Figure 2.2). These ingredients are briefly discussed in the subsequent subsections.

Figure 2.2: Basic Ingredients for In-Network Data Aggregation

2.4.1 Routing Protocols

An efficiently designed routing protocol is the most important requirement for in-network data aggregation (Fasolo et al., 2007; Mahdi, Abdul Wahab, Idris, Abu Znaid, et al., 2016; Xu et al., 2015). Different forwarding patterns are required in data aggregation unlike in conventional routing. If data aggregation is performed to reduce energy usage, then the packets should be routed to the destination depending on their content, and the next hop should be selected to enhance in-network aggregation. This type of routing is termed as data-centric routing. In a data-centric model, metrics that consider aggregation point locations, energy efficiency, traffic load, type of data, information priority are used during the search for relay nodes by a source node. Section 2.5 presents a detailed overview of the routing protocols for in-network data aggregation.

2.4.2 Data Aggregation Functions

Combining data from different nodes is an important purpose of the in-network aggregation mechanism. Various types of aggregation functions are utilized in data aggregation (Akkaya & Younis, 2005; Mantri, Prasad, & Prasad, 2014), and most of them are associated with a particular sensing application. However, they can be classified as follows according to their characteristics:

- **Lossy and lossless:** Data can be compressed or merged by aggregation functions based on a lossy or lossless scheme. However, data originality cannot be retained if the data are aggregated via lossy means. Additionally, compressing values may result in losing precision. By contrast, data preservation is guaranteed in lossless transmission, that is, the complete reconstruction of the values at the receiving end is possible.

University of Malaya

- **Duplicate sensitive and duplicate insensitive:** Several copies of the same information can be received at the intermediate node. In these cases, the same data may be aggregated multiple times. In the duplicate sensitive approach, the resultant value depends on the number of times the same value is considered. Otherwise, the aggregation function is considered duplicate insensitive. For example, the aggregation function that takes the minimum value is duplicate insensitive, whereas a duplicate sensitive function considers the average value.

Additional requirements should be met by efficient aggregation functions. Particularly, computational and energy constraints should be considered to enable the implementation of these aggregation functions by using basic operations. These considerations are important in the design of routing protocols and aggregation functions.

2.4.3 Data Representation Algorithms

The entire generated or received information may not be necessarily stored by a node in its internal buffer because of storage constraints. Thus, the storage, deletion, compression, or transmission of the data should be performed. A feasible mode of information representation is required in these operations (Shrivastava, Buragohain, Agrawal, & Suri, 2004; Zheng et al., 2015). The differences in the requirements among applications may cause variations in the corresponding data structure. Although a common data structure is used by all nodes, adaptability should be observed to conform to characteristics that are either location- or node-specific. Moreover, distributed coding techniques, which compress data by using correlation knowledge, are currently being used in data compression and representation (Srisooksai, Keamarungsi, Lamsrichan, & Araki, 2012; Xiong, Liveris, & Cheng, 2004).

University of Malaya

2.4.4 Data Aggregation Scheduling Algorithms

The aggregation procedure is scheduled for a specific aggregation function or data structure by data aggregation scheduling algorithms. Accuracy of data, latency, energy consumption, and timeliness of aggregated data are critically considered in these scheduling algorithms (Bagaa, Challal, Ksentini, Derhab, & Badache, 2014; Suriyachai, Roedig, & Scott, 2012). Two types of scheduling algorithms for data aggregation are used depending on the waiting time (Bachir, Dohler, Watteyne, & Leung, 2010; Park, Lee, & Yoo, 2015):

- **Unslotted data aggregation scheduling:** This approach presents a continuous and predefined waiting time. The data are received from the downstream nodes at the upper level during the waiting period. When the waiting time expires, the aggregation and forwarding of the received data are performed. This mechanism is executed at the routing layer, and CSMA-based mechanisms, which are contention-based MAC protocols, are utilized by the protocols in this approach.
- **Slotted data aggregation scheduling:** In this approach, the waiting time is subdivided into a set of slots with gaps to avoid signal interference. The last slot is selected for forwarding the aggregation result, whereas the other slots are tasked to receive the data from the downstream nodes. The gaps between the slots allow for the switching off a node transceiver. This process is performed on the MAC and routing layers. Usually, a contention-free MAC protocol, such as TDMA, is used in this case. Therefore, the perfect synchronization of the nodes is realized by the protocols. The simulation results from the literature prove that slotted data aggregation achieves less latency than unslotted data aggregation.

University of Malaya

2.5 Routing Protocols for In-Network Data Aggregation in WSNs

The studies conducted on in-network aggregation have targeted the issues of packet forwarding to facilitate the in-network aggregation of data. The main objective of these studies is to modify existing routing protocols to perform data aggregation. Numerous protocols using hierarchical structures have been proposed. Examples include tree-based routing protocols in which the sink node is the root (Madden, Franklin, Hellerstein, & Hong, 2002). However, many complex tree construction approaches have also presented. In addition to tree structure-based protocols, cluster-, chain-, and grid-based protocols have also been employed for in-network data aggregation (Mahajan & Mahotra, 2011; Naeimi et al., 2012; Leandro Aparecido Villas et al., 2013; Wang, Chiang, Hsieh, & Chen, 2013). Structure-free and hybrid in-network data aggregation routing protocols have also been proposed for specific applications (Bagaa, Derhab, Lasla, Ouadjaout, & Badache, 2012; Fan, Liu, & Sinha, 2007; Yousefi, Yeganeh, Alinaghipour, & Movaghar, 2012). In the following subsections, every type of routing protocol is expounded by discussing its important concepts in addition to its advantages and disadvantages. Moreover, the routing protocols for in-network data aggregation in a variety of network topologies are classified and several related studies are indicated in Figure 2.3. Finally, we compare the different in-network aggregation routing protocols in WSNs by using a few significant metrics, the design objectives and limitations of these routing protocols are presented in Table 2.2.

University of Malaya

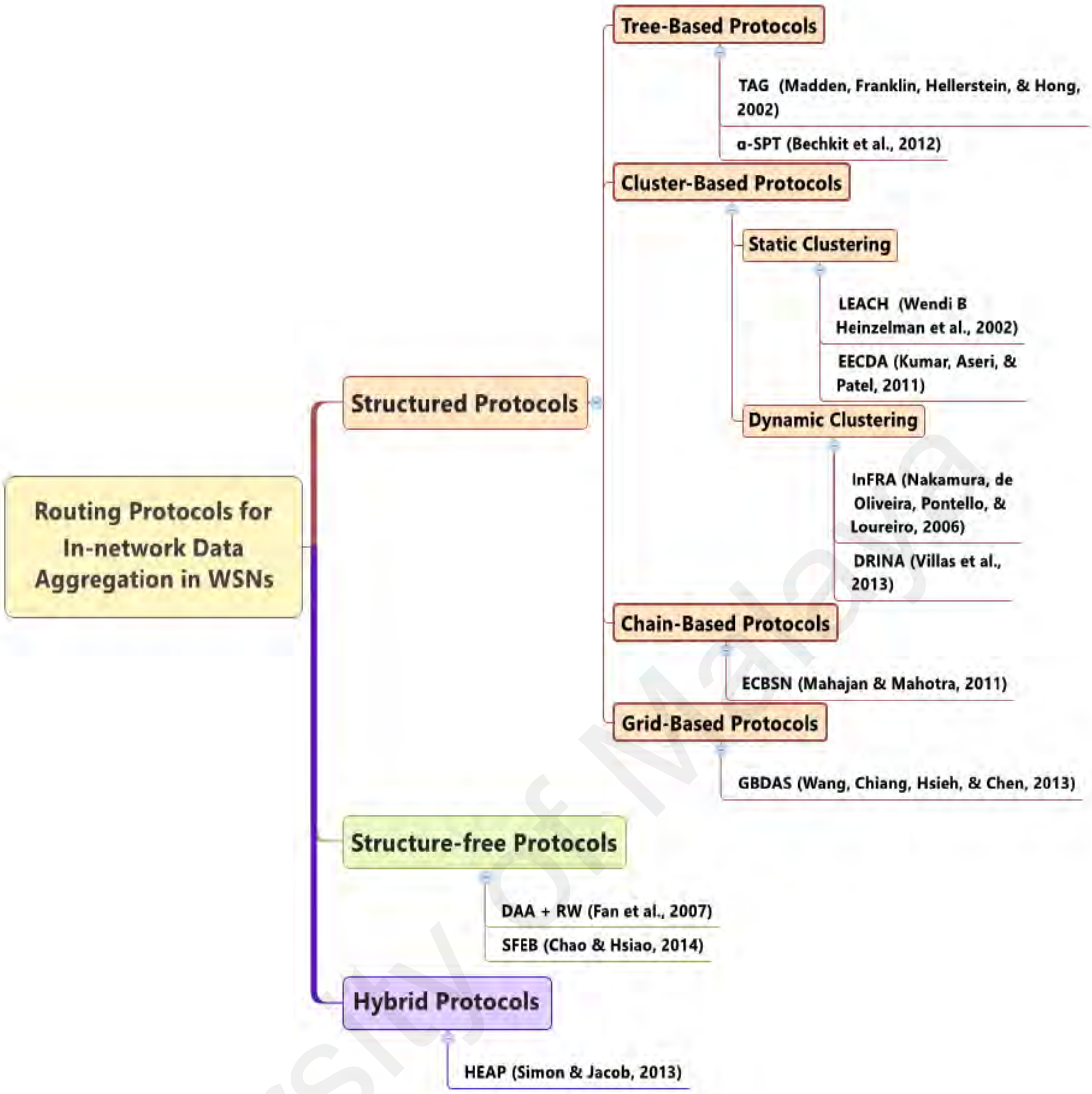


Figure 2.3: Classification of Routing Protocols for In-Network Data Aggregation in WSNs

2.5.1 Structured Protocols for In-Network Data Aggregation

The routing protocols for in-network data aggregation are surveyed in this section along with their advantages and disadvantages. These routing protocols are primarily developed for structured WSNs.

2.5.1.1 Tree-based protocols

In the literature, many early protocols employ tree-based routing strategies for data aggregation. For instance, many studies have used the simple topology based on parent and child association for tree-based data aggregation techniques (Castelluccia, Chan, Mykletun, & Tsudik, 2009). The data sent by the children are aggregated by the parent node, which in turn sends the aggregated data to its own parent node. In fact, the tree-based approach is a simple approach to aggregating data by constructing the conventional shortest path routing tree (SPT) for the data that are directed from the source to the sink node (Krishnamachari et al., 2002). However, tree-based data aggregation approaches have the following limitations:

(i) Tree-based techniques provide a simple approach to aggregate data, but this approach results in high latency because data aggregation is not performed until the packets have arrived at the parent node or grandparent node.

(ii) The possibility that the data are not aggregated near the event of interest is high because any two nodes that sense the same event might have different parent nodes. As a result, efficiency of data aggregation is reduced as the data is transmitted over a long path to the grandparent node.

(iii) Tree-based data aggregation schemes require a substantial number of control messages to construct and update the routing tree and consequently consume more energy.

(iv) Tree construction is inflexible because it depends on the assumption that the source nodes in the network are fixed and predetermined.

(v) The main drawback of tree-based protocols is packet loss due to bad channel links. In this case, the entire aggregated data from the children nodes are lost.

Although tree-based structures are costly to maintain and prone to damage due to limited network strength, they are still used in designing optimal data aggregation protocols at the intermediate nodes and energy efficient networks. For example, in the study of (Ding, Cheng, & Xue, 2003), the routing paths are organized to balance energy consumption evenly and optimally and aggregate data at the intermediate nodes

University of Malaya

simultaneously. In the studies of (Luo, Zhu, Wu, & Chen, 2011; Shan, Chen, Luo, Zhu, & Wu, 2014), the data-centric routing approaches uses the SPT routing protocol and ensures energy awareness at the parent nodes. Details of the protocols belonging to this category follow.

(a) Tiny AGgregation service (TAG):

Tiny AGgregation Service (TAG) service was introduced in the studies of (Madden, Franklin, et al., 2002; Madden, Szewczyk, Franklin, & Culler, 2002), where data aggregation was implemented on a real-world testbed. TAG service falls under the category of tree-based aggregation approach and aggregates data by using periodic traffic patterns. Every tree node gathers and sends data periodically to its parent. In addition, every node selects a time interval during which the packets are received from its children nodes. This time interval is included in the query forwarded to the children. After receiving the query, the child forwards a similar query to its children but sets the receiving time interval ahead than that in the query from its parent. Apparently, the interval length should be sufficient to allow all the children to transmit and process the relevant data. However, the interval duration is specific to the environment, and extensive delays result from long time intervals. Different tree levels with different time intervals are used in node assignment for timing in TAG Service to allow the bottom nodes of the tree to initiate data transmission. However, in case of link or device failures or dynamic topologies, TAG service may exhibit inefficiency similar to other tree-based approaches. Moreover, the failures at the intermediate nodes affect the concerned subtrees and cause disconnections. A change in topology induces the TAG service to reorganize the tree structure. This process results in high network overhead and increased energy consumption.

(b) Weighted Shortest Path Tree (α-SPT):

Generally, the SPT topology is adopted in WSNs because each sensor node in SPT approaches the root with the minimum hop count. However, a randomly constructed SPT may not increase the network lifetime. The authors of (Bechkit et al., 2012) utilized a weighted cost function to construct a new weighted shortest path tree (α-SPT). The load on important links is mitigated by assigning weights according to a geometric sequence. A 17% improvement was observed in the lifetime of the network by using α-SPT compared with the shortest path spanning tree strategy. In α-SPT, each link is assigned a weight depending on the length of the path to the root. High weights are possessed by the links that are close to the root. This strategy increases the lifetime of a network by considering the weights of the links for load balancing, which is not observed in a randomly constructed SPT.

2.5.1.2 Cluster-based protocols

Similar to tree-based protocols, cluster-based protocols are also used in hierarchically organized networks (Liu, 2012; Naeimi et al., 2012). The cluster-based protocols are extensively used hierarchical data aggregation. They can efficiently manage data, reduce communication overhead, enable improved 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. Every cluster consists of nodes that are within the vicinity of each other, and nodes are grouped as per important considerations. A cluster node may be designated either as a cluster head (CH) or as a cluster member (CM). First, the data from the member nodes are collected by the CH. Then, the data are aggregated and forwarded to the upstream node (Abbasi & Younis, 2007; Zhu, Shen, & Yum, 2009). Cluster-based protocols have the following advantages (Bouabdallah, Rivero-Angeles, & Sericola, 2009; O. Younis, Krunz, & Ramasubramanian, 2006):

University of Malaya

(i) Enhancement of bandwidth utilization to reduce useful energy consumption (i.e., minimization of collisions caused by channel contention).

(ii) Minimization of overhead to reduce wasteful energy consumption.

(iii) Maximization of the lifetime of the network by occasionally adopting a balanced approach to energy consumption and load distribution (for example, the CH and its corresponding members associated change during the cluster reformation process).

(iv) Prevention of long-distance transmission among nodes to boost resource utilization and lower energy consumption.

In a clustered network, cost is classified as intracluster or intercluster cost. The communication cost from the nodes within a cluster to the CH is considered intracluster cost, whereas the one from the CH to the BS is intercluster cost (Singh & Sharma, 2015). Scalability and effectiveness enhancements of a cluster are validated using the clustering cost. The drawbacks can be determined by the 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 random topological change of the network, 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 are 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, Yamani, et al., 2016).

- **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 (Wendi Beth Heinzelman, 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 formed reactively within the vicinity of the event-sensing nodes. Once the event is located, a specific sensor node is selected as the CH (ideally the node with the maximum energy or adjacent to the event), whereas the other event-sensing nodes are designated as the member nodes (Jain, Saini, & Bhooshan, 2014). The main benefit of this approach is that only the participating nodes are involved in the aggregation of the data. Therefore, dynamic clustering conserves the energy of the idle nodes (S. Park, 2006; Villas et al., 2013).

In the following subsections, four clustering routing protocols in WSNs are analyzed in detail, and their advantages and disadvantages are highlighted.

University of Malaya

(a) 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 of becoming 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.

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

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 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; 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 fairly shared 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 in accordance with their allocated time slots.

Although LEACH is the simplest hierarchical protocol and can reduce energy consumption in a WSN, it poses a number of problems, including the following (Abbasi & Younis, 2007; J. Chen, 2011; Liu, 2012; 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 to construct 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 the intercluster and intracluster communications; thus, LEACH is not a scalable routing protocol.

Several modified versions of the original LEACH protocol 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,

University of Malaya

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

(b) Energy Efficient Clustering and Data Aggregation (EECDA):

Energy-efficient clustering and data aggregation (EECDA) is designed for heterogeneous WSNs, in which the nodes have different energy levels; this protocol also presents such features as data aggregation and energy efficient routing in WSNs (Kumar, Aseri, & Patel, 2011). The noteworthy objective of this protocol is to present innovative mechanisms for the selection of the aggregation path and CH. Single-hop communication is used between the cluster nodes and CH, and the probability for a node to become the CH is optimally determined. In this protocol, the nodes are categorized as normal, advanced, or super. During the CH election, the initial energy of each node is compared with that of the normal node, and the weighted election probability is measured. Therefore, the same node is not elected for transmission repeatedly, and network stability and energy management are consequently improved. After the CH election, the aggregation path with the maximum cumulative remaining energy is selected rather than the one with the minimal energy consumption. Simulations were conducted to compare the network lifetime of EECDA with LEACH, EEHCA, and EDGA (Kumar et al., 2011). EECDA exhibited better performance than the other cluster-based approaches by increasing the network lifetime and leveraging the heterogeneity of the nodes. Furthermore, the stability period (time period from the start of the operation until the death of the first alive node) of EECDA was longer compared with the other protocols. Network stability and lifetime are improved by EECDA but at the cost of complex mathematical calculation. Moreover, the selection of the aggregation path and CH in large-scale networks requires a massive amount of messages to be exchanged; thus, the communication cost is high in EECDA.

University of Malaya

(c) Information Fusion-based Role Assignment (InFRA):

In (Nakamura et al., 2006) the authors proposed the reactive algorithm information fusion-based role assignment (InFRA) for event-based WSNs. The roles of sink, collaborator, coordinator, and relay are assigned when any type of event occurs. In this protocol, clusters are formed when similar events are detected by various nodes. Then, the coordinator aggregates the data from all the collaborators and sends the event data to the sink in a multi-hop manner. InFRA generates the SPT linking all the source nodes to the sink to enable intercluster data aggregation. Moreover, InFRA implements a role-migration policy. This policy asserts that the role of coordinator is transferred from one node to another node so that the energy load is distributed evenly among the nodes in the cluster. Furthermore, InFRA uses intracluster and intercluster data aggregation schemes.

The disadvantages of InFRA are as follows: (i) the construction and maintenance of the routing structure incur large overhead and result in poor network scalability for InFRA. (ii) Each time a new event is detected, the information of the event is broadcast throughout the network to notify other nodes, and the paths from the available coordinators to the sink node are updated. These processes are costly, and they limit network scalability. (iii) InFRA is unsuitable for frequently occurring events or encountering events with short durations.

(d) Data Routing for In-Network Aggregation (DRINA):

In (Leandro Aparecido Villas et al., 2013) the data routing for in-network aggregation (DRINA) is proposed to overcome the disadvantages of InFRA. The protocol was designed to maximize the advantages of in-network aggregation. In DRINA, an event is sensed by multiple nodes. These nodes form a cluster. Then, the information is sent to the selected CH so that aggregation can be performed, and the tree links the CHs to the sink node. Primarily, the sink node is considered the root of the

University of Malaya

constructed tree, and the distance between a node and the sink node is called the hop distance. Upon the occurrence of an event, the path between the CH and sink node is labeled as the "established route." Moreover, the hop distances of the nodes are updated in DRINA to determine the shortest distance between the event nodes and a node in the established route. A greedy incremental tree (GIT) is constructed by the hop distance metric in the network. This GIT construction exhibits satisfactory performance in in-network aggregation.

The main goals of DRINA are to minimize the number of messages in the setup phase of the routing tree and to maximize overlapping routes. The routes with the highest aggregation rates are selected to conduct reliable data transmission, and the minimum number of control packets is required for the establishment of the routes. However, DRINA presents the following disadvantages: (i) a heavy load is exerted on the nodes on the previously constructed path, and this lack of load balance causes such nodes to expire prematurely. (ii) Correlated events are ignored because of the assumption that the data from differing event areas could be aggregated adequately. (iii) The data have to be routed sometimes over the longer paths; as a result, the total energy depleted increases. (iv) DRINA is unsuitable for events that occur for long time periods because the routes are static and the energies of the nodes that belong to the routing structure are rapidly consumed.

2.5.1.3 Chain-based protocols

In cluster-based sensor networks, data aggregation is performed at the CHs, which receive information from the sensor nodes. When the distances between the CHs and sensor nodes are large, a significant amount of energy may be used by the nodes for communication. The key idea behind chain-based data aggregation is that energy efficiency can be enhanced if sensors transmit only to close neighbors. Chain-based data aggregation involves transmitting data to the closest neighbor of every sensor node

University of Malaya

(Tang, You, Guo, Guo, & Ma, 2012). However, this technique has certain shortcomings, particularly in a single long chain (Ahn, Kim, Sim, Youn, & Song, 2011). The literature contains many chain-based protocols, such as energy chain-based sensor network (ECBSN).

Hierarchical chains comprising two layers are adopted in ECBSN (Mahajan & Mahotra, 2011). The lower layer is composed of multiple chains using power-efficient gathering in sensor information systems (PEGASIS) (Lindsey & Raghavendra, 2002). In this layer, the node with the highest residual energy in every chain is selected as the leader of that chain. Correspondingly, the upper layer chain is formed by linking all the leaders in the lower layer, and the node that has the shortest distance to the BS is selected as the leader of the upper layer chain. Once the data have been gathered by the leader of every lower layer chain, they are transmitted to the leader of the upper layer, and the leader subsequently sends the data to the BS. Although the general goal of ECBSN is to provide energy-efficient and long-lasting paths, data redundancy is not considered in ECBSN and global knowledge of all the nodes is required at the BS.

2.5.1.4 Grid-based protocols

Grid-based protocols involve the division of the sensor fields into regions, which are in turn divided into grids, and a set of sensors are assigned as data aggregators in fixed regions. Moreover, direct transmission occurs from the sensors to the data aggregator in every grid. Thus, no direct communication occurs among sensors present within the grids. The applications that require adaptability to the mobility of the event and dynamic changes, such as weather forecasting and military surveillance applications, prefer grid-based data aggregation (Gungor, Lu, & Hancke, 2010; Rajagopalan & Varshney, 2006).

Grid-based data aggregation scheme (GBDAS) is another in-network data aggregation protocol in WSNs. It partitions the sensor field into 2-D logical grids of

University of Malaya

cells (Wang et al., 2013). In GBDAS, the cell head is the node with the highest remaining energy in the cell, and it is responsible for collecting and aggregating its own data and those of the other sensor nodes. A chain composed of all the cell heads is constructed, and the aggregated data move from one head to another until the data reach the BS. However, the collected data should be aggregated by all of the nodes; thus, the delay is remarkable, particularly at the end node. Furthermore, this approach is infeasible for large-scale networks, in which the BS is located distant.

2.5.2 Structure-Free Protocols for In-network Data Aggregation

In structure-free networks, all the nodes share equal responsibility and battery status in the considered region. The structure-free networks use data-centric routing data aggregation. Crisis management applications extensively use structure-free networks as the durations of the events in these applications vary from a few hours to a few days. Moreover, no energy is expended in forming any structure. The average delay and overhead related to structure maintenance are also significantly reduced, and the network is robust to node failures. Therefore, applications that face the challenges of energy limitation and time sensitivity can benefit from these protocols because of the provisions for fault tolerance and energy management. Examples of crises for management include fire development, safe zone discovery, gas diffusion, and natural disasters; these protocols can also be used in tracking the location of intrusion robots and persons who need to be rescued (Chao & Hsiao, 2014; Yadav & Yadav, 2016). Two structure-free protocols are described in the following subsections, with emphasis on their salient characteristics.

(a) Data-Aware Anycast and Randomized Waiting (DAA + RW):

In (Fan et al., 2007) the authors presented an event-driven reporting strategy called data-aware anycast and randomized waiting (DAA + RW) for aggregating data in

University of Malaya

structure-free WSNs . In the DAA + RW strategy, anycast request-to-send is sent to determine the next hop node and report the sensed data to the BS. If the sensor nodes have sensed the same information related to the event or if they are near the BS, then these sensor nodes are prioritized to generate anycast clear-to-send. This process improves aggregation efficiency. In the DAA + RW strategy, the number of transmissions is reduced by adopting a RW scheme. Consequently, every node that needs to send data has to wait a random amount of time before initiating transmission. Reasonable aggregation is realized if a node close to the BS waits for a long time. Nevertheless, in this strategy, not all packets may be aggregated.

(b) Structure-Free and Energy-Balanced data aggregation (SFEB):

In structure-free and energy-balanced data aggregation protocol (SFEB) (Chao & Hsiao, 2014), energy-consumption efficiency and data-gathering efficiency are enhanced by two-phase data aggregation with dynamic aggregator selection. In the first phase, primary aggregators (PAs) and secondary aggregators (SAs) are selected among the sensor nodes on the basis of their positions upon the occurrence of an event. Subsequently, the aggregators are paired (i.e., PA/SA pairs). In the proposed strategy, the nodes near the BS are selected as aggregators, whereas the other sensor nodes are forwarding nodes. Early aggregation and forwarding node selection are achieved using this strategy. In the second phase, the packets collected by the aggregators are sent to the BS; the number of transmitted data packets is approximately half that of control packets. The aggregated data sent by the PAs include the packets sent by the reporting sensor nodes. If the sensor nodes fail to transmit data to the aggregators, then the sensor nodes directly send their packets to the BS. In the SFEB strategy, self-configuration, localization, and synchronization of the sensor nodes are assumed.

Compared with DAA + RW, SFEB exhibits relatively good performance in terms of energy efficiency because of the significant decrease not only in the data packets but also in the control packets. However, delay and computational complexity are the main drawbacks of SFEB. Numerous complex mathematical formulations are involved in the aggregator selection method, and applying it to resource-constrained WSN nodes is unrealistic.

2.5.3 Hybrid Protocols for In-network Data Aggregation

WSN applications that present low delay tolerance and require low energy and control overhead can employ hybrid protocols for aggregating data. The pros of structure-free and structured protocols are combined in this class of in-network data aggregation protocols. A hybrid energy efficient aggregation protocol (HEAP) is an example of a hybrid protocol. It combines dynamic and static data aggregation strategies for reducing energy consumption and guaranteeing the transmission of data in large-scale WSNs (Simon & Jacob, 2013). HEAP was designed to be suitable for various applications, such as periodic monitoring, on-demand data transmission, and event detection. In HEAP, the nodes are divided into zones. The node with rich resources is selected as the CH in every zone. For applications requiring the aggregation of data from all of the nodes, static aggregation is implemented using the tree structure. By contrast, dynamic aggregation is implemented when the event is detected by nodes that belong to different clusters. This strategy requires the nodes to wait for a random time, depending on the expiration time of the packets. All packets from neighboring nodes with same event ID can be aggregated without considering the zones. Complex mathematical calculations are not included in the protocol, and data aggregation is performed with the minimum number of nodes; consequently, less delay and high energy efficiency are yielded. However, this protocol is unsuitable for bursty traffic as it presents a waiting time.

University of Malaya

Table 2.2: Comparison of Different In-network Aggregation Routing Protocols

Scheme	Route structure	Design objective	Aggregation nodes	Load balancing	Scalability	Overhead	Energy efficiency	Limitations
TAG	Tree	Aggregate data with a periodic traffic pattern	Aggregator node	No	Low	High	Low	Requires extra exchanges of messages to construct and maintain the tree
α-SPT	Tree	Increase network lifetime	Aggregator node	Yes	Moderate	High	Low	Presents computational complexity and high memory requirement
LEACH	Static cluster	Maximize network lifetime	CH	Yes	Low	Moderate	Moderate	Low scalability and good CH distribution cannot be guaranteed
EECDA	Static cluster	Improve performance in terms of network lifetime and stability	CH	Yes	Low	Very high	Low	Unsuitable for large networks and exhibits computational complexity
InFRA	Dynamic cluster	Maximize overlapping routes	CH and intermediate nodes	No	Low	Very high	Moderate	Poor scalability and high cost
DRINA	Dynamic cluster	Maximize overlapping routes and minimize control overhead	CH and intermediate nodes	No	Very high	Low	High	Lacks load balance and inefficient for dynamic routes
ECBSN	Chain	Provide energy-efficient and long-lasting paths	Leader nodes	Yes	Low	Moderate	High	Does not consider data redundancy and requires global knowledge of all the nodes at the sink
GBDAS	Grid	Maximize the lifetimes of nodes	Cell head and chain leader	Yes	Low	Moderate	High	Low scalability and high delay, particularly at the end nodes
DAA + RW	Structure free	Eliminate control overhead and data redundancy	Intermediate nodes	No	Very high	Low	Moderate	Not all packets may be aggregated, and the aggregator node may not have sufficient energy required to transmit packets to the BS
SFEB	Structure free	Reduce energy consumption	Aggregator nodes	Yes	Very high	High	Moderate	High delay and computational cost, and requires position information
HEAP	Hybrid	Energy-efficient data aggregation for large-scale networks	Zone cluster heads	Yes	Very high	Low	Very high	High delay and inappropriate for bursty data

University of Malaya

2.6 Data Correlation Protocols in WSNs

In routine WSNs, the sensor nodes in charge of data gathering from sensing regions are scattered redundantly as per the desired applications concerned. Generally, sensor node redundancy addresses two objectives. First, the memory, computational capability, and communication range limitations necessitates the introduction of a certain degree of redundancy to satisfy the QoS requirements. Second, WSNs are usually deployed in far-flung and dangerous regions where human access is risky and difficult. Thus, a backup system in the form of redundant nodes is necessary to prevent node failures. However, substantial but unnecessary energy consumption is incurred during the processing and forwarding of sensed information to the BS because numerous nodes sense the same information. Data correlation techniques are utilized in the data collection and aggregation processes to reduce the number of data transmissions and avoid the superfluous energy consumption (Aminian, Akbari, & Sabaei, 2013). Many correlation-based techniques have been presented for WSNs (Bai & Jamalipour, 2008a; H. Chen, Mineno, Obashi, Kokogawa, & Mizuno, 2007; Yoon & Shahabi, 2007).

In the literature, data correlation protocols are classified into three main types, namely, spatial correlation, temporal correlation, and spatio-temporal correlation. In this section, several data correlation protocols are discussed. Figure 2.4 illustrates the data-correlation protocols according to their types. Moreover, the important features, design objectives and limitations of the spatial and/or temporal data correlation protocols proposed in the literature are presented in Table 2.3.

University of Malaya

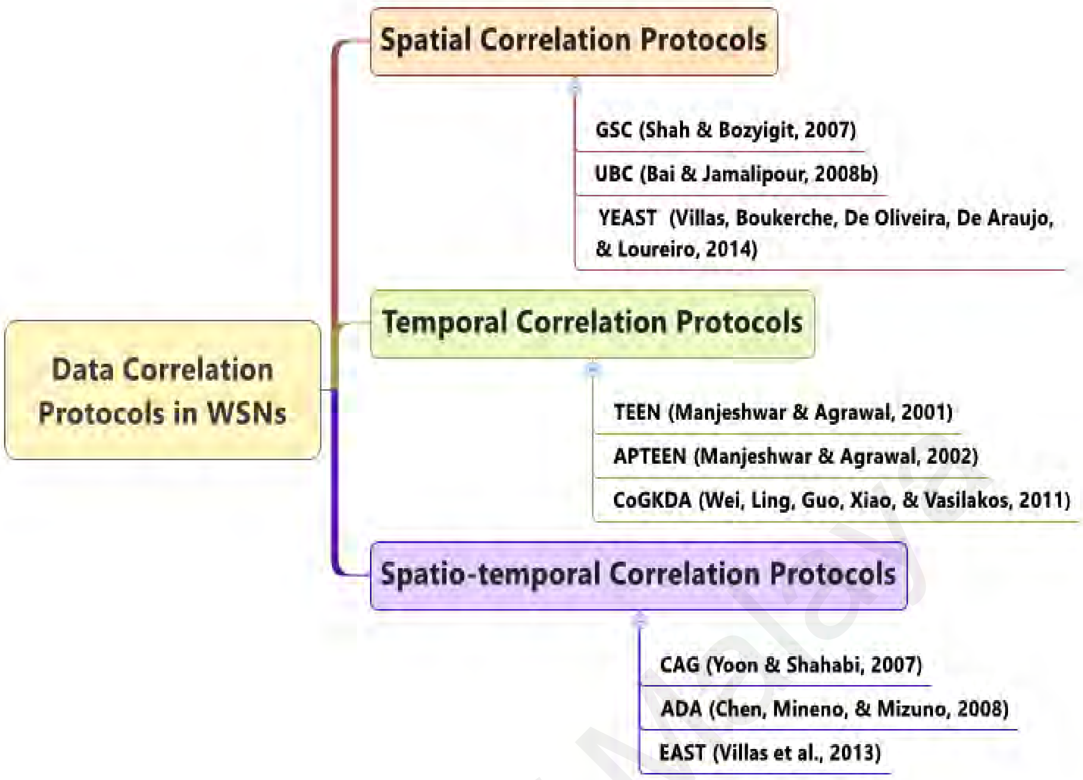


Figure 2.4: Classification of Data Correlation Protocol in WSNs

2.6.1 Spatial Correlation Protocols

The data sensed by geographically close nodes show spatial correlation. This situation may be avoided by selecting few nodes to represent the event that has been detected and to transmit relevant data to the sink. Sensed information about the event is reported to the sink node by these representative nodes only (Tsai & Huang, 2014; Villas, Boukerche, De Oliveira, De Araujo, & Loureiro, 2014). Figure 2.5 shows the temperature readings of two geographically close nodes in two days to present an example of spatial correlation. The data sensed by nodes 1 and 25 shows the same variance curves.

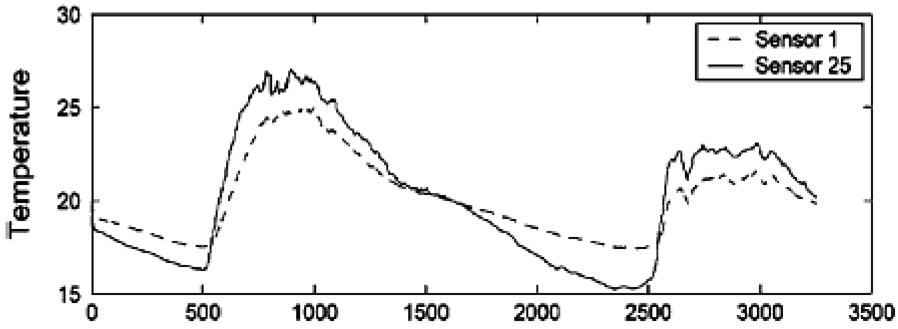


Figure 2.5: Temperature Variations Observed by Two Sensor Nodes

Source: (Pan, Gao, Gao, & Liu, 2014)

The subsequent subsections describe several spatial correlation protocols and discuss their key features, merits, and demerits.

(a) Gridiron Spatial Correlation (GSC):

Shah and Bozyigit proposed, gridiron spatial correlation (GSC) for spatial correlation in WSNs (Shah & Bozyigit, 2007). The correlation region is dynamically changed by GSC. This process allows this protocol to adapt to the reliability requirements. GSC uses the squared rectangles to form the correlation regions and assumes that spatial correlation exists among the nodes present in these rectangles. Moreover, the redundant and nearby resources are identified by the CH. Energy level and closeness are also considered to turn off the activity of the nodes. However, the control strategy in GSC is not implemented on multi-hop members, that is, it can only show desirable results if the communication radius of the CH is larger than the radius of the event.

(b) UnBalanced Clustering (UBC):

This protocol investigates the importance of setting the optimal cluster radius in accordance with the aggregation features. The spatial correlation of the generated data decreases along the spatial field in real-world scenarios. Thus, the work in (Bai &

Jamalipour, 2008b) focused on finding the optimal cluster radius to achieve an efficient aggregation rate. Moreover, a linear transform in the form of a discrete cosine transform (DCT) is used so that the correlation can be exploited as an independent coefficient. In the DCT, statistically spatially dependent data are mapped into a set of independent coefficients. The statistical results of (Bai & Jamalipour, 2008b) showed that an exponential graph of aggregation rate (Ag_a) versus cluster radius (R) is obtained. The aggregation rate (Ag_a) is approximated by

$$Ag_a = (1 - r\_min)e^{-\pi R^2 \rho d/\delta} + r\_min \qquad (2.2)$$

Where ρd represents the node density of the field, δ represents a constant factor that controls the dropping speed of the aggregation rate, and r_min represents the minimum aggregation convergence point attained from the statistical results. When the aggregation model in Equation 2.2 is used, variability is observed in the optimal cluster radius for different areas of the network fields. Thus, the network is divided into unbalanced clusters in accordance with the calculated optimal cluster radius for every region. The simulation of (Bai & Jamalipour, 2008b) revealed that the average energy consumption of the network for different cluster numbers of equal-size clustering protocols is significantly improved.

(c) dYnamic and scalablE tree Aware of Spatial correlaTion (YEAST):

In (Villas et al., 2014) the authors presented dynamic and scalable tree aware of spatial correlation (YEAST), which considers the importance of spatial correlation, that is, the area where similar events are detected by the nodes. YEAST aims at maximizing data aggregation along the communication path and controlling the energy usage by the nodes by eliminating unnecessary notifications. Particularly, adaptive correlation areas are defined by YEAST to reduce data redundancy. In YEAST, a set of representative nodes are selected in every correlation region and the collected data are reported to the

University of Malaya

coordinator. Geographical information is also used to route data to the sink and perform in-network aggregation. Furthermore, the position of the node is used to establish a function for the estimation of the measurement values of member nodes. However, YEAST has two restrictions: (i) the absolute position of the nodes should be known by the network nodes, and (ii) the uniformity of the physical variable that has been sensed must be within the correlation area. In addition, the source phenomenon of the field is considered by YEAST. However, in realistic situations, large transmission delays can be encountered because of the mobility of the physical event. Consequently, the aggregation rate obtained is poor, and cluster-based data aggregation is inappropriate for dynamic applications. In YEAST, performing exaggerated routes overlapping regardless of the actual energy resources of the nodes and the traffic load could result in the premature death of the backbone nodes.

2.6.2 Temporal Correlation Protocols

Temporal correlation protocols use the correlation of sensor readings with the previous reading; such correlation is known as temporal correlation. A substantial temporal correlation is present among consecutive observations of every node (Das, Misra, Wolfinger, & Obaidat, 2016; Kandukuri, Lebreton, Lorion, Murad, & Lan-Sun-Luk, 2016). For example, smart meters periodically transmit power usage data. However, if the previous power usage is higher than the new one with respect to the specific threshold value for temporal correlation, then the transmission of the new power usage is suppressed; otherwise, data transmission occurs. An example is depicted in Figure 2.6, at 0 o'clock, power usage data are transmitted, whereas at 1 o'clock, the difference value is transmitted. If the threshold value is lower than the difference of the two readings, then suppression occurs. In Figure 2.6, when the threshold (τ) is 10, the left node does not transmit at 1 o'clock, whereas the right node transmits the difference value of 13 (Choi & Chae, 2014).

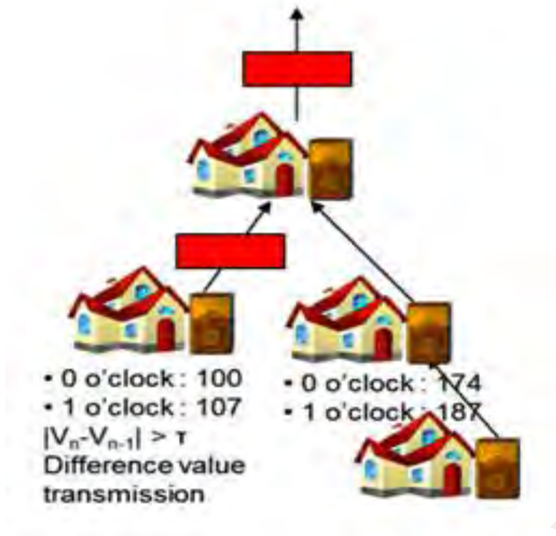


Figure 2.6: Smart Meters of Power Usage

Source: (Choi & Chae, 2014)

The subsequent subsections describe some temporal correlation protocols and highlight their key features, strengths, and weaknesses.

(a) Threshold sensitive Energy Efficient sensor Network (TEEN):

A few applications, such as environmental monitoring, only require variations in physical parameters. However, proactive clustering protocols, such as LEACH, present a detailed picture of the entire sensing area in every round. Thus, an energy efficient and reactive protocol named threshold sensitive energy efficient sensor network protocol (TEEN) was proposed by (Manjeshwar & Agrawal, 2001). Two values, namely, hard threshold (H_T) and soft threshold (S_T), are defined in the TEEN protocol to control data transmission and exploit the temporal coherency of sensor readings for suppressing data redundancy and reducing energy consumption. However, two conditions must be fulfilled before sensed data are sent by the transmitter. First, the H_T value should be lower than the current value of the sensed attribute, and second, the difference between the previously recorded value and the new sensed value should be higher than the S_T value. Thus, the desired value of the sensed attribute is constrained by H_T to reduce the number of transmissions. Moreover, in case of a minimal change in the sensed value, the transmission of the sensed data is eliminated by S_T. If the sensed attributes suddenly change, then TEEN adapts to them by employing the hierarchical strategy in addition to the data-centric approach. Thus, TEEN is highly suitable for time-critical applications. However, the downside of TEEN is that no data can be transmitted to the user from the network if the threshold values are not attained.

(b) Adaptive Periodic Threshold-sensitive Energy Efficient sensor Network (APTEEN):

The shortcomings of TEEN were addressed in (Manjeshwar & Agrawal, 2002) by proposing the adaptive periodic threshold-sensitive energy efficient sensor network (APTEEN) protocol. This protocol provides the user with increased control over the performance of the network. APTEEN is a hybrid protocol that integrates the merits of reactive and proactive protocols. The time intervals used to send data are adjustable, and adaptation is exhibited in case of sudden changes in the attributes. The APTEEN protocol is implemented by broadcasting the count time (T_C) and threshold values of the CHs. Moreover, the maximum periods between two successive reports generated by the nodes are defined. If the nodes are not triggered to send data when the threshold values are not exceeded, then the sensed data are sent every T_C interval. Although the operation of APTEEN results in high energy-consumption efficiency and increased network lifetime compared with TEEN, APTEEN has a high implementation complexity and increased data transmissions. However, flexibility and a complete network picture is provided by APTEEN.

(c) Combined Grey model and Kalman filter Data Aggregation (CoGKDA):

The authors of (Wei, Ling, Guo, Xiao, & Vasilakos, 2011) proposed a prediction-based data aggregation algorithm for reducing redundancy in transmitting data. The algorithm focuses on predicting a data series and minimizing the overhead related to temporal data correlation while transmitting data. Two models of data aggregation, namely, grey model-based data aggregation and Kalman filter-based data aggregation are combined into CoGKDA and used to achieve the desired accuracy level. Grey model-based data aggregation requires only a few data items for rapid modeling, whereas Kalman filter-based data aggregation processes data with noisy measurements. Thus, CoGKDA enhances prediction accuracy, scalability, and adaptability to dynamically changing parameters. The basic feature of this algorithm is the synchronicity of the prediction of the data series in the sink and sensor nodes. Alternatively, in the initial stage, the acceptable prediction error threshold and cumulative error threshold (accumulation of the error in continuous prediction) are broadcast by the sink node to other sensor nodes depending on the desired accuracy level. Then, the sensed data of all the nodes are estimated by the sink node by utilizing the data received at the current period and from the previous data sensing period. At the other end, the same prediction procedure is implemented by every sensor node by using the same data sequence. The predicted values are saved in a queue. Thus, a comparison between the sensed value and the predicted value is performed by every node in the subsequent round. If the error between the two values is higher than the prediction error threshold, then the sensed value is sent by the node. In the meantime, the predicted value is used by the sink if no sensed value is received from the sensor node in that round. However, the protocol encounters the following challenges due to continuous successful predictions: distinguishing between successful prediction and node failure is the reason for skipping data transmission and high cumulative error. This problem can be addressed by using an additional threshold value for the number of continuous and successful predictions. The introduction of this threshold value ensures that the sensor readings are sent by the sensor node after the threshold value is exceeded. This protocol significantly reduces redundancy in data transmission aside from eliminating transmission overhead. However, the trade-off between overhead due to communication and reduction in concurrent error should be determined to achieve the desired accuracy level.

2.6.3 Spatio-temporal Correlation Protocols

Spatio-temporal correlation protocols utilize correlation readings of nodes in both temporal and spatial dimensions. A high degree of correlation is usually present in the data sensed by geographically close nodes (spatial correlation) or/and sensed during a certain time period (temporal correlation). This observation is used by many protocols to benefit from spatially and temporally correlated sensor data.

(a) Cluster AGgregation (CAG):

The work in (Yoon & Shahabi, 2007) presented clustered aggregation (CAG) protocol. This protocol groups nodes that sense the same values into clusters (spatial correlation). No change is enforced in these clusters as far as the sensed values remain within a threshold over a time period (temporal correlation). CAG disregards data redundancy rapidly generates summary of data distribution. These features lead to remarkable energy savings. Moreover, CAG has two operation modes: interactive and streaming. The interactive mode involves the generation of only one set of responses for a query, whereas the streaming mode involves a periodic generation of responses to a query. Only the spatial correlation of data is exploited in the interactive mode, whereas both temporal and spatial data correlations are exploited in the streaming mode. In CAG, the estimation of the cluster size helps in assigning suitable weights to the CH values during the computation of the aggregates. Furthermore, these two modes facilitate the efficient computation of accurate results and ensure that the results are within the thresholds provided by the user, regardless of the data distribution. Finally, global communications are avoided in CAG, and clusters utilizing local communications are adjusted to reduce transmissions substantially.

(b) Adaptive Data Aggregation (ADA):

In (H. Chen, Mineno, & Mizuno, 2008) the authors presented adaptive data aggregation (ADA) for clustered WSNs. The aggregation rate at the CHs and the frequency of reporting at the sensor nodes control the defined degrees of spatial and temporal aggregation, respectively. In ADA, the current scheme state, which depends on the observed reliability, determines the spatial and temporal aggregation degrees. The major operation of ADA occurs at the sink node, but certain parts of the operation are also performed at the sensor nodes and CHs. ADA focuses on infusing adaptability to dynamic network conditions. A huge amount of energy is consumed in following the CH model of ADA. Additionally, the power consumption and network scalability are disregarded in the implementation of this mechanism.

(c) Efficient Data Collection Aware of Spatio-Temporal Correlation (EAST):

In (Leandro A Villas et al., 2013) the Efficient data collection aware of spatio-temporal correlation (EAST) protocol is proposed to conducts real-time data collection depending on the temporal and spatial correlations. The EAST scheme involves the two-leveled grouping of sensors by spatial correlation, whereas the representative nodes and leaders conduct temporal suppression. The spatial distance criterion or Euclidean distance between sensors is used to form exclusive correlated regions. In EAST, the sensed region is divided into "event areas," which are further divided into "correlation regions." An event area is managed by a coordinator node, whereas a correlation region is represented by a representative node because a single reading within the region is sufficient to represent the entire region. This approach can efficiently aggregate event-driven data, but the complexity increases for densely deployed WSNs. In addition, this protocol is unsuitable for dynamic applications. The EAST protocol uses the multi-hop scheme to send data to the destination and disregards any parameters related to energy consumption and congestion level in the selection of route nodes. As a result, a node failure resulting from energy depletion affects the capability of the protocol to forward packets successfully to the destination and leads to unstable network structure.

University of Malaya

Table 2.3: Brief Comparison of Different Data Correlation Protocols

Scheme	Route structure	Design objective	Spatial correlation	Temporal correlation	Scalability	Overhead	Limitations
GSC	Cluster	Remove data redundancy	Yes	No	Low	High	Not applicable to multi-hop members
UBC	Cluster	Determine the optimal cluster radius to achieve efficient data aggregation rate	Yes	No	High	High	High computational complexity and additional memory requirement
YEAST	Cluster and straight line segments	Minimize control overhead and maximize overlap routes	Yes	No	Very High	Low	Requires node position information and neglects the actual energy resources of nodes as well as the traffic load in the selection of route nodes
TEEN	Cluster	Decrease transmissions by constraining the required value of the sensed attribute	No	Yes	Very High	Low	No data received by the user in case of failure in reaching thresholds
APTEEN	Cluster	Increase flexibility and provide a detailed network picture	No	Yes	High	Moderate	Complex implementation and increased data transmissions
CoGKDA	Tree-based cluster	Provide a prediction mechanism for data series and reduce overhead incurred due to data transmission	No	Yes	Moderate	High	Excessive cumulative error
CAG	Cluster	Remove data redundancy	Yes	Yes	Moderate	Very High	Data-centric maintenance
ADA	Cluster	Attain the required reliability of the aggregated data at the sink node while decreasing the imposed cost on the sensor nodes and CHs	Yes	Yes	Low	High	Poor scalability and high cost
EAST	Cluster and straight line segments	Eliminate redundant notifications and maximize overlap routes	Yes	Yes	Very High	Moderate	Increases complexity for a large-scale and dense WSN and disregards any parameters related to energy consumption and congestion level in the selection of route nodes

University of Malaya

2.7 Chapter Summary

This chapter provided a background of sensor network technology and its evolution as well as briefly described the unique features, challenges, and requirements of WSNs. Moreover, in-network data aggregation was defined. In-network data aggregation protocols (i.e., routing protocols, data aggregation functions, data aggregation scheduling algorithms, and data representation algorithms) in WSNs were also reviewed in detail. Every protocol adopts an energy-saving mechanism to prolong the network lifetime. For this reason, using aggregation protocols is vital to reduce the number of data transmissions and energy consumption. The routing protocols for in-network data aggregation in WSNs were reviewed and classified. The main concepts and representative protocols of each type of routing protocol were discussed. Furthermore, a comparison between the different routing protocols was presented, with their main advantages and limitations highlighted. Finally, existing data correlation protocols and their connection to the present study were comprehensively reviewed. These protocols were categorized as spatial, temporal, or spatio-temporal correlation protocols. The routing and data correlation protocols discussed are summarized in Tables 2.2 and 2.3, respectively.

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

University of Malaya

CHAPTER 3: RESEARCH METHODOLOGY

3.1 Introduction

This chapter presents the general methodology used in the design and implementation of the proposed schemes in this thesis. First, systematic description of research methodology is presented, followed by discussion of various stages of structural modules of proposed methodology along with details of implementation. Discussion is also carried out on system model components used in this study. Finally, this chapter defines metrics considered for performance evaluation.

3.2 Research Methodology

A systematic organization of methodologies used in this work is conducted by categorizing proposed schemes into various stages and detailing implementation features of every stage. This organization clarifies structural components of proposed techniques and their integration to achieve research aims. Figure 3.1 highlights main research stages followed in the present study; they are discussed as follows:

3.2.1 Review of Literature

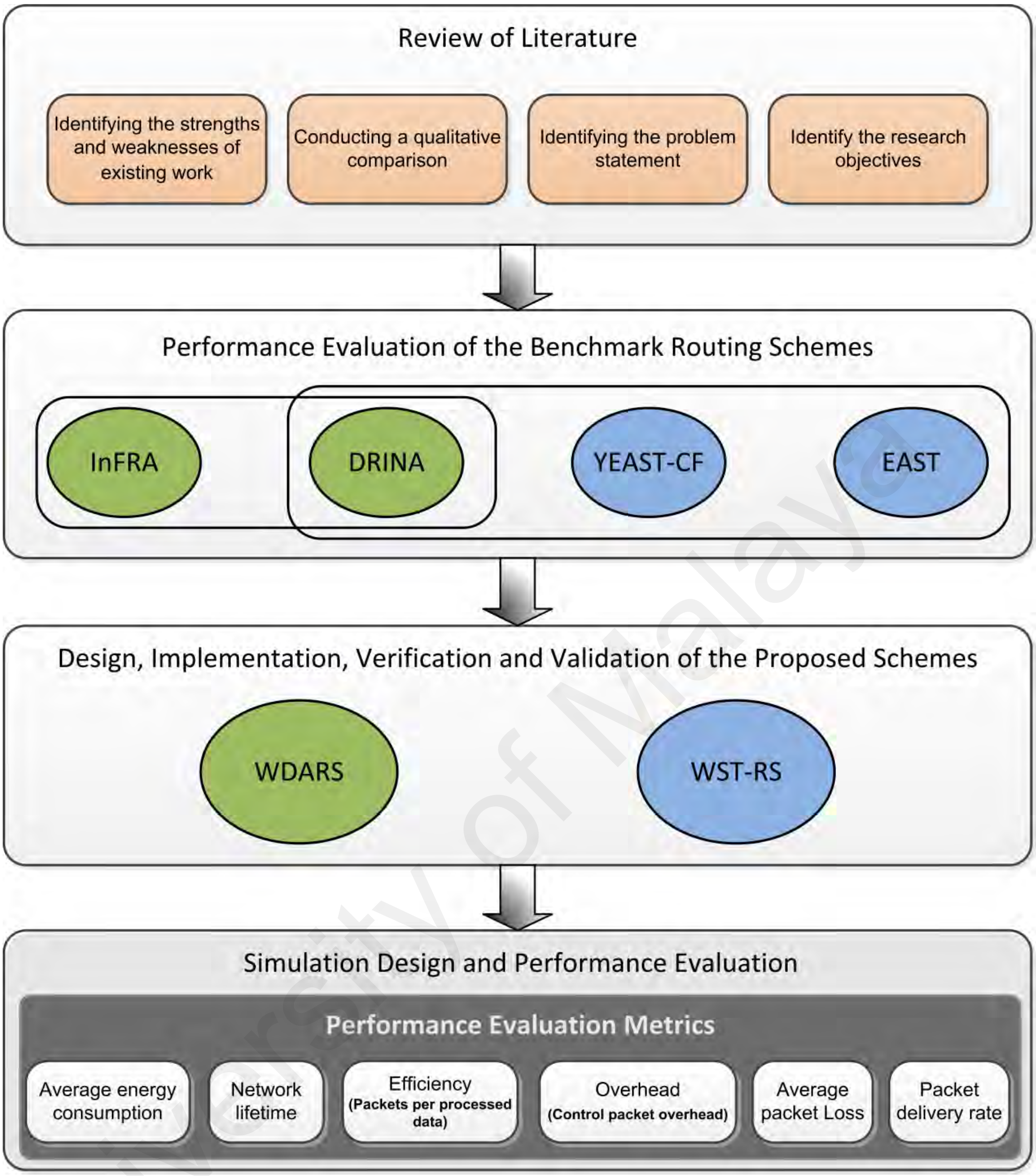
This stage reviews relevant and credible state-of-the-art studies on key issues of routing in wireless sensor networks (WSNs) to investigate support provided by routing protocol to save energy, thereby prolonging network lifetime. In WSNs, energy depletion causes failure of nodes and restricts their ability to forward packets to destinations. Redundant data sensed by nodes present additional constraint, which can affect design of routing schemes. As presented earlier in Chapter 2, most protocols proposed in literature attempted to solve several problems affecting routing in WSNs by exploiting specific network structure and combining correlated information. However, many drawbacks are confronted by existing routing strategies. Hence, efficient performance is not observed. Furthermore, the research on routing combined with

University of Malaya

spatial-temporal correlation mechanism and data aggregation in WSNs has previously uncovered new opportunities and challenges.

However, in Chapter 2, different in-network data aggregation and correlation awareness routing protocols are discussed and classified with special assertion on their route structure. According to routing classification in Chapter 2, the benchmark protocols are selected from the hierarchical structures categories to investigate and evaluate the performance of proposed schemes. The main objective of the benchmark studies (InFRA, DRINA, YEAST-CF, and EAST) is to target the issues of packet forwarding to facilitate the in-network aggregation of data in event-driven WSNs. These protocols are reactive cluster-based routing protocols which aim to increase aggregation rate by maximizing the formation of overlapping routes. The routes with the highest aggregation rates are selected to conduct data transmission, and the minimum number of control packets is required for the establishment of the routes. Moreover, YEAST-CF and EAST protocols aim to maximizing data aggregation along the communication path and controlling the energy usage by the nodes by eliminating unnecessary notifications. Therefore, in this research, Information-Fusion-based Role Assignment (InFRA) ,Data Routing for In-Network Aggregation (DRINA), dYnamic and scalablE tree Aware of Spatial correlaTion with variation of Closest First (YEAST-CF), and Efficient Data Collection Aware of Spatio-Temporal Correlation (EAST) protocols were selected depending on their required features and performance analysis of most existing routing protocols in WSNs.

University of Malaya



University of Malaya

Figure 3.1: Main Stages of Proposed Research Methodology

3.2.2 Performance Evaluation of Benchmark Routing Schemes

DRINA and EAST were implemented using MATLAB with IEEE 802.15.4 as medium access control (MAC) layer protocol. Features and mechanisms of DRINA and EAST were evaluated as previously explained in detail in Chapter 2. DRINA and EAST were validated using network simulator SinalGo version v.0.75.3 (Leandro A Villas et al., 2013; Leandro Aparecido Villas et al., 2013), wherein different scenarios were proposed to verify and to validate DRINA and EAST protocols and to compare them with other known routing protocols. After implementation of DRINA and EAST in MATLAB, several simulation tests were performed to verify and to validate their functionalities and to ensure correctness of implementation. Obtained *xls* file from simulator was also analyzed.

For example, node with hop to tree number 5 which is labeled in black in Figure 3.2, and it represents cluster head of the second event; this node attempts to access the path to multi-hop relay (MHR) nodes of routes established by previous nodes labeled with zero. The other routing paths between these two events are divided into three groups with different colors based on multi-criteria node weight metric: green nodes are lightweight, blue nodes are medium weight, and red nodes are heavy weight. In DRINA algorithm, multi-hop path only uses hop number to previously established path as link cost metric. Similarly, distance serves as link cost metric for EAST algorithm. This condition can lead to increased energy consumption of specific intermediate nodes and path congestion issues. These algorithms force paths from events occurring far apart from each other to overlap at the nearest overlap point and aggregate data from two distant events. This aggregation results in poor data formation and overloading in overlapping paths, thereby causing imbalance of energy consumption in networks and increasing the number of dead nodes. Total energy consumption is boosted by data transmission over lengthy paths.

Prior to development of new schemes, several simulations related to work presented in this research were used as basis and were executed to evaluate performance of DRINA and EAST. Information-Fusion-based Role Assignment (InFRA) and dYnamic and scalablE tree Aware of Spatial correlaTion with variation of Closest First (YEAST-CF) routing protocols were used at this stage for comparative performance analysis of basic and developed schemes. In this context, obtained results from assessment can

University of Malaya

show advantages and limitations of DRINA and EAST protocols according to comparison with InFRA and YEAST-CF, respectively, in terms of considered performance evaluation metrics. Based on this stage of research methodology, drawbacks of DRINA and EAST protocols were analyzed and identified. In this research work, these schemes were subsequently used as benchmarks for development of schemes under different scenarios.

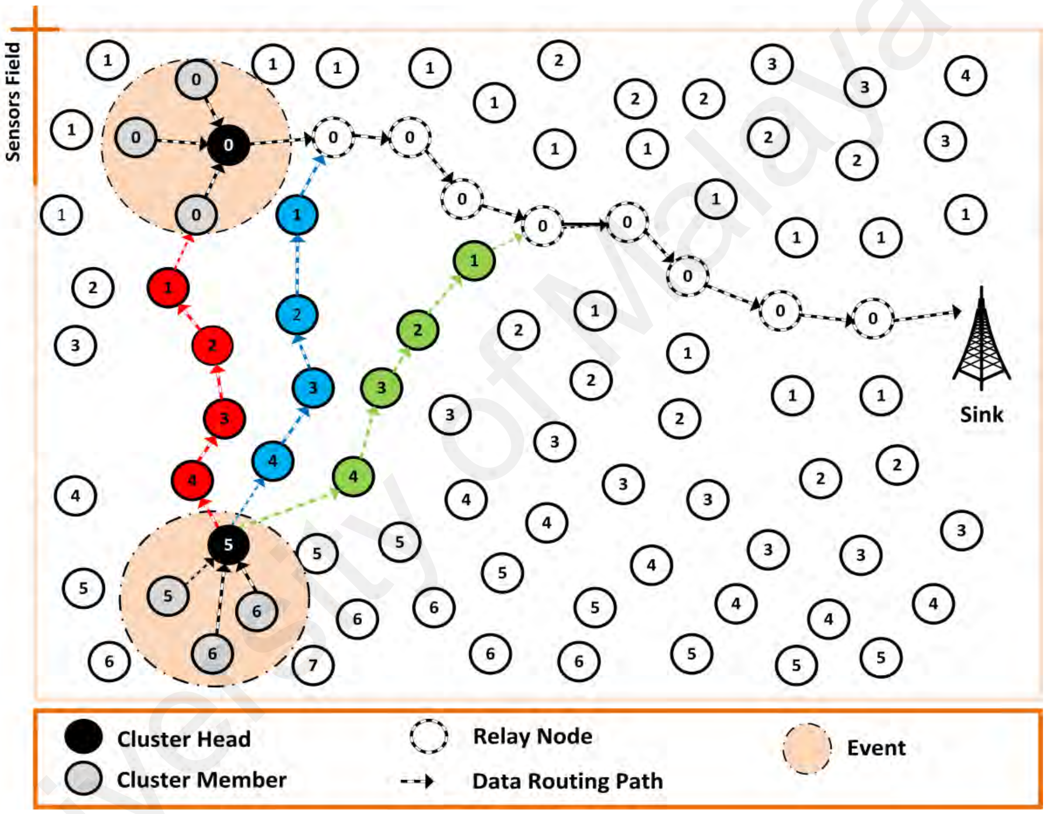


Figure 3.2: Validation Example of Building a New Path in Benchmark Routing Schemes

3.2.3 Design, Implementation, Verification and Validation of Proposed Schemes

This work proposes new energy-efficient and load-balanced routing schemes for in-network data aggregation in event-based WSNs. The proposed routing schemes aim to decrease energy consumption and optimally balance loads. This goal can be achieved by considering several node-related metrics, which reflect node status in terms of residual energy (E_{res}), available buffer memory level (B_{ava}), distance between node and hop tree in hops (HtT), distance between node and sink in hops (HtS), and node position (P_r) to

University of Malaya

determine routing decisions depending on link quality. The proposed schemes decrease communication overhead and data exchange by exploiting correlations in data readings of detected events. At this stage, methodology used is divided into three sub-stages: design, implementation, and validation.

3.2.3.1 Design of proposed schemes

This sub-stage oversaw development of weighted data aggregation routing scheme (WDARS) and weight aware spatial-temporal correlation routing scheme (WST-RS) for efficient data collection and aggregation. This process was conducted by modifying few functionalities of DRINA and EAST protocols to select the most stable nodes as cluster head, relay, and representative nodes for data collection, aggregation, and routing. WDARS selects an ideal point for overlapping of routes, resulting in fewer relay nodes. WST-RS scheme adaptively adjusts itself per node density in event areas and application accuracy requirement; such variables were not considered by the proposed solutions. WDARS ranks nodes based on their distance (in hops) from hop tree, distance (in hops) from sink, residual energy, and accessible buffer size. By contrast, WST-RS focuses on node positions to calculate physical distance, residual energy, and accessible buffer size. As shown in Figure 3.3, in designing proposed schemes, methodology is divided into four phases: discovery of node broadcast region and hop tree building between sensor nodes and sink, formation of clusters and election of cluster head, formation of spatial correlated regions and application of temporal correlation mechanism, and data routing based on multi-criteria node weight metric stage.

University of Malaya

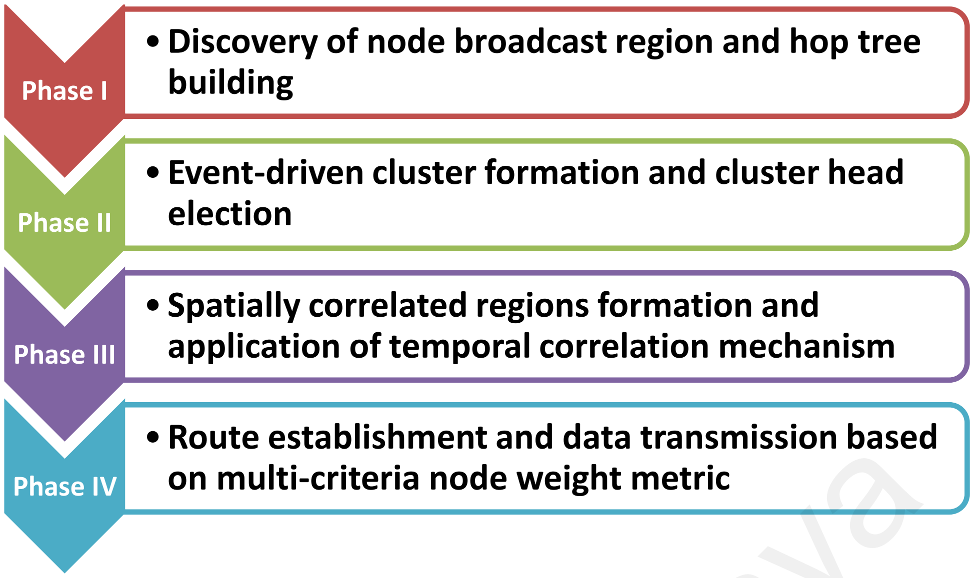


Figure 3.3: Design Phases of Proposed Schemes

Phase I: Discovery of node broadcast region and hop tree building:

Input of initialization phase is a set of sensor nodes with random deployment in pre-determined sensor field. Discovery procedure of broadcast region of nodes is established by discovering other neighboring nodes and calculating required metrics for newly introduced multi-criteria node weight (MCNW). Nodes assign initial weights to each link per values of proposed MCNW and save it in the table for their neighbors. Each node utilizes these weights to identify its neighbors as possible parents within its radio frequency (broadcast) region. MCNW comprises metrics related to energy and congestion degree, distances (in hops) to established paths and sink, and node position. These metrics are utilized by a new link cost function which measures link quality and selects the most efficient and reliable paths for data transmission. In this phase, calculated metrics are utilized in route computation, cluster head selection, representative node selection, and proper point determination for routes overlapping in successive phases. At the end of this phase, tree topology is joined by all network nodes, wherein root nodes of tree are sink nodes.

Phase II: Event-driven cluster formation and cluster head election:

As a result of dynamic cluster architecture, a cluster is reactively formed within proximity of event sensing nodes. Once an event is detected, the next step in design methodology is formation of dynamic cluster architecture and cluster head selection. In this phase, a specific sensor node is selected as cluster head (ideally the lightweight), and remaining sensor nodes become cluster member nodes. The main advantage of this approach is that only required nodes become active in sensing event data, thereby conserving energy expenditure of nodes outside event ranges. Data aggregation is constantly performed adjacent to events. Thus, data may be perfectly aggregated. Finally, every member node remembers its cluster head, and every cluster head is liable to collect information from associated member nodes and forward aggregated data toward sinks using MHR nodes.

Phase III: Spatially correlated regions formation and application of temporal correlation mechanism:

In this phase, a model of spatial-temporal correlation is performed between observations of every sensor node; correlation was exploited by second proposed scheme (WST-RS) to decrease communication and data exchange. The proposed spatial-temporal correlation model is based on observation that sensor readings in monitoring area exhibit high correlation for geographically close sensors (spatially correlated) and high correlation for specific periods of time (temporally correlated).

Spatial correlation is executed by decomposing entire areas of sensors per event into grid of spatially correlated regions and selecting single representative nodes at each spatially correlated region to perform data collection on behalf of all nodes in the same correlated region. Representative node selection is based on distance to center of correlated regions and residual energy of nodes. In this methodology, initial size of spatially correlated regions is defined per node density in event areas and evaluated by

University of Malaya

application according to desired accuracy. When size of spatial correlated region is equal to zero, spatial correlation is absent between sensor nodes, and all nodes in cluster are kept in sensing status. Otherwise, each cluster node determines spatially correlated region to which it belongs. The node should be aware of its position, central position of events, and size of correlated region for computations.

After forming spatially correlated region and selecting representative nodes, proposed methodology is applied for temporal correlation. Sensor nodes store additional information, which are utilized for temporal tolerance threshold to perform temporal correlation among successive sensor data readings. Stored information is based on the role of nodes in routing arrangement of sensors (i.e., representative node or cluster head). Representative nodes only store the last reported reading, that is, the last successfully sent reading to cluster heads by sensors, whereas cluster head node maintains last data reported from its cluster representative node. The proposed temporal correlation design phase compares new readings against last reported reading. Reporting of new value is made toward upper level nodes when error between these readings is greater than temporal tolerance threshold. Otherwise, new readings are suppressed. In case of suppression, the node applying temporal correlation mechanism is dependent on last reported data stored by upper level nodes, which are used to supply the missing groups. Temporal tolerance threshold value ranges from zero to any positive number. This threshold value is also application-specific and utilized by sink nodes to specify temporal correlation tolerance for collected data of events.

Phase IV: Route establishment and data transmission based on multi-criteria node weight metric:

In this phase, the methodology utilizes weights, which were created during the first phase for routing tree formation, in the neighbors table. Each node alerts its neighbors for possible parent nodes within its radio range. This condition implies that each node can use information in the neighbors table to send data packets to sink nodes. However, to even energy dissipation and to avoid congestion delay caused by data collisions, adopted methodology monitors remaining energy level and accessible buffer memory of nodes in MHR set. When values are lower than set limits, then new routing path formation is initiated. The methodology considered by proposed schemes successfully maximizes routes overlapping via ideal aggregation point while simultaneously ensuring that data transmission across lightweight routes is achieved with desired reliability in WSNs.

3.2.3.2 Implementation of proposed schemes

At this sub-stage of our work, WDARS and WST-RS are developed for efficient data collection and aggregation, as shown in Figure 3.4. Simulation tool used in this study was MATLAB, whereas IEEE 802.15.4 at MAC layer was used for carrying out simulations. 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 interactive environment for algorithm development, data analysis, and plotting capabilities.

University of Malaya

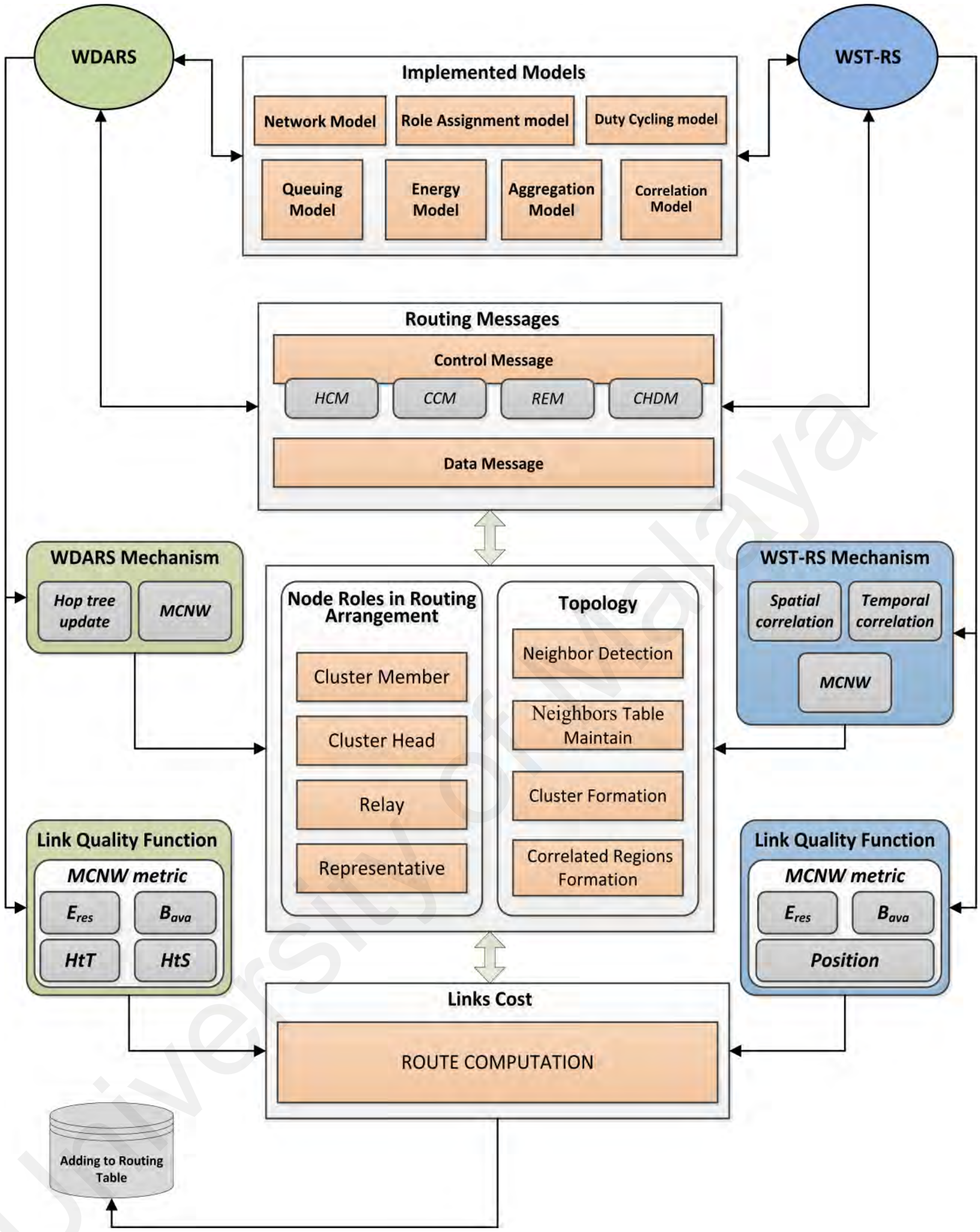


Figure 3.4: Architecture of Implemented Proposed Schemes

3.2.3.3 Verification and validation of proposed schemes

Verification confirms that the proposed schemes are correctly converted from pseudo code to functional application. In this section, we verify and validate implementation of proposed routing schemes by performing extensive simulation tests. Initially, we verify functionalities of MCNW for route computation, cluster head, and representative nodes election to ensure that it will provide the same expected results using simulation approach. This step further tests MCNW's behavior in different simulation scenarios and investigates the period during which desired connectivity can be maintained by the network. Different evaluation metrics were used to assess data aggregation rate, communication overhead, routing tree quality, and load balancing. Such assessment can provide evidence that the proposed scheme satisfies design requirement and provides high aggregation rate, avoiding nodes with limited energy resources and high congestion degree. This test also optimally balances load in networks and efficiently exploits data correlation in readings of sensor nodes to decrease data communication. Chapters 4 and 5 respectively discuss further details regarding implementation, verification, validation, and performance analysis of WDARS and WST-RS.

3.2.4 Simulation Design and Performance Evaluation

At this point of research, experiments are performed using MATLAB R2013a in a computing environment with 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. Initially discussed techniques are evaluated based on performance. A comparison of proposed schemes was carried out depending on specific performance metrics related to proposed design aims. Subsequent sections provide more details of this step; such details include system models and considered performance metrics.

3.3 System Models

Various models were implemented to evaluate the proposed routing methods. Hence, a brief description of the models used is provided in this section to clarify the concept. Table 3.3 at the end of this section summarizes the common network models utilized in this work.

University of Malaya

3.3.1 Network Model

From a viewpoint of graph theory, a wireless sensor network can be modeled by an undirected graph G (V, E), where network nodes' set is denoted by V and E is a set of arcs; $E \subset V \times V$ represents the set of communication links between a pair of nodes. Every node $\upsilon \in V$ can conduct direct communication with neighbor nodes that lie in its communication range. For nodes which are not adjacent neighbors, routing protocol is used for packet transmission. Packet transmission from node υ to the destination node can be done by selecting one of the probable paths in the graph G (V, E). The average hop length E is increased with the increase in number of nodes' set V. As a result, performance of the routing protocol is affected. Hence, network size is considered as an important parameter in the evaluation of routing protocols in WSNs. To study the impact of network size on performance of the network using proposed schemes, number of nodes was varied in this study

3.3.2 Duty Cycling and Sleep Scheduling Model

There are various important notification rounds for every network node in the proposed methodologies. Two periods, period 1 and period 2, are included in every round. During the first period, in which the proposed routing schemes play an important part, packet transmission within the communication range is performed by every sensor node. During the second period, detection and sensing of any occurring event is done by the network sensor nodes. There are two different working modes for every node in period 2:

- **Sensing mode:** All probable events can be detected by the node during this period resulting in high energy consumption.
- **Sleeping mode:** During this mode no message transmission activity is performed as sensing components of the node are switched off. Much less

University of Malaya

energy is consumed during sleeping mode than sensing mode. Hence, there is a decrease in overall network energy consumption if many nodes are in sleeping phase.

The features and differences of both periods, i.e. sensing mode and the sleeping mode are given in Table 3.1. Switching of sensor status between two modes is possible.

Table 3.1: The Characteristics of Duty Cycling Phases

Functions	Phase 1	Phase 2	
		Sleeping	Sensing
Timer	√	√	√
Sensing	×	×	√
CPU	√	×	√
Transmitting	√	×	×
Receiving	√	×	×

3.3.3 Energy Consumption Model

Various WSN protocols have been presented in literature that assumed certain radio features in transmission and receiving phases. In this work a well formulated first-order radio model is used without loss of generality (W. R. Heinzelman, Chandrakasan, & Balakrishnan, 2000). According to the energy consumption model in Figure 3.5, the communication energy consumed to transmit an ℓ-bit packet over a distance d by the radio can be described by the following equation:

$$E_{Tx}(\ell, d) = \ell \times E_{node(Tx)} + \ell \times \varepsilon_{op-amp} \times d^{\eta} \qquad (3.1)$$

Furthermore, a sensor node is responsible in forwarding data of other sensor nodes. In this procedure, the sensor nodes receive data packets and their energy is not affected by variation of distance between communication pairs. Therefore, the energy required for receiving an ℓ-bit packet over a distance d is given by the following equation:

University of Malaya

$$E_{Rx}(\ell) = \ell \times E_{node(Rx)} \qquad (3.2)$$

The change in distance between communication sensor nodes affects transmitter energy $E_{Tx}(\ell,d)$ only whereas receiver energy is not related to it and stays constant. Therefore, the equation 3.2 can be rewritten as follows:

$$E_{TX}(\ell,d) = \begin{cases} \ell \times E_{elec} + \ell \times E_{fs} \times d^2, & if\ d < d_0 \\ \ell \times E_{elec} + \ell \times E_{mp} \times d^4, & if\ d \geq d_0 \end{cases} \qquad (3.3)$$

Where, E_{elec} denotes the dissipated energy that is consumed to run the transmitter and receiver (E_{TX} and E_{RX}, respectively). E_{elec} is based on different features, such as modulation, digital coding, filtering, and spreading of signals. Distance between transmitter and receiver is denoted by d, whereas d_0 is calculated as $d_0 = sqrt(E_{fs}/E_{mp})$. E_{mp} and E_{fs} depend on the distance between the transmitter and the receiver and the transmitter amplifier model. If d is greater than d_0, then the multipath model (d^4) is used. Otherwise, the free space model (d^2) is used to measure the dissipated energy (Elbhiri, El Fkihi, Saadane, & Aboutajdine, 2010).

The sensing structure of each sensor node is activated and data is collected from the surroundings by consuming sensing energy denoted by E_S. It is assumed that a constant amount of energy $E_{node(sensing)}$ is dissipated to sense one bit. Hence, the overall energy dissipation for ℓ- bits is given by the following equation:

$$E_S = E_{node(sensing)} \times \ell \qquad (3.4)$$

Finally, the sensor network has been structured into distributed clusters. The sensor nodes not only forward the data but also contribute in making key decisions regarding the network operation. Furthermore, the computation of data aggregation consumes extra energy $E_{node(aggregation)}$, but it relatively low in comparison to the energy dissipated during communication (Raghunathan, Schurgers, Park, & Srivastava, 2002). It is

University of Malaya

assumed that a constant amount of energy $E_{node(aggregation)}$ is dissipated to aggregate one bit. Hence, the overall energy dissipation for ℓ-bits is given by the following equation:

$$E_{Aggre} = E_{node(aggregation)} \times \ell \qquad (3.5)$$

For simplicity, we assume the energy dissipated for receiving, computation and sensing are the same, and equal to 50 nJ/bit. Table 3.2 defines the meaning of different energy terms in this work and their typical values.

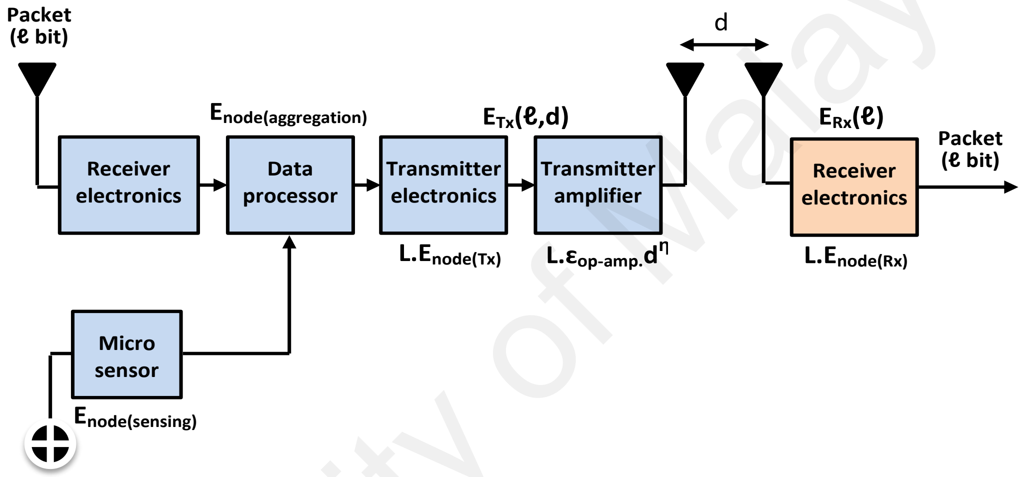


Figure 3.5: Energy Consumption Model

Table 3.2: Energy Consumption Parameters

Term	Meaning
$E_{node(Tx)}$	Energy consumed in transmitter electronics per bit (nJ/bit)
$E_{node(Rx)}$	Energy consumed in receiver electronics per bit (taken to be 50nJ/bit)
$E_{node(sensing)}$	Energy consumed in sensing one bit (taken to be 50 nJ/bit)
$E_{node(aggregation)}$	Energy consumed for aggregating one bit (taken to be 50 nJ/bit/signal)
$\mathcal{E}_{op\text{-}amp}$	Energy consumed in the transmitter amplifier for either a free-space channel or a multi-path fading channel (taken to be 10 pJ/bit/m^2 and 0.0013 pJ/bit/m^4 respectively)
d	Distance that the data has to travel
ℓ	Data packet size in bits
η	Path loss exponent with range between 2 and 6

3.3.4 Role Assignment Model

For setting up routing infrastructure, following roles are considered by the adopted methodology in proposed schemes:

- **Cluster Member (CM):** This node is responsible for the discovery of an event and forwarding the gathered data to the cluster head.
- **Cluster Head (CH):** Cluster head is responsible for event detection and performs the data aggregation. Then, the aggregated data are transmitted towards the sink.
- **Relay:** A node whose duty is to forward the received data towards its next possible hop. In some cases, relay nodes represent a data aggregation point when the data paths are overlapped on it.
- **Sink:** A collection of nodes or personal computers having high computational energy and processing capability. The sink is liable to receive all the data from the cluster head and other member nodes.
- **Representative Node (RN):** A node whose duty is to detect and report an event to a cluster head. It is representative of all spatial correlated region nodes with same readings.

All network nodes except the sink act as relay nodes during the time when there is no detection of any event. When at least one node detects an event, the proposed routing scheme starts and role assignment takes place. The role assignment procedure depends on the node status and its position in the event.

3.3.5 Queuing Model

Before the data packets are passed to the data link layer, they are temporarily buffered in queues by the network layer due to limited data link capacity. Generally,

University of Malaya

different outgoing interfaces have separate queues which are maintained by the network protocol. For every outgoing interface, multiple queues are also maintained which have different priorities. The inspiration of queuing model employed in this study is obtained from (Akkaya & Younis, 2003). Packet assignment to relevant queues is done depending on their priorities. The primary type of priority queue is identified as the drop tail queue.

3.3.6 Data Aggregation Model

The assumption that sensing devices of each node are homogenous and single packet generation of ℓ bits by every node is used in this study. Let $\mathcal{P}_{i-incoming}$ be the generated packet in node i which contains locally sensed traffic at node i ($\mathcal{P}_i$) and the sum of packets flowing in from neighboring nodes $\mathcal{P}_j$ which monitor similar events, and is represented by:

$$\mathrm{P}_{i-incoming} = \mathrm{P}_i + \sum_{j\in S_j}^{k} \mathrm{P}_i = (k+1)\times \ell, \quad i = 1, \ldots, N. \qquad (3.6)$$

Where, S_j (j = 1, . . . , k) represents set of neighboring nodes detecting similar events. The total amount of data going out from node i is denoted by $\mathcal{P}_{i-outgoing}$ and given by the following equation:

$$\lambda_{\,i-outgoing} = \alpha \times \mathrm{P}_{\,i-incoming} = \delta \times (k+1) \times R \qquad (3.7)$$

Where, δ denotes the aggregation factor ranging from $(\frac{1}{k+1})$ to 1. When δ = $(\frac{1}{k+1})$, multiple data packets have been aggregated into a single packet. This condition is also referred to as perfect aggregation. When δ = 1, total number of outgoing and generated packets is same as aggregation has not been performed by the node. Hence, at node i the number of packet transmissions τ_i is given by:

University of Malaya

$$\tau_i = \lceil \delta \times (k + 1) \rceil packets \qquad (3.8)$$

Where ⌈ ⌉ specifies the ceiling of $\delta \times (k+1)$.

Table 3.3: Brief Description of the Various System Models

Model	Description
Network Model	Modeling the network topology as an undirected graph *G (V, E)*, *V* represents the set of network nodes and *E* is a set of communication links between a pair of nodes if they are within transmission range of each other.
Duty Cycling and Sleep Scheduling Model	Every sensor node takes several working rounds and each round is composed of two periods (period 1 and period 2). *Period 1:* in which the proposed routing schemes play an important part by transmitting the sensed data. *Period 2:* in which the detection and sensing of any occurring event is done by the network.
Energy Consumption Model	The energy consumption model can be categorized into three energy consumption submodules: communication, computation, and sensing. Among these activities, the communication submodule consumes a significant amount of energy in the sensor nodes. It depending on the distance between source and destination node. In this research work, we utilize the radio transceiver model as in (W. R. Heinzelman et al., 2000).
Role Assignment Model	Specifying the required actions to the nodes in the routing arrangement. In proposed schemes the roles such as cluster member, cluster head, relay, sink, and representative node are assigned when any event takes place.
Queuing Model	The queuing model employed in this study is obtained from (Akkaya & Younis, 2003) in which the packet assignment to relevant queues is done depending on their priorities. Queuing of packets takes place till the availability of buffer space. The queue is checked on arrival of a new packet. If it is full, the packet is dropped.
Data Aggregation Model	Data aggregation can be application specific and can adopt various forms. However, the adopted data aggregation model in this study is a basic version of data aggregation. The node waited and aggregated the data for a fixed time whenever it received its first data packet at the aggregation points. This time point is known as the data aggregation time, which is before the data are transmitted to the sink node. This time was set to 150 ms for all the compared protocols. This delay could be minimized by setting the degree of aggregation (He et al., 2006) to five packets, that is, if five packets are received by the aggregation node before the expiration of the data aggregation time, then the aggregated packet is transmitted by the aggregation node to the sink node.

University of Malaya

3.4 Performance Evaluation Metrics

By carrying out intense simulations, 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:

- **Energy consumption:** This metric gives information about the total energy consumed in sensing, processing and transmission 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.
- **Network lifetime:** The time during which the network is able to carry 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 dies. In our performance evaluation, the former definition is used as sensor network lifetime.
- **Packet delivery rate:** This metric is defined as the number of packets successfully received at the sink node. Routing tree's quality is determined by this metric (lower packet delivery rate indicates higher built tree aggregation rate).
- **Efficiency (packets per processed data):** This metric is defined as the rate between the total data and control packets transmitted and the number of packets received by the sink.
- **Average packet loss:** This metric is defined as the percentage of packets lost with respect to packets sent. In this research, we calculate the average of packet loss between aggregation point and the sink node.

University of Malaya

- **Overhead (control packet overhead):** This metric is indicated to the routing tree formation cost which includes the total number of control packets used to create the clusters and set up the routing parameters.

3.5 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 (InFRA, DRINA, YEAST-CF, and EAST) to the design and the implementation the proposed schemes (WDARS and WST-RS). The research methodology flowchart showed the sequence of the research phases and presented information on the connections between the structural components of every phase. The different system models implemented in the simulation were described. Finally, the performance evaluation metrics considered to study the performance of the proposed schemes were defined. The next chapter presents the proposed WDARS routing scheme with its performance evaluation.

University of Malaya

CHAPTER 4: WEIGHTED DATA AGGREGATION ROUTING SCHEME (WDARS)

4.1 Introduction

Event-driven networks, such as WSNs, generate a substantial amount of data that should be routed via multi-hops to the sink node. Therefore, routing protocols play a significant part in gathering and forwarding data in WSNs. In-network data aggregation is a strategy for optimizing a routing task in WSNs. In in-network data aggregation, the processing capability of intermediate sensor nodes along the routing paths is utilized (see (Gasser, 2014) for more details). This approach reduces a significant number of bytes that are transmitted during the network operation by aggregating data at the intermediate nodes to allow for bandwidth and energy savings. The issues of redundancy and numbers of transmissions are reduced by employing in-network data aggregation.

In event-driven WSNs, the monitoring capability deteriorates when the overlapping paths of uncorrelated events perform extensive data aggregation. Consequently, network performance is not improved. In addition, inefficient in-network data aggregation neglects the network state and causes the early energy depletion of multi-hop relay (MHR) nodes and the uneven structure of the network because of the excessive amount of dead nodes. Therefore, a balance between optimizing data aggregation and link cost is necessary.

In this chapter, an energy-efficient and load-balanced routing scheme for in-network data aggregation is presented. This scheme is referred to as weighted data aggregation routing scheme (hereinafter referred to as WDARS). This routing scheme is a modification of DRINA. WDARS aims to provide an efficient and reliable multiple-hop routing to the destination on the basis of the quality of the links between the source and destination. In view of this goal, a link cost function is introduced to assess the quality

University of Malaya

of the links by considering the new multi-criteria node weight (MCNW) metric, in which both energy efficiency and load balancing are considered. WDARS uses a hop tree to attain the optimal data aggregation. The node weight is considered in constructing and updating the hop tree to achieve dynamic behavior for event-driven WSNs. Moreover, WDARS determines the ideal point for routes overlapping by considering the distance to the established routing path by previous events and the distance to the sink node. The proposed WDARS was evaluated and validated by simulation, and the results were compared with those of InFRA and DRINA by using performance metrics for dense static networks.

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

4.2 Sensor Network Model and Basic Assumptions

The important aspects of the network and the assumptions used in the proposed scheme are as follows:

- Sensor nodes are randomly deployed all over the sensor field.
- Once the sensor nodes are deployed, they become stationary and are capable of performing different routing roles, such as CH and normal sensor node.
- All sensors exhibit static and homogeneous behavior in terms of storage, processing, battery power, sensing, and communication capabilities.
- The distance between two nodes can be computed in hops, depending on the received hop configuration message. Therefore, a location-tracking system, such as GPS, is not required.
- Sensor nodes can compute their residual energy and available buffer memory.

University of Malaya

- The sensing procedure is performed by every node when significant events are observed.
- A sensing and communication distance area exists in a circular field. Two sensors within the same range are considered neighbors.
- Every node possesses a unique ID, and it can forward data at any time during the network operation.
- A symmetric wireless radio channel is considered for modeling so that the energy required transmitting data from sensor node S_i to sensor node S_j is equal to the energy used for transmitting data from sensor node S_j to sensor node S_i.
- Finally, a single BS is considered distant from the sensor field and connected to the power supply, whereas the sensor nodes are non-rechargeable and expire after their energies are exhausted.

4.3 Description of WDARS Scheme

The developed WDARS routing scheme introduces appropriate modifications to solve the issues related to energy and load balancing in DRINA. In WDARS, certain functions of DRINA are retained and leveraged, whereas other functions are significantly modified. Like DRINA, WDARS is a routing protocol for in-network data aggregation. It uses MHR routing to send packets from the source to the destination. It also maintains a routing tree that incorporates the newly established route by updating the value of the hop tree of a node to maximize routes overlapping to promote in-network data aggregation.

Both energy efficiency and load balancing are enhanced in the developed WDARS routing scheme by introducing the following new mechanisms to modify the main functions of the conventional DRINA:

University of Malaya

(i) A new MCNW performance metric is introduced. It comprises four routing metrics, namely, residual battery, available buffer memory, distance (in hops) to the hop tree, and distance (in hops) to the sink node.

(ii) The link cost is computed using the developed MCNW metric. Furthermore, the structures of the hop configuration message (HCM) and cluster configuration message (CCM) were modified to include the MCNW metric in the routing tree formation and CH selection processes. Then, the routing algorithm determines the best path to the destination while optimizing the data aggregation. The best path is derived using the ideal aggregation point according to the new developed link metric instead of the hop count, which was used in the conventional DRINA.

The proposed protocol forms a completely distributed cluster and an efficient routing tree with optimal energy conservation and congestion avoidance. Figure 4.1 depicts the suggested approach and the roles used in the routing arrangement.

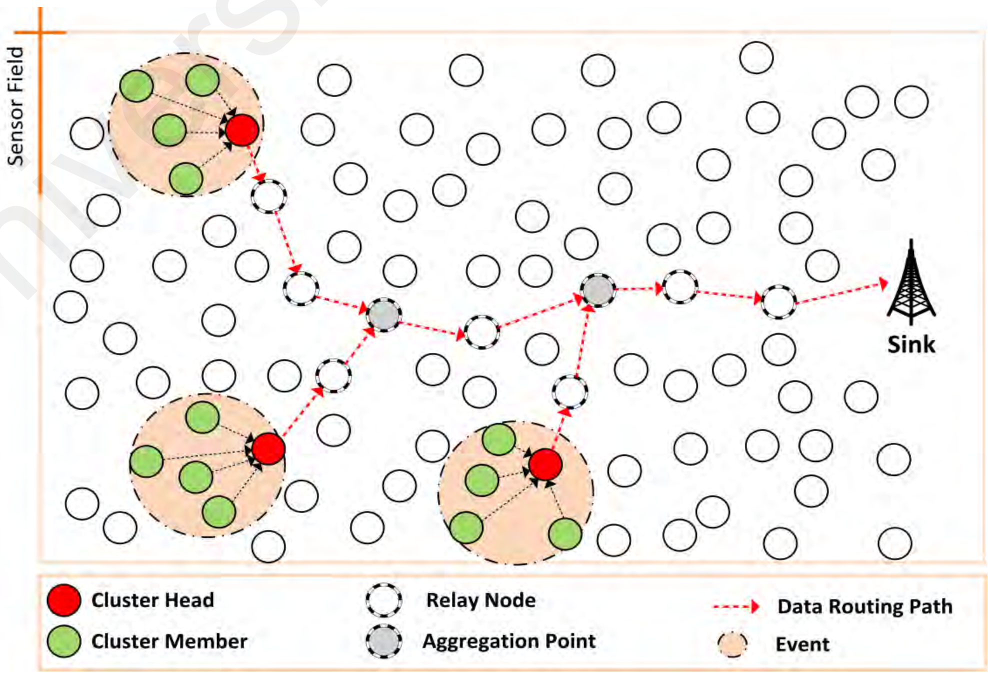


Figure 4.1: Diagram of the Network Model of WDARS for Routing to the Sink Node

University of Malaya

Other primary and secondary functions of DRINA, namely, dynamic cluster formation, hop tree updating, and route repair mechanism, were retained. Thus, in this section, the design components of WDARS are described in conjunction with the proposed modifications. However, before these components are described, the developed MCNW metric is first explained in the following subsection.

4.3.1 MCNW Metric for WDARS

The utilized MCNW metric for node performance in the proposed WDARS routing scheme is based on four metrics, namely, residual energy (E_{res}), available buffer memory size (B_{ava}), hop to sink (*HtS*), and hop to tree (*HtT*). The MCNW performance metric represents the status of the nodes in terms of their energy resource, congestion level, and its distance (in hops) from a path established by previous events and distance (in hops) to the sink node. For estimating the MCNW for every node *i*, four individual node weights. These weights depend on the involved node-related metrics, namely, residual energy-based node weight (NW_E_{res}), available buffer-based node weight (NW_B_{ava}), hop to sink-based node weight (*NW_HtS*), and hop to tree-based node weight (*NW_HtT*).

(a) Residual energy-based node weight (NW_E_{res}):

The E_{res} of a node *i* indicate residual energy of the battery attached to that node at a specific instant. This parameter is derived from the battery model. The node weight based on E_{res} is calculated using Equation 4.1. The node whose remaining energy is high corresponds to lightweight NW_E_{res}, which reduces the probability of energy exhaustion.

$$NW\_E_{res}(i) = \left(1 - \frac{E_{res}(i)}{E_{init}(i)}\right)^2 \qquad (4.1)$$

University of Malaya

Where $E_{res}(i)$ is the residual energy of node i at an instant and $E_{init}(i)$ is the initial battery energy level of the node, which refers to the maximum battery capacity. Equation 4.1 suggests that the result approaches 1 when the remaining energy of node i decreases. Conversely, the resulting node weight approaches 0, and the node weight decreases when the remaining energy is high. Furthermore, if the node energy does not change (i.e., the same as the initial energy), then 0 energy weight is obtained.

(b) Available buffer-based node weight (*NW_B_{ava}*):

The accessible buffer size represents the residual memory space, which can be used to store the sensed data during the time a node is waiting to get serviced. Data buffering occurs when a node receives data whose amount exceeds the amount of data it can forward. However, if no buffer space is available at the node, then the data is dropped, and congestion ensues. Thus, each node should be aware of the buffer size of its neighbors to conduct a reliable packet transmission and to avoid congestion among the nodes. The node weight depending on the available buffer size is denoted by NW_B_{ava}, which is calculated by Equation 4.2. The node with a high buffer size corresponds to lightweight NW_B_{ava}, which leads to minimal congestion and packet loss.

$$NW\_B_{ava}(i) = \left(1 - \frac{B_{ava}(i)}{B_{total}(i)}\right)^2 \qquad (4.2)$$

Where $B_{ava}(i)$ is the available buffer memory of node i at an instant and $B_{total}(i)$ is the node's total buffer size, which refers to the maximum buffer capacity. As suggested by Equation 4.2, the node weight is approximate to 0 when the available buffer memory is large, whereas the node weight approaches 1 when the buffer size is exhausted.

(c) Hop to sink-based node weight (*NW_HtS*):

The hop to sink is the distance (in hops) from the node to the sink. The candidates for the next hop may be one hop closer or farther from the sink, or their hop distances from the sender can be same because the information of only one-hop neighbor is used in the proposed scheme. The value of *HtS* metric is used by the neighbors of node i to determine the weight of node i on the basis of the distance (hop count) to the sink. This node weight is determined by Equation 4.3. The node with a low number of hops to the sink node corresponds to lightweight *NW_HtS*, which leads to low energy consumption.

$$NW\_HtS(i) = \frac{\Big(\big(HtS(i) - HtS(n)\big) + 1\Big)}{HtS(i)} \qquad (4.3)$$

Where *HtS*(*i*) is the minimum number of hops between the node *i* and the sink node and *HtS*(*n*) is the number of hops between the neighbor of node *i* and the sink node. Equation 4.3 suggests that 0 weight is obtained if node *i* is one hop closer than its neighbor to the sink. Conversely, the weight is 2 if node *i* is one hop farther than its neighbor from the sink. The weight of node *i* is 1 if the hop distances of node *i* and its neighbor are the same.

(d) Hop to tree-based node weight (*NW_HtT*):

In addition to hop to sink, each the hop to tree parameter of the node is used as a metric. The metric requires that the minimum number of hops between the node and the path established by previous events should be maintained. At the beginning of tree formation, the same values are assigned to *HtT* and *HtS*. The value of the *HtT* parameter changes immediately after the first event is detected. It then continues to change with the occurrence of new events. On the contrary, the value of *HtS* remains unchanged in every node. The *HtT* of any node may change because of the occurrence of the following events:

University of Malaya

(i) When a node is included in the backbone structure (MHR nodes), that is, the *HtT* of the sink node and the other nodes belonging to the backbone structure are zero.

(ii) When an HCM received by a node provides more accurate information about the distance.

In addition, candidates for the next hop may be one hop closer or farther from the path that has already been established, or their hop distances from the sender can be same as only the information of one-hop neighbors is used in the proposed scheme. The value of *HtT* metric is used by the neighbors of node *i* to determine the weight of node *i*, weight on the basis of the distance (number of hops) to the established path. This node weight is calculated by Equation 4.4. The node with a low number of hops to the destination corresponds to lightweight *NW_HtT*, which leads to maximum routes overlapping and high aggregation rate.

$$NW_{HtT(i)} = \frac{\left(\left(HtT(i) - HtT(n)\right) + 1\right)}{HtT(i)} \qquad (4.4)$$

Where *HtT(i)* is the minimum number of hops between node *i* and the path established by previous events. The *HtT(n)* is the number of hops between the neighbor of node *i* and the established path. As suggested in Equation 4.4, if node *i* is one hop closer than its neighbor to the established path, then its weight is 0. If node *i* is farther than its neighbor from the established path, then its weight is 2. If the hop distances of node *i* and its neighbor are the same, then the weight of node *i* is 1.

4.3.2 MCNW Weight Calculation in WDARS

Based on the four specific node weights discussed in Section 4.3.1, a composite MCNW weight of node *i* is constantly and independently measured in accordance with a normalized weighted additive utility function (NWAUF; (Malakooti & Thomas, 2006). The NWAUF weight depends on the values satisfying the normalizing criteria and weights of importance that range from 0 to 1. The final MCNW weight of node *i* is estimated by Equation 4.5.

$$NW(i) = W_1 \times NW_{E_{res}(i)} + W_2 \times NW_{B\,ava}(i) + W_3 \times NW_{HtS(i)} + W_4 \times NW\_HtT(i) \tag{4.5}$$

Where *NW_Eres(i)*, *NW_Bava(i)*, *NW_HtS(i)*, and *NW_HtT(i)* are the node weights of the node-related metrics of MCNW; W_1, W_2, W_3, and W_4 are the normalized weight factors of the node weights. The sum of the normalized weight factors, which denote the importance of the components of the MCNW metric, is equal to 1. Equal weights can be assigned to the different metrics according to the additive combination rule (Farooq & Tang Jung, 2013). In this study, the normalized weight factors of E_{res} and B_{ava} metrics are set to 0.2. The normalized weight factors of *HtS* and *HtT* metrics are relatively high compared with those of the first two metrics, and they are set to 0.3.

4.3.3 Details of WDARS

The proposed scheme in this chapter comprises three phases. The first phase involves the broadcast of the region discovery by the nodes and the establishment of a hop tree between the sensor nodes and sink node. The second phase begins as soon as a node senses any event. In this phase, clusters are formed, and CHs are selected. The third phase includes route establishment, data aggregation, and routing.

4.3.3.1 Discovery of node broadcast region and hop tree building (Phase I)

The input in the initialization phase is a set of nodes deployed in a pre-determined sensor field. Each node identifies its neighbors, which are possible parents, within its radio frequency (broadcast) region. It also determines its hop distance to the sink, residual energy, and available buffer size. The algorithm responsible for initialization begins by broadcasting an HCM from the sink to all the sensors within its communication range (Step 1 of Algorithm 4.1). In addition to the common message fields, the HCM contains five key parameters, namely, *Node-ID*, *Type*, *HtT*, *HtS*, E_{res}, and B_{ava}, which are defined in Table 4.1.

Table 4.1: Header of the HCM for WDARS

No.	Parameter	Description
1	*Node-ID*	ID of the node that transmitted/retransmitted the HCM
2	*HtT*	Distance from the node to the hop tree (in hops)
3	*HtS*	Distance from the node to the sink (in hops)
4	E_{res}	Residual energy of the node
5	B_{ava}	Available buffer memory size of the node

As mentioned in Section 4.3.1, the same values are assigned to *HtT* and *HtS* at the beginning of the tree formation. The *HtT* initial value for the sink node is 0 and infinity for the other nodes when the hop tree begins to form. An actual value of the node energy is used, and the available buffer memory size of a node is considered maximum. Once the neighboring nodes of the sink receive the HCM (Step 2 of Algorithm 4.1), the node performs the following tasks:

(i) Verify whether the value of *HtT* in the HCM message is lesser than its *HtT* value (Step 3.1 - Algorithm 4.1) to guarantee that each node records the minimum number of hops to the sink.

(ii) Depending on the validity of the condition in (i), the node maintains the information of its neighbors whose HCMs are received in its neighbors table (Step 3.1.1 of Algorithm 4.1).

University of Malaya

(iii) The node also verifies whether the value of *First_Sending* is true (Step 3.1. 2 of Algorithm 4.1). If the value of *First_Sending* is true, then sensor node increases the values of *HtT* and *HtS* by one in a sensor node. Then, the sensor node computes their remaining energy after one complete transmission and updates the E_{res} field. Moreover, it computes the obtainable buffer size and updates the B_{ava} field and finally circulates the HCM to other neighbors, as shown in Algorithm 4.1 (Steps 3.1.2.1 to 3.1.2.3). Otherwise, if the condition of *First_Sending* is false, that is, the HCM has already been sent by the node.

(iv) If the condition in Step 3.1 of Algorithm 4.1 is false, then the HCM message will be dropped (Step 3.3 of Algorithm 4.1), which indicates that the stored *HtT* in the node provides more accurate information to the sink.

(v) The node also updates the routing table (Steps 3.4 and 3.5 of Algorithm 4.1) by using the MCNW metrics to compute the weights of their next-hop neighbors and to select the lightweight node as its next hop, depending on Equation 4.5.

This process continues until the tree topology is formed by all the network nodes. The sink node is the root node of the tree. Figure 4.2 shows the hop tree building process, in which the numbers in the sensors indicate the *HtT* values, which increase as it moves away from the sink.

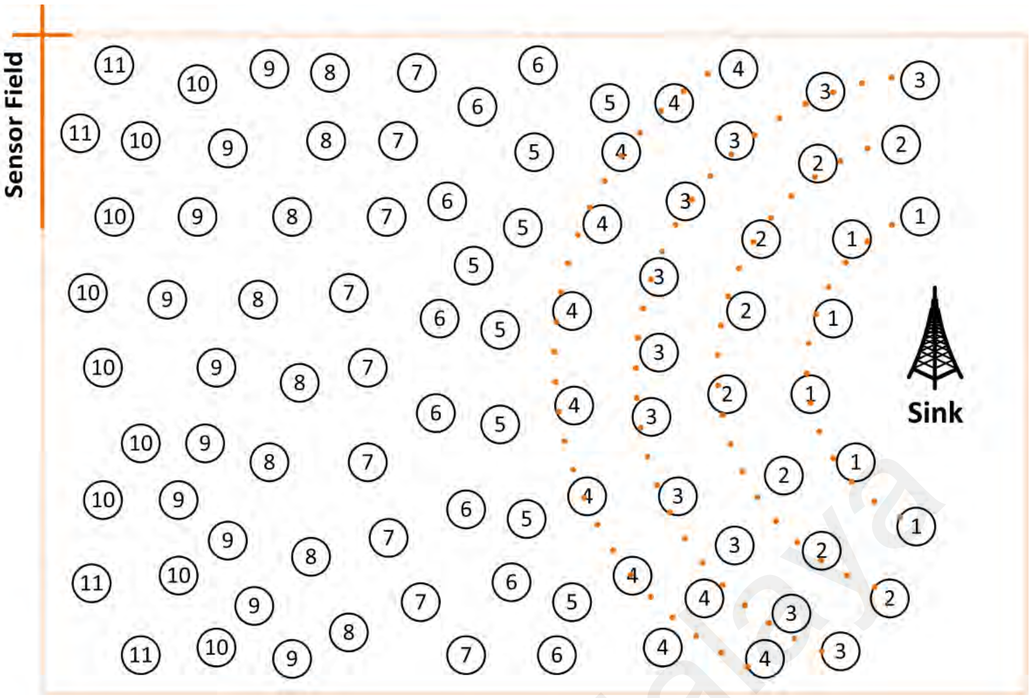


Figure 4.2: Hop-Tree Building Process

Algorithm 4.1: Discovery of node broadcast region and hop tree building

---

Step 1: *The sink node broadcasts the initialization message HCM;*
Step 2: *"N" is the set of network nodes that receive the HCM such that x ∈ N;*
Step 3: ***Foreach*** *x ∈ N*
  Step 3.1: ***If*** *HtT(HCM) < HtT(x) ← true* ***then***
    Step 3.1.1: *Node x maintain its Neighbors_table with insert (Node-ID(HCM), HtT(HCM), HtS(HCM), E_{res}(HCM), and B_{ava}(HCM));*
    Step 3.1.2: ***If*** *First_Sending(x) ← true* ***then***
      Step 3.1.2.1: *HtT(x) ← HtT(HCM) + 1;*
      Step 3.1.2.2: *HtS(x) ← HtS(HCM) + 1;*
      Step 3.1.2.3: *Update the HCM;*
        Step 3.1.2.3.1: *Node-ID(HCM) ← Node-ID(x);*
        Step 3.1.2.3.2: *HtT(HCM) ← HtT(x);*
        Step 3.1.2.3.3: *HtS(HCM) ← HtS(x);*
        Step 3.1.2.3.4: *E_{res}(HCM) ← E_{res}(x);*
        Step 3.1.2.3.5: *B_{ava}(HCM) ← B_{ava}(x);*
        Step 3.1.2.3.6: *Broadcast(HCM);*
    Step 3.1.3: ***End If****;*
  Step 3.2: ***End If****;*
  Step 3.3: ***Else*** *Discard HCM message;*
  Step 3.4: *Node x computes the NW according to MCNW metrics for each node in Neighbors_table(x);*
  Step 3.5: *NextHop(x) ← Node-ID(Bestneighbor);*
  // Bestneighbor = The neighbor with the lowest NW.
Step 4: ***End For****;*
Step 5: ***End****.*

University of Malaya

4.3.3.2 Event-driven cluster formation and CH election (Phase II)

In this phase, a dynamic cluster architecture is formed. Once the event is located, all the nodes that detected the event will be an input in the cluster formation algorithm. At the end, a specific sensor node will be selected as the CH by the algorithm, whereas the nodes that lie within the range of the CH will be designated as member nodes. In this process, any node that has sensed the event takes the role of CH. Then, all the event nodes propagate their information by a cluster configuration message (CCM) (Step 1 of Algorithm 4.2). If a node receives a CCM that provides more accurate information regarding the distance in hops to the sink node (for the first event) or already established path (for the successive events), the node will set its role to CM and retransmits the received CCM. Otherwise, the node will discard the received CCM, and after a specific time interval, the node broadcasts a Declaration Message as CH (CHDM) with its *Node_ID* to its cluster members. The CH is selected by considering the hop to tree metric (Step 4 of Algorithm 4.2). In case two or more nodes have the same *HtT*, the node with the higher E_{res} is considered eligible (Step 4.3 of Algorithm 4.2). Finally, the member nodes remember their CH, and all the event detection reports are directly sent to the CH. Figure 4.3 shows an example of event cluster formation.

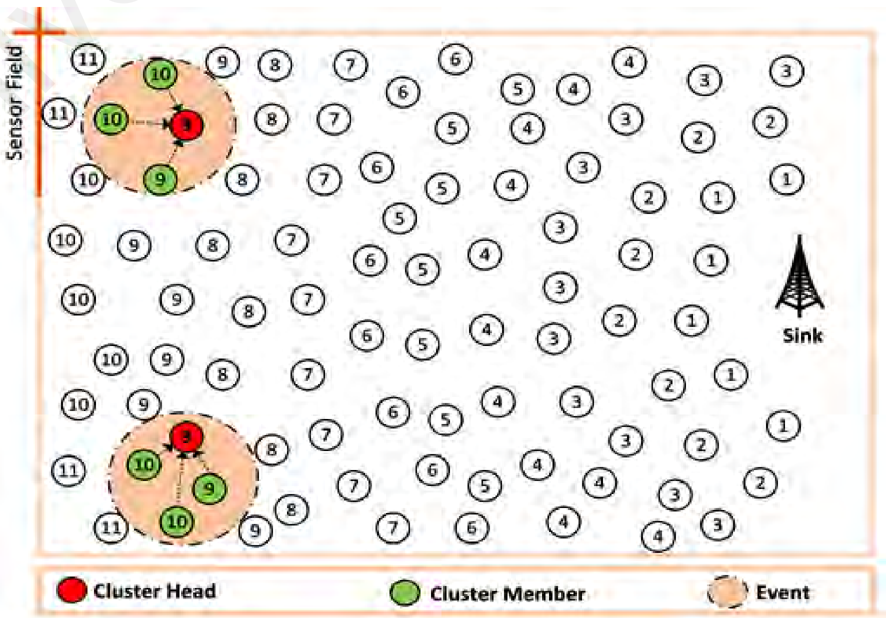


Figure 4.3: Example of Event Cluster Formation

University of Malaya

Algorithm 4.2: Event-driven cluster formation and CH election

Input: *S //S is the set of nodes that detected the event.*
Output: *u // u is a node in set S voted as the cluster head.*
Step 1: ***Foreach*** *x* ∈ *S do*
  Step 1.1: *Role(x)* ← *CH;*
  Step 1.2: *Node x create and updates the CCM;*
  Step 1.3: *Node x broadcasts CCM;*
Step 2: ***End For****;*
Step 3: "N" *is the set of nodes in a network that received the CCM and detected the event.*
Step 4: ***Foreach*** *u* ∈ N
  Step 4.1: ***If*** *HtT(u) > HtT (CCM)* ***then***
    Step 4.1.1: *Role (u)* ← *CM*;
    Step 4.1.2: *Node u retransmits the CCM received from x;*
  Step 4.2: ***End If****;*
  Step 4.3: ***ElseIf*** *HtT (u) == HtT(CCM)* **&&** $NW\_E_{res}(u) > NW\_E_{res}(CCM)$ ***then***
    Step 4.3.1: *Role (u)* ← *CM;*
    Step 4.3.2: *Node u retransmits the CCM received from x;*
  Step 4.4: ***End If****;*
  Step 4.5: ***Else*** *Node u discards the CCM received from x;*
Step 5: ***End For****;*
Step 6: ***End****.*

4.3.3.3 Route establishment and data transmission based on MCNW (Phase III)

In this phase, a group of CH, which was elected in the previous phase, is considered the input. The new route that would transfer the event data will be the output of this process. Thus, the routing tree formation is based on the saved MCNW weights in the neighbors table, which was created in the first phase. Each node is well aware of all its neighbors and can use the information in the neighbors table to send data packets to the sink node.

First, the CH is responsible for the routing tree formation and routing packets of the new event to the sink (Step 1 of Algorithm 4.3). The CH will check if its *HtT* is zero, that is, it is a part of the backbone of the hop tree; thus, creating a fresh route as the new backbone of the hop tree is not required (Step 2 of Algorithm 4.3).

WDARS keeps track of the remaining energy level and accessible buffer memory of the nodes in the backbone to acquire an even energy distribution and to avoid congestion delay, which is caused by data collisions. If both weights of both parameters exceed the set weight limit, then a new routing path formation is initiated, as shown in (Step 3 of Algorithm 4.3). During the reformation process, the neighboring node that has the lowest *HtT* and *HtS* as well as highest E_{res} and B_{ava} among the candidate nodes is selected as the alternative next hop. Furthermore, in the reformation of the routing path, the threshold weight ($Th_{reshold\_w}$) factors for NW_E_{res} and NW_B_{ava} in every node slightly increase if no suitable node can be found.

The CH then creates a route establishment message (REM) and sends it to its next hop (Step 4 of Algorithm 4.3). If the REM is received by the next-hop node, then the next-hop node will retransmit the message and initiate the process of updating the hop tree (Steps 5 and 6 of Algorithm 4.3). These steps are repeated until the sink node is reached or the node that participated in a previously constructed route is discovered. The routes are created by selecting the best neighbor in every hop.

The hop tree should be updated so that all source nodes can be connected via the lightweight paths, the data aggregation can be optimized, and the energy load can be balanced in the succeeding events. In the proposed scheme, the *HtT*, *HtS*, E_{res}, and B_{ava} values are updated at each node to fulfill these objectives. The correlation among data is important in data aggregation; a higher degree of correlation generates better aggregation results. The spatial distance between nodes determines the spatial correlation between the data sensed by different nodes. Thus, if two events are close, then the sensed data are highly correlated, whereas if the events are far apart, then the sensed data exhibit a low degree of correlation. Usually, data aggregation is not efficient for events that are far apart. In WDARS, the purpose of using MCNW weights, particularly those of the *HtT* and *HtS* metrics in route establishment is to achieve the

University of Malaya

ideal point for routes overlapping of different events. When events are close, the data paths overlap rapidly at the nearest ideal point for aggregating the data of the events. Conversely, when the events are far apart, data routes overlapping occurs at the point where the lightweight path to reach the destination will be selected, and data transmission over long tracks is avoided.

Algorithm 4.3: Route establishment and data transmission based on MCNW metric

Step 1: *The cluster head v of the new event begins to form the routing tree;*
Step 2: ***If*** *$HtT(v) == 0$* ***then***
  Step 2.1: *The cluster head v start send data to its next hop;*
  // Route formation is no longer needed; node v is already a part of the backbone.
Step 3: ***If*** *NW_E_{res} (NextHop (v))* **&&** *NW_B_{ava} (NextHop (v)) > $Th_{reshold\_w}$* ***then***
  Step 3.1: *Cluster head v finds a new next hop with the lowest NW that satisfies the predefined weight limits by using the neighbors_table;*
Step 4: *The cluster head v sends an REM to its next hop;*
Step 5: ***Repeat***
  Step 5.1: *Node x is next hop of v that received REM;*
  Step 5.2: *$HtT(x) \leftarrow 0$;*
  // Node x becomes a part of the new routing structure.
  Step 5.3: *$Role(x) \leftarrow Relay$;*
  Step 5.4: ***If*** *$NW\_E_{res}(NextHop(x))$* **&&** *$NW\_B_{ava}(NextHop(x)) > Th_{reshold\_w}$* ***then***
    Step 5.4.1: *Node x finds a new next hop with the lowest NW that satisfies the predefined limits by exploiting the neighbors table;*
  Step 5.5: *Node x sends an REM to its next hop;*
  Step 5.6: *Node x broadcasts the HCM with HtT=0;*
Step 6: ***Until*** *The sink node or a node belonging to the routing structure of the previous event is found;*
Step 7: ***End***.

Figure 4.4(a) shows the new path and updated hop tree formation based on *HtT* only in DRINA proposed by (L. Villas et al., 2013). The DRINA protocol forces the paths from the events that occur far apart to overlap and aggregate the data from two consistent events. This technique results in poor data formation; overlapping path overload, which causes an imbalance of energy consumption in the network; and dead node increase. Moreover, data transmission over longer path increases the total energy consumption. In WDARS, routes overlapping depend on both *HtT* and *HtS* in addition

University of Malaya

to the node status reflected by E_{res} and B_{ava}, as shown in Figure 4.4(b). As shown, a route overlapping is linearly related to the distance between the two events and guarantees data transmission over the lightweight route.

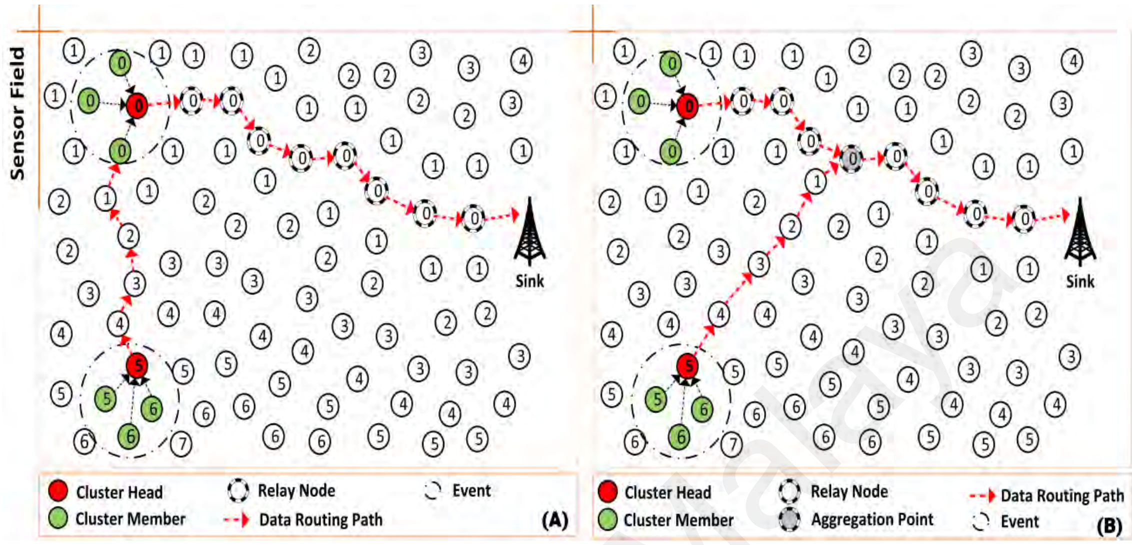


Figure 4.4: A Comparison of the New Path Formation Processes of (A) DRINA and (B) WDARS

4.4 Performance Evaluation

The proposed WDARS was evaluated with various network test cases in MATLAB environment. The results of the simulation experiments were also analyzed using several performance metrics to assess the capability and the efficiency of the proposed scheme. WDARS was compared with the DRINA and InFRA protocols, which were also implemented in MATLAB to ensure that all schemes were run on the same platform and under the same conditions and simulation parameters. Furthermore, WDARS was tested and validated to prove its effectiveness on promoting energy efficiency and load balancing.

4.4.1 Simulation Setup and Scenario Assumptions

The proposed model was simulated using MATLAB with IEEE 802.15.4 as the MAC layer protocol. The simulation environment was set up in a manner that would satisfy the multi-hop routing requirements and clearly present the effects of the

University of Malaya

proposed enhancements to DRINA. This simulation was based on a WSN in an event-based environment with an area of 500 m × 500 m. The nodes were randomly placed, If two sensors are within the range of each other, they are considered neighbors. A circular event space with random position, time, and event duration was considered. The number of nodes varied from 100 to 300, with a step size of 50 nodes. The sensor field, communication range, and event radius were kept constant. There is one base station in the setup that gathers data from the sensor nodes and is positioned outside the monitoring area. The sensors have energy limitation as they are operated by battery, whereas there is no resource limitation at the sink. If any sensor's energy depletes, it is considered as a dead sensor.

Moreover, we have used the reactive sensors clustering hierarchy within the vicinity of the event-sensing nodes. After the cluster head has received first data packet, it waits and aggregates data for a fixed time known as data aggregation time before transmitting it to the sink. The time is set to 150 ms for every protocol. This delay has been minimized by configuring a value known as degree of aggregation (He et al., 2006) to five packets. This implies that if five packets are received by the aggregation node before expiration of data aggregation time, the packet is transmitted by it to the sink node. As the number of nodes increased, the simulation time also gradually increased. The other parameters applied to the simulation are listed in Table 4.2.

University of Malaya

Table 4.2: Settings of Simulation Parameters for WDARS

Parameter	Type/Value
Channel	Wireless channel
Antenna	Omnidirectional
Underlying MAC protocol	IEEE 802.15.4
Sink node	One with fixed coordinates
Shape of monitoring region	Square
Size of monitoring region	500 m × 500 m
Number of sensor nodes	100, 150, 200, 250, and 300
Topology	Tree-based dynamic cluster
Initial node energy	2 J
Number of events	3
Event radius	80 m
Communication radius	80 m
Simulation time	3000 s
Data packet size	1024 bytes
Control packet size	56 bytes
Network threshold	0.1 of nodes be alive

4.4.2 Experimental Results and Discussion

Several comprehensive simulations were performed by varying certain parameters to study the behavior of the proposed WDARS. The results were analyzed and compared with InFRA and DRINA. The following performance metrics were used average energy consumption, network lifetime, efficiency (packet per processed data), packet delivery rate, average packet loss, and control packet overhead.

Figure 4.5 presents a comparison of the energy consumption levels of the nodes for various network node quantities among WDARS, DRINA, and InFRA protocols. The InFRA protocol recorded the highest energy consumption, followed by DRINA and WDARS protocols. The highest energy consumption in InFRA resulted from the transmission of more control packets throughout the event detection to inform the nodes to update the routes from the existing coordinators to the sink node. Consequently, the communication cost of InFRA was the highest. In DRINA, longer paths were formed because of the unwanted overlapping of distant and uncorrelated events. Therefore,

University of Malaya

significant energy was wasted by forwarding data over longer paths. On the contrary, WDARS consumed the least energy by considering the remaining energy (E_{res}) of the nodes to stabilize the energy consumption among the nodes. Moreover, the *HtT* and *HtS* weights, which were considered in the MCNW metric, played a significant role in saving energy by selecting the appropriate overlapping points. Consequently, a balanced trade-off between data aggregation and link cost was achieved. The total energy consumption averages of all the tested protocols are depicted in Figure 4.6. The energy consumption average from the node initial energy of the proposed WDARS (11.65%) was lower than those of DRINA (21.75%) and InFRA (35.71%).

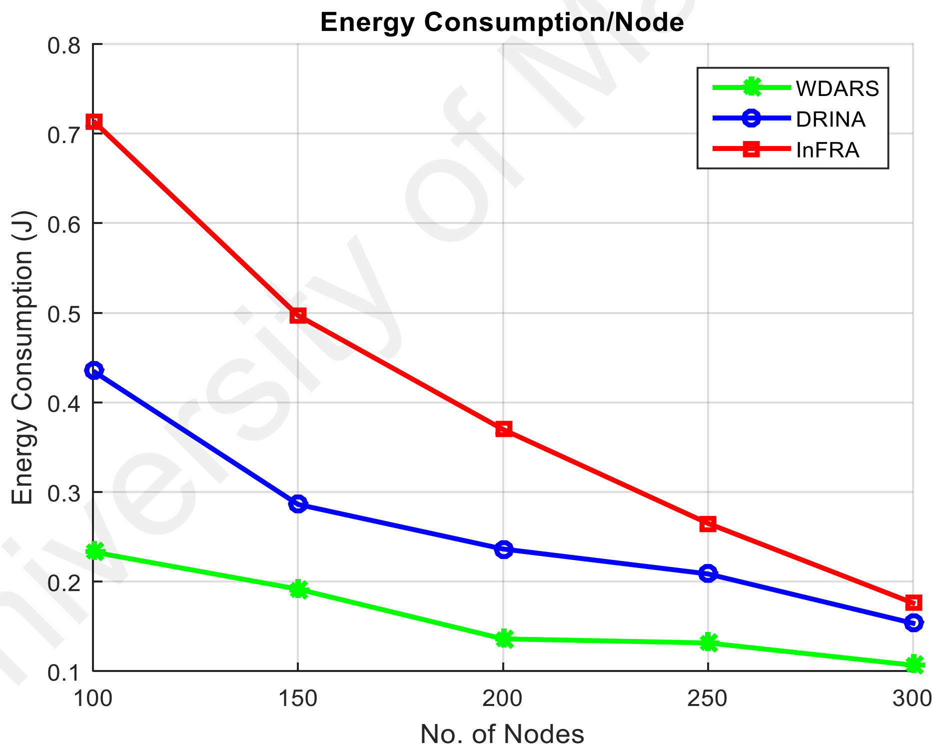


Figure 4.5: Comparison of the Average Energy Consumption Levels for Different Node Quantities

University of Malaya

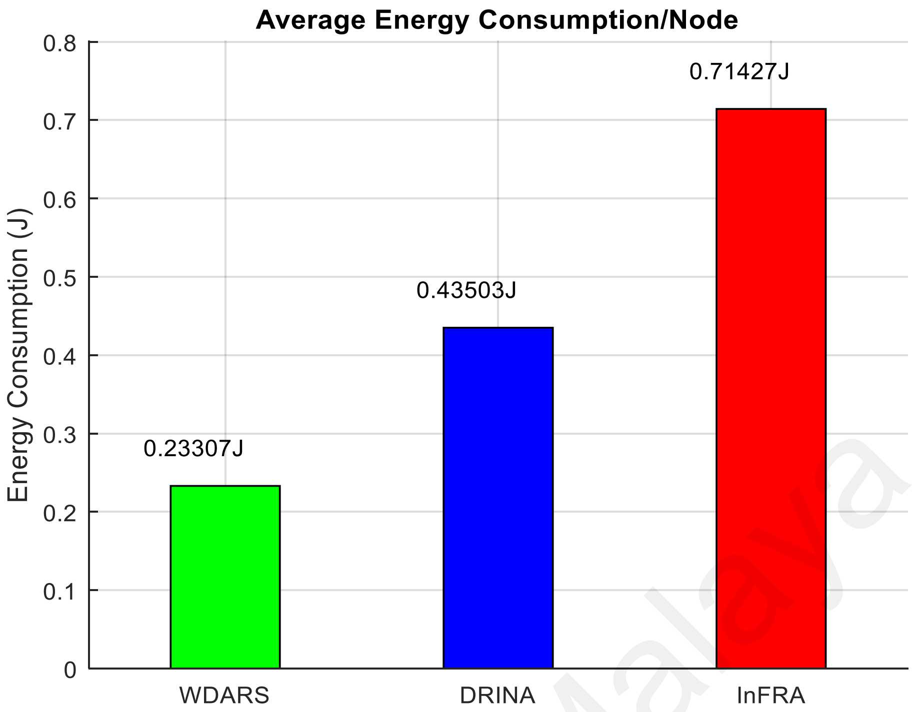


Figure 4.6: Comparison of the Total Average Energy Consumption Levels among WDARS, DRINA, and InFRA

The network lifetime, which is highly dependent on the routing protocol, is influenced by two factors, namely, energy consumed over time and initial node energy. If E_i denotes the initial energy capacity of a node and e_i denotes the energy consumed by each node, then the working period of the node can be expressed as (t_i=E_i/e_i). The information related to the network lifetime is reflected by t_i. The simulation results showed that WDARS maintained the longest network lifetime regardless of the number of nodes in the network. As shown in Figure 4.7, the network lifetime of all the protocols increased as the number of nodes increased in a fixed sensor field size. The lifetime of the network using the WDARS protocol exceeded those of the networks using DRINA and InFRA for all node quantities. At the minimum node quantity, the network lifetime of WDARS increased by up to 2 h compared with that for DRINA and 2.2 h compared with that of InFRA. At a node quantity of 300, the network lifetime of WDARS was 13.3 h and 13.9 h higher than those of DRINA and InFRA, respectively. The superiority of WDARS in terms of the network lifetime could be attributed to energy-efficient and load-balancing mechanisms adopted by WDARS. The effect of the scheme seemed to improve with an increase in node quantities.

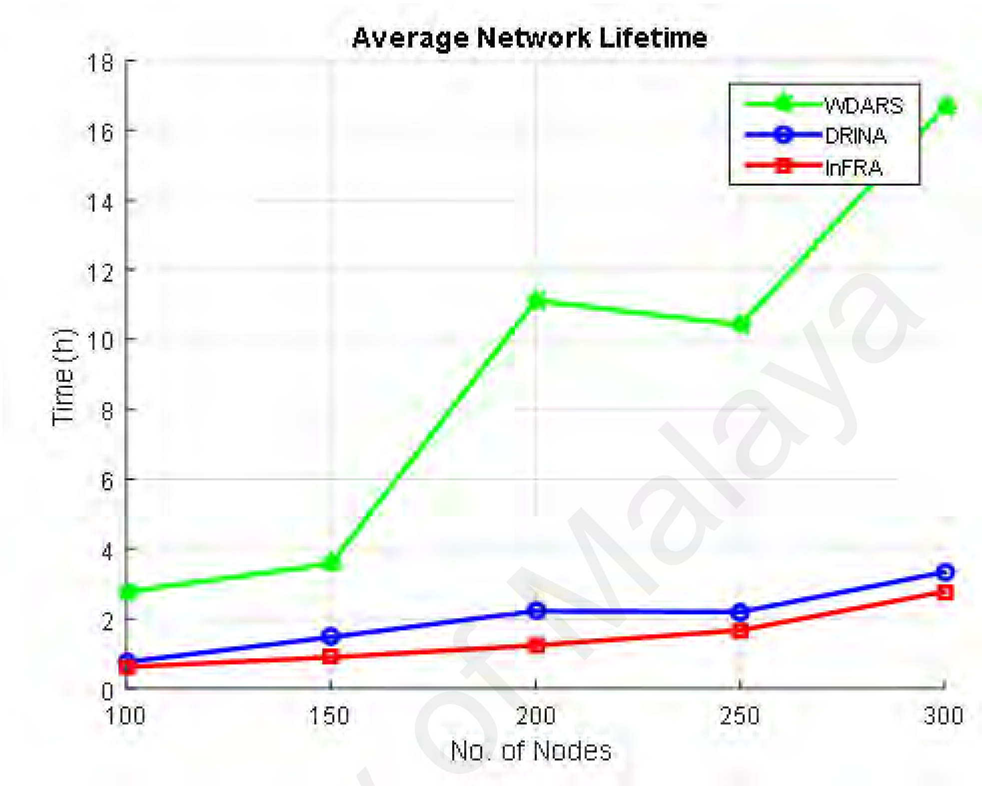


Figure 4.7: Comparison of the Network Lifetimes for Different Node Quantities

Figure 4.8 shows the efficiency of WDARS, which effectively decreased the packet per processed data and outperformed both InFRA and DRINA in all the experiments regardless of the node quantity. Compared with DRINA (InFRA), WDARS achieved 9.09% (37.50%) efficiency improvement at a node quantity of 100 and 10.31% (56.52%) at a node quantity of 300. The outstanding performance of WDARS is due to the fact that its design requires a relatively low number of control packets to establish and maintain a routing tree. Moreover, the data aggregation quality achieved by the routing tree built by WDARS was higher than those achieved by the routing trees constructed by InFRA and DRINA.

University of Malaya

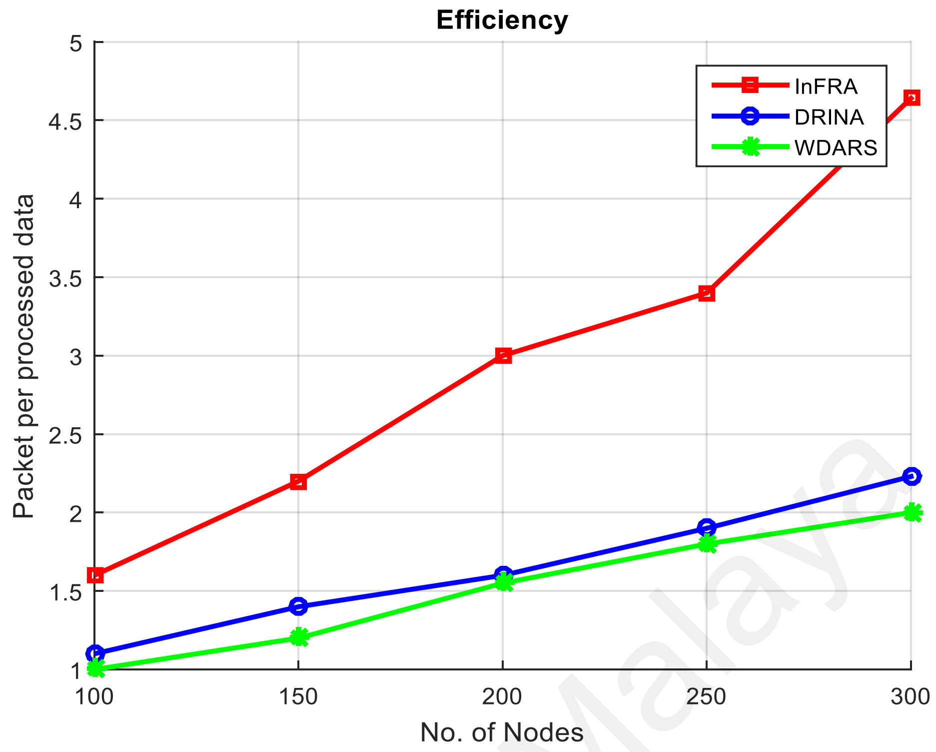


Figure 4.8: Comparison of the Efficiency Levels for Different Node Quantities

Furthermore, packet delivery rate was used to quantify the quality of the routing tree built by the tested protocols. A low packet delivery rate implied a high aggregation rate of the built tree. As shown in Figure 4.9, the packet delivery rates of all protocols increased as the node quantity increased in a fixed sensor field size (high network density). At the lowest node quantity, the packet delivery rate of WDARS was lower than those of DRINA and InFRA by11.23% and 41.46%, respectively. At the highest node quantity, the packet delivery rate of WDARS was 20.69% and 48.31% lower than those of DRINA and InFRA, respectively. This result is mainly due to the selection of an ideal point for routes overlapping in WDARS. As a result, the relay nodes in the routing tree of the propose protocol are less than those in the routing trees of DRINA and InFRA.

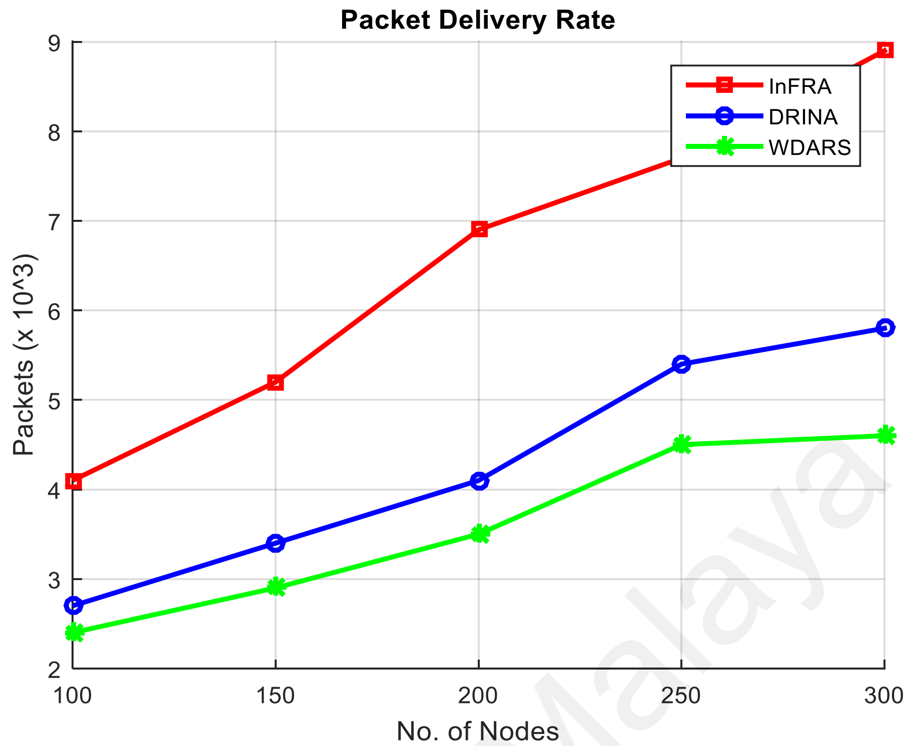


Figure 4.9: Comparison of the Packet Delivery Rates for Different Node Quantities

Another significant metric is packet loss. The numbers of packets lost for different numbers of nodes were determined. Collisions and full queues cause packets to drop at the destination node. The resending of these packets increases energy consumption, reduces throughput (channel is blocked for a short time with every dropped packet), and increases delay (a packet arrives at a later time). As shown in Figure 4.10, the packet loss increased with increasing number of nodes. Therefore, packet loss is directly related to the number of sensor nodes. The packets loss of WDARS and DRINA were generally comparable. The minimum packet loss rates (0.02%–0.05%) were generated by WDARS for different numbers of nodes. By contrast, the average packet loss rates of the two other protocols ranged from 0.03% to 1.0%. The performance of InFRA deteriorated as the number of sensor nodes increased. The high packet loss of this protocol is not only due to its centralized operation but also to the broadcasting of the information of an event all over the network to notify other nodes and the updating of

University of Malaya

the paths from the available CHs to the sink node every time a new event is sensed. These processes are costly and limit network scalability.

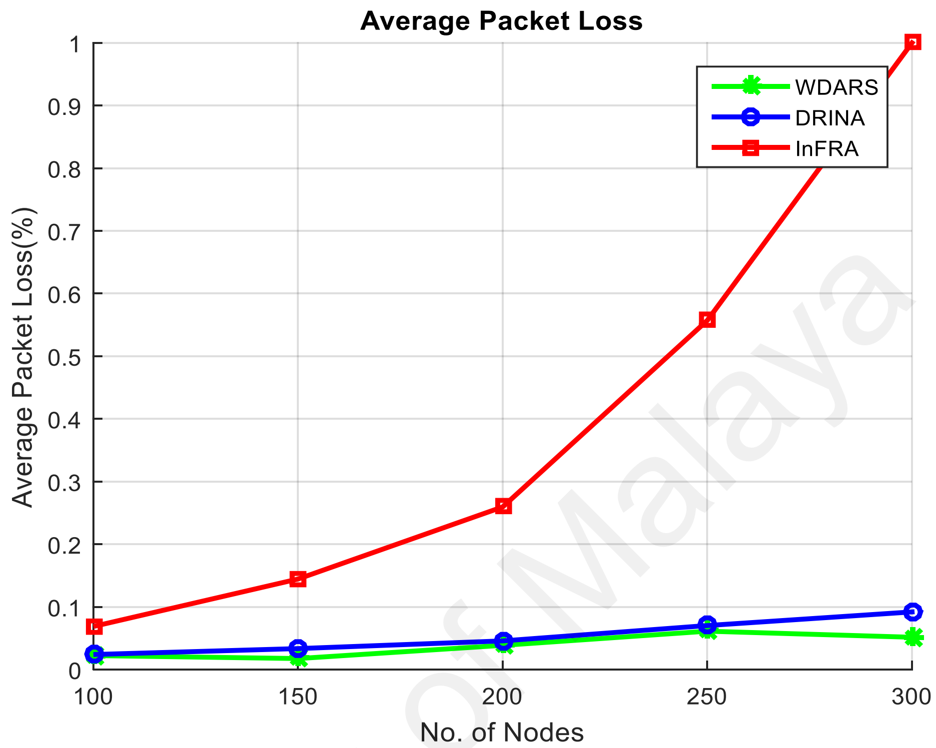


Figure 4.10: Comparison of the Average Packet Loss Rates for Different Node Quantities

Moreover, for a node quantity of 100, the net improvement of WDARS in terms of control packet overhead was 37.90% compared with InFRA and approximately 8.40% compared with DRINA, as shown in Figure 4.11. For a node quantity of 300, the net improvement of WDARS was 57.92% and 9.48% compared with InFRA and DRINA, respectively. WDARS exhibited significantly lower overhead than InFRA. This result is ascribed to the local computation of routes at the CH in WDARS. Moreover, WDARS builds its routing tree with fewer relay nodes than InFRA and DRINA does. By contrast, InFRA scheme requires the configuration of several transmission routes.

University of Malaya

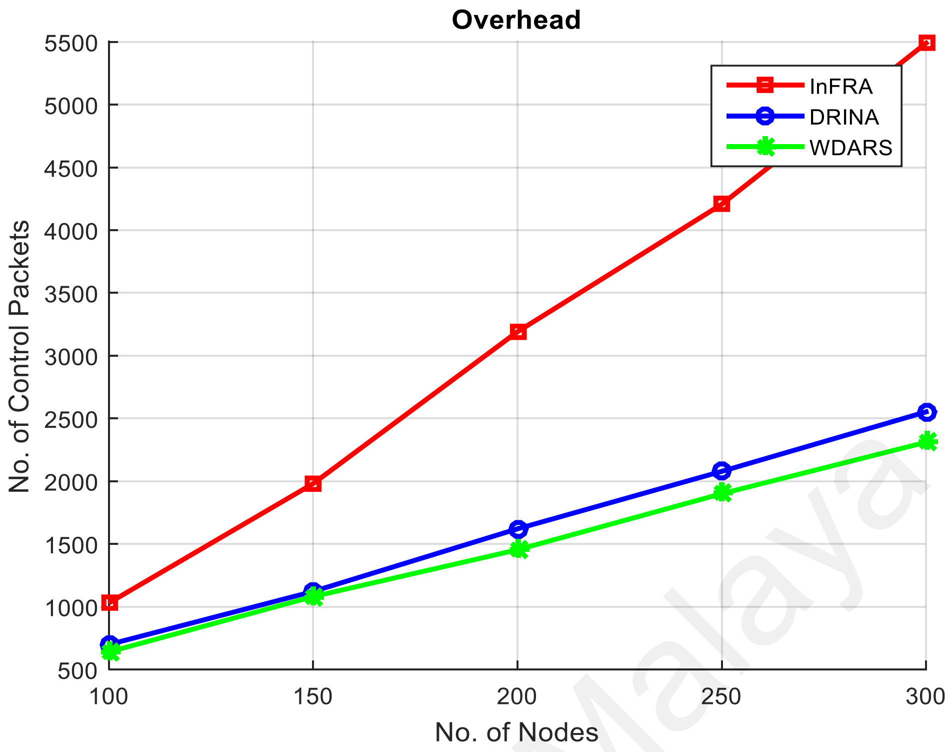


Figure 4.11: Overhead Comparison for Different Node Quantities

4.5 Chapter Summary

This chapter presented the structural design, functions, and the simulation results of the proposed WDARS routing scheme in WSNs. WDARS is an in-network data aggregation multi-hop routing scheme that aims to enhance the energy efficiency and load balancing in WSNs. NWAUF was used to estimate the MCNW metric for every node on the basis of the remaining energy, available buffer memory size, hop to tree, and hop to sink metrics. The MCNW metric was utilized during the route computation in the link quality measurement, and routes overlapping were optimized by determining the ideal aggregation point while simultaneously ensuring the data transmission across a lightweight route in WSNs. The simulation results obtained from the performance evaluation of the proposed WDARS were compared with those of DRINA and InFRA schemes. The simulation results showed that the proposed WDARS outperformed DRINA and InFRA in terms of average energy consumption, network lifetime, packets

University of Malaya

per processed data, packet delivery rate, average packet loss, and control overhead, particularly in dense networks.

University of Malaya

CHAPTER 5: WEIGHT AWARE SPATIAL–TEMPORAL CORRELATION ROUTING SCHEME (WST-RS)

5.1 Introduction

WSNs are event-based systems in nature that constantly monitor physical environment. The sensor nodes in WSNs perform collective sensing, which is realized by densely deploying these nodes. Collective sensing mitigates energy and computational issues. Furthermore, the cooperative nature of WSNs renders these sensor networks advantageous over traditional sensing in terms of coverage and communication. However, the deployment of high-density sensor nodes usually creates a possibility to produce spatially and temporally correlated data when sensing an event. Spatial and temporal correlations are based on the premise that sensor nodes that are geographically close to each other have similar readings as that of their previous ones. Such correlations create considerable data redundancy and consequently increase communication cost in WSNs. Therefore, in this study, the spatial and temporal correlations among the sensor observations are exploited to develop an efficient and feasible communication protocol for event-driven WSNs. The protocol aims to optimize the communication and energy efficiency of WSNs.

In this chapter, we present the weight-aware spatial-temporal correlation routing scheme (hereinafter referred to as WST-RS) for efficient data collection and aggregation in WSNs. WST-RS is developed by exploiting both correlations (spatial and temporal) of a detected event and extending the EAST protocol to decrease redundant communication and data exchange. In the proposed WST-RS scheme, an event area is divided into spatially correlated regions, and the CH and representative nodes implement a temporal suppression strategy. A representative value is generated by the CH for the entire cluster on the basis of the data received from the representative nodes, thereby forming a subset of all the nodes that are sensing the same event.

University of Malaya

Moreover, WST-RS considers different metrics of the developed MCNW in selecting the CHs, representative nodes, and route computation process. In WST-RS, the MCNW metric considers the position of a node; energy status of a node, which denotes the stability of the node; and available buffer memory size. The estimated value of the MCNW is utilized by the nodes to assess the neighboring nodes and link quality between the source and destination pairs and to construct the multi-hop route to the destination effectively. Moreover, an important characteristic of WST-RS is its capability to alter itself adaptively depending on the node density in the event area and node weight. These variables are usually ignored by existing protocols. Consequently, the energy consumption is reduced in WST-RS by eliminating redundant notifications and optimally balancing the load among the sensor nodes.

In Chapter 3, the theoretical concepts of WST-RS were presented. In this chapter, the design architecture of WST-RS and the simulation results of its performance evaluation are discussed in detail.

5.2 Sensor Network Model and Basic Assumptions

The key network features and assumptions of the proposed scheme highlighted in this section are summarized as follows:

- Random distribution of *n* sensor nodes is performed in a two-dimensional surveillance area.
- The sensor nodes are homogeneous in terms of their communication and processing capabilities as well as energy budget.
- The single sink node of the sensor is linked to the power supply.
- The sensor nodes are static and energy constrained. They continue their operation until their respective node energies are consumed.

- The location-aware feature of the sensor nodes requires them to obtain location information via GPS or localization algorithms.
- A unique ID is assigned to every network node, and all the nodes participate in the communication procedure by forwarding data.
- At any time, the sensor node is capable of calculating its remaining energy and accessible buffer size while waiting to be serviced.
- The radio link is symmetric; thus, the energy consumed for transmitting data from node *A* to node *B* is the same as that for transmitting data from node *B* to node *A*.
- All the nodes adhere to the Boolean disk coverage model. This model implies that the sensing radius is constant and the sensing area is represented by a disk whose center is at the spatial position of the sensor node. The node senses all the events occurring within the disk, whereas those that occur outside the disk are not sensed. This sensing model is commonly called the omnidirectional sensing model (Ghosh & Das, 2008).
- Communication between sensor nodes is only possible if the communication range (R_c) is larger than the distance between them.
- The transmission power of every sensor node is controlled on the basis of the distance of the desired sensor node.

5.3 Description of WST-RS Scheme

In this section, we define the developed MCNW metric for the proposed scheme, the spatial and temporal correlation models, and the structure of WST-RS for efficient data collection and aggregation in WSNs. The key attribute of WST-RS is its consideration of both spatial and temporal correlations among sensor nodes in WSNs. The proposed WST-RS rely on some of the functions of EAST protocol, to handle the data redundancy resulting from spatial and temporal correlations in WSNs, which was discussed in Chapter 2 of this thesis. Handling data redundancy is essential to overcome the challenges incurred by sensing and forwarding redundant data in highly dense WSNs. WST-RS possesses certain desirable functions of the EAST protocol, namely, neighbor position discovery, dynamic cluster formation, and routes overlapping. However, other functions of EAST scheme are extended in WST-RS. These functions include the calculation of the size of spatially correlated regions. The additional information stored in the sensor nodes, which can be used in temporal suppression. The other extended features are representative node and CH selection, MHR node selection, and route computation based on MCNW metric estimation. In the following subsections, the developed MCNW metric and the spatial-temporal correlation model are described initially before describing in detail the proposed WST-RS protocol.

5.3.1 MCNW Metric for WST-RS

The MCNW of a node in WST-RS is mainly based on three node-related metrics, namely, residual energy (E_{res}), available buffer memory size (B_{ava}), and node position with respect to the destination (P_r). In the estimation of MCNW, each node i uses the values of the three metrics to calculate the corresponding node weights, namely, the residual energy-based node weight (NW_E_{res}), available buffer-based node weight (NW_B_{ava}), and position-based node weight (NW_P_r).

The residual energy-based node weight (NW_E_{res}) and available buffer-based node weight (NW_B_{ava}) in WST-RS are the same as those in WDARS. Details of these node weights are discussed in Sections 4.3.1(a) and (b); thus, we will discuss in this section only the position-based node weight (NW_P_r).

The position metric favors the nodes that have made the greatest progress P_r toward the destination. Therefore, the next-hop candidate is the node closest to the destination because the proposed scheme requires the storage of the information of only the nodes within its communication range R_c. The value of the node position is used by the neighbors of node i to determine the weight of node i with respect to its distance (physical distance) to the destination. This process is illustrated in Figure 5.1, in which the position metric selects node i over node j.

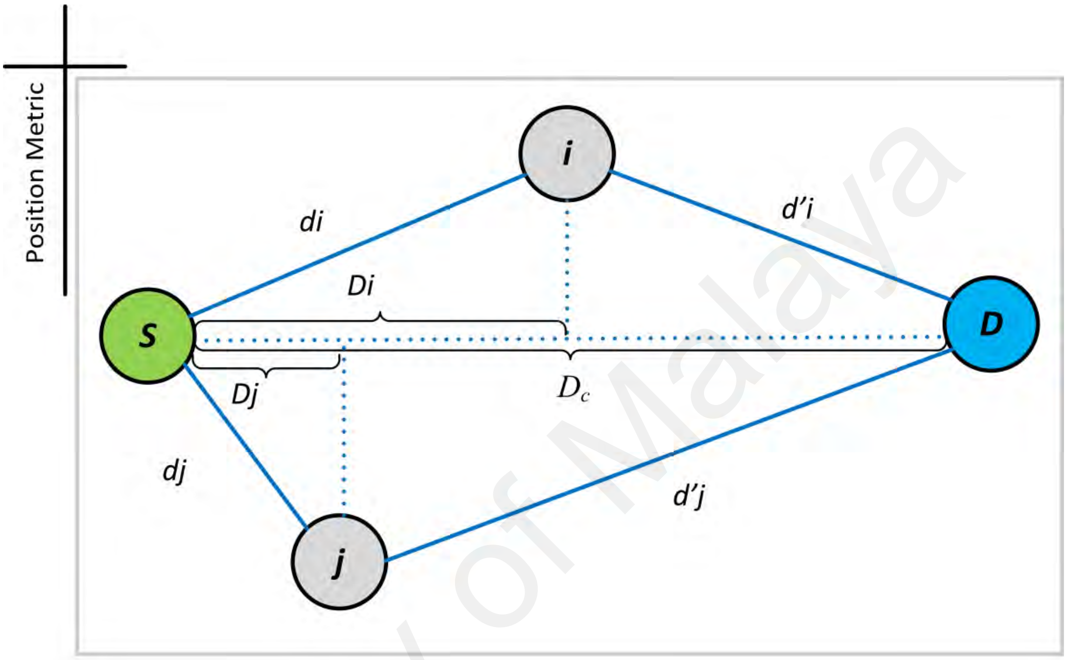


Figure 5.1: Graphical Representation of the Position Metric

However, as regards the distance of each neighbor to the destination, the distance metric and its corresponding node weight are quantified using Equations 5.1 and 5.2, respectively.

$$P_r(i) = \left(\frac{D_c + d(i)^2 - d'(i)^2}{2 \times D_c} \right) \qquad (5.1)$$

$$NW\_P_r(i) = \left(1 - \frac{P_r}{D_c} \right) \qquad (5.2)$$

Where D_c is the shortest distance between the source and destination; d and d' are the distances that separate the intermediate node from the source and destination, respectively, as depicted in Figure 5.1. These distances are calculated on the basis of the

University of Malaya

positions of the source and destination. Therefore, the nodes near to the destination are prioritized, which is ensured by the position metric. As demonstrated in Equation 5.2, the shortest distance to the destination minimizes the value of $NW\_P_r(i)$.

5.3.2 MCNW Weight Calculation in WST-RS

A composite MCNW weight of node i is constantly and independently measured using the three specific node weights discussed in Section 5.3.1 in accordance with a NWAUF (Malakooti & Thomas, 2006). The MCNW weight is dependent on the values satisfying normalizing criteria and weights of importance that range from 0 to 1. The final MCNW weight of node i is estimated by Equation 5.5.

$$NW(i) = W_1 \times NW\_E_{res}(i) + W_2 \times NW\_B_{ava}(i) + W_3 \times NW\_P_r(i) \qquad (5.3)$$

Where $NW\_E_{res}(i)$, $NW\_B_{ava}(i)$, and $NW\_P_r(i)$ are the node weights of the node-related metrics of MCNW; W_1, W_2, and W_3 are the normalized weight factors of the node weights. The normalized weight factors, whose sum is equal to 1, specify the significance of the components of the MCNW metric. Equal weights can be assigned to the different metrics according to the additive combination rule (Farooq & Tang Jung, 2013). In this study, the normalized weight factors of the E_{res} *and* B_{ava} metrics are set to 0.3, whereas that of the P_r metric is set to 0.4, which is relatively high than the normalized weight factors of the first two metrics.

5.3.3 Spatial-Temporal Correlation Model

This subsection presents the model of the spatial-temporal correlation between the observations of sensor nodes. The proposed spatial-temporal correlation model is established on the premise that data correlation occurs in most WSN applications when the nodes that sensed the data are geographically close (spatially correlated) to each other or when the data are sensed at a certain period of time (temporally correlated). In

University of Malaya

this study, a novel spatial-temporal correlation model is proposed to reduce the effect of the high correlation between the data from densely correlated sensors. The proposed spatial-temporal correlation model is used in the routing protocol to decrease communication and data exchanges.

Definition 1 (spatially correlated region): Spatially correlated region, which is denoted by *SCR*, is defined as a region in which the readings of a sensor nodes are similar for an application. Thus, the region can be represented by a single reading report. The size of the spatial correlated region differs according to the application requirements and event characteristics. In this study, the initial size of the spatially correlated region is defined on the basis of the node density in the event area by the CH. The spatially correlated region is fine-tuned by resizing it depending on the response generated by the sink node. The size of the spatially correlated region is calculated on the basis of redundancy and reliability.

Close and redundant sources in the vicinity are identified by the CH, which then divides the event area into spatially correlated regions by considering the node density in the zone. The spatially correlated region size of an event area is computed by assuming that n nodes observe the event source, R_c is the communication radius of the nodes, and R_e is the maximum radius of the event. The dimension of the spatially correlated region (*SCR_D*) is described by Equation 5.4, and the event area is decomposed into spatially correlated regions of size *SCR_Size*, which is equal to (*SCR_D* × *SCR_D*).

$$SCR\_D = 2R_c/Node\_Dens \qquad (5.4)$$

Where *Node_Dens* is defined as the node density in the event zone:

$$Node\_Dens = \sqrt{\frac{2(R_e)^2}{n}} \qquad (5.5)$$

Correlation regions are shaped as rectangles with spatially correlated nodes lying in the rectangle. These nodes are assumed to detect similar values, as shown in Figure 5.2. The maximum size of the spatially correlated region is the length of the right triangle. Therefore, the maximum size of the spatially correlated region is $SCR\_Size = R_c \cos 45°$ because the sensor node communication range (R_c) is the hypotenuse. This consideration is imperative to guarantee that the communication among all the nodes in the same spatially correlated region is possible. Hence, the communication radius of the spatially correlated region may vary between zero and ($R_c \cos 45°$).

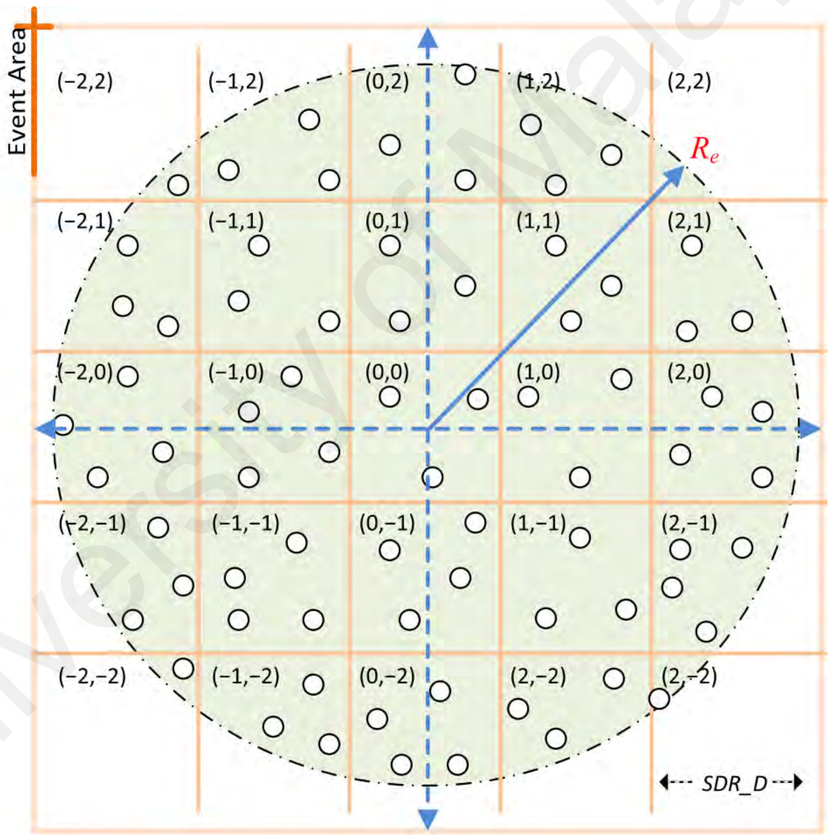


Figure 5.2: Computation of the Size of the Spatially Correlated Regions in a Cluster Zone

Definition 2 (temporally correlated data): The sensor node must keep additional information to perform temporal correlation among successive data readings. The kept information is based on the roles in the routing arrangement of the sensor (i.e., representative node or CH). Only the previous reported reading is stored by the

University of Malaya

representative nodes. This reading is defined as the last reading successfully sent to the CH by the sensor node. The CH node then keeps the last data reported from its cluster representative node.

At the representative node level, when a new reading is available (x_i), it is compared with the last reported reading (x_{i-1}). The new reading is then reported to the CH. If the error between these readings is greater than the temporal tolerance threshold ($Th_{reshold\_tt}$), the percentage of the temporal error between these two successive readings is calculated by Equation 5.6; otherwise, the value of the new reading is suppressed:

$$Error = \frac{|x_{i-1} - x_i|}{x_i} \times 100\% \qquad (5.6)$$

At the CH level, the received data from various representative nodes are combined to calculate the aggregation result, which is then forwarded to the sink node through the upper neighboring node. If complete data are not received by the CH from one of its representative nodes, then the CH utilizes the last reported data from that node. The CH considers its own reading, and it is aggregated with a reading of a group of representative nodes, regardless of its temporal error value. This process is carried out because only the aggregated value of the group is changed by aggregation and the size of the aggregation result is not increased. At this point, the CH takes an old aggregation result (the one calculated from the old data of all cluster representatives and its own old reading) and compares it against the newly generated aggregation result. The tuple from the final aggregation result is eliminated if the aggregation value has not changed because of it. Application of $Th_{reshold\_tt}$ = 0 at the aggregation level at the CH is equivalent to an elimination. An entirely empty aggregation result or that in which some groups are missing compared with the old aggregation result is provided through this process. In both cases, the CH depends on the fact that the missing groups are included in the last reported data stored in the sink node.

The value of $Th_{reshold\_tt}$ ranges from 0 to any positive number that is not infinity. This application-specific value is utilized by the sink node to identify the temporal correlation tolerance for the collected data of the event. The value of $Th_{reshold\_tt}$ signifies the degree of tolerance of the sensor nodes for variations in its readings. For instance, if the application specifies $Th_{reshold\_tt}$ = 5%, then only those readings that differ by more than 5% from the previously reported readings are reported.

5.3.4 Details of WST-RS

The main objective of the proposed WST-RS is to minimize the energy consumption in sensor networks by removing redundant readings from the sensor nodes in the event region. WST-RS utilizes the spatial and temporal correlations to reduce the information transmitted by individual nodes (communication cost controls the usage of power in sensor networks) and to enhance data routing quality when a number of the sensor readings cannot be propagated within a given time constraint in the network.

In such a scenario, a spatial correlation mechanism is applied by the CHs in their respective areas by dividing its clustered region into correlated regions and then selecting a representative node in every correlated area on the basis of the node's proximity to the center of the correlation area and amount of residual energy. Moreover, applying the temporal correlation at the representative node and CH levels allows for the transmission of readings of the sensor nodes on the condition that the difference between the new reading and the last recorded reading is greater than the specified threshold.

Moreover, the proposed WST-RS builds a fully distributed cluster and an efficient routing tree that allow for maximum energy conservation and congestion avoidance. This scheme connects all the sensor nodes that detected the event to the sink node and maximizes data aggregation while optimally balancing the load among nodes. The

University of Malaya

structure and roles that are used in the routing arrangement of WST-RS are illustrated in Figure 5.3.

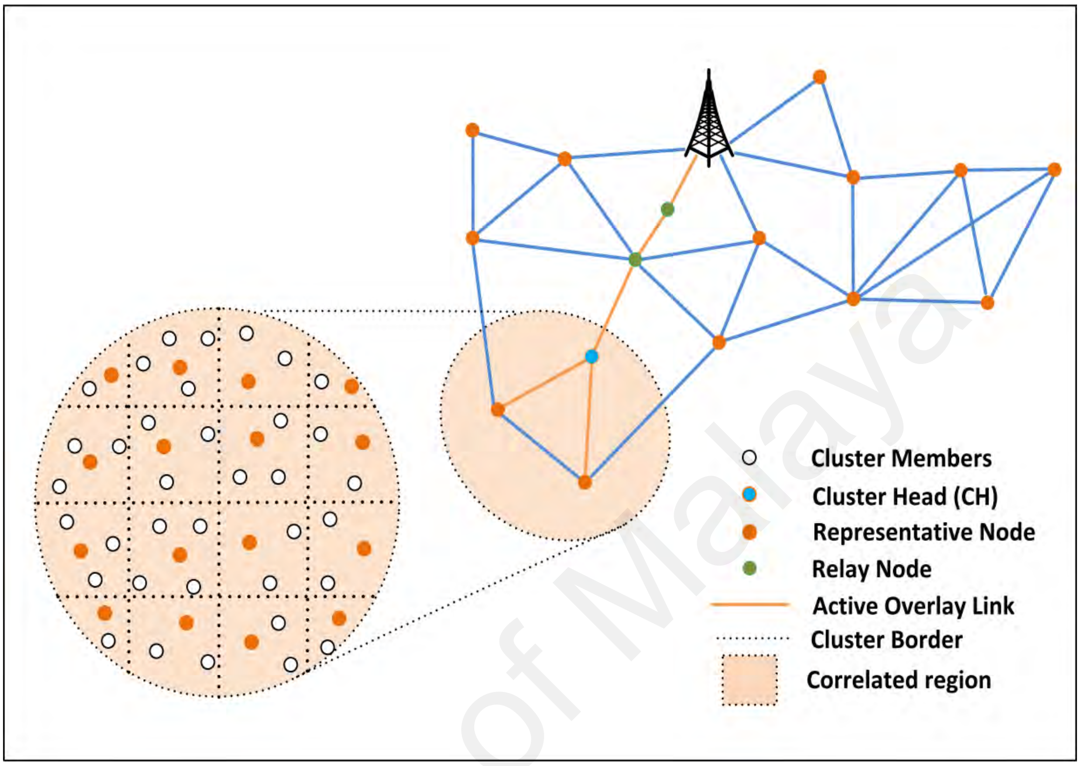


Figure 5.3: Network Model Diagram of WST-RS for Routing to the Sink Node

WST-RS consists of four phases. The first phase involves discovery of the broadcast region of the nodes. The second phase starts as soon as any event is sensed by a node. This phase involves cluster formation and CH election. The third phase includes the formation of the spatially correlated region and the selection of the representative node in each correlated region. The fourth phase involves the application of the presented temporal correlation strategy and data transmission based on the MCNW metric.

5.3.4.1 Node broadcast region discovery (Phase I)

In this phase, the nodes are assumed to be randomly deployed in the sensing region. The procedure for the discovery of the broadcast region of the nodes is established and described in Algorithm 5.1. This algorithm is used to discover neighboring nodes and calculate the required metrics for MCNW, through which the most suitable neighbor is selected as the next hop. An HCM in addition to the general packet fields comprising of five key parameters (Table 5.1) is broadcast from the sink to initiate the algorithm. These parameters include the node ID (*Node_ID*), node coordinates (*Node_Coord*), sink coordinates (*Sink_Coord*), residual energy (E_{res}), and available buffer size (B_{ava}).

Table 5.1: Header of HCM for WST-RS

No.	Parameter	Descriptions
1	*Node_ID*	ID of the node that transmitted/retransmitted the initialization packet
2	*Node_Coord*	Node coordinates (X_n, Y_n) that relays the initialization packet
3	*Sink_Coord*	Sink node coordinates (X_s, Y_s)
4	E_{res}	Residual energy of the node
5	B_{ava}	Available buffer memory size of the node

When the HCM is received by the sensor nodes, the neighbor information is maintained in the neighbors table for the selection of the next hop. Initially, the neighbors table is empty, and it is initialized by storing the received information in the HCM. The entire procedure is as follows. First, the sink node sends the HCM to all the nodes within its communication range. Upon receiving the HCM, each node within the communication range of the sink node maintains the set of neighbors by creating a new entry in its neighbors table for HCM source. If the value of *First_Sending* is true, then the node updates the *Node_ID* as well as the *Node_Coord*. Then, the node calculates its remaining energy after one transmission to update the E_{res} field, calculates its accessible buffer size, and updates the B_{ava} field of the HCM. Finally, the node propagates the HCM to other neighbors. Otherwise, if the condition of *First_Sending* is false, that is, the HCM has already been sent by the node, then the node will only update the neighbors table.

University of Malaya

The node sorts the list of the next-hop candidates in its neighbors table according to node weight score (*NW*), which was based on the proposed MCNW metric. The most suitable candidate, which is the node with the lowest *NW* among all the neighboring nodes of the current forwarder, is selected for the next forwarding. Accordingly, the data of the clusters can be transmitted to the sink node via the most energy-efficient, load-balanced, and reliable path.

The neighbors table is constructed at every node by only one transmission in the proposed protocol. Furthermore, determining the entire path from the source to the destination is not required. Only the neighbors of every node should be determined. Therefore, the proposed scheme requires less memory because the routing table only maintains the data from its neighbors. Moreover, the proposed scheme is easily implemented because of the absence of complex computations. This scheme prolongs the network lifetime and reduces the load on the nodes.

Algorithm 5.1: Node broadcast region discovery

Step 1: *The sink node broadcasts the initialization message HCM;*
Step 2: "*N*" *is the set of nodes in a network that receives HCM such that* $x \in N$;
Step 3: ***Foreach*** $x \in N$
  Step 3.1: *Node x maintain its Neighbors_table with insert (Node_ID(HCM), Node_Coord(HCM),* E_{res}*(HCM), and* B_{ava}*(HCM));*
  Step 3.2: ***If*** *First_Sending(x)* ← *true* ***then***
    Step 3.2.1: *Update the HCM;*
      Step 3.2.1.1: *Node_ID(HCM)* ← *Node_ID(x);*
      Step 3.2.1.2: *Node_Coord (HCM)* ← *Node_Coord(x);*
      Step 3.2.1.3: E_{res}*(HCM)* ← E_{res}*(x);*
      Step 3.2.1.4: B_{ava}*(HCM)* ← B_{ava}*(x);*
    Step 3.2.2: *Broadcast(HCM);*
  Step 3.3: ***End If****;*
  Step 3.4: *Node x computes the NW according to the MCNW metrics for each node in Neighbor_table(x);*
  Step 3.5: *NextHop(x)* ← *Node_ID (Bestneighbor);*
  // Bestneighbor = The neighbor with the lowest NW
Step 4: ***End For****;*
Step 5: ***End****.*

University of Malaya

5.3.4.2 Event-driven cluster formation and CH election (Phase II)

In phase II, a dynamic cluster architecture is formed. Once an event is detected, the CH election algorithm is initialized, and all the nodes that detected the event are now qualified to be CHs and as such run the CH procedure. This process is described in Algorithm 5.2. Any node that sensed the event can assume the role of a CH. Then, all event nodes broadcast their information by a CCM. This CCM contains information about the node ID and position. In case a node receives a CCM that provides more accurate information regarding the physical distance to the destination, the node sets its role to CM and retransmits the received CCM. Otherwise, the node will discard the received CCM, and after a specific time interval, the node broadcasts a CHDM with its *Node_ID* and *Node_Coord* to its cluster members. When two or more nodes have the same distance to the destination, the node with the higher E_{res} is considered eligible to be the CH. The CH is selected on the basis of its position with respect to the first event so that the CH will be the nearest node to the sink. For the successive events, the CHs will be the nodes that are nearest to the previously created routes by the CHs of the previous events. The member nodes can now remember their CH, and every CH is responsible for collecting information from the member nodes and forwarding the aggregated data to the sink.

Algorithm 5.2: Event-driven cluster formation and CH election

---

Input: S *// S is the set of nodes that detected the event.*
Output: *u // u is a node in set S elected as the CH.*
Step 1: ***Foreach*** *x* ∈ *S do*
  Step 1.1: *Role(x)* ← *CH;*
  Step 1.2: *Node x create and update the CCM;*
  Step 1.3: *Node x broadcast CCM;*
Step 2: ***End For****;*
Step 3: *"N" is the set of nodes in a network that received CCM and detected the event;*
Step 4: ***Foreach*** *u* ∈ *N*
  Step 4.1: ***If*** *Distance to destination(u) > Distance to destination(CCM)* ***then***
    Step 4.1.1: *Role (u)* ← *Member Node;*

University of Malaya

Step 4.1.2: *Node u retransmits the CCM received from x;*

Step 4.2: ***End If****;*

Step 4.3: ***ElseIf*** *Distance to destination(u) == Distance to destination(CCM)* ***&&*** *$E_{res}(u) < E_{res}(CCM)$* ***then***

Step 4.3.1: *Role (u) ← Member Node;*

Step 4.3.2: *Node u retransmits the CCM received from x;*

Step 4.4: ***End If****;*

Step 4.5: ***Else*** *Node u discards the CCM received from x;*

Step 5: ***End****.*

5.3.4.3 Spatially correlated regions formation (Phase III)

In the third phase, the spatially correlated regions are formed to define the correlation degree among sensor nodes in the event region. In WST-RS, the spatially correlated regions (*SCRs*) formation starts after the cluster phase is completed. The proposed method for spatial correlation is executed, and the cluster zone is decomposed into a grid of rectangles (*SCR_D* × *SCR_D*) of spatially correlated regions based on the definition 1 discussed in Section 5.3.3. Each region is denoted by an ordered pair *SCR_OP* (X_{SCR}, Y_{SCR}), as depicted in Figure 5.4.

After the spatially correlated region construction has been completed, every node checks the spatially correlated region to which it belongs in the event region. If *SCR_D* = 0, then spatial correlation is not observed between sensor nodes and all cluster nodes are maintained in a sensing status. Otherwise, each cluster node determines the ordered pair (*SCR_OP*) of the spatially correlated region to which it belongs. For this computation, the node should be aware of its position (X_n, Y_n), the center point of the event (X_e, Y_e), and the size of the spatially correlated region (*SCR_Size*).

Before the value of *SCR_OP* is calculated, the node needs to determine the quadrant in which its spatially correlated region lies in the event coordinator system. The value of *SCR_Quad* (X_Q, Y_Q) is calculated by Equations 5.7 and 5.8:

$$\mathrm{X_Q} = \frac{(X_n - X_e)}{0.5 \times SCR\_Size} \qquad (5.7)$$

$$Y_Q = \frac{(Y_n - Y_e)}{0.5 \times SCR\_Size} \quad (5.8)$$

The spatially correlated region lies in four regions: in any of the four quadrants or along the axes in either the positive or negative sign. Each warrants a different calculation of the *SCR_OP* (X_{SCR}- and Y_{SCR}-*axes*), as shown in Equations 5.9 and 5.10:

$$X_{SCR} = \begin{cases} \left\lfloor \frac{(X_n - X_e) - (0.5 \times SCR\_Size)}{SCR\_Size} \right\rfloor + 1 & if\ X_Q > 1 \\ \left\lfloor \frac{(X_n - X_e) + (0.5 \times SCR\_Size)}{SCR\_Size} \right\rfloor - 1 & if\ X_Q < -1 \\ 0 & Otherwise \end{cases} \quad (5.9)$$

$$Y_{SCR} = \begin{cases} \left\lfloor \frac{(Y_n - Y_e) - (0.5 \times SCR\_Size)}{SCR\_Size} \right\rfloor + 1 & if\ X_Q > 1 \\ \left\lfloor \frac{(Y_n - Y_e) + (0.5 \times SCR\_Size)}{SCR\_Size} \right\rfloor - 1 & if\ X_Q < -1 \\ 0 & Otherwise \end{cases} \quad (5.10)$$

The purpose of our spatial correlation strategy is to select a single node in every spatially correlated region. To achieve this, we propose to select the active node based on its distance from the center of the spatially correlated region and the residual energy of the node. For example, let *SCR_Center* (X_{CSCR}- and Y_{CSCR}-*axes*) be the center of the spatially correlated region. We then measure the Euclidean distance (*Ed*) of each node lying in the spatially correlated region to the point from the *SCR_Center* (X_{CSCR}, Y_{CSCR}) using Equation 5.11. The closest node is selected as an active member, which we refer to as the representative node, whereas the other nodes simultaneously turn their radios and sensing units off to conserve energy. As depicted in Figure 5.4, the node shaded in

University of Malaya

red (N1) is selected as the representative node because its distance to the center of the spatially correlated region is shorter. The representative node of the spatially correlated region reports the sensed data on behalf of all the nodes of the spatially correlated region.

$$Ed = \sqrt{(X_{CSCR} - X_n)^2 + (Y_{CSCR} - Y_n)^2} \qquad (5.11)$$

The appointed representative nodes should have sufficient energy to perform these tasks as represented by $E_{res}(RN) > Th_{reshold\_Eres}$ because the representative node has to witness, process, and route the event information to the CH. The threshold level is expected to be slightly decreased by the CH if no suitable representative node can be found.

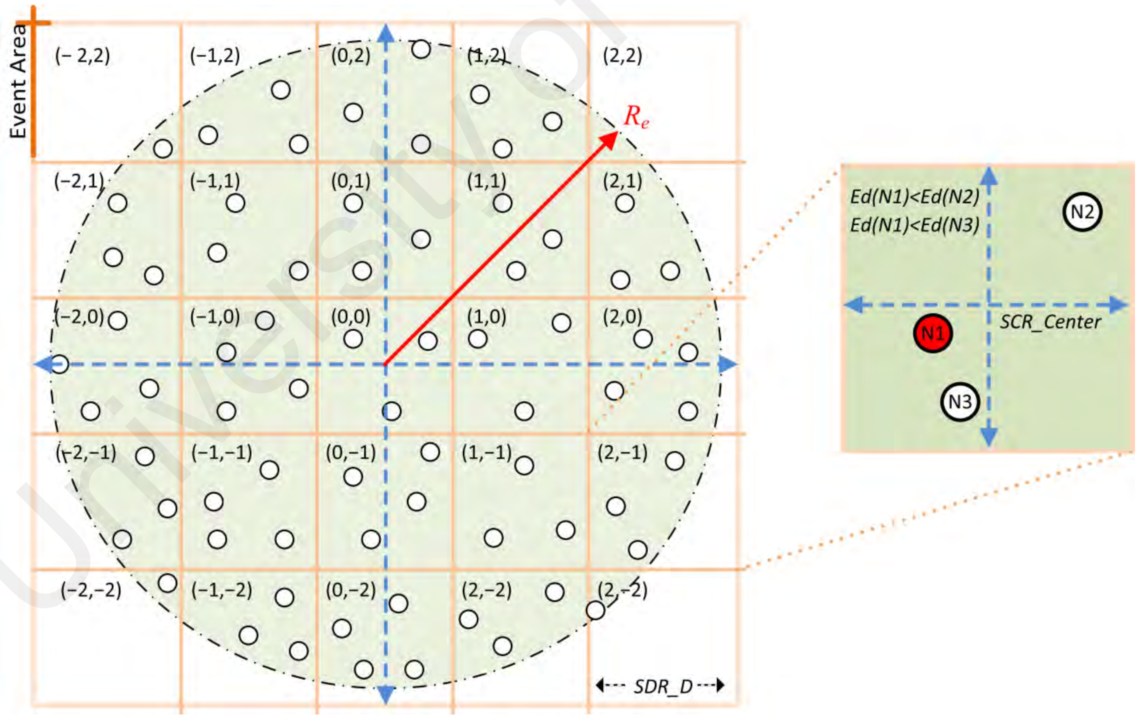


Figure 5.4: Formation Mechanism of the Spatially Correlated Region

As mentioned earlier in section 3.3.2, a normal working procedure of a sensor node comprises several major notification rounds, and each of these rounds includes two periods, which are denoted by period 1 and period 2. The WST-RS scheme utilizes the

University of Malaya

first period of the notification round to determine the current status of whether the representative node should continue as the sensing node. The CHs use the residual energy threshold to determine whether the representative node should be kept in the sensing mode in the next notification round or should its status be changed to sleeping mode and another node be selected as a new representative node. However, an updated list of representative nodes is broadcast by the CHs during this time. The list of representative nodes is only revised if the energy of the current representative nodes becomes less or equal to some residual energy threshold value. The alternation of the representative node in every spatially correlated region is conducted to balance the energy consumption of the nodes that are reporting data and extend the network lifetime. Finally, algorithm 5.3 describes the steps for the spatially correlated region formation phase.

Algorithm 5.3: Spatially correlated regions formation

Input: ***S*** *// S is the set of cluster nodes.*
Output: *N, R // N is the set of SCRs, R is the set of representative node.*
Step 1: *The CH computes the spatial correlated region dimension SCR_D;*
// Using equation (5.4).
Step 2: *CH decomposes the cluster zone to (SCR_D × SCR_D) rectangles;*
Step 3: ***Foreach*** *x* ∈ S *do*
  Step 3.1: ***If*** *SCR_D ≠ 0* ***then***
    Step 3.1.1: *Node x determines the quadrant SCR_Quad (X_Q, Y_Q) in which its spatial correlated region lies;*
    // Using equations (5.7) & (5.8).
    Step 3.1.2: *Node x calculates the ordered pair SCR_OP (X_{SCR}, Y_{SCR}) of its spatial correlated region;*
    // Using equations (5.9) & (5.10).
    Step 3.1.3: *Node x updates its role based on the broadcasted list from CH;*
  Step 3.2: ***End If****;*
  Step 3.3: ***Else***
    Step 3.3.1: *Role(x) ← Representative Node;*
Step 4: ***End For****;*
Step 5: ***End****.*

University of Malaya

5.3.4.4 Routing establishment and data transmission based on MCNW metric with consideration of temporal correlation (Phase IV)

In event-driven sensing applications, consecutive observations exhibit temporal correlation to a certain degree. Therefore, a temporal correlation mechanism is introduced in this phase to exploit this temporal correlation completely while the data are transmitted by the sensor nodes.

Once the spatially correlated region formation and representative node selection are completed in the previous phases, the representative nodes bear the monitoring responsibility in the network and then decide which data to forward to the CH by using the proposed temporal correlation mechanism. To achieve this, the CH gathers and aggregates the data from the representative nodes and then exploits the temporal correlation mechanism based on the definition 2 discussed in Section 5.3.3 to determine what data should be transmitted to the sink node. Algorithm 5.4 describes the proposed procedure for this phase of WST-RS. This procedure utilizes the MCNW, which is saved in the neighbors table in phase I to find the energy efficient route with the desired load balancing.

The CH begins by determining if it is a part of the backbone of the previous event route. If yes, then creating a new route is not required. Otherwise, the CH starts its task by creating a new route by selecting the best neighbor (i.e., the neighbor with the lowest *NW*) as its next hop. This process is necessary to ensure that data of the event data are sent to the sink node via the most reliable and energy efficient route. However, the selection of the best neighbor is performed in two phases.

On the occurrence of the first event, the neighboring node with a position that leads to the highest progress toward the sink is selected as the relay node, as shown in Figure 5.5(A). By contrast, the best neighbor is considered to have a position that leads to the highest progress toward the nearest node that belongs to a route created by the previous event, as depicted in Figure 5.5 (B). Both cases notably consider the residual energy level and accessible buffer memory for the node selection. This process is observed to increase the aggregation points, thereby ensuring that their occurrence is close to the events.

Once the sink is notified of the occurrence of a new event in the network, it adds the identifier of the CH and its position to its table (CHs_Table). The position of the CH is then sent to all CHs by the sink node. Each CH updates its lists of CHs on receiving such a message and then uses this information to create a routing tree. Thus, the CH of the successive event will create their multi-hop routing tree to the existing routing tree. This process is conducted by selecting the nodes with the lightweight route at each hop. The process of updating the list of CHs will maximize the route overlap for efficient network data aggregation.

Algorithm 5.4: Routing establishment and data transmission based on MCNW metric with consideration of temporal correlation

---

Step 1: *The set of nodes S that sensed the event and begins to transfer the data;*
Step 2: ***Foreach*** *$x \in S$ do*
  Step 2.1: ***If*** *Role(x) == Representative Node* **&&** *Error(x_i) > $Th_{reshold\_tt}$* ***then***
  // x_i is a new reading of node x.
    Step 2.1.1: *Node x send the value of new reading to CH;*
  Step 2.2: ***ElseIf*** *Role(x) == Representative Node* **&&** *Error(x_i) < $Th_{reshold\_tt}$* ***then***
    Step 2.2.1: *Node x suppresses the value of the new reading;*
  Step 2.3: ***Else***
    Step 2.3.1: *Node x processing received readings;*
    Step 2.3.2: ***If*** *Aggregation(x_i) = Aggregation(x_{i-1})* ***then***
      Step 2.3.2.1: *Node x suppresses the value of the Aggregation(x_i);*
    Step 2.3.3: ***Else*** *Node x starts to establish route by sending REM to its next hop;*
Step 3: ***Repeat***
  Step 3.1: *Node u is next hop of node x that received REM;*
  Step 3.2: *Role(u) $\leftarrow$ Relay;*
  Step 3.3: ***If*** *NW_E_{res}(NextHop(u))* **&&** *NW_B_{ava}(NextHop(u)) > $Th_{reshold\_w}$* ***then***
    Step 3.3.1: *Node x finds a new next hop with less NW that satisfies the predefined limits by exploiting the neighbor's table;*
  Step 3.4: *Node u send route establishment message REM to its next hop;*
Step 4: ***Until*** *finds the sink node or a node that belongs to the routing structure for the previous event;*
Step 5: ***End***.

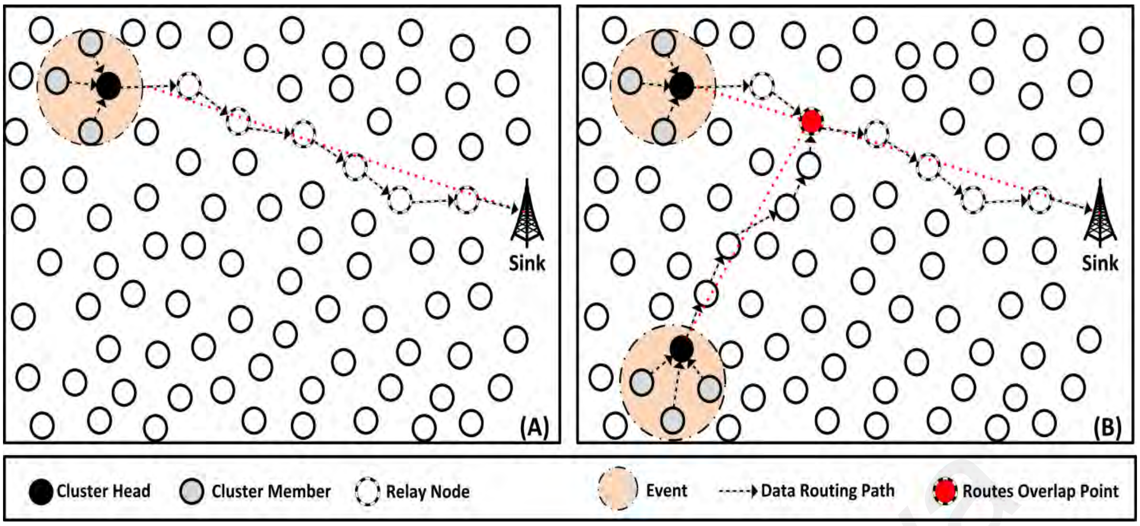


Figure 5.5: Example of Establishing New Routes in WST-RS

5.4 Performance Evaluation

In this section, the experimental setup, simulation results, and discussion of the results of the performance evaluation of WST-RS are presented. The performance of the WST-RS was evaluated by conducting various simulation experiments. Several performance metrics were used to assess the capability and efficiency of the proposed scheme under different WSN environmental conditions. The performance of the proposed WST-RS is evaluated by comparing its simulation results with those of three notable dynamic cluster-based and multi-hop routing algorithms, namely, DRINA (hop update based), YEAST-CF (spatial correlation based), and EAST (spatial and temporal correlation based).

5.4.1 Simulation Setup and Scenario Assumptions

The simulation scenarios are generated specifically for event-driven traffic generation model in WSNs where sensors transmit huge amount of data during the events occurring. Simulation experiments were performed using MATLAB with IEEE 802.15.4 as the MAC layer protocol. The random deployment of sensor nodes is carried out in the sensor field. The sensing and communication ranges of each sensor are fixed, 30m and 35m respectively. In the experiments, events were assumed to occur randomly

University of Malaya

in the entire network area. The first event occurred at time 1,000 s, whereas all other events began at a uniformly distributed random time between 1,000–14,000 s. In addition, the positions at which these events occurred are random. Table 5.2 shows the parameters used in the simulation experiments. The parameters are the standards used in practice. However, in some simulations, certain parameters may vary; these parameters are set in ranges in Table 5.2 as well.

Table 5.2: Settings of Simulation Parameters for WST-RS

Parameters	Type/Value
Channel	Wireless channel
Antenna	Omnidirectional
Underlying MAC protocol	IEEE 802.15.4
Sink node	One with fixed coordinates
Deployment area	1,000 m × 1,000 m (Scenario I) 700 m × 700 m (Scenario II)
Number of sensor nodes	500–3000
Topology	Dynamic cluster
Node density	18.5
Initial node energy	2 J
Notification rate	1/min
No. of events	1–6
Event duration	600 s
Event radius	50 m
Sensing radius	30 m
Communication radius	35 m
Simulation time	14,000 s
Data packet size	512 bytes
Control packet size	32 bytes

University of Malaya

5.4.2 Experimental Results and Discussion

Several extensive simulations and analyses were performed on certain parameters to study the behavior of the proposed WST-RS. The performances of WST-RS, DRINA, YEAST-CF, and EAST were compared using two test cases (Scenarios I and II). The two scenarios were used to show the effects of varying simulation parameters on the performance. The following six metrics that were used to assess the performance of the proposed scheme: energy consumption, network lifetime, packet delivery rate, efficiency (packets per processed data), average packet loss, and overhead (control packet overhead).

5.4.2.1 Scenario I: Impact of node density

For the evaluation of the impact of node density, the sensor field, communication range, and event radius were fixed while the number of nodes varied from 500 to 3000, with a step size of 500 nodes. The nodes were randomly deployed. The experiments were performed for node densities given by $\mu = n \pi R_c^2 / M$, where n is the number of nodes, M is the sensor field, and R_c is the transmission radius. The number of nodes was varied, whereas the latter two parameters were kept constant. In addition, the number of events in this scenario was set to three, which were randomly placed in the sensor field.

When the density varied, WST-RS still outperformed DRINA, YEAST-CF, and EAST. As depicted in Figure 5.6, when the network density increased, the consumed energy of the four algorithms significantly decreased, but the average energy consumption of WST-RS was the lowest. Thus, less energy is spent by WST-RS for processing data packets generated by the source nodes. Consequently, the average energy consumption of WST-RS was 55.32%, 36.60%, and 19.23% lower than those of DRINA, YEAST-CF, and EAST, respectively. Such superiority is attributed to the capability of WST-RS to work in a distributed manner to construct and maintain the correlation between sensor nodes and their readings adaptively. This feature allows the nodes that sensed the event to remove redundant notifications. Hence, decrease the energy consumption for processing and transmission of these redundant notifications.

Network lifetime is defined as the time before the death of the first network node. As shown in Figure 5.7, the average network lifetime of WST-RS was 3.08, 2.18, and 1.88 times longer than those of DRINA, YEAST-CF, and EAST, respectively. This result is due to the energy saved by WST-RS and the load balancing that occurs when the node density is high and the network size is stable. The approach of the energy balance adopted by the WST-RS appears to be beneficial with increased of network density

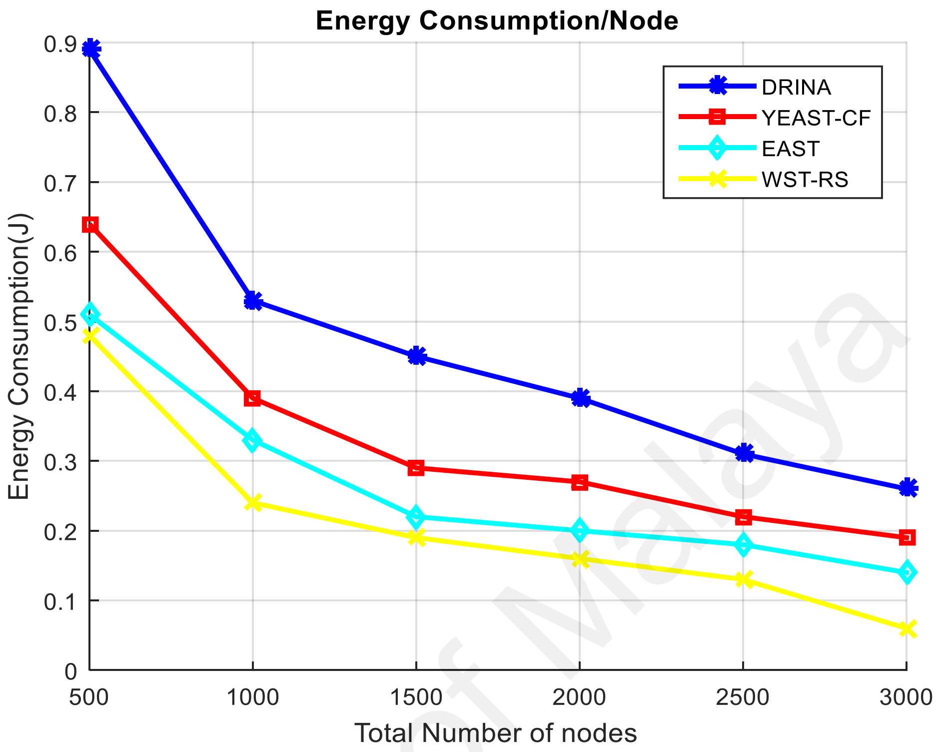


Figure 5.6: Comparison of the Energy Consumption Levels in Scenario I

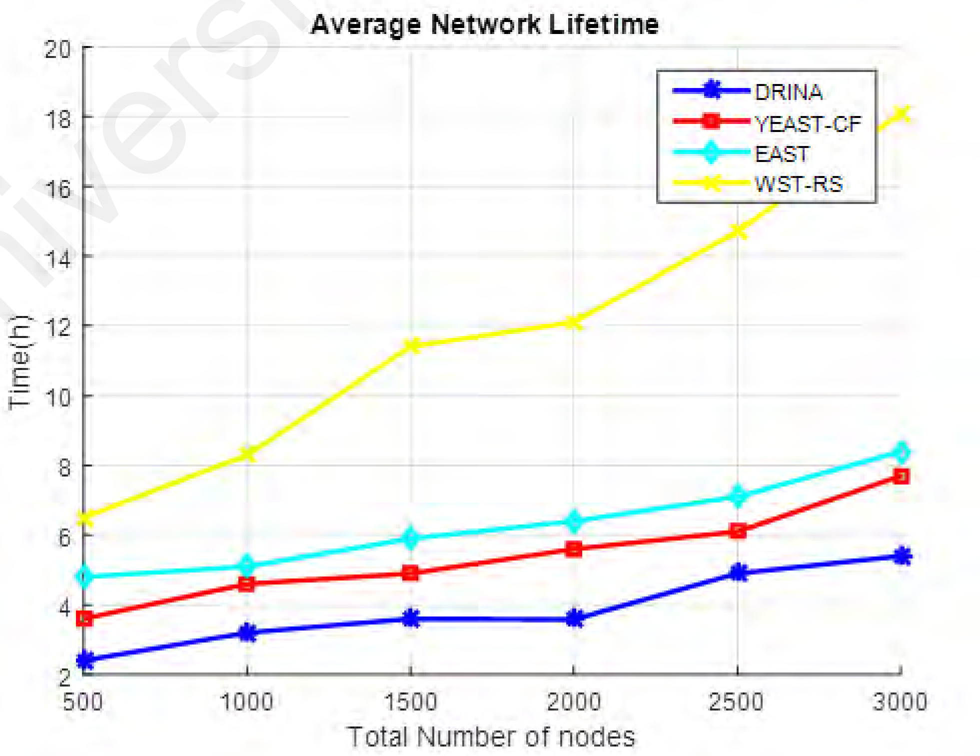


Figure 5.7: Comparison of the Average Network Lifetimes in Scenario I

University of Malaya

Figure 5.8, shows the efficiency (packets per processed data) of WST-RS, DRINA, YEAST-CF and EAST in different network density. The WST-RS presents 49.84%, 13.07%, and 4.38% increased efficiency compared with DRINA, YEAST-CF, and EAST, respectively. The efficiency improvement of WST-RS algorithm is attributed to its consideration of the spatial and temporal correlations of the data as well as the relatively low number of control messages required to establish the routing tree.

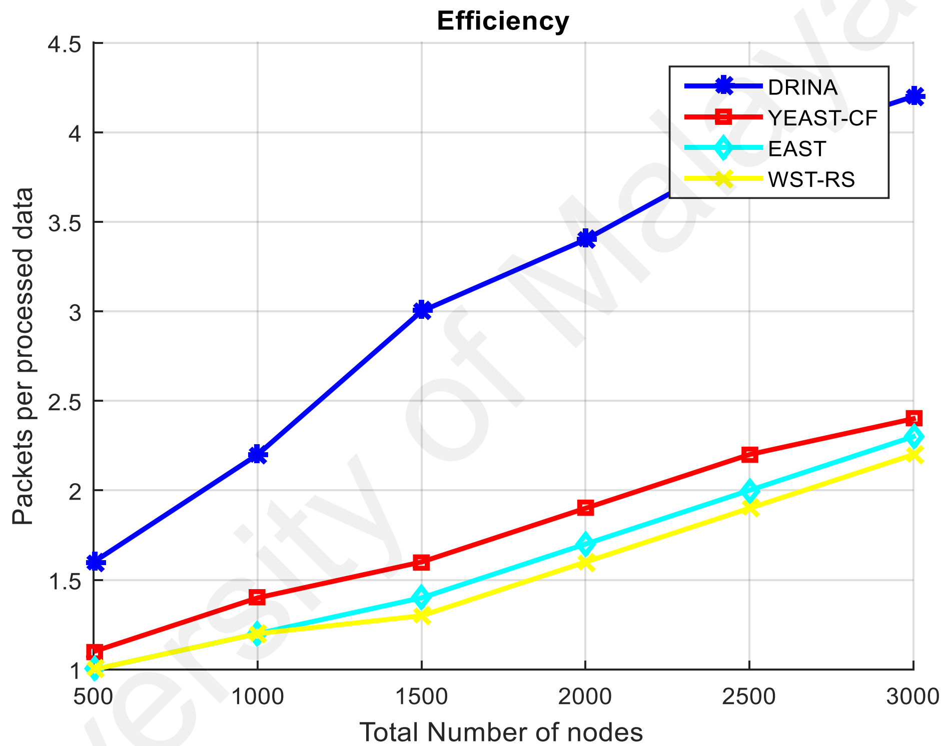


Figure 5.8: Comparison of the Efficiency Levels in Scenario I

Another evaluation metric that indicates the quality of the routing tree built by the algorithms is measured by the packet delivery rate. A lower packet delivery rate indicates higher built tree aggregation rate. The data aggregation quality of the routing tree established by WST-RS was better than those of the routing trees built by DRINA, YEAST-CF, and EAST, as shown in Figure 5.9. Utilizing the MCNW metric, WST-RS has the advantage of using fewer MHR nodes in its routing tree compared with DRINA, YEAST-CF, and EAST.

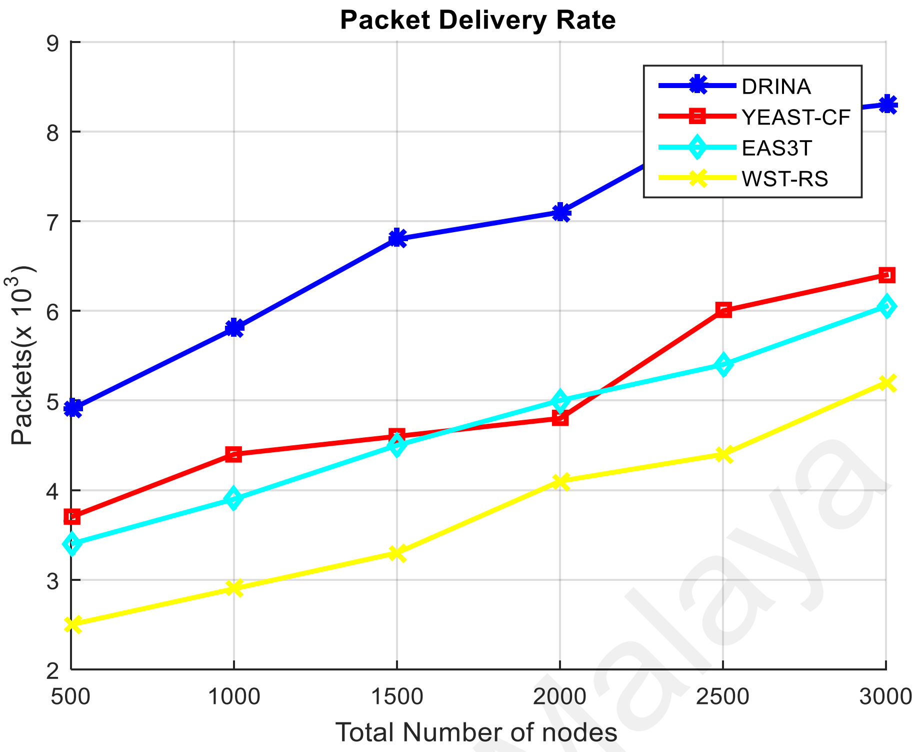


Figure 5.9: Comparison of the Packet Delivery Rates in Scenario I

The number of packets lost due to collisions and full queues at the destination node was also analyzed to compare WST-RS, DRINA, YEAST-CF, and EAST (Figure 5.10). WST-RS lost the least number of packets for different network densities, exhibiting packet loss rates ranging from 0.03% to 0.07%. By contrast, the other protocols generated 0.04%–0.21% average packet loss rates. This result is because WST-RS continuously tracks the available buffer memory sizes of the CH and backbone nodes.

Finally, Figure 5.11 shows the number of control messages required to create and maintain the routing structure. As shown, the control packet overhead of all the routing protocols increased with increasing node density. Furthermore, the experimental results in Scenario 1 showed that the numbers of control packets required in WST-RS, YEAST-CF, and EAST were relatively similar and were all less than that in DRINA. WST-RS specifically requires on average only 41.24% of the control packets used by DRINA to build a routing structure.

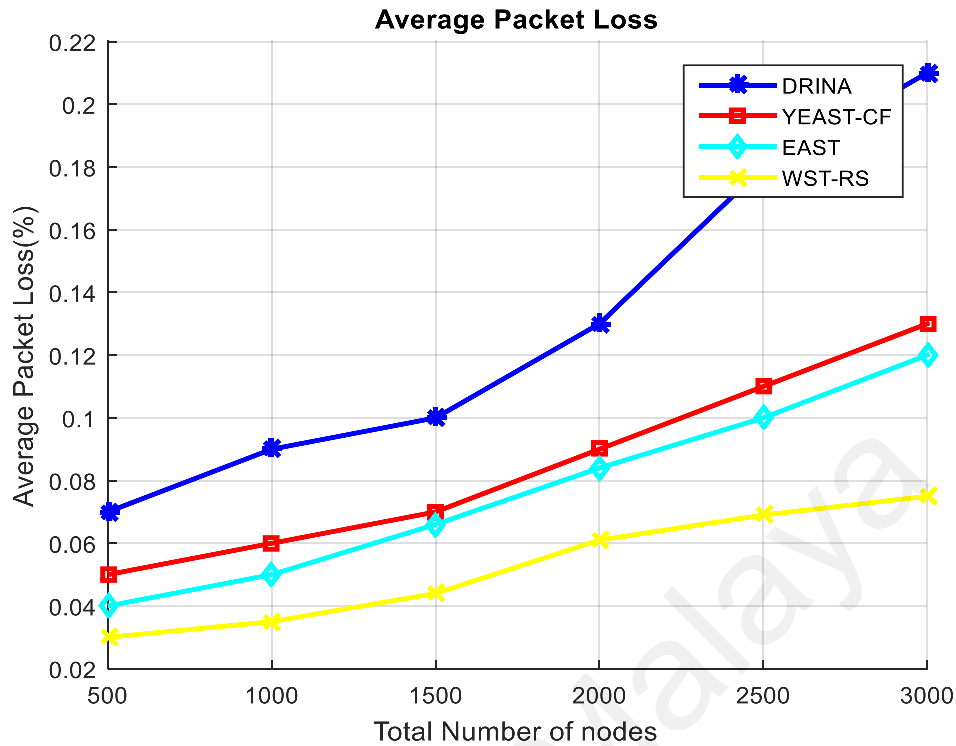


Figure 5.10: Comparison of the Average Packet Loss Rates in Scenario I

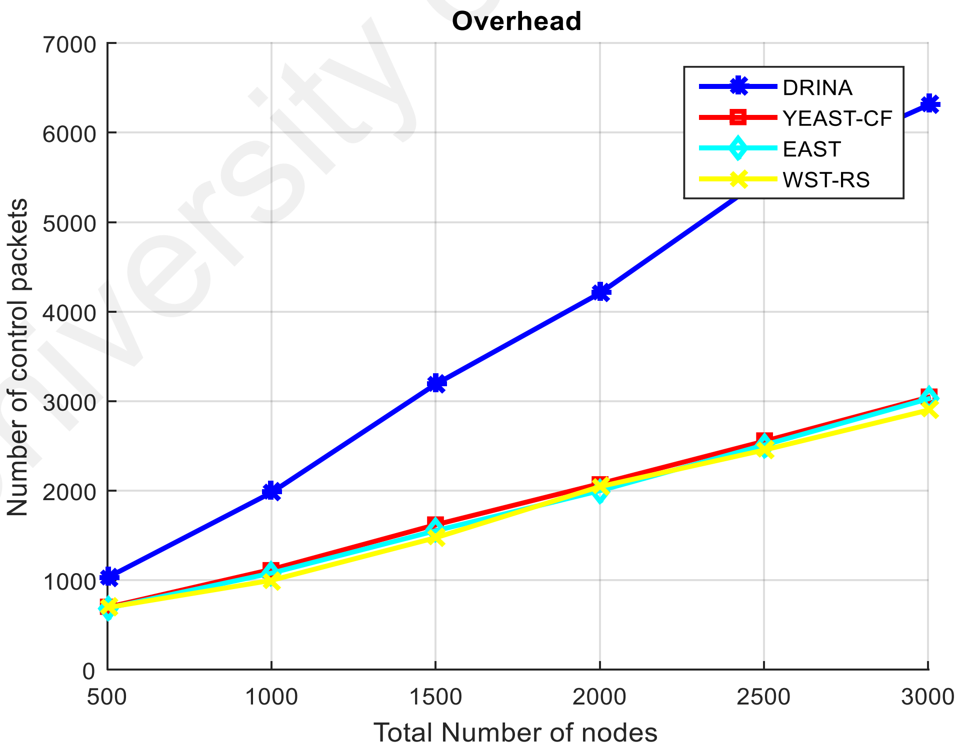


Figure 5.11: Overhead Comparison in Scenario I

University of Malaya

5.4.2.2 Scenario II: Impact of event scalability

In this scenario, we wanted to evaluate the behavior of the proposed WST-RS with increasing number of simultaneous events. For this evaluation, we simulated 1,500 nodes in a network with a fixed size of 700 m × 700 m and then increased the number of simultaneous events from 1 to 6. The events were randomly placed in the sensor field.

First, the ratio of energy consumed by all nodes for sensing, processing and transmission the generated data packets to the total number of network nodes is compared. The consumption levels of all the compared algorithms were low when the number of events was low, as shown in Figure 5.12. The performance of the proposed WST-RS was better than those of the other compared protocols in terms of energy consumption throughout the simulation for different numbers of concurrent events. The nodes in the proposed WST-RS recorded the lowest energy consumption, which accounted for only 22% of the initial energy. By contrast, the energy consumption levels of DRINA, YEAST-CF, and EAST were 48%, 33%, ad 28% of the initial energy. The proposed methodology of WST-RS allows the nodes that sensed the event to remove redundant notifications. Hence, decrease the energy consumption for processing and transmission of these redundant notifications.

An analysis of the network lifetime was also performed with respect to simulation time. The network lifetime curves of all schemes were declining with the increasing event number, as shown in Figure 5.13. This result is due to the increase in load as the number of events increases. The time when the first node died in WST-RS was significantly later than those in DRINA, YEAST-CF, and EAST. The lifetime of the network using WST-RS was nearly 3.04 times longer than that of DRINA, 2.27 times longer than that of YEAST-CF, and 1.77 times longer than that of EAST. These results can be attributed to the utilization of the developed MCNW metric by WST-RS to

University of Malaya

establish and update the routing tree and stabilize energy consumption and other resources.

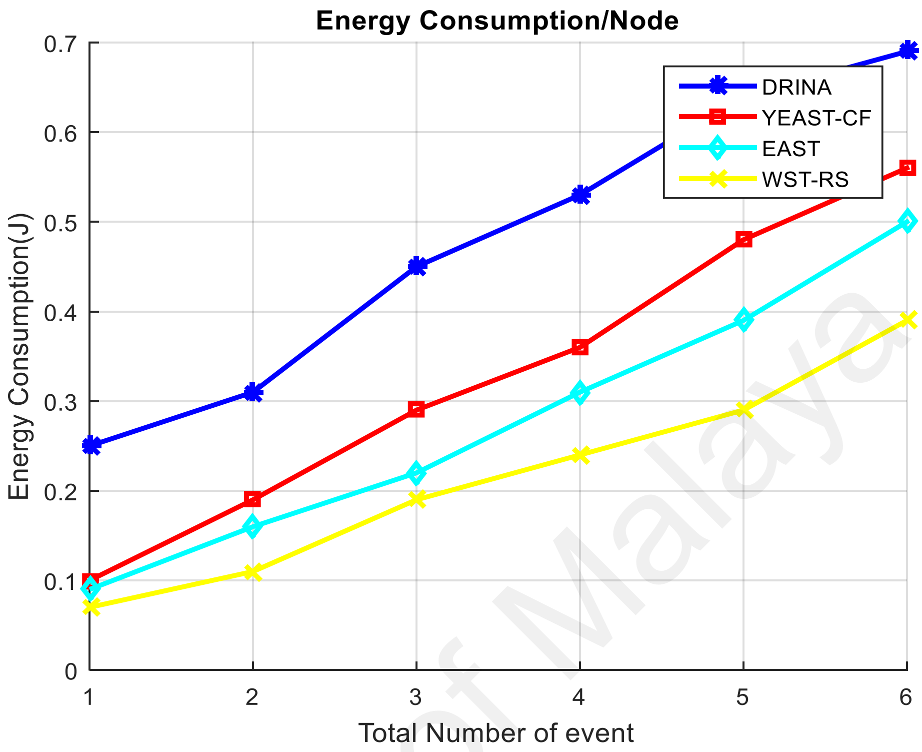


Figure 5.12: Comparison of the Energy Consumption Levels in Scenario II

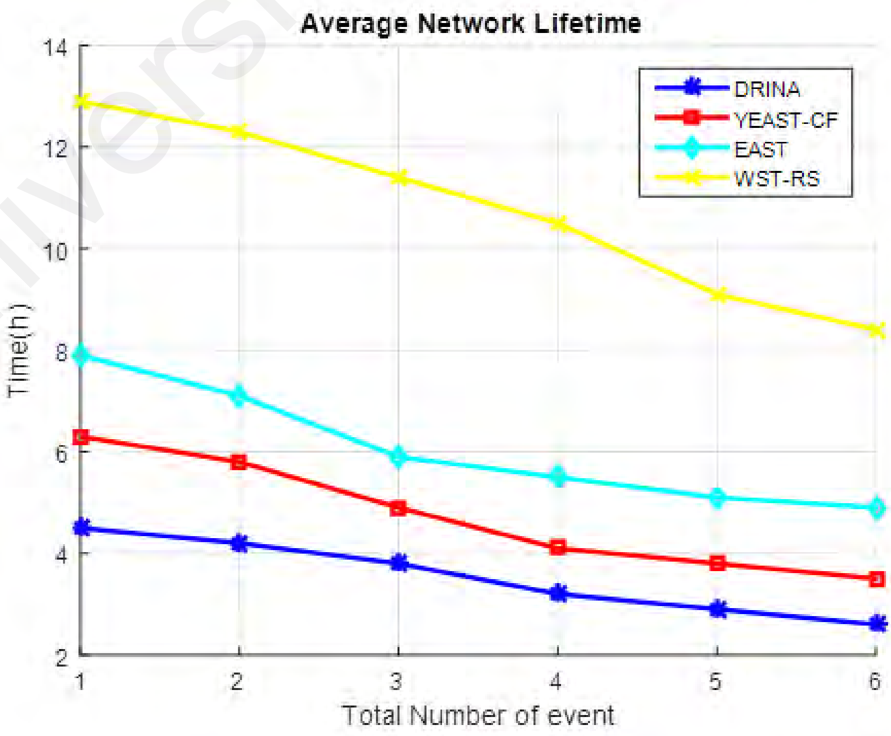


Figure 5.13: Comparison of the Average Network Lifetimes in Scenario II

University of Malaya

Figure 5.14 clearly shows that the proposed WST-RS was more efficient than DRINA, YEAST-CF, and EAST protocols by 48.50%, 16.26%, and 8.85%, respectively. The efficiency of a routing protocol is also reflected by the total packets sent to the sink node. The WST-RS scheme takes into consideration the convergence of events data in space and time.

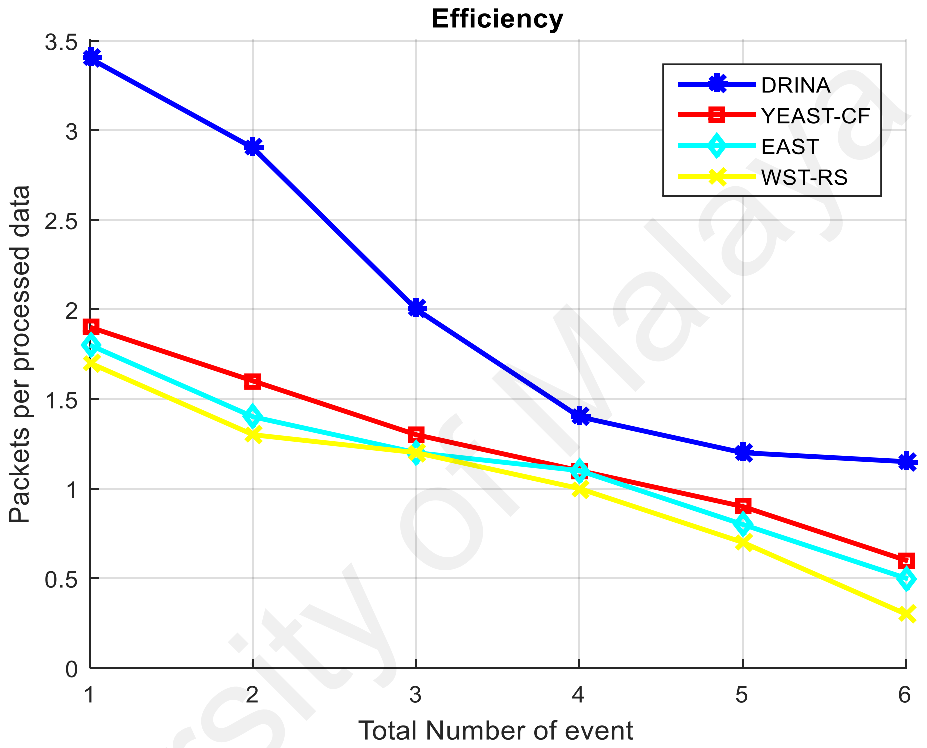


Figure 5.14: Comparison of the Efficiency Levels in Scenario II

Figure 5.15 show that the WST-RS obtained a relatively high data aggregation rate across different numbers of concurrent events. The WST-RS scheme connects all the sensor nodes that detected the event to the sink node via minimal relay nodes tree. The perfect aggregation results were achieved via minimal relay nodes tree.

University of Malaya

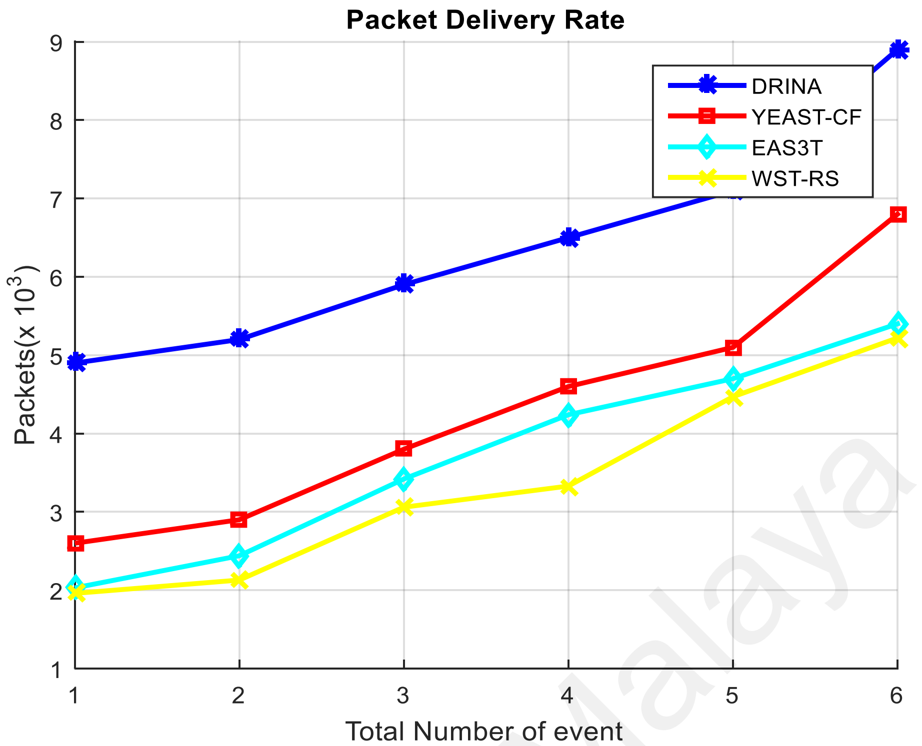


Figure 5.15: Comparison of the Packet Delivery Rates in Scenario II

In Scenario I, the performance of the proposed WST-RS was also highlighted by comparing it with existing protocols in terms of the average packet loss rate as a function of the number of simultaneous events. As shown in Figure 5.16, the average packet loss rates of all the protocols increased with an increase in the number of sensed events. Therefore, a direct relationship between packet loss and number of sensed events was confirmed. However, WST-RS generated the minimum average packet loss rates, which ranged from 0.02% to 0.06%, for different numbers of events. By contrast, the other compared protocols generated 0.04% to 0.31% average packet loss rates. The superiority of the WST-RS is mainly due to the fact that the proposed protocol considers the current congestion level of the node in selecting MHR nodes.

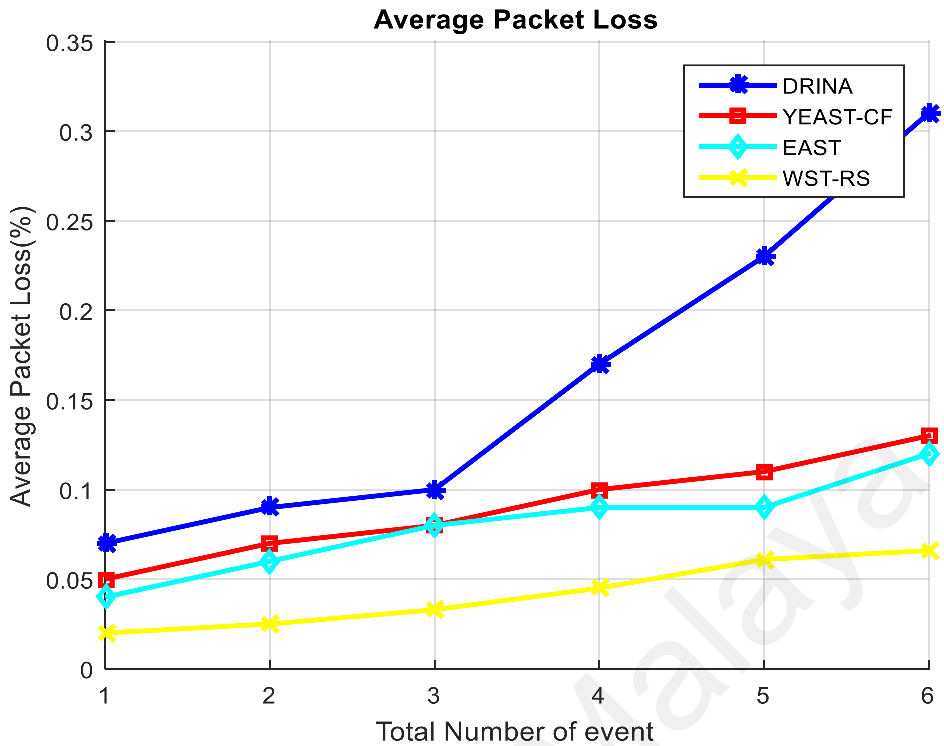


Figure 5.16: Comparison of the Average Packet Loss Rates in Scenario II

Finally, the control packet overhead parameter was used to analyze and compare the performances of the compared protocols. The overhead variations of all the routing protocols are shown in Figure 5.17. The obtained results showed that an increase in the number of events affected the control packet overhead. WST-RS showed a slightly higher performance than the YEAST-CF and EAST protocols. This improved performance of WST-RS is mainly attributed to its use of few relay nodes, which account for low control overhead in general.

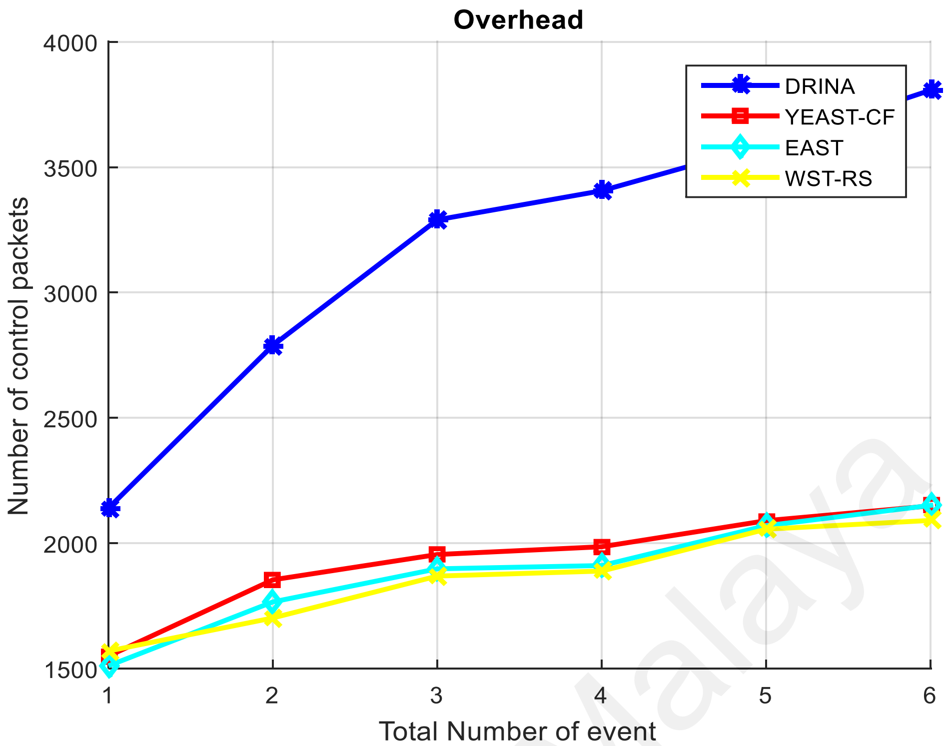


Figure 5.17: Overhead Comparison in Scenario II

5.5 Chapter Summary

This chapter discussed the structural components and modules of the proposed WST-RS. Its salient features of route computation based on the developed MCNW metric and capability to exploit spatial-temporal correlation in WSNs were emphasized. In the first phase of WST-RS, the implemented model returns the values of the metrics (i.e., the residual battery, available buffer memory size, and position of the node) considered in the MCNW metric. These metrics are then exploited by WST-RS to construct a dynamic route for improving the quality of the routing tree and optimizing data aggregation during communication. Moreover, these metrics are also utilized by the link cost function to assess the quality of links and select the most stable path to the destination by avoiding the nodes with high congestion levels and low remaining battery energy. The capability of the proposed WST-RS to work in a distributed manner to construct and maintain the correlation between sensor nodes and their readings adaptively was also exhibited. This feature allows for the efficient energy consumption of the nodes that sensed the event by removing redundant notifications.

Two main scenarios related to network density and event scalability were considered in evaluating the effects of different parameters on the performance of the proposed WST-RS. The evaluation results of WST-RS were compared with those of DRINA (hop based), YEAST-CF (spatial correlation based), and EAST (spatial and temporal correlation based) routing schemes by using several performance metrics related to energy efficiency and load balancing. The proposed WST-RS exhibited the best performance results among the compared routing schemes. Such superiority is attributed to the capability of WST-RS to reduce energy consumption, the number of dropped packets, the number of packets per processed data, and the number of control packets. Moreover, WST-RS increases the network lifetime and data aggregation rate. Consequently, WST-RS enables the system to achieve a high reliability in different WSN scenarios. In the next chapter, the conclusions of this thesis, along with recommendations for future work, are presented.

University of Malaya

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 the academe, mainly due to the decentralization and dynamism of WSNs. However, WSNs are often described by their constrained resources and topology management complexity. Hence, routing techniques are expected to adapt to the characteristics of these networks. Numerous routing schemes have been successfully proposed that adapt to these characteristics. However, other challenging problems are associated with these routing schemes, such as data redundancy and non-balanced load distribution that lead to energy resource exhaustion in nodes. The well-known DRINA and EAST are routing protocols for in-network data aggregation that are expected to be effective solutions to the aforementioned routing problems of residing redundancy in sensor node data through maximizing the rate of routes overlapping for high aggregation. Furthermore, EAST utilizes the merits associated with spatial and temporal correlations to reduce data communication and save energy. However, both DRINA and EAST routing schemes did not take into consideration energy and traffic load balancing simultaneously while making routing decisions. Hence, this study aimed to make the best utilization of the advantage associated with in-network aggregation routing to decrease data communication. This approach will at the same time consider multiple metrics that relate to energy, congestion level, and distance to the destination during MHR nodes selection and route computation. The aims are to decrease the energy consumption of the nodes and effectively balance the load distribution among the sensors. The proposed schemes start by considering node-related metrics to set the willingness of nodes to become a part of the MHR nodes. Next, the schemes proceed by estimating the MCNW metric that is exploited by link cost assessment function to ascertain the efficiency and quality of links and selecting a path to the destination that will lead to finding the ideal point for routes overlapping. The second proposed scheme exploits the spatial and temporal correlation among sensor data in a distributed manner and then utilizes a new mechanism for calculating the spatial and temporal convergences.

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 routing issues related to energy consumption, load balancing and data redundancy are all tackled in this research. Four objectives are defined in Section 1.5 in order to achieve the mentioned main goal of this thesis. The following discussions demonstrate how each one of the reached objectives is met in the research study.

Objective 1: *To investigate the existing in-network data aggregation and correlation awareness routing protocols:* This objective has been achieved by investigating the related literature of in-network data aggregation routing protocols and data correlation algorithms using quantitative analysis. Several representative in-network data aggregation and data correlation routing protocols are analyzed, and their key features, merits, and demerits are highlighted. Based on the review of the existing routing protocols and the results obtained from their performance evaluation, the issues related to energy consumption, load balancing, and data redundancy became clearer. The structure of each protocol is summarized and discussed in chapter 2.

Objective 2: *To propose an innovative in-network data aggregation routing scheme for energy-efficient and load-balanced routing:* The second objective of this research work focuses on providing a step-by-step procedure for presenting an innovative in-network data aggregation routing scheme. This routing scheme was developed as a weighted in-network data aggregation routing scheme (WDARS) for energy-efficient and load-balanced routing in WSNs. It considers energy and load balancing awareness metrics to reduce the energy consumption and balance the load distribution among sensor nodes, as well as help improve the network lifetime, especially in a large-scale environment. Unlike the DRINA scheme, the proposed WDARS scheme exploits multi-metrics related to energy and load balancing to help select the set of MHR nodes between source-destination pairs. This proposed scheme makes use of the normalized weighted additive utility function (NWAUF) function in estimating the MCNW metric associated with each node with respect to residual energy capacity, available buffer memory size, hop to sink, and hop to tree. These metrics were utilized by a link cost function to measure the quality of links when route computation has been completed. The core and auxiliary functionalities of the proposed WDARS scheme is discussed in detail in Chapter 4.

Objective 3: *To propose a novel routing scheme for efficient data collection and aggregation that considers both spatial and temporal correlations to lower communication and data exchange:* Sensors data correlation creates significant data redundancy, which leads to unnecessary network resource utilization in the processing and forwarding of the redundant data to base station. Therefore, developing an efficient and feasible routing scheme with the capability of exploiting both spatial and temporal correlations among the sensor observations was the third objective of this thesis. This routing scheme was proposed as weight aware spatial-temporal correlation routing scheme (referred to as WST-RS). The WST-RS was proposed to extend network lifetime by applying a spatial-temporal correlation mechanism to a detected event and to decrease data redundant communication exchange. The WST-RS benefits from some functionalities of the well-known EAST algorithm with extended functionalities to work in a distributed manner and achieve better energy consumption. This outcome was achieved by adaptively constructing and maintaining an ideal correlation among sensor nodes and their readings. Moreover, the WST-RS adapts the mechanism for MHR nodes set selection based on MCNR metric estimation to ensure the stability of nodes in terms of available energy and available buffer size. Such stability allows the best path selection to the destination using the link cost assessment function. A comprehensive discussion of the WST-RS design architecture is provided in Chapter 5.

Objective 4: *To evaluate the developed schemes with different simulation scenarios and evaluation metrics:* The final objective is met by conducting several extensive simulation experiments for testing and evaluating the proposed schemes under different scenarios. The simulated work was executed 5–10 times, and the average of the readings was used to generate the graphical results. The efficiency of the WDARS in terms of energy and load balancing was evaluated in Section 4.4. The WDARS advantage and capability to select nodes that have higher remaining energy, lower congestion level, and less hops toward the destination for constructing routing tree to the destination outperform the conventional schemes. The WDARS has the ability to find the ideal point for routes overlapping by considering the distance to the established routing path by previous events and the distance to the sink. 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 average energy consumption, network lifetime, packet per processed data, packet delivery rate, average packet loss, and control overhead. The empirical experimental results show that WDARS scheme outperforms DRINA and InFRA in terms of energy consumption by 11.65% as compared to 21.75% and 35.71% respectively. Furthermore, the first node to die in WDARS is much longer than that in InFRA and DRINA, nearly 7.42 times longer than InFRA and 4.25 times longer than DRINA.

Correspondingly, with the integration of energy efficiency, load balancing, and spatial-temporal correlation awareness techniques, the developed WST-RS has likewise succeeded in decreasing the communication of data, avoiding the nodes with low residual energy capacity and high congestion level during MHR nodes selection and route computation, thus preserving the energy and network lifetime in large-scale WSN scenarios. The experiment results showed that the WST-RS scheme outperforms DRINA, YEAST-CF and EAST schemes in terms of energy consumption. The nodes in the proposed WST-RS scheme recorded the lowest energy consumption with 16.25% from the initial energy as compared to the 35.79%, 24.83%, and 20.58% for DRINA, YEAST-CF, and EAST respectively. The first node to die in WST-RS is much longer than that in DRINA, YEAST-CF, and EAST, nearly 3.06 times longer than DRINA, and 2.22 times longer than YEAST-CF, and 1.82 times longer than EAST. A detailed discussion of the performance evaluation and comparison of the WST-RS with EAST, YEAST-CF, and DRINA schemes is provided in Section 5.4.

6.3 Recommendations for Future Work

Two in-network data aggregation routing schemes centered on energy efficiency with load balancing for multi-hop WSNs are proposed in this research. However, other unsolved issues still need to be addressed in the aspect of routing protocol for in-network data aggregation, which may be considered for further research. These unsolved issues include the following:

(i) With time restrictions, implementing and deploying sensors programmed with WDARS or WST-RS in a real-life environmental monitoring context are not possible. Therefore, an implementation and evaluation of their performance in a real-life still need to be validated.

University of Malaya

(ii) Several parameters were used in the proposed schemes. Further investigation to find the optimum values of these parameters using optimization techniques aims to aid in deciding the best weightage based on multiple objectives.

(iii) The waiting time for aggregator nodes can also be investigated by taking into consideration different application needs. Therefore, new techniques may be required to control the delay from waiting time for aggregator nodes in multi-criterion.

6.4 Limitations of this Work

The WDARS and WST-RS schemes proposed in this work were not implemented in a real network deployment scenario. The implementation, validation, and performance evaluation of these schemes were done by simulation. Thus, the input values and parameters used were based on assumptions from previous studies. Furthermore, both WDARS and WST-RS are based on the DRINA and EAST implementations, and as such can only be used for event-based WSNs.

University of Malaya

REFERENCES

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

AbdelSalam, Hady S, & Olariu, Stephan. (2009). *A lightweight skeleton construction algorithm for self-organizing sensor networks.* Paper presented at the 2009 IEEE International Conference on Communications.

Abdulsalam, Hanady M, & Kamel, Layla 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.

Abid, Besem, Nguyen, Tien Trung, & Seba, Hamida. (2015). New data aggregation approach for time-constrained wireless sensor networks. *The Journal of Supercomputing, 71*(5), 1678-1693.

Adya, Atul, Bahl, Paramvir, Padhye, Jitendra, Wolman, Alec, & Zhou, Lidong. (2004). *A multi-radio unification protocol for IEEE 802.11 wireless networks.* Paper presented at the Broadband Networks, 2004. BroadNets 2004. Proceedings. First International Conference on.

Ahn, Kyu Sung, Kim, Dae Gun, Sim, Back Sun, Youn, Hee Yong, & Song, Ohyoung. (2011). *Balanced chain-based routing protocol (BCBRP) for energy efficient wireless sensor networks.* Paper presented at the Parallel and Distributed Processing with Applications Workshops (ISPAW), 2011 Ninth IEEE International Symposium on.

Akkaya, Kemal, & Younis, Mohamed. (2003). *An energy-aware QoS routing protocol for wireless sensor networks.* Paper presented at the Distributed Computing Systems Workshops, 2003. Proceedings. 23rd International Conference on.

Akkaya, Kemal, & Younis, Mohamed. (2005). A survey on routing protocols for wireless sensor networks. *Ad hoc networks, 3*(3), 325-349.

Akyildiz, Ian F, Su, Weilian, Sankarasubramaniam, Yogesh, & Cayirci, Erdal. (2002). A survey on sensor networks. *IEEE communications magazine, 40*(8), 102-114.

Aminian, Mehdi, Akbari, Mohammad K, & Sabaei, Masoud. (2013). A rate-distortion based aggregation method using spatial correlation for wireless sensor networks. *Wireless personal communications, 71*(3), 1837-1877.

Bachir, Abdelmalik, Dohler, Mischa, Watteyne, Thomas, & Leung, Kin K. (2010). MAC essentials for wireless sensor networks. *IEEE Communications Surveys & Tutorials, 12*(2), 222-248.

Bagaa, Miloud, Challal, Yacine, Ksentini, Adlen, Derhab, Abdelouahid, & Badache, Nadjib. (2014). Data aggregation scheduling algorithms in wireless sensor

University of Malaya

networks: Solutions and challenges. *IEEE Communications Surveys & Tutorials, 16*(3), 1339-1368.

Bagaa, Miloud, Derhab, Abdelouahid, Lasla, Noureddine, Ouadjaout, Abdelraouf, & Badache, Nadjib. (2012). *Semi-structured and unstructured data aggregation scheduling in wireless sensor networks.* Paper presented at the INFOCOM, 2012 Proceedings IEEE.

Bai, Fan, & Jamalipour, Abbas. (2008a). *3D-DCT data aggregation technique for regularly deployed wireless sensor networks.* Paper presented at the 2008 IEEE International Conference on Communications.

Bai, Fan, & Jamalipour, Abbas. (2008b). *Performance evaluation of optimal sized cluster based wireless sensor networks with correlated data aggregation consideration.* Paper presented at the 2008 33rd IEEE Conference on Local Computer Networks (LCN).

Bechkit, Walid, Koudil, Mouloud, Challal, Yacine, Bouabdallah, Abdelmadjid, Souici, Brahim, & Benatchba, Karima. (2012). *A new weighted shortest path tree for convergecast traffic routing in WSN.* Paper presented at the Computers and Communications (ISCC), 2012 IEEE Symposium on.

Borzi, Skog , & Baranauckas. (2007). Bridge Collapse in Minneapolis Kills at Least 7, *The New York Times* http://www.nytimes.com/2007/08/02/us/02bridge.html.

Bouabdallah, Nizar, Rivero-Angeles, Mario E, & Sericola, Bruno. (2009). Continuous monitoring using event-driven reporting for cluster-based wireless sensor networks. *IEEE Transactions on Vehicular Technology, 58*(7), 3460-3479.

Buttyán, Levente, Gessner, Dennis, Hessler, Alban, & Langendörfer, Peter. (2010). Application of wireless sensor networks in critical infrastructure protection: challenges and design options [security and privacy in emerging wireless networks]. *IEEE Wireless Communications, 17*(5), 44-49.

Castelluccia, Claude, Chan, Aldar CF, Mykletun, Einar, & Tsudik, Gene. (2009). Efficient and provably secure aggregation of encrypted data in wireless sensor networks. *ACM Transactions on Sensor Networks (TOSN), 5*(3), 20.

Chao, Chih-Min, & Hsiao, Tzu-Ying. (2014). Design of structure-free and energy-balanced data aggregation in wireless sensor networks. *Journal of Network and Computer Applications, 37*, 229-239.

Chen, Huifang, Mineno, Hiroshi, & Mizuno, Tadanori. (2008). Adaptive data aggregation scheme in clustered wireless sensor networks. *Computer Communications, 31*(15), 3579-3585.

Chen, Huifang, Mineno, Hiroshi, Obashi, Yoshitsugu, Kokogawa, Tomohiro, & Mizuno, Tadanori. (2007). *Adaptive data aggregation for clustered wireless sensor networks.* Paper presented at the International Conference on Ubiquitous Intelligence and Computing.

University of Malaya

Chen, Jie. (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, Wei-Peng, Hou, Jennifer C, & Sha, Lui. (2004). Dynamic clustering for acoustic target tracking in wireless sensor networks. *IEEE transactions on mobile computing, 3*(3), 258-271.

Cheng, Bo-Chao, Liao, Guo-Tan, Tseng, Ryh-Yuh, & Hsu, Ping-Hai. (2012). Network lifetime bounds for hierarchical wireless sensor networks in the presence of energy constraints. *Computer Networks, 56*(2), 820-831.

Choi, Kyung, & Chae, Kijoon. (2014). *Data aggregation using temporal and spatial correlations in Advanced Metering Infrastructure.* Paper presented at the The International Conference on Information Networking 2014 (ICOIN2014).

Das, Sankar Narayan, Misra, Sudip, Wolfinger, Bernd E, & Obaidat, Mohammad S. (2016). Temporal-Correlation-Aware Dynamic Self-Management of Wireless Sensor Networks. *IEEE Transactions on Industrial Informatics, 12*(6), 2127-2138.

Ding, Min, Cheng, Xiuzhen, & Xue, Guoliang. (2003). *Aggregation tree construction in sensor networks.* Paper presented at the Vehicular Technology Conference, 2003. VTC 2003-Fall. 2003 IEEE 58th.

Draves, Richard, Padhye, Jitendra, & Zill, Brian. (2004). *Routing in multi-radio, multi-hop wireless mesh networks.* Paper presented at the Proceedings of the 10th annual international conference on Mobile computing and networking.

Efthymiou, Charilaos, Nikoletseas, Sotiris, & Rolim, Jose. (2006). Energy balanced data propagation in wireless sensor networks. *Wireless Networks, 12*(6), 691-707.

Elbhiri, Brahim, El Fkihi, Sanaa, Saadane, Rachid, & Aboutajdine, Driss. (2010). *Clustering in wireless sensor networks based on near optimal bi-partitions.* Paper presented at the Next Generation Internet (NGI), 2010 6th EURO-NF Conference on.

Fan, Kai-Wei, Liu, Sha, & Sinha, Prasun. (2007). Structure-free data aggregation in sensor networks. *IEEE Transactions on Mobile Computing, 6*(8), 929-942.

Farooq, Hasan, & Tang Jung, Low. (2013). Energy, traffic load, and link quality aware Ad Hoc routing protocol for wireless sensor network based smart metering infrastructure. *International Journal of Distributed Sensor Networks, 2013*.

Fasolo, Elena, Rossi, Michele, Widmer, Jorg, & Zorzi, Michele. (2007). In-network aggregation techniques for wireless sensor networks: a survey. *IEEE Wireless Communications, 14*(2), 70-87.

Gasser, Marc. (2014). In-network aggregation techniques in wireless sensor networks. *Internet Economics, 8*, 7.

University of Malaya

Ghosh, Amitabha, & Das, Sajal K. (2008). Coverage and connectivity issues in wireless sensor networks: A survey. *Pervasive and Mobile Computing, 4*(3), 303-334.

Gielow, Fernando, Jakllari, Gentian, Nogueira, Michele, & Santos, Aldri. (2015). Data similarity aware dynamic node clustering in wireless sensor networks. *Ad Hoc Networks, 24*, 29-45.

Gungor, Vehbi C, Lu, Bin, & Hancke, Gerhard P. (2010). Opportunities and challenges of wireless sensor networks in smart grid. *IEEE transactions on industrial electronics, 57*(10), 3557-3564.

Halgamuge, Malka N, Zukerman, Moshe, Ramamohanarao, Kotagiri, & Vu, Hai Le. (2009). An estimation of sensor energy consumption. *Progress In Electromagnetics Research B*.

He, Tian, Gu, Lin, Luo, Liqian, Yan, Ting, Stankovic, John A, & Son, Sang Hyuk. (2006). *An overview of data aggregation architecture for real-time tracking with sensor networks.* Paper presented at the Parallel and Distributed Processing Symposium, 2006. IPDPS 2006. 20th International.

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

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

Heinzelman, Wendi Rabiner, Chandrakasan, Anantha, & Balakrishnan, Hari. (2000). *Energy-efficient communication protocol for wireless microsensor networks.* Paper presented at the System sciences, 2000. Proceedings of the 33rd annual Hawaii international conference on.

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

Hu, Xiaoya, Wang, Bingwen, & Ji, Han. (2013). A wireless sensor network-based structural health monitoring system for highway bridges. *Computer-Aided Civil and Infrastructure Engineering, 28*(3), 193-209.

Jaigirdar, Fariha Tasmin, & Islam, Mohammad Mahfuzul. (2016). *A new cost-effective approach for battlefield surveillance in wireless sensor networks.* Paper presented at the 2016 International Conference on Networking Systems and Security (NSysS).

Jain, Tapan Kumar, Saini, Davinder Singh, & Bhooshan, Sunil Vidya. (2014). Cluster head selection in a homogeneous wireless sensor network ensuring full connectivity with minimum isolated nodes. *Journal of Sensors, 2014*.

Kandukuri, Somasekhar, Lebreton, Jean, Lorion, Richard, Murad, Nour, & Lan-Sun-Luk, Jean Daniel. (2016). *Energy-efficient data aggregation techniques for*

University of Malaya

exploiting spatio-temporal correlations in wireless sensor networks. Paper presented at the 2016 Wireless Telecommunications Symposium (WTS).

Karl, Holger, & Willig, Andreas. (2007). *Protocols and architectures for wireless sensor networks*: John Wiley & Sons.

Krishnamachari, L, Estrin, Deborah, & Wicker, Stephen. (2002). *The impact of data aggregation in wireless sensor networks.* Paper presented at the Distributed Computing Systems Workshops, 2002. Proceedings. 22nd International Conference on.

Kumar, Dilip, Aseri, Trilok C, & Patel, RB. (2011). EECDA: energy efficient clustering and data aggregation protocol for heterogeneous wireless sensor networks. *International Journal of Computers Communications & Control, 6*(1), 113-124.

Lai, Wei Kuang, Fan, Chung Shuo, & Lin, Lin Yan. (2012). Arranging cluster sizes and transmission ranges for wireless sensor networks. *Information Sciences, 183*(1), 117-131.

Lazarescu, Mihai T. (2017). Wireless Sensor Networks for the Internet of Things: Barriers and Synergies *Components and Services for IoT Platforms* (pp. 155-186): Springer.

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

Lindsey, Stephanie, & Raghavendra, Cauligi S. (2002). *PEGASIS: Power-efficient gathering in sensor information systems.* Paper presented at the Aerospace conference proceedings, 2002. IEEE.

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

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

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.

Luo, Dijun, Zhu, Xiaojun, Wu, Xiaobing, & Chen, Guihai. (2011). *Maximizing lifetime for the shortest path aggregation tree in wireless sensor networks.* Paper presented at the INFOCOM, 2011 Proceedings IEEE.

Madden, Samuel, Franklin, Michael J, Hellerstein, Joseph M, & Hong, Wei. (2002). TAG: A tiny aggregation service for ad-hoc sensor networks. *ACM SIGOPS Operating Systems Review, 36*(SI), 131-146.

Madden, Samuel, Szewczyk, Robert, Franklin, Michael J, & Culler, David. (2002). *Supporting aggregate queries over ad-hoc wireless sensor networks.* Paper

University of Malaya

presented at the Mobile Computing Systems and Applications, 2002. Proceedings Fourth IEEE Workshop on.

Mahajan, Shilpa, & Mahotra, Jyoteesh. (2011). A novel chain based wireless data sensor network (ECBSN) technique. *Int. J. Comput. Sci. Telecommun, 2*, 83-87.

Mahdi, Omar Adil, Abdul Wahab, Ainuddin Wahid, Idris, Idna, Yamani, Mohd, Khan, Suleman, Al-Mayouf, Yusor Rafid Bahar, & Guizani, Nadra. (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, Omar Adil, Abdul Wahab, Ainuddin Wahid, Idris, Mohd Yamani Idna, Abu Znaid, Ammar, Al-Mayouf, Yusor Rafid Bahar, & Khan, Suleman. (2016). WDARS: A Weighted Data Aggregation Routing Strategy with Minimum Link Cost in Event-Driven WSNs. *Journal of Sensors, 2016*, 12. doi: 10.1155/2016/3428730

Malakooti, Behnam, & Thomas, Ivan. (2006). *A distributed composite multiple criteria routing using distance vector.* Paper presented at the Networking, Sensing and Control, 2006. ICNSC'06. Proceedings of the 2006 IEEE International Conference on.

Manjeshwar, Arati, & Agrawal, Dharma P. (2001). *TEEN: ARouting Protocol for Enhanced Efficiency in Wireless Sensor Networks.* Paper presented at the IPDPS.

Manjeshwar, Arati, & Agrawal, Dharma P. (2002). *APTEEN: A Hybrid Protocol for Efficient Routing and Comprehensive Information Retrieval in Wireless Sensor Networks.* Paper presented at the Ipdps.

Mantri, Dnyaneshwar, Prasad, Neeli Rashmi, & Prasad, Ramjee. (2014). Two Tier Cluster Based Data Aggregation (TTCDA) for Efficient Bandwidth Utilization in Wireless Sensor Network. *Wireless personal communications, 75*(4), 2589-2606.

Naeimi, Soroush, Ghafghazi, Hamidreza, Chow, Chee-Onn, & Ishii, Hiroshi. (2012). A survey on the taxonomy of cluster-based routing protocols for homogeneous wireless sensor networks. *Sensors, 12*(6), 7350-7409.

Nakamura, Eduardo Freire, de Oliveira, Horacio ABF, Pontello, Luciana F, & Loureiro, Antonio Alfredo Ferreira. (2006). *On demand role assignment for event-detection in sensor networks.* Paper presented at the 11th IEEE Symposium on Computers and Communications (ISCC'06).

Pan, Liqiang, Gao, Huijun, Gao, Hong, & Liu, Yong. (2014). A spatial correlation based adaptive missing data estimation algorithm in wireless sensor networks. *International Journal of Wireless Information Networks, 21*(4), 280-289.

Pantazis, Nikolaos A, Nikolidakis, Stefanos A, & Vergados, Dimitrios D. (2013). Energy-efficient routing protocols in wireless sensor networks: A survey. *IEEE Communications Surveys & Tutorials, 15*(2), 551-591.

University of Malaya

Park, Junyoung, Lee, Sunggu, & Yoo, Sungjoo. (2015). Time slot assignment for convergecast in wireless sensor networks. *Journal of Parallel and Distributed Computing, 83*, 70-82.

Patel, Rutusha, & Kanawade, Shailaja. (2017). *Deployment of Data Aggregation Technique in Wireless Sensor Network.* Paper presented at the Proceedings of the International Conference on Data Engineering and Communication Technology.

Pukhrambam, Priyanka, Bhattacharjee, Sanghita, & Das, Himanish Shekhar. (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.

Raghunathan, Vijay, Schurgers, Curt, Park, Sung, & Srivastava, Mani B. (2002). Energy-aware wireless microsensor networks. *Signal Processing Magazine, IEEE, 19*(2), 40-50.

Rajagopalan, Ramesh, & Varshney, Pramod K. (2006). Data aggregation techniques in sensor networks: A survey.

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

Rawat, Priyanka, Singh, Kamal Deep, Chaouchi, Hakima, & Bonnin, Jean Marie. (2014). Wireless sensor networks: a survey on recent developments and potential synergies. *The Journal of supercomputing, 68*(1), 1-48.

Sadler, Christopher M, & Martonosi, Margaret. (2006). *Data compression algorithms for energy-constrained devices in delay tolerant networks.* Paper presented at the Proceedings of the 4th international conference on Embedded networked sensor systems.

Shah, Ghalib A, & Bozyigit, Muslim. (2007). *Exploiting energy-aware spatial correlation in wireless sensor networks.* Paper presented at the 2007 2nd International Conference on Communication Systems Software and Middleware.

Shan, Mengfan, Chen, Guihai, Luo, Dijun, Zhu, Xiaojun, & Wu, Xiaobing. (2014). Building maximum lifetime shortest path data aggregation trees in wireless sensor networks. *ACM Transactions on Sensor Networks (TOSN), 11*(1), 11.

Shen, Chien-Chung, Srisathapornphat, Chavalit, & Jaikaeo, Chaiporn. (2001). Sensor information networking architecture and applications. *IEEE Personal communications, 8*(4), 52-59.

Shrivastava, Nisheeth, Buragohain, Chiranjeeb, Agrawal, Divyakant, & Suri, Subhash. (2004). *Medians and beyond: new aggregation techniques for sensor networks.* Paper presented at the Proceedings of the 2nd international conference on Embedded networked sensor systems.

University of Malaya

Simon, Santhosh, & Jacob, K Paulose. (2013). HEAP: hybrid energy-efficient aggregation protocol for large scale wireless sensor networks. *International Journal Of Computers & Technology, 4*(2c2), 713-721.

Singh, Santar Pal, & Sharma, SC. (2015). A survey on cluster based routing protocols in wireless sensor networks. *Procedia computer science, 45*, 687-695.

Sohraby, Kazem, Minoli, Daniel, & Znati, Taieb. (2007). *Wireless sensor networks: technology, protocols, and applications*: John Wiley & Sons.

Srisooksai, Tossaporn, Keamarungsi, Kamol, Lamsrichan, Poonlap, & Araki, Kiyomichi. (2012). Practical data compression in wireless sensor networks: A survey. *Journal of Network and Computer Applications, 35*(1), 37-59.

Suriyachai, Petcharat, Roedig, Utz, & Scott, Andrew. (2012). A survey of MAC protocols for mission-critical applications in wireless sensor networks. *IEEE Communications Surveys & Tutorials, 14*(2), 240-264.

Tang, Feilong, You, Ilsun, Guo, Song, Guo, Minyi, & Ma, Yonggong. (2012). A chain-cluster based routing algorithm for wireless sensor networks. *journal of intelligent manufacturing, 23*(4), 1305-1313.

Tsai, Ming-Hui, & Huang, Yueh-Min. (2014). A sub-clustering algorithm based on spatial data correlation for energy conservation in wireless sensor networks. *Sensors, 14*(11), 21858-21871.

Villas, Leandro A, Boukerche, Azzedine, De Oliveira, Horacio ABF, De Araujo, Regina B, & Loureiro, Antonio AF. (2014). A spatial correlation aware algorithm to perform efficient data collection in wireless sensor networks. *Ad Hoc Networks, 12*, 69-85.

Villas, Leandro A, Boukerche, Azzedine, Guidoni, Daniel L, De Oliveira, Horacio ABF, De Araujo, Regina Borges, & Loureiro, Antonio AF. (2013). An energy-aware spatio-temporal correlation mechanism to perform efficient data collection in wireless sensor networks. *Computer Communications, 36*(9), 1054-1066.

Villas, Leandro Aparecido, Boukerche, Azzedine, Ramos, Heitor Soares, de Oliveira, Horacio AB Fernandes, de Araujo, Regina Borges, & Loureiro, Antonio Alfredo Ferreira. (2013). DRINA: a lightweight and reliable routing approach for in-network aggregation in wireless sensor networks. *IEEE Transactions on Computers, 62*(4), 676-689.

Villas, Leandro, Boukerche, Azzedine, Ramos, Heitor S, de Oliveira, Horacio ABF, de Araujo, Regina Borges, & Loureiro, Antonio. (2013). DRINA: a lightweight and reliable routing approach for in-network aggregation in wireless sensor networks. *Computers, IEEE Transactions on, 62*(4), 676-689.

Wan, Shaohua, Zhang, Yudong, & Chen, Jia. (2016). On the Construction of Data Aggregation Tree With Maximizing Lifetime in Large-Scale Wireless Sensor Networks. *IEEE Sensors Journal, 16*(20), 7433-7440.

University of Malaya

Wang, Neng-Chung, Chiang, Yung-Kuei, Hsieh, Chih-Hung, & Chen, Young-Long. (2013). Grid-Based Data Aggregation for Wireless Sensor Networks. *J. Adv. Comput. Netw, 1*.

Wei, Guiyi, Ling, Yun, Guo, Binfeng, Xiao, Bin, & Vasilakos, Athanasios V. (2011). Prediction-based data aggregation in wireless sensor networks: Combining grey model and Kalman Filter. *Computer Communications, 34*(6), 793-802.

Wheeler, Andrew. (2007). Commercial applications of wireless sensor networks using ZigBee. *IEEE Communications Magazine, 45*(4), 70-77.

Xiangning, Fan, & Yulin, Song. (2007). *Improvement on LEACH protocol of wireless sensor network.* Paper presented at the Sensor Technologies and Applications, 2007. SensorComm 2007. International Conference on.

Xiaoyan, Mo. (2006). *Study and design on cluster routing protocols of wireless sensor networks.* Ph. D. Dissertation. Zhejiang University, Hangzhou, China.

Xiong, Zixiang, Liveris, Angelos D, & Cheng, Samuel. (2004). Distributed source coding for sensor networks. *IEEE Signal Processing Magazine, 21*(5), 80-94.

Xu, Xi, Ansari, Rashid, Khokhar, Ashfaq, & Vasilakos, Athanasios V. (2015). Hierarchical data aggregation using compressive sensing (HDACS) in WSNs. *ACM Transactions on Sensor Networks (TOSN), 11*(3), 45.

Yadav, Sarika, & Yadav, Rama Shankar. (2016). A review on energy efficient protocols in wireless sensor networks. *Wireless Networks, 22*(1), 335-350.

Yassein, M Bani, 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.

Yoon, Sunhee, & Shahabi, Cyrus. (2007). The Clustered AGgregation (CAG) technique leveraging spatial and temporal correlations in wireless sensor networks. *ACM Transactions on Sensor Networks (TOSN), 3*(1), 3.

Younis, Mohamed, Senturk, Izzet F, Akkaya, Kemal, Lee, Sookyoung, & Senel, Fatih. (2014). Topology management techniques for tolerating node failures in wireless sensor networks: A survey. *Computer Networks, 58*, 254-283.

Younis, Mohamed, Youssef, Moustafa, & Arisha, Khaled. (2003). Energy-aware management for cluster-based sensor networks. *Computer networks, 43*(5), 649-668.

Younis, Ossama, & Fahmy, Sonia. (2004). HEED: a hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks. *IEEE Transactions on mobile computing, 3*(4), 366-379.

Younis, Ossama, Krunz, Marwan, & Ramasubramanian, Srinivasan. (2006). Node clustering in wireless sensor networks: recent developments and deployment challenges. *IEEE network, 20*(3), 20-25.

University of Malaya

Yousefi, Hamed, Yeganeh, Mohammad Hossein, Alinaghipour, Naser, & Movaghar, Ali. (2012). Structure-free real-time data aggregation in wireless sensor networks. *Computer Communications, 35*(9), 1132-1140.

Zhang, De-Gan. (2012). A new approach and system for attentive mobile learning based on seamless migration. *Applied Intelligence, 36*(1), 75-89.

Zheng, Haifeng, Yang, Feng, Tian, Xiaohua, Gan, Xiaoying, Wang, Xinbing, & Xiao, Shilin. (2015). Data gathering with compressive sensing in wireless sensor networks: a random walk based approach. *IEEE Transactions on Parallel and Distributed Systems, 26*(1), 35-44.

Zhu, Xiaorong, Shen, Lianfeng, & Yum, Tak-Shing Peter. (2009). Hausdorff clustering and minimum energy routing for wireless sensor networks. *IEEE transactions on vehicular technology, 58*(2), 990-997.

Zungeru, Adamu Murtala, Ang, Li-Minn, & Seng, Kah Phooi. (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.

University of Malaya

LIST OF PUBLICATIONS AND PAPERS PRESENTED

Articles on Research Topic

1. **Omar Adil Mahdi**, Ainuddin Wahid Abdul Wahab, Mohd Yamani Idna Idris, Ammar Abu Znaid, Yusor Rafid Bahar Al-Mayouf, and Suleman Khan, "WDARS: A Weighted Data Aggregation Routing Strategy with Minimum Link Cost in Event-Driven WSNs," *Journal of Sensors*, vol. 2016, Article ID 3428730, 12 pages, 2016. (*ISI-Cited Publication*). doi:10.1155/2016/3428730.
2. **Omar Adil Mahdi**, Ainuddin Wahid Abdul Wahab, Mohd. Yamani Idna Idris, Ammar M. A. Abu znaid, Suleman Khan, Yusor Rafid Bahar Al-Mayouf, and Nadra Guizani. (2016). A comparison study on node clustering techniques used in target tracking WSNs for efficient data aggregation. *Wirel. Commun. Mob. Comput.*, 16: 2663–2676. (*ISI-Cited Publication*). doi: 10.1002/wcm.2715.

Conference Proceedings on Research Topic

1. **Omar Adil Mahdi** ; Ainuddin Wahid Abdul Wahab ; Mohd Yamani Idna Idris ; Ammar Abu Znaid ; Suleman Khan, and Yusor Rafid Bahar Al-Mayouf. " ESAM: Endocrine inspired Sensor Activation Mechanism for multi-target tracking in WSNs ", Proc. SPIE 9902, *Fourth International Conference on Wireless and Optical Communications*, 99020B (October 7, 2016); doi:10.1117/12.2262089; http://dx.doi.org/10.1117/12.2262089

Articles in Collaboration with Group Members

1. Ammar M. A. Abu, Mohd. Yamani Idna Idris, Ainuddin Wahid Abdul Wahab, Liana Khamis Qabajeh, and **Omar Adil Mahdi**. (2016). Low communication cost (LCC) scheme for localizing mobile wireless sensor networks. *Wireless Networks*, 1-11. (*ISI-Cited Publication*) doi: 10.1007/s11276-015-1187-6

University of Malaya

2. Suleman Khan, Abdullah Gani, Ainuddin Wahid Abdul Wahab, Salman Iqbal, Ahmed Abdelaziz, **Omar Adil Mahdi**, and *et al.*, "Towards an Applicability of Current Network Forensics for Cloud Networks: A SWOT Analysis," in *IEEE Access*, vol. 4, no., pp. 9800-9820, 2016. (*ISI-Cited Publication*) doi:10.1109/ACCESS.2016.2631543

3. Yusor Rafid Bahar Al-Mayouf, Nor Fadzilah Abdullah, Mahamod Ismail, Salih M Al-Qaraawi, **Omar Adil Mahdi**, Suleman Khan. (2016). Evaluation of efficient vehicular ad hoc networks based on a maximum distance routing algorithm. *EURASIP Journal on Wireless Communications and Networking, 2016*(1), 265. (*ISI-Cited Publication*) doi: 10.1186/s13638-016-0760-8

4. Yusor Rafid Bahar Al-Mayouf, Mahamod Ismail, Nor Fadzilah Abdullah, Ainuddin Wahid Abdul Wahab, **Omar Adil Mahdi**, Suleman Khan, Kim-Kwang Raymond Choo. (2016) Efficient and Stable Routing Algorithm Based on User Mobility and Node Density in Urban Vehicular Network. *PLoS ONE* 11(11): e0165966. (*ISI-Cited Publication*) doi:10.1371/journal.pone.0165966

Submitted Article on Research Topic:

1. WST-RS: Weight Aware Spatial-Temporal Correlation Routing Strategy for Efficient Data Collection and Aggregation in WSNs. Ad Hoc Networks, *(ISI journal, IF: 3.178 , Q1, Under review).*

Seminars

1. Postgraduate Research Excellence Symposium (PGRes) Held in Faculty of Computer Science and Information Technology, Universiti Malaya. Malaysia. May, 2014.

University of Malaya

2. Postgraduate Research Excellence Symposium (PGRes) Held in Faculty of Computer Science and Information Technology, Universiti Malaya. Malaysia. June, 2015.

3. Postgraduate Research Excellence Symposium (PGRes) Held in Faculty of Computer Science and Information Technology, Universiti Malaya. Malaysia. June, 2016.

University of Malaya