



**HAL**  
open science

# Spatio-Temporal Adaptive Sampling Techniques for Energy Conservation in Wireless Sensor Networks

Somasekhar Kandukuri

► **To cite this version:**

Somasekhar Kandukuri. Spatio-Temporal Adaptive Sampling Techniques for Energy Conservation in Wireless Sensor Networks. Performance [cs.PF]. Université de la Réunion, 2016. English. NNT : 2016LARE0021 . tel-01481346

**HAL Id: tel-01481346**

**<https://theses.hal.science/tel-01481346>**

Submitted on 2 Mar 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Doctoral Thesis in  
Computer Networks & Signal Processing



# **Spatio-Temporal Adaptive Sampling Techniques for Energy Conservation in Wireless Sensor Networks**

Somasekhar Reddy Kandukuri

Supervisor: Professor Jean Daniel Lan-Sun-Luk

Co-Supervisors: Dr. Nour Mohammad Murad and Dr. Richard Lorion

Laboratory of Energy, Electronics and Process ( $LE^2P$ )

Thesis Defended on October 7th, 2016

Jury-Members: Prof. Abdelhamid Mellouk, President and  
Reviewer

Prof. Jean Marc Thiriet, Reviewer

Dr. Denis Genon-Catalot, Examiner

Prof. Jean Daniel Lan-Sun-Luk, Supervisor

Dr. Richard Lorion, Co-supervisor

Dr. Nour Mohammad Murad, Co-supervisor

*”Dedicated To my Parents”*

*” We live on an island surrounded by a sea of ignorance. As our island of knowledge grows, so does the shore of our ignorance”. - John Archibald Wheeler*

## Abstract

Wireless sensor networks (WSNs) technology have been demonstrated to be a useful measurement system for numerous both indoor and outdoor applications. There is a vast amount of applications that are operating with WSN technology, such as environmental monitoring, for forest fire detection, weather forecasting, water supplies, air pollution monitoring, natural resource protection, etc,.. The further application domains of elderly care, industrial, structural buildings monitoring are also more active application areas in WSNs. The independence nature of WSNs from the existing infrastructure. Virtually, the WSNs can be deployed in any sort of location, and provide the sensor samples accordingly in both time and space. On the contrast, the manual deployments can only be achievable at a high cost-effective nature and involve significant work. In real-world applications, the operation of wireless sensor networks can only be maintained, if certain challenges are overcome. The lifetime limitation of the distributed sensor nodes is amongst these challenges, in order to achieve the energy optimization. The propositions to the solution of these challenges have been an objective of this thesis.

The analysis of the contributions in wireless sensor networks, provides to the application domain of periodic and delay-tolerant applications. The restrictions, which are the limited lifetime resources of devices, these issues are identified at the system level of sensor nodes. In general, the lifetime of a sensor node is highly affected on the lifetime of the total sensor network. It has been addressed with respect to the energy efficiency of nodes, as well as the utilization of energy-efficient approaches, which have been specifically used to design for the nodes module operability. With respect to the energy efficiency, an analysis of the literature study has been performed on the contributions to the energy consumption modules of the node, thus, leads to the desire to minimize node redundant data transmissions in both time and space. The energy-efficient routing and data reduction techniques are presented, which features energy efficiency and nodes lifetime as a key attribute by utilizing the drawbacks of literature studies in both routing and data reduction.

Moreover, a discussion for the usage of single-hop sensor networks based on topological, data-centric, hierarchical and location-based is made. Among these methods, a hierarchical cluster-based network model while following the data-centric approach with its focus on low communication overhead is presented. However, the communication link and the signal strength between any sensor devices must be good. While there is a great amount of research being conducted in routing, and the adaptation and integration of existed routing techniques helps to reduce the time, in order not to develop their own topology for maintaining the network operations. However, it is not the case in data driven. As the data redundancy in sensor nodes cannot be negligible, thus, affecting the nodes lifetime. In order to exploit the node redundancies and suppress the redundant data transmissions at individual node level. A new data pre-filtration method has been introduced with its focus on reducing the signal bandwidth as well as the redundant transmissions, as a basis for periodic and delay-tolerant

operations. While the contributions are related to each other, in order to follow and achieve the reduction of redundant data transmissions, thus, increases the lifetime of node as well as the network.

In summary, the contributions which have been presented in this thesis, address the system lifetime, exploitation of redundant and correlated data messages, and then the sensor node in terms of usability. The considerations have led to the simple data redundancy and correlated algorithms based on hierarchical based clustering, yet efficient to tolerate both the spatio-temporal redundancies and their correlations. Furthermore, a multihop sensor network for the implementation of propositions with more features, both the analytical proofs and at the software level, have been proposed.

## Résumé

La technologie des réseaux de capteurs sans fil (WSN) démontre qu'elle peut être très utile dans de nombreuses applications. Ainsi chaque jour voit émerger de nouvelles réalisations dans la surveillance de notre environnement comme la détection des feux de forêt, l'approvisionnement en eau, la surveillance de la pollution de l'air... Les champs d'applications couvrent aussi des domaines émergents et sensibles pour la population avec les soins aux personnes âgées ou les patients récemment opérés dans le cadre de l'hospitalisation à domicile. L'indépendance des architectures RCSFs par rapport aux infrastructures existantes permet aux RCSFs d'être déployées dans presque tous les sites afin de fournir des informations temporelles et spatiales. Dans les déploiements opérationnels le bon fonctionnement de l'architecture des réseaux de capteurs sans fil ne peut être garanti que si certains défis sont surmontés. La minisisation de l'énergie consommée en fait partie. La limitation de la durée de vie des noeuds de capteurs est fortement couplée à l'autonomie de la batterie et donc à l'optimisation énergétique des noeuds du réseau. Nous présenterons plusieurs propositions à ces problèmes dans le cadre de cette thèse.

Les contributions récentes dans les réseaux de capteurs sans fil présentent des travaux dans le domaine des applications de mesures périodiques ou tolérantes aux délais. Les restrictions, liées à la durée de vie limitée des dispositifs, sont identifiées au niveau des fonctions des noeuds du réseau. La durée de vie d'un noeud du réseau est aussi fortement affectée par l'architecture et la durée de vie globale du réseau de capteurs. La problématique énergétique a été traitée localement avec l'amélioration du rendement énergétique des noeuds, ainsi que par l'utilisation des approches économie en énergie lors de la conception et de fonctionnement du module. En ce qui concerne l'efficacité énergétique globale, une étude bibliographique a été effectuée et conduit à la minimisation des échanges et à la suppression des transmissions redondantes à la fois temporelles et spatiales. L'amélioration des techniques de routage et de réduction de données est présentée comme un élément clé de l'optimisation de l'efficacité énergétique des réseaux de capteurs.

Ainsi, un état de l'art concernant l'utilisation des réseaux de capteurs hiérarchiques géo-localisés à 2 couches et centré sur les données a été réalisé. Parmi ces méthodes, le modèle de réseau à base de clusters hiérarchiques centré sur les données mettant l'accent sur la réduction des coûts de transmission a été étudié. Il prend en considération la qualité de la communication entre les noeuds. Bien qu'il y ait un nombre important de recherches menées dans le routage, l'amélioration des protocoles de routage est abordée dans un premier temps. Il permet d'optimiser les temps de transmissions et de garantir le fonctionnement du réseau. Il n'en est pas de même dans le cas des réseaux commandés par les données. Comme la redondance des données dans les noeuds de capteurs peut ne pas être négligeable, affectant ainsi la durée de vie des noeuds, il est nécessaire d'analyser et supprimer les redondances de données au niveau du noeud du réseau. C'est pourquoi, dans une deuxième partie, un procédé de préfiltrage des données a été introduit en mettant l'accent sur la réduction de la bande

passante ainsi que la suppression des transmissions redondantes dans les réseaux de capteurs périodiques et tolérants aux délais. Les deux propositions contribuent à la réduction des transmissions de données redondantes, ce qui augmente la durée de vie du noeud, ainsi que du réseau global.

En résumé, les contributions qui ont été présentées dans cette thèse, abordent la durée de vie globale du réseau, l'exploitation des messages de données redondantes et corrélées et enfin le fonctionnement noeud lui même. Les travaux ont conduit à la réalisation d'algorithmes de routage hiérarchiques et de filtrage permettant la suppression des redondances. Ils s'appuient sur les corrélations spatio-temporelles des données mesurées. Enfin, une implémentation de ce réseau de capteurs multi-sauts intégrant ces nouvelles fonctionnalités est proposée.

## Acknowledgements

First and foremost, i would like to thank my supervisor Professor Jean Daniel Lan-Sun-Luk, who has given me the continuous support and encouragement during throughout the thesis. I am very grateful for your guidance, and trust, but also the impact, which driven me to complete the thesis that the way it did.

I would like to express my sincere gratitude to my co-supervisor Nour Mohammad Murad, who has given me the opportunity to conduct this thesis work and who has supported me during the three years of this research. I am also thankful for your guidance and encouragement throughout the thesis work.

I am very thankful to my co-supervisor Richard Lorion, who has given me the immense confidence to continue this work, who has supported and given me the valuable and technical discussions over the years. I am indebted for your support and comments, but also the freedom, which made my work develop that the way it presented.

I am extremely thankful to Denis Genon-Catalot, who has given me the opportunity to conduct this research work in LCIS lab, Valence. I am very grateful for your guidance, encouragement and freedom during the period, but also, who has been involved in both annual CSTs and the final thesis review. In particular, I like to thank my jury members Professor Jean Marc Thiriet and Professor Abdelhamid Mellouk, who has taken the time to review this work.

I am also thankful to my lab director Jean Pierre Chabriat, who has given me the continuous support and for providing the necessary resources. I thank Frédéric Alicalapa, who has given me the encouragement and support during initial days of the thesis. I also like to thank Patrick Jeanty, and Kelly Grondin-D'anna, who has supported me in the administrative parts throughout the thesis.

A special thanks to Mickaël Lebreton, who is one of the co-authors of several papers, who has involved in the great technical discussions over the years. I am very glad to have worked together with you and also for the amount of papers that we manage to present. I also thank Laurent Chane Kuang Sang, Eric, Sebastian, and Kevin from IUT-RT department for creating a great fun working environment.

I like to thank all my colleagues at the department of Physics and RT from both LE<sup>2</sup>P and LCIS laboratories. In particular, Gregory, Medhi and Youness from LCIS lab, and Max La McNace, Li Peng, Li Qi, Jérôme, Chao Tang and Farid from LE<sup>2</sup>P lab for having a valuable talks and for the interesting discussions. I also like to thank Avinash, Mekala, Mahesh and Sayuj from other laboratories, with whom i used to hang out sometimes and had a great fun times. I want to express my sincere gratitude to every close friend of mine, who supports me and there for me directly and indirectly during the every greatest hour of need.

Finally, and most importantly, a special thanks to my parents, my brother, and my wife for balancing my life. I am grateful to each and every one of you for being there, and that for your consistence support.

## Acronyms

<b>ACK</b>	Acknowledgment
<b>ADC</b>	Analog-to-Digital Converter
<b>AR</b>	Auto-Regressive
<b>ARMA</b>	Auto-Regressive Moving Average
<b>CS</b>	Compression-based Sensing
<b>CSMA</b>	Carrier Sense Multiple Access
<b>CTS</b>	Clear-to-Send
<b>DAWF</b>	Data Aggregative Window Function
<b>DCT</b>	Discrete Fourier Transform
<b>DGM</b>	Direct Graph Model
<b>DPM</b>	Dynamic Power Management
<b>DPM</b>	Dynamic Probabilistic Model
<b>DSC</b>	Distributed Source Coding
<b>DSP</b>	Digital Signal Processor
<b>FIFO</b>	First In First Out
<b>FPGA</b>	Field Programmable Gate Array
<b>FT</b>	Fourier Transform
<b>GPRS</b>	General Packet Radio Service
<b>GSM</b>	Global System for Mobile Communications
<b>GUI</b>	Graphical User Interface
<b>IEEE</b>	Institute of Electrical and Electronic Engineers
<b>LGM</b>	Linear Graph Model
<b>LLC</b>	Logical Link Control
<b>LPM</b>	Low Power Module

<b>LQI</b>	Link Quality Indicator
<b>MA</b>	Moving Average
<b>MAC</b>	Medium Access Control
<b>MAE</b>	Mean Absolute Error
<b>MCU</b>	Micro Controller Unit
<b>MRM</b>	Multi-path Ray Model
<b>MTU</b>	Maximum Transmission Unit
<b>NCDM</b>	Non-Correlated Data Message
<b>NRDM</b>	Non-Redundant Data Message
<b>RAM</b>	Random Access Memory
<b>RF</b>	Radio Frequency
<b>RSSI</b>	Received Signal Strength Indicator
<b>RTS</b>	Request-to-Send
<b>RV</b>	Relative Variation
<b>Runicast</b>	Reliable Unicast
<b>RX</b>	Receiver
<b>SDR</b>	Spatial Data Redundancies
<b>SR</b>	Spatial Redundancies
<b>TDMA</b>	Time Division Multiple Access
<b>tct</b>	Threshold Coherency Tolerance
<b>TDR</b>	Temporal Data Redundancies
<b>TR</b>	Temporal Redundancies
<b>UDGM</b>	Unit Disk Graph Model
<b>WSN</b>	Wireless Sensor Network

## **Publications**

1. Somasekhar Kandukuri, Jean Lebreton, Denis Genon-Catalot, Nour Murad, and Richard Lorion, "Energy-Efficient Data Window Aggregation Techniques for Energy Saving in Wireless Sensor Networks", in proceedings of the 21st IEEE Symposium on Computers and Communications (ISCC 2016), pp. 1-6, IEEE, June 2016.
2. Somasekhar Kandukuri, Jean Lebreton, Nour Murad, Richard Lorion, and Jean-Daniel Lan-Sun-Luk, "Energy-Efficient Cluster-Based Protocol using An Adaptive Data Aggregative Window Function (A-DAWF) for Wireless Sensor Networks", in Proceedings of the 17th IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM 2016), pp. 1-4, IEEE, June 2016. (Acceptance rate for full papers 21%)
3. Somasekhar Kandukuri, Jean Lebreton, Richard Lorion, Nour Murad, and Jean-Daniel Lan-Sun-Luk, "Energy-Efficient Data Aggregation Techniques for Exploiting Spatio-Temporal Correlations in Wireless Sensor Networks", in proceedings of the 15th Annual IEEE Wireless Telecommunications Symposium (WTS 2016), pp. 1-6, IEEE, April 2016.
4. Somasekhar Kandukuri, Nour Murad, and Richard Lorion, "Cluster-Head Techniques for Single-hop Routing protocol in Energy-Efficient Wireless Sensor Networks", IoP Conference Series, pp. 1-4, February 2016.
5. Somasekhar Kandukuri, Nour Murad, and Richard Lorion, "A single-hop clustering and energy efficient protocol for wireless sensor networks", IEEE Radio and Antenna Days of the Indian Ocean (RADIO), pp. 1-2, Sept 2015.

### **Other Related Papers**

The following papers are not directly related to this dissertation, yet the contained information that have indirectly influenced in this thesis.

1. Jean Mickael Lebreton, Somasekhar Kandukuri, Nour Murad, and Richard Lorion, "Interference Evaluation of Wi-Fi Devices over Wake-Up Radio in wireless sensor networks", IEEE Radio and Antenna Days of the Indian Ocean (RADIO), pp. 1-2, Oct 2016.
2. Jean Mickael Lebreton, Somasekhar Kandukuri, Nour Murad, and Richard Lorion, "An energy-efficient duty-cycled wake-up radio protocol for avoiding overhearing in wireless sensor networks", in Wireless Sensor Network publication, pp. 1-16, July 2016.

3. Jean Lebreton, Nour Murad, Somasekhar Kandukuri, and Richard Lorion, "An Energy-Efficient Addressing Mechanism for Wake-up Radio Systems", in proceedings of the 15th Annual Wireless Telecommunication Symposium (WTS 2016), IEEE, pp. 1-6, April 2016.

# Table Of Contents

<b>Abstract</b>	<b>i</b>
Résumé	iii
Acknowledgements	v
<b>Acronyms</b>	<b>vi</b>
<b>Publications</b>	<b>viii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Predictive Monitoring . . . . .	2
1.2 Energy Efficiency . . . . .	2
1.2.1 Routing Protocols . . . . .	3
1.2.2 Data-Driven Approaches . . . . .	4
1.2.3 Duty-Cycling Approaches . . . . .	5
1.3 Wireless Sensor Networks Topology . . . . .	5
1.3.1 Single-hop WSNs . . . . .	5
1.3.2 Multi-hop WSNs . . . . .	6
1.4 Problem Formulation and Efficiency . . . . .	6
1.5 Thesis Contributions and Outline . . . . .	11
<b>2 Energy Efficiency in Wireless Sensor Networks</b>	<b>13</b>
2.1 Wireless Sensor Network Lifetime . . . . .	13
2.2 Hardware Platform Choices in Sensor Networks . . . . .	14
2.3 Energy Efficiency over WSNs . . . . .	17
2.4 Routing and Data-Driven . . . . .	18
2.5 Data Reduction Techniques for Energy-Efficiency in WSNs . . . . .	18
2.5.1 Accuracy in Restored Data . . . . .	20
2.5.2 Source Coding and Network Coding . . . . .	21
2.5.3 Discrete Cosine and Fourier Transformation in Data Compression . . . . .	21
2.6 Necessity of Data Reduction in Distributed Sensor Networks . . . . .	22
2.7 Compression-Based Sensing in Distributed WSNs . . . . .	23
2.7.1 In-network Processing . . . . .	24

2.7.2	Distributed Source Coding . . . . .	25
2.7.3	Compressive Sensing (CS) Theory . . . . .	25
2.8	Data Prediction in Distributed WSNs . . . . .	26
2.9	Energy-Efficient Data Acquisition in WSNs . . . . .	28
2.10	Summary . . . . .	30
<b>3</b>	<b>Single-hop WSNs: Energy-Efficient Routing Approaches</b>	<b>31</b>
3.1	Introduction . . . . .	31
3.2	Motivation . . . . .	33
3.3	Background and Related Works . . . . .	33
3.4	Radio Energy Consumption Model . . . . .	44
3.5	SCEEP Algorithm Designs and Specifications . . . . .	45
3.6	Simulation Results . . . . .	50
3.7	Background Study of COOJA/Contiki . . . . .	53
3.7.1	Rime Protocol Stack . . . . .	54
3.7.2	Cross-level Network in COOJA . . . . .	54
3.8	Design Decisions and Specifications in COOJA . . . . .	55
3.9	Summary . . . . .	56
<b>4</b>	<b>Single-hop WSNs: A Data Window Aggregation Techniques</b>	<b>57</b>
4.1	Introduction . . . . .	57
4.2	Related Work . . . . .	59
4.3	System Model Design and Specifications . . . . .	61
4.3.1	Preliminaries . . . . .	64
4.3.2	Learning Phase . . . . .	65
4.3.3	Adaptive Phase . . . . .	65
4.4	Algorithmic Approach for Spatio-Temporal Redundancies . . . . .	68
4.5	Design Implementations in COOJA . . . . .	70
4.5.1	Design Decisions of DAWF Integration in COOJA Simulator . . . . .	70
4.5.2	Adaptive Compression and Prediction-Based Mechanism . . . . .	72
4.6	Experimental Study Case and Results . . . . .	73
4.6.1	Temporal Performance Analysis . . . . .	74
4.6.2	Experimental Study Case Based on Intel Data sets . . . . .	74
4.6.3	Experimental Study Case Based on LCIS DREAL Data sets . . . . .	82
4.6.4	Spatial Performance Analysis . . . . .	91
4.6.5	Signal Reconstruction . . . . .	96
4.7	DAWF Temporal Evaluations in COOJA Cross-level Rime Network . . . . .	100

4.8	Summary . . . . .	104
<b>5</b>	<b>Conclusions and Perspectives</b>	<b>105</b>
5.1	Contributions from this thesis . . . . .	107
5.2	Perspectives . . . . .	109
	<b>Bibliography</b>	<b>111</b>
<b>A</b>	<b>Multi-hop WSNs: A Multi-tier Heterogeneous Cluster-based Approach</b>	<b>123</b>
A.1	Hybrid Energy-Efficient and Throughput Protocol for Multi-Hop WSNs	123
A.1.1	Algorithm 1 . . . . .	127
A.1.2	Algorithm 2 . . . . .	128
A.2	Discussion . . . . .	132
A.3	Summary . . . . .	132
<b>B</b>	<b>DAWF Experimental Study Case in COOJA Interface</b>	<b>135</b>
B.1	DAWF Propositions Based on Real-World Datasets . . . . .	135

## List Of Figures

1.1	Energy conservation methods that are related to this thesis . . . . .	3
1.2	A List of Data-driven approaches that are involved in this dissertation	4
1.3	A typical architecture of Single-hop Wireless Sensor Network. . . . .	7
1.4	A typical architecture of Multi-hop Wireless Sensor Network. . . . .	8
2.1	An architecture of typical sensor nodes in Wireless Sensor Networks. Nodes typically possess with the combination of sensing, processing and communication capabilities, and they are powered by energy sources (e.g., battery) . . . . .	15
3.1	First order radio energy consumption model. . . . .	44
3.2	Set-up phase of the network algorithm . . . . .	48
3.3	Steady-state phase of the network algorithm . . . . .	49
3.4	Number of alive nodes over time. . . . .	51
3.5	Total number of data bits received at the CH. . . . .	52
3.6	Number of depleted nodes over time. . . . .	52
3.7	COOJA/Contiki system overview . . . . .	53
3.8	COOJA Rime Network Protocol Stack . . . . .	54
4.1	A cluster-based model diagram of Spatio-Temporal sampling through DAWF algorithm. . . . .	63
4.2	A typical structure of a cluster-based star topology, between sensor nodes and CH. . . . .	64
4.3	Learning Phase in Sensor Nodes . . . . .	66
4.4	Adaptive phase in sensor nodes . . . . .	67
4.5	Structure of established Rime cross-level network stack . . . . .	71
4.6	Rime Packet header fields with DAWF $w_M$ payload message size. . .	71
4.7	Sensor Nodes Deployment Map of Intel Berkeley Research Lab [105]: the highlighted area has chosen for our experiments. . . . .	75
4.8	DAWF with a fixed $w_M$ at various thresholds of $Th$ . . . . .	76
4.9	DAWF with a fixed $Th$ at various window sizes $w_M$ . . . . .	76
4.10	Number of window data transmissions (TX) over the varied window sizes of time. . . . .	78
4.11	Total number of reduced transmissions (%) over the windows. . . . .	79
4.12	Mean Absolute Error (MAE) cost over the window sizes. . . . .	79
4.13	CH A-DAWF among nodes spatial data with different thresholds. . . .	80

4.14	Energy cost comparative analysis between TiNA protocol with $tct$ , and DAWF at the window size of 10 . . . . .	81
4.15	DREAL Deployment of Indoor Wireless Sensor Network. . . . .	83
4.16	Temperature datasets through DAWF performances with various thresholds $Ths$ over different $w_N$ sizes of temporal data. . . . .	83
4.17	Humidity datasets through DAWF performances with various thresholds $Ths$ over different $w_N$ sizes of temporal data. . . . .	84
4.18	Sensor node DAWF performances of solar temperature over various window sizes. . . . .	85
4.19	Temperature: Sensor node temporal reduced transmission metric performances over various thresholds and the various window sizes . . .	85
4.20	Humidity: Sensor node temporal reduced transmission metric performances over various thresholds and the various window sizes . . .	86
4.21	Solar: Sensor node temporal reduced transmission metric performances over various thresholds and the various window sizes . . . . .	87
4.22	Trade-off between DAWF first algorithm and second algorithm . . . .	88
4.23	A comparative analysis of Ambient temperature phenomena between TiNA protocol with relative $tct$ , DAWF and A-DAWF algorithms with the window size of 5 . . . . .	89
4.24	A comparative analysis of Ambient temperature phenomena between TiNA protocol with relative $tct$ , DAWF and A-DAWF algorithms with the window size of 10 . . . . .	89
4.25	$R_{TX}$ comparative analysis of Solar temperature phenomena between TiNA protocol with relative $tct$ , DAWF and A-DAWF algorithms with the window size of 5 . . . . .	90
4.26	$R_{TX}$ comparative analysis of Solar temperature phenomena between TiNA protocol with relative $tct$ , DAWF and A-DAWF algorithms with the window size of 10 . . . . .	91
4.27	Humidity Phenomena: CH node performances over various window sizes. . . . .	92
4.28	Temperature Phenomena: CH node performances of over various the window size. . . . .	93
4.29	Solar Phenomena: CH node performances of over various thresholds through the window sizes. . . . .	93
4.30	CH DAWF $R_{TX}$ metric performances of nodes temperature over various thresholds through the window sizes. . . . .	94
4.31	CH DAWF reduced transmission metric performances of humidity over various thresholds through the window sizes. . . . .	95
4.32	CH DAWF reduced transmission metric performances of solar temperature over various thresholds through the window sizes. . . . .	95

4.33	MAE metric cost of node 1 at Interpolation over various windows and at an absolute threshold. In this scenario, we present the prediction error rates through MAE metric, while performing the interpolation of signal reconstruction for the considered amount of phenomena at various window sizes. . . . .	97
4.34	MAE metric cost of node 1 at Interpolation over various thresholds at an absolute window size. In this scenario, we present the prediction error rates through MAE metric, while performing the interpolation of signal reconstruction for the considered amount of phenomena at various threshold values. . . . .	97
4.35	Original and interpolated signals of ambient temperature from node 1.	99
4.36	Original and interpolated signals of ambient temperature from node 1.	99
4.37	Original and interpolated signals of humidity sensor of node 1. . . . .	100
4.38	Z1 Mote Light Sensor from COOJA. . . . .	101
4.39	Tmote Sky COOJA node energy consumption (milli Joules) over time. While carrying a window dataset of payload data size 10 bytes. . . . .	102
4.40	Influence of other parameters on Radio energy consumption of TmoteSky in COOJA. . . . .	103
4.41	Influence of other parameters on CPU and LPM modes energy consumption of TmoteSky in COOJA. . . . .	104
A.1	Cluster-Head and Member nodes communication. . . . .	126
A.2	Route-leaders, member nodes and non-member nodes Communication.	126
A.3	Network nodes and Route-leaders communication. . . . .	127
A.4	Data transmissions between RL and CH nodes . . . . .	127
A.5	Broadcast and unicast message exchanges between CM and all cluster nodes. . . . .	129
A.6	Data exchange between Current CH via other neighboring CH nodes to sink. . . . .	130
A.7	Transmission Distances for Single-hop and Multi-hops. . . . .	132
B.1	DAWF Sky Platform Unit Disk Graph Model Graphical Interface . . . .	136
B.2	DAWF Sky Platform Multi-Path Ray Model Graphical Interface. . . .	137
B.3	DAWF Z1 Platform Unit Disk Graph Model Graphical Interface. . . .	138
B.4	DAWF Z1 Platform Multi-Path Ray Model Graphical Interface. . . .	139
B.5	A Snippet of Intel Reseach Lab Indoor Datasets. . . . .	140
B.6	A Snippet of DREAL Datasets from LCIS Lab. . . . .	140

## List Of Tables

2.1	Hardware platform specification choices of the nodes in sensor networks.	17
3.1	A list of well-known routing protocols in WSNs. . . . .	38
4.1	DAWF performances at various thresholds and window sizes. . . . .	77
4.2	Energy consumption cost at TX mode of TAG, TiNA and DAWF protocols. . . . .	81
4.3	Module parameters for the implementation of DAWF data reduction algorithm. . . . .	101
A.1	A list of Protocol Stack Considerations. . . . .	130
A.2	A list of Simulation Parameters. . . . .	131
A.3	ContikiMAC Energy rates. . . . .	131

# Chapter 1

## Introduction

Wireless Sensor Networks (WSNs) have been evolved as a new technology of monitoring systems in this modern world. WSN interfaces with small or large amount of sensor devices, which are capable of sensing and process the simple performing tasks by using single-hop or multi-hop communication methods. Unlike the wired networks, wireless sensor networks can be easily deployed even where the sensing is unfathomable to process. In real-world scenarios, the usage of WSN technology emerging itself into the many application areas, such as risk monitoring in habitat, tracking and surveillance, health-care, forest fire-detection, structural, landscape and volcano monitoring.

In many application scenarios, wireless sensor networks proved that it can be easily deployable compared to the wired systems, hence the sensor devices are energy constrained and can have limited battery resources. While some sensor nodes are energy-harvesting devices, but majority of them are still being used as battery operated devices in the network. Since, battery replaceable or rechargeable is difficult to change during its network operation. However, network systems should be considered the significance of energy-efficient mechanisms to maximize the network lifetime.

The size of the network deployment always depends upon the application requirements. The classification of big or small networks with moderate transmission ranges has presented in the well-known ExScal deployment for large scale sensor networks [1], which also proven that ExScal had high connectivity with the help of multihop routing. There are always two typical sensor network systems are active in research, such as single-hop and multi-hop sensor networks.

However, multi-hop network systems are always not necessary for where that small amount of traffic or nodes communication to happen. For instance, Unmanned Aerial Vehicles (UAV) have been studied, such as Flying Gridswarms or the indoor UltraSwarm [2]. These network systems consist of a small amount of vehicles with transmission ranges on the same order as the coverage area, thus the multi-hop com-

---

<sup>1</sup>The remainder of this thesis document, the terms that are used as wireless sensor node, sensor node, node, member node, cluster node will be treated as synonymously, and always refer to the wireless sensor nodes.

munication is not required in these sorts of systems. Nevertheless, such networks can also be optimized for energy efficiency.

In these monitoring systems, the traditional aspect of wireless sensor networks as large, well-connected both single-hop and multi-hop networks of small static or dynamic sensor nodes. Whether the network has less number of nodes or low connectivity during its operational state, thus the energy-efficient strategies for predicting the nodes energy and data statistics for different application areas. One of such is monitoring of risk in both indoor and outdoor environments. Furthermore, WSN can also have identical tasks and functionalities, internally, which can work for both the networks of homogeneous and heterogeneous systems.

## 1.1. Predictive Monitoring

Although, predictive monitoring can evaluate any kind of sensor network environments, such as event or data detection, energy prediction, positions and routing. The foundation of predictive monitoring is the prediction of network events, which enables a better understanding for network observation, in order for conserving the network and node energies, and then the data from general message losses. Predictive monitoring also has a universal feature to be involved in all sorts of WSN application domains, such as environmental monitoring, elderly health-care monitoring, object tracking and detection, volcano monitoring and industrial. Multiple known protocols exist for prediction [3, 4, 5, 6]. However most of them are limited to their specificity for predicting certain parameters like estimated energy consumption or residual energy [7], nodes short-level prediction for data aggregation to be reported to the network, and early node event prediction than commence.

## 1.2. Energy Efficiency

Before the discussion of different sensor network categorizations that we study, this thesis explains an overview of proposed energy efficient strategies that can be developed with these systems.

The energy efficient strategies for sensor network can be categorized as in Fig. 1.1. Adopting the taxonomy used from the literature survey [4], and then modified according to this thesis research involvements. Data-driven approaches involve to reduce the amount of sensor gathered data that to be transmitted over the network. In general use cases, communication module in sensor nodes consumes more energy than the data processing and sensing modules [8]. But the context will be different in sensor nodes if the sensing module contains multiple sensors than one or two. On contrary, routing approaches involve to reduce the communication costs over hops and maintaining the proper reliable links between the hops, in order for finding the short distances between the hops. On the other hand, prediction of nodes event or residual energy ap-

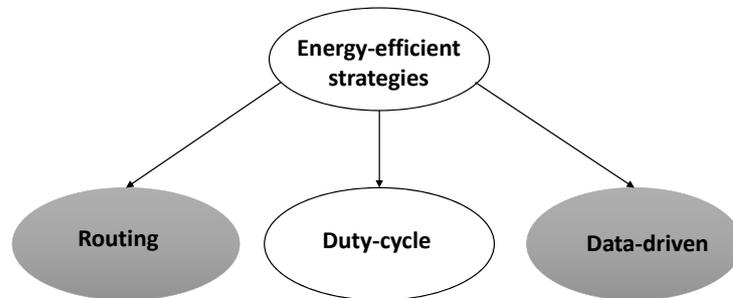


Figure 1.1 – Energy conservation methods that are related to this thesis

proaches involve to predict the depletion of nodes for replacing the nodes or waking up additional sleeper nodes to join in the network, in order to maximize its lifetime. Duty-cycling approaches involve for controlling radio transceiver operational states to be either awake, active, sleep, and turnoff when no communication is needed. Thus, we consider the energy-efficient approaches of mobility to make a possible extension of our future work. As our thesis especially focuses on energy-efficient strategies in data-driven and routing mechanisms rather than duty-cycling approaches.

### 1.2.1. Routing Protocols

The energy-efficient strategies of routing approaches are not only for finding the shortest path metrics to process the data, they can also save the energy, and then reduce the computational load while finding the shortest paths towards the destination, thus helps to maximize the network lifetime. There is numerous routing protocols studied in the literature survey [9, 10], where we further categorized and briefly described various well-known routing protocols in chapter 3 and section 3.3.1.

There has been an extensive research conducted in the routing area of wireless sensor networks, such as energy efficiency in reliable data transmissions, reliability in links, scalability and adaptability in sensor networks and, etc.,. However, there are still many issues which remain unsolved to meet the requirements of different applications. One such, the energy efficiency in routing for generic applications is a major challenge to achieve, because the considerations vary from application to the application.

### 1.2.2. Data-Driven Approaches

The first part of energy efficient strategy is that we consider the data-driven approaches. In a typical sensor network, computational module is a second highest energy consumer after the communication module [11]. The energy efficient strategies of data-approaches can be categorized as shown in Fig. 1.2. The data-driven approaches re-

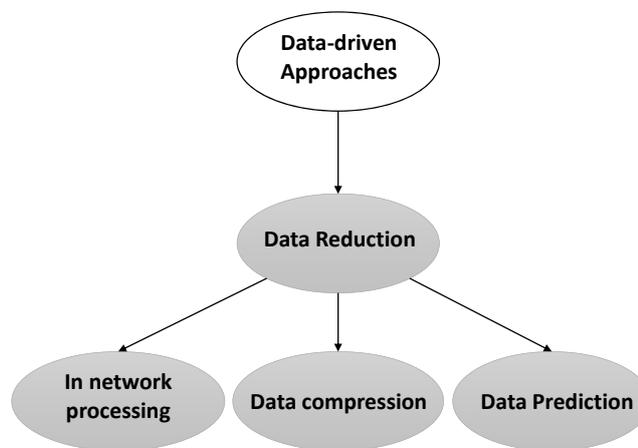


Figure 1.2 – A List of Data-driven approaches that are involved in this dissertation

duce the amount of data samples before processing to the network. They often rely on the predictability of the monitored physical phenomena, which are thus reliably representable through the mathematical models. However, One of the most well-studied data-driven methods is data aggregation, such approaches are particularly relevant to our research work. The basic principle of this idea is to move the computation away from a remote processor and onto the local nodes to the physical phenomena, thus trading off communication for local computation [12]. For instance, many deployed sensing autonomous nodes record and report the unprocessed data to a remote site, as studied in the Volcan Reventador sensor network [13]. While doing some processing, combining, or filtering of the physical phenomena locally on the nodes, and the amount of data is communicated to the remote sites also can be reduced, along with the consumed energy. To achieve a greater degree of in-network processing, data can be routed where more data aggregation can occur, instead of a shortest path metric, and is known as data-centric routing [11].

### 1.2.3. Duty-Cycling Approaches

The energy-efficient strategies of duty cycling is another major energy conservation technique to be considered in this work, however, we have considered to use some existed duty-cycled mechanisms while designing the original framework of this thesis, such as ContikiMAC, X-MAC and LPM. Duty-cycling works with two basic principles such as sleep or wake-up schemes to reduce the radio energy consumption. Duty-cycling has further categorized into two, namely as topology control and power management. In topology control, node redundancy is exploited such that only a select subset of nodes are awake at a time. Node redundancy viewed in terms of either geographic location or network connectivity. Location-driven topology control schemes have been studied in [14, 15], and connectivity-driven topology control schemes are studied in [16, 17].

In power management schemes, nodes distributively implement a sleep/wake-up cycles. Power management can be viewed as a complementary technique to the topology control. It implemented in different layers of network architecture. Some wake-up schemes run top of a medium access control (MAC) protocol [18, 19]. There are some other wake-up schemes which implemented directly into the MAC protocols [20, 21].

## 1.3. Wireless Sensor Networks Topology

In this section, we describe the classification of sensor network models that we study, and for each network model that has been stated the approaches are most applicable to be energy efficient. In order to summarize the differences between type of networks that has been applied towards the propositions. The few network parameters that we consider are varying the number of nodes  $N$ , remaining energy of the nodes  $E_i$  and the transmission range  $r$ , and the total network deployed area. A typical network architecture of the propositions are classified in both the Figures 1.3 and 1.4.

### 1.3.1. Single-hop WSNs

This section describes a single-hop clustering structure, which is targeted for small scale sensor network. In this architecture, nodes have two sorts of transmission ranges, which are termed as short-range for the communication between regular sensor nodes (SNs) and cluster-head (CH) nodes, and long-range for the communication link between CH nodes and sink or base station (BS). Because the larger transmission range helps the single-hop CH nodes can be communicated directly to the network. These single-hop network topologies are commonly assumed in wireless sensor network research area, and the differences of both the single hop and multihop networks has been well presented in [22].

Most of the energy conservation approaches are designed with both duty-cycling

and data-driven approaches for different single-hop WSNs, as shown in Figure 1.3, which are briefly explained in chapter 3. As an example, our work presents for the delay-tolerant WSN applications that works within the framework of an energy-efficient clustering approach in single-hop WSNs. This application is a network energy prediction system that consists of a limited amount of static sensor nodes. The proposition targets to design the approach towards a system employing certain amount of energy-constrained sensor nodes. In this work, we assume a distributed homogeneous computing network architecture, in which the network energy prediction algorithm and data-driven aggregation approach is computed on nodes to process the data at every round of time in seconds.

### **1.3.2. Multi-hop WSNs**

This thesis also presents a typical multi-hop WSN scenario for large scale sensor networks, as shown in Figure 1.4. Due to the network scalability issue in single-hop WSNs. This dissertation further focuses on multi-hop WSNs as well, in order to adapt the propositions to achieve scalability and efficiency. In some cases, nodes at different locations cannot be able to communicate with the network, directly. Instead, intermediate nodes are required to relay data, like forming multi-hop communication routes. These multi-hop topologies commonly used in sensor network research, as described in ExScal deployment [1]. Sensor networks are usually envisioned as collection of small devices with relatively short-range radios, and few of the leader nodes, which can have long-range radios, in order to be well connected with the network as a large.

Most of energy conservation approaches are designed with both short and long range radio sensors for multi-hop WSNs. As an example, the proposition structure of Figure. 1.4 presents for the large-scale WSN applications that works within the framework of an energy conservation approach. In general, continuous monitoring systems require limited node processing capabilities to be communicated with multiple other neighboring nodes. Thus, the proposition challenges to adapt this approach to a system employing many energy-constrained sensor nodes. The energy-efficient strategy handles this challenge most directly through a data-driven in-network processing approach. In this case, we assume a distributed heterogeneous network model, in which the algorithm is computed on nodes local to the detected event. This approach helps to reduce a great amount of communication over-heads as compared to the single-hop WSNs or a centralized system, in which all of the data can be reduced and transmitted by leader nodes to a remote site for processing.

## **1.4. Problem Formulation and Efficiency**

The main contribution of this thesis has been presented as to maximize the wireless sensor network lifetime without losing the energy and reliability in various application systems. For instance, if any sort of sensor network is not being deployed as an energy

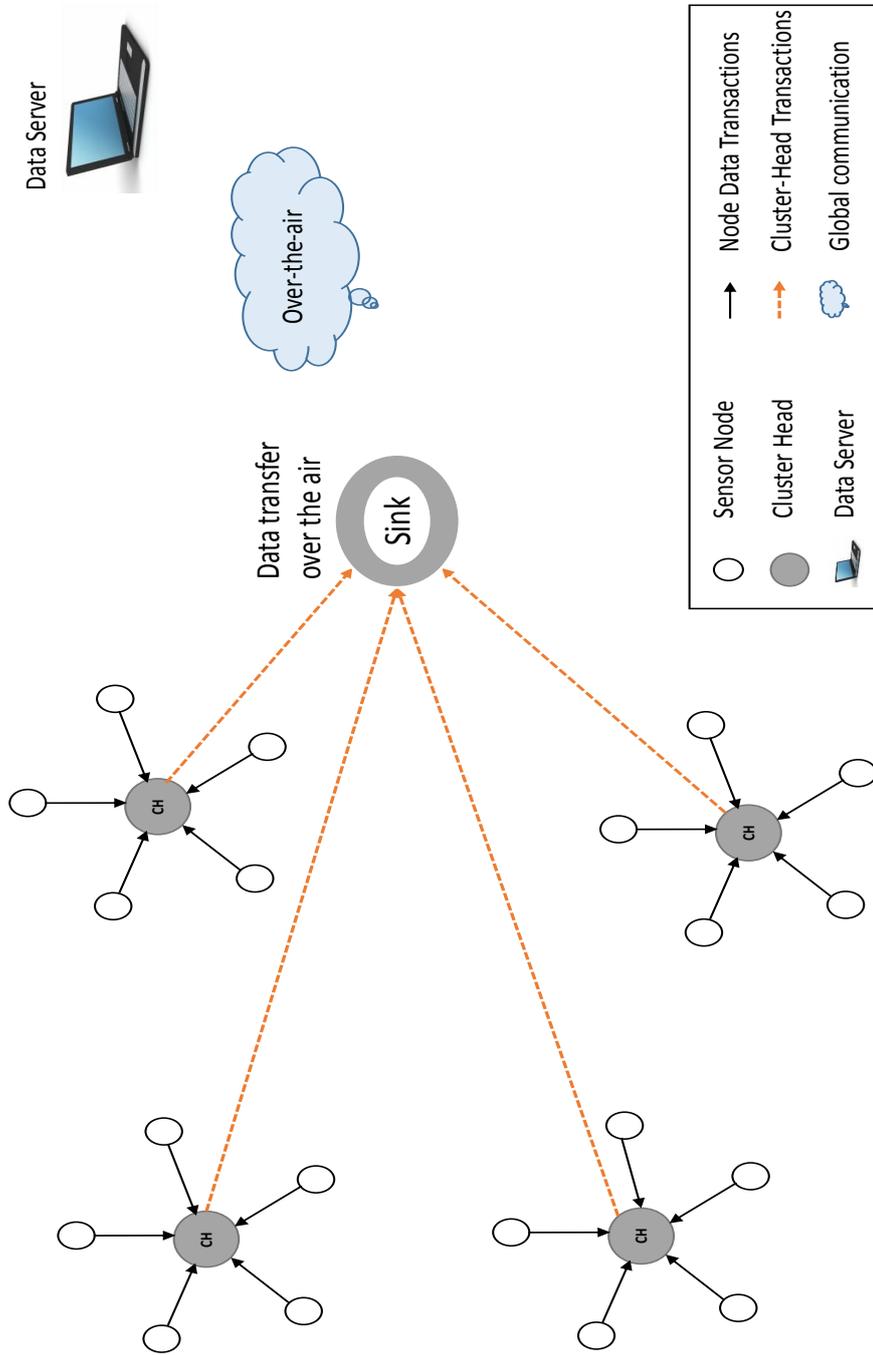


Figure 1.3 – A typical architecture of Single-hop Wireless Sensor Network.

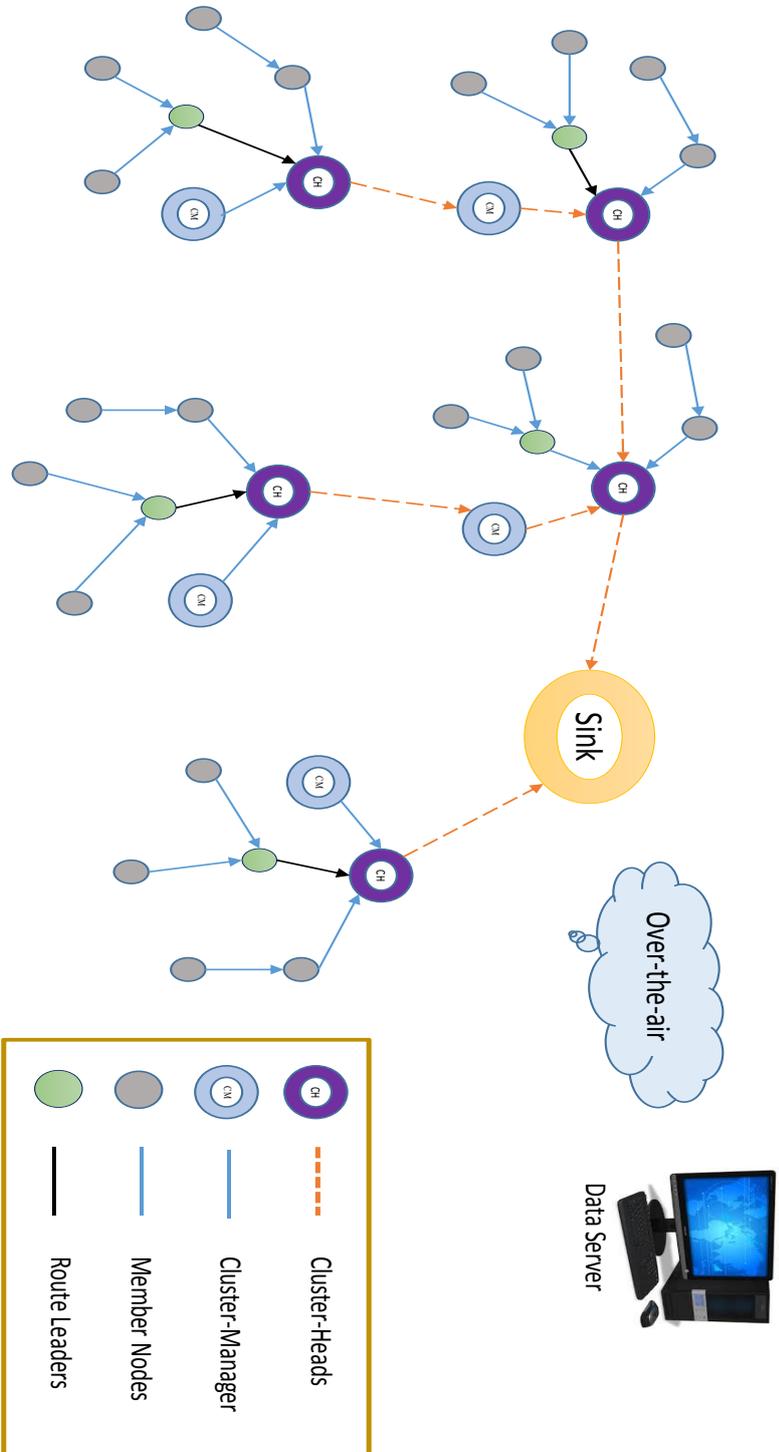


Figure 1.4 – A typical architecture of Multi-hop Wireless Sensor Network.

optimized network, then the system failures occur often in that application area. In order to achieve the energy efficiency problem in distributed sensor networks, node must be designed as an autonomous and intelligent device to make a decision by itself before processing any data or gather the neighboring nodes data towards the base station (BS).

This thesis mainly focuses on routing and data redundancy techniques in WSNs using cluster-based topology and in-network aggregation that are based on spatio-temporal sampling. In this regard, this dissertation aims to model, analyze, design and develop the routing and data redundancy WSN protocols. As both the combination of routing and data-driven approaches are in order for delivering the better performances within the network, in terms of saving nodes energy and maintain a good communication links among the nodes and data reliability. Generally, network lifetime relies on nodes lifetime in WSNs, within this context, we first address the various issues in both routing and data-driven approaches, and then present our propositions along with each other, as they are further briefly described in the following sections.

The nodes are constrained with several limited hardware capabilities, such as computational capacity, memory, and energy. In order to achieve the better resource preservations in computation and communication modules of the nodes. Node must be an efficient device to handle the resources by itself, especially in terms of handling the data sensing, pre-processing of data, and then transmissions. Nodes are supposed to be optimized, because of their limited battery resources. Due to the proposition considerations, nodes send significant data only every time, unless the window bandwidth contain all the non-redundant information, then the window will be transmitted as one packet. As studied in chapter 2 and chapter 4 that the literature studies used either compressed data forwarding method or prediction-based method or carrying the maximum payload dataset, which termed as maximum transmission unit (MTU) of CC2420's RF transceiver, limited upto 127 bytes of data [23]. In this regard, we follow the basic principle of Tiny Aggregation (TAG) data service for TinyDB, thus an efficient data transmission protocol must be used for data payload as much larger as possible into one packet over the time duration of  $i$ . Thus reduces the required number of regular transmissions in sensor nodes as well as conserves the energy.

In the contrast, we present and adapt the data window concept from one of our earlier research works [24], in order to tolerate the window evaluations over short *EPOCH DURATIONS*, which we termed as  $r$  round seconds of the window in SNs and  $R$  round seconds in CH nodes. This dissertation deals with the following research integrations based on the propositions and their requirements.

- Pre-filtration data window protocol which carries a viable solution for transmission of necessary sensor information in WSNs.
- The data window protocol, hence, provides the possible pre-processing integration through the COOJA/Contiki developments within the network.
- Although, data window protocol offers all sort of data pre-filtering services,

such as aggregation, compression or prediction-based, as well as the integration in such like small databases, as TinyDB and Antelope.

The above points are covered up by chapter 3 and 4, which focus on the both cluster-based routing and efficient data transmission protocol through cross-level network of COOJA/Rime. Our objective is to conserve the energy in sensor nodes and extend all nodes lifetime of the network. The main issue in sensor nodes are that they transmit or push the data transmissions based on their sensor detections irrespective of the content evaluations, thus make nodes to spent their energy for redundant data transmissions as well, which leads the nodes to drain out their energy much earlier than the expected lifetime period, and then the nodes can also be depleted. In order to suppress or reduce the non-significant data transmissions, we present a DAWF spatio-temporal protocol in sensor nodes for exploiting temporal redundancies as well as the spatial redundancies and correlations in CH nodes.

In this scenario, each node does the pre-filtration, and pushes their data transmissions while monitoring their sensor readings over the short window intervals. This method not only does the pre-filtration of data, it also reduces the total bandwidth of the signal, and transmits the data accordingly. The proposed solution is a less computationally resource constrained method for the sensor nodes. As a result, it does not hold any data storage tables for longer evaluations, DAWF only uses shorter window evaluation method to exploit the data redundancies over the period, and clears the window data on the basis of First-In-First-Out (FIFO) queue method during every successful window data transmission, and then reuse the window at every instance time  $t$  of sensory detections.

However, the parameters of window sizes ( $w_M, w_Q$ ), threshold ( $tct, sct$ ) values, and their considerations can vary from phenomena to phenomena and their homogeneity to heterogeneity nature as well as from one application to another application. Nonetheless, the proposition has the ability to adapt itself towards the exploitation of either redundancies within the nodes or the spatio-temporal correlations among the nodes based on the application requirements. Although, we show that, with our simple and special combination of data reduction methods, it is feasible to recover the originally acquired data without losing the accuracy in data, as presented in chapter 4. Nodes are chosen as, normal sensor nodes and the CH nodes as super nodes, which held with high computational and processing powers. Since sink node is a highly resourced device in the network, which does the signal reconstruction based on its received data sets among the network, however, enabling the signal reconstruction mode is a strictly network application choice.

The use case can be expanded with many other different considerations in multi-hop networks, where the method can act for exploiting both spatio-temporal redundancies and correlations over the source areas. As the multi-hop routing novel approach has been designed and presented in chapter 5 according to the DAWF mechanism needs.

## 1.5. Thesis Contributions and Outline

In this thesis dissertation, both the node lifetime restriction and the system usability problem have been addressed, in order to reduce the operational costs of wireless sensor networks. As part of this research problem have been previously targeted, and is too broad to be solved as a whole. The following list are credited as main contributions from this research work.

1. Study, analyze, model, and design a single-hop cluster-based method on the basis of hierarchical routing protocol (LEACH), in order to maximize the network lifetime based on its current CHs remaining energy. (Paper 5)
2. Cluster-Head Techniques for Single-hop Routing protocol in Energy-Efficient Wireless Sensor Networks. (Paper 4)
3. Design and development of an energy-efficient cluster-based protocol using an Adaptive Data Aggregative Window Function (A-DAWF) for exploiting temporal data redundancies in Wireless Sensor Networks (WSNs). (Paper 3)
4. Design and development of Energy-Efficient Data Aggregative Window Function (DAWF) mechanism for exploiting spatio-temporal data redundancies in Wireless Sensor Networks. (Paper 2)
5. Design and development of Energy-Efficient Data Aggregative Window Function (DAWF) protocol for exploiting spatio-temporal data redundancies in different physical quantities of Wireless Sensor Networks. (Paper 1)

The remainder of this dissertation is as categorized into four chapters, and the organization is as follows.

### Chapter 2

The chapter starts with the introduction and a background study of the energy-efficiency problem through various methods over wireless sensor networks. It then offers the related work of routing and data-driven discussion for the literature survey, relating to some well-known protocols of different distributed and centralized methods in both single-hop and multi-hop WSNs. The routing study of literature described more elaborately in chapter 3 and their discussion and summary of the chapter. While the second section of chapter 3 deals with various data-driven approaches, as follows the taxonomy of data-driven approaches from [8]. After an analysis of energy-efficiency towards the lifetime constraints in WSNs, and the contributions of routing and data-driven techniques in sensor nodes and their tasks to the energy consumption of the system. A sensor node protocol design by using efficient routing hierarchical, and data-driven techniques with the core on energy-efficiency is presented. Furthermore, various data-driven methodologies and their impact in terms of pros and cons

are discussed. We then propose and develop various energy-efficient techniques in the following chapters.

### **Chapter 3**

Chapter 3 focuses on the background study of efficient routing protocols and related works in WSNs. Although, the chapter presents various routing literature studies and made a decision of their impact over WSNs in both distributed and centralized systems. But, the chapter mainly related to hierarchical clustering networks rather all the routing techniques. It then proposes a system design for energy-efficient routing protocol based on LEACH CH algorithm while comparing the simulation results with other well-known protocols. Nevertheless, it also provides a discussion and summary of the various well-known routing protocols, which were also listed in a table based on energy, reliability and delay metrics.

### **Chapter 4**

Chapter 4 deals with the data-driven techniques over wireless sensor networks. The data reduction methodology is one of the major energy conservative field in WSNs, and the computational module is a second highest energy consumer module in the sensor networks. Many earlier known aggregation techniques only focused on combining all of the data among nodes and then carrying by relay or master nodes towards the BS. On the other hand, many well-known data reduction protocols are presented based on either spatial correlation findings or temporal prediction-based networks. There is a very limited amount of work available based on spatio-temporal techniques for WSNS, even though they have their own set of limitations due to the vast application requirements. Thus, this thesis focuses on to present a data reduction novel approach, especially, it considers to design for exploiting spatio-temporal redundancies and correlations by utilizing the data window aggregation mechanism over distributed sensor networks. The proposition has been particularly designed for targeting the both periodic and delay-tolerant applications. This chapter also explores the designs with the combination of sing-hop distributed clustering network.

### **Chapter 5**

In this chapter, thesis contributions are summarized and discussed, and then conclusions are given based on the experimental study case results that are made.

## Chapter 2

# Energy Efficiency in Wireless Sensor Networks

The background study of this chapter focuses on both routing and data-driven of WSNs. The field of WSN is too broad to deal with, and this thesis primarily describes the study of it focuses in, and then it explains the literature survey of various data reduction techniques. Then presents a case study and analysis of the lifetime constraints, as well as influencing factors in relation to hybrid energy-efficient techniques on a sensor node level, in order for extending the nodes lifetime. The contributions of this thesis presents the energy-efficient protocols with the combination of both routing and data reduction techniques.

### 2.1. Wireless Sensor Network Lifetime

Over the years, WSNs enable both the small and large scale networks for autonomous sensing. The scale of the network varies on the basis of application needs. Theoretically, a large number of nodes are deployed over a wide area, especially for remote area or harsh location, which can organize themselves and send their sensor data back to the user without any further turnout. A key element for doing so the nodes to achieve this task that the nodes should always have a proper energy supply, as well as the system must always be energy-efficient during the task. In order to avoid the disruptions of initial advantages offered by the system, when the system demands for user to revisit the deployment site often, for exchange or replenish the systems batteries. Due to the importance of the system lifetime at the application level, lifetime parameter has been used widely in the designing part at all levels of the sensor networks. On the other note, several definitions of the WSN lifetime has been presented in the literature and an extensive coverage of these classes is available in [25].

The primary classification made by separating the sensor node lifetime out of wireless sensor network lifetime. Although, these both are typically relied on each other.

The liveliness of the sensor node can be determined by its operational status of the

task at a given time. The lifetime of a sensor node usually predefined by its energy sources (either battery or other power sources) and its energy consumption over the operational period. Thus, the energy resources are the only limiting factor towards the nodes lifetime. However, the operational disruptions occur based on different operational factors during its original lifetime of a sensor node, these lead to tolerate additional consideration for resisting the operational disruptions.

Comparing to the determination of WSN lifetime, it can not easily be determined without considering as many application constraints as into the account. Typical some other definitions are either based on the coverage of the monitored area or the network connectivity. With the combination and extension of these factors might also be considered into the definition of a sensor network lifetime [25]. Since the nodes are building blocks of network, and their lifetimes should vary from application to application, which further are connected to each other.

The WSN lifetime scenario definitions are specifically categorized based on their operating levels, such as *n-of-n*, *k-of-n*, and *m-in-k-of-n* scenarios. Among these, the first two are straightforward solutions, where *n-of-n* means all the nodes should be alive, and in the case of *k-of-n*, at least  $k$  nodes have to be alive out of total  $n$  nodes. Nevertheless, if the solution is not met since more nodes are dead than are alive, then the sensor network lifetime as well considered as being completely disconnected. Therefore, different nodes can have different significance in their tasks to the network that they are operating in. In comparison, end node has lesser significance than a regular node which regularly forwards the data of other nodes to the network. These cases are integrated in *m-in-k-of-n* scenario, in which they divide  $m$  critical nodes (e.g., routers) and  $n$  non-critical nodes, where at least  $m$  critical nodes are required for a task execution out of  $k$  remaining nodes.

The WSN lifetime is an ultimate parameter of interest, because of its application dependency. The remainder of this chapter addresses the node lifetime and the link and contributions of this thesis for maximizing the nodes lifetime. However, WSN lifetime can be addressed indirectly, because of its link between node and network lifetime. Thus limit the possibilities of extending node and network lifetime while relating to this thesis contributions, which then presents in both chapter 3 and 4 to maximize the network lifetime.

## 2.2. Hardware Platform Choices in Sensor Networks

The initial proposition of the sensor nodes architecture from WSNs has not changed extensively. It mainly contains the modules of computation, communication, sensing and power management among all. In majority of cases, these functionalities can be used as same other than the application-specific tasks, which requires some additional resources. An abstract overview of general sensor node hardware architecture is given in figure 2.1.

The computation module of the node among others has several other tasks to hold

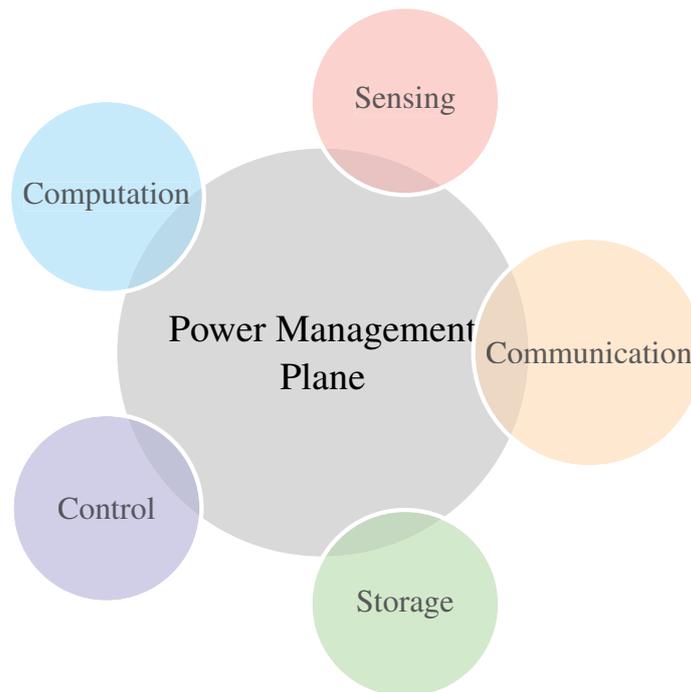


Figure 2.1 – An architecture of typical sensor nodes in Wireless Sensor Networks. Nodes typically possess with the combination of sensing, processing and communication capabilities, and they are powered by energy sources (e.g., battery)

with. Because it combines the control and storage unit. It then also controls the other components that are within the platform, which can process and stores the data, and then provides an interface to the end user. Thus, the implementation of computational module is truly an application dependency. However, most of the nodes implement the computation module on the basis of low-power micro-controller. Microcontrollers are easily programmable and have low power demands, they can also integrate a wide variety of different hardware modules, which can be very useful for the accomplishment of the node tasks. Some of the popularly known devices include as Atmel's ATmega series [26, 27], Texas Instruments MSP430 [28],[29], PIC controllers from Microchip [30], ARM Cortex M3 from ST Microelectronics [31]. For the high-end processing applications, Field Programmable Gate Arrays (FPGAs) and Digital Signal Processors (DSPs) can also be used as co-processors [32].

The active operation in microcontroller consumes a considerable amount of power (i.e., in general the order of hundreds of  $\mu\text{A MHz}^{-1}$ ). Most of the microcontrollers contain with a series of operating modes, which then helps system to conserve the energy that whenever the activity is not needed. In order to facilitate this operation, the microcontroller must trigger to be woken up again for resuming the normal operation when needed.

Like the computational module, the communication implementation also being a dependent to certain degree on the application. Moreover, majority of the commu-

nication systems use similar featured communication devices, namely as low-power RF transceiver, which typically operates with the license free ISM bands. In the early stages of research in WSN, RF has been used as a popular communication method. The recent advancements in communication standard developments, IEEE 802.15.4 protocol with the compatible transceivers become a de-facto communication standard. This is because its world-wide free licensing frequency band, as well as its grown interest in commercialization.

The communication module can be assigned for the establishment of a link between individual sensor nodes. This communication link is being used, in order to exchange or share the information or data messages between nodes. And, particularly to some sort of remote applications, in order to propagate data between individual sensors to the data collection point (i.e. network sink). In addition to this, there might be a global communication required to implement, which creates a link between local sensor network and outside world and, it thus provides a universal accessing service to the monitoring system. The typical global communication modules are namely WiFi [33], long-range communications [13] and GSM/GPRS [34]. However, Network with the number of nodes that are equipped with such modules are strictly limited, since both the price and energy consumption are excessive to maintain these sort of systems.

Even though local communication system in the nodes, communicating remains costly for these sort of resource constrained devices. Even sometimes when node operating with idle mode or only listening to surrounding noise, the energy consumption of the communication module is huge. Therefore, the communication module of sensor nodes should turnoff their radio whenever it's possible. However, in order to deal with efficient way of communication handling problem or preserving the energy at other modules, the appropriate energy-efficient protocol must be applied. The communication and computation modules parameters or considerations might vary based on their resource requirements of the application. But, the most application-specific part of a sensor node is sensing module. Because, a given particular application can be assigned to monitor a certain physical quantities or detect specific required events. This, leads to specific task-oriented sensors which have the ability to fulfill its application demands. Although, the type of sensor for WSN application can be imaginable, but the choice of sensors that are found in the literature is rather limited. The majority of existed literature has been conducted with the consideration of low-power sensors, with a typical example of temperature sensing. Because of the limited application-oriented research, and the difficulties in performance comparison when there is other underlying assumptions. Thus leads to the assumptions within the community that do not hold true. The typical negligible energy consumption module is sensing. This might also true, because there are many power hungry sensors that have higher energy consumption than the communication module, which requires many warm up times in case of finding accuracy in detected readings. The SAQnet measurement system has also proved that energy consumption of various power hungry sensors consumed more energy than the communication module [12].

As mentioned above, the standard node can be equipped with all those four mod-

ules with some additional features on the basis of application requirements.

The communication module is an obligatory device to manage the communication services among sensor nodes, as given the explanation in [35]. Hence, the amount of sensors or actuators that are being used in sensor node may vary on the basis of its application scenario. Therefore, the targeted device lifetime of the system and its application scenario, must be predefined by power supply source.

In general, sensor nodes are resource constrained devices and limited, in terms of battery or other power sources, storage memory, computational capabilities [36]. The typical hardware platform choices of the nodes and its specifications are highlighted in 2.1, (e.g., MICA2dot, Mica2, MicaZ, TmoteSky/TelosB, Z1(Zoltera), IRIS), and few other recent ones are OPAL, .NOW/eMote. However, the power consumption level at each nodes module vary based on their hardware platform resources, for instance, as it has shown some of microcontrollers current draw operating states are listed in the table 2.1.

Types	Mica2/2Dot/Z	TmoteSky/TelosB	Z1	.NOW/eMote
$\mu$ controller	ATmega128L	MSP430F1611	MSP430F2617	Cortex M3
Architecture	8-bit	16-bit	16-bit	32-bit
SRAM[kB]	8	10	10	96
Flash[kB]	64-256	48	92	128-1024
Current Draw				
Active[mA MHz <sup>-1</sup> ]	1	0.5	0.5	0.214
Sleep[ $\mu$ A]	5-100	1.1-75	1.1-75	0.3-0.9
Off[nA]	250	100	100	120

Table 2.1 – Hardware platform specification choices of the nodes in sensor networks.

## 2.3. Energy Efficiency over WSNs

In order to ensure the system longevity in network lifetime, energy saving methods should be applied on the nodes. With the help of energy saving protocols, node enhances itself to manage the power resources of all the system modules at whenever the task required to do.

Like the nodes system architecture modules, energy efficiency has also been diverged into each and every module of the sensor node. The dynamic power management (DPM) method is presented in [37] by implementing the different operation modes, such as full active, idle, and sleep. Usually, the most power consumer mode is when the system being active, which defines for all the sensor activities, namely sending, listening, and data gathering or collecting. DPM also presented that sleep mode consumes very little amount of micro joules ( $\mu$ J) energy rather millijoules (mJ).

Energy efficient protocols in WSNs have been contributed immensely over more than a decade in numerous directions of the WSN applications. Although, energy-

efficient protocols become a greater asset of WSNs to extend the network lifetime at some extent level compared to the predefined lifetime by its power sources. Due to the wideness of WSN and limited resources of nodes, many protocols work well within their own considered parameters and for their applications only.

As discussed in the previous chapter, this thesis further explains related to this thesis. The energy-efficiency is much concerned issue in WSNs, and its too broad to cover in total. We then further narrow down the subject towards to this thesis topic. The below sections explain a background and an insight into the problem definition that is targeted by this thesis.

## **2.4. Routing and Data-Driven**

More than a decade, there has been an extensive research conducted in the area of both routing and data-driven over many different WSN application scenarios. The energy-efficient protocols in the literature survey of [8] presented several classifications of routing and data-driven approaches. We further briefly explained the classification of routing protocols and its summary and problem findings in the chapter 3. Although, an extensive research study has been investigated in routing, and then proposed the improvements over the literature which is directly related to this thesis, but the contributions are limited to this area. The further investigations of this thesis has been conducted in the field of data-driven for WSNs. As the relation between routing and data-driven lies on each other, since the sensing information needs to be aggregated or compressed by a data-driven method, and then the aggregated data can be routed towards the BS while following the in-network processing. However, the propositions of this thesis are designed based on both routing and data reduction techniques. In chapter 4, the DAWF propositions are followed along with the routing, in order for network management and similar other tasks which are not directly implying to the sensed data.

The basic principle of routing is to prolong the networks lifetime by managing routing discovery while guaranteeing the reliable data communication among the nodes. The design factors in routing are influenced by many challenges to achieve, such as reliability, energy, and delay. The classifications of network structure in routing is described [38]. Thus, this thesis propositions have been possessed in hierarchical networks. The further methodological study of routing which is directly related to the propositions that have been surveyed in chapter 3.

## **2.5. Data Reduction Techniques for Energy-Efficiency in WSNs**

The second part of this thesis scope shifts on the literature survey for data gathering and disseminating in WSN. In particular, this thesis initially designed for sing-hop

WSNs, and we assume that the routing for network management in data window aggregation is single-hop clustering, since the sensor nodes (SNs) are distributed, which can forward their filtered data to their corresponding cluster-heads (CH). The purpose of window aggregation is to possess a simple pre-filtration methods at node level, latter in CHs as well. In these assumptions, it forms like a static-clustering, but the head nodes are considered as super nodes, which have strong resources, in terms of memory, computation, power and storage capabilities.

The initial target is to design window pre-filtration techniques on single-hop networks, because the network complexity or communication overhead cost in multi-hop/anycast is high. Nevertheless, the routing and data-driven techniques diverge in two different ways, and there is no direct link between them in WSNs, unless the system is being used for in-network data processing [39], which are helpful for data fusion among hop-by-hop or at head nodes towards the sink. However, this is obvious, if the system applies a data-collection principle like spanning tree model by Tan et al [40], where node may have a data overload problem when the data being carried away through hop-by-hop towards the sink, which then collects the data at sink level. This leads to the overloaded nodes run out of battery, which then completely lost their connectivity and the data as well. if too many nodes are overloaded in the network, then it is not feasible to recover network connectivity. In this case, the network should be partitioned among the overloaded nodes. Therefore, network portioning also become another challenge to achieve in multihop networks, where the sensed data become inaccessible during the network partition, and they cannot be delivered to the sink.

As the radio transmission is a main energy drainer of low power sensor nodes, it is essential to suppress the redundant data transmissions at node level. This prospect not only preserves the energy at radio transmission, it can also helps to save the receiver's total system energy for that iteration, which enhances the nodes energy levels immensely across the network as well as the network lifetime. In this section, we categorize the subjects of this thesis that posses in data-driven on different criteria. Latter, we narrow down our study in the field of data-driven to the categories, that are most related to this thesis that are presented in chapter 4. It thus provide an insight, which exploits a problem definition that are targeted in this thesis.

There is a great deal of research being conducted in the area of data-driven. However, the study turns into several subcategories, which are briefly distinguished and explained in [8]. The data redundancy problem in WSNs is widely studied area in the recent past literature works, it still draws a huge attention due to the extensive application roles and their unstoppable growth in WSNs.

The data redundancy techniques can be sub-classified, namely as *in-network* aggregation, *compression* based, and *prediction* based or *forecasting* based.

1. **In-network processing:** Mostly rely on data aggregation, such as computing the average of certain data level at relay nodes between the end devices and sink node. Thus, it manages the data reduction among the nodes while traversing the network towards the sink or BS. However, most of the appropriate in-network

aggregation techniques are application-specific, and can be followed to its application needs, as such that [39].

Moreover, pure aggregation functions are non-reversible, which is not a feasible solution to retrieve the original data. For instance, the aggregated data packet  $p_a$  among the nodes at an intermediate node towards the sink, as soon as it received at the sink node, then it is not possible to reconstruct the complete signal of  $p_a$ .

2. **Compression-based:** Data compression techniques can be used for encoding the information by nodes and decoding the received data at the sink. [41, 42, 43, 44]. Forwarding compressed data from the nodes to either relay nodes or towards the sink, can able to partly or completely reconstruct the original data signal from the compressed signal.
3. **Prediction-based:** Prediction can be used as model parameters towards the sensed phenomena. Thus the queries can be sent by the model instead of original phenomena. The further classification of prediction-based techniques is given in [8].

The data-driven principles that are being considered throughout this thesis with the structure of data redundancy techniques. However, in-network processing has the least priority in our data redundancy techniques, somehow in-term principle has been used to forward information through the head nodes towards the sink.

Let consider  $S_{raw}$  be a list of finite generated sampling data at given time instances  $t$  by node, and sensed values can be recorded by a buffer. Thus, the data compression-based signal transform function of  $f$ , draws a raw sampling of  $S_{raw}$  as an input by the smaller dataset, while applying the window filtration to this considered amount of smaller dataset, which then derives it as after the reduction of its original bandwidth (BW), which then indicates it as

$$BW_{reduced} = f(BW_{original}), \text{ such that } |BW_{reduced}| < |BW_{original}| \quad (2.1)$$

Nevertheless, the inverse transform function can be denoted as to retrieve the original signal, where as

$$BW_{original} = f^{-1}(BW_{reduced}) \quad (2.2)$$

### 2.5.1. Accuracy in Restored Data

The accuracy in data or quality of data that being sent is another metric parameter to measure the delivered data quality in data redundancy schemes, that how efficient that the mechanism is delivering the performances. It then further classified into two categories, namely data accuracy in compression without any loss (loss-less) and data inaccuracy in compression.

- **Accurate data compression (Loss-less):** This sort of compression is usually used in sending of large size of sensitive data as a single archived file (e.g., software application files or medical image files), and it is also commercially used for compressing the software's. In this case, none of the data is lost while applying compression and decompression to the acquired data.
- **Inaccurate data compression (Lossy):** Lossy compression means certain detailed level of data can be lost even after executing the decompression technique, because of compression. For instance, the image and video processing compression methods, such as JPEG2000 [45].
- **Unrecoverable compression:** Unrecoverable compression states that recovery of original data is irreversible, and decompression cannot be applied. For instance, in-network processing, where the relay nodes send averaged data among the nodes to sink, which means it only received the one final average value by relay among all nodes original data.

Since, the distributed propositions of this thesis are designed under the consideration of lossy redundancy, thus it is proven in chapter 4 that the signal reconstruction is effective, yet some unimportant data is being omitted. The trade off also shown between the delivered data to the sink and the transmission cost over the network.

### 2.5.2. Source Coding and Network Coding

Data compression mostly achieves in a distributed fashion, however compression or redundant techniques in WSNs demand an application-specific implementations. Although, compression algorithm usually happens on smaller chunks of data rather than the total available data sets, which is not into the memory of computing machine. In this regard, it will be easier to observe more correlations in raw data, if the compression algorithm has an access to whole collected data that are fit into the available memory at once. Though, it is not possible to hold that amount of memory and storage capabilities, which requires at least in the order of hundreds of mega or gigabytes. Practically, compression algorithms go with much lesser chunks of data sizes and process each one of them differently.

Compression technique in WSNs is a resource constrained and intensive operation, because it fetches the data of each and every sensed value from sensor node, thus leads to the resource intensity and often need to be forwarded through the hops and can be available to the base station.

### 2.5.3. Discrete Cosine and Fourier Transformation in Data Compression

Often, the transformation functions are being highly used in compression techniques for huge sets of data. It can be used either way of both compression and decompress-

sion, transform function at compression for reducing the amount of non-related data that being sent, and the inverse transform function can be applied in order to retrieve the original data. A linear transformation on discrete data is often performed and it can be represented by a matrix multiplication.

Discrete Cosine Transform (DCT) and Fourier Transform (FT) are widely popularized transformation techniques that are highly used in linear transformations for compressing the signals. Wavelet transform is another sort of transformation that is commonly used in compression graphs [46]. In transform compression, most of the signals are recorded from physical phenomena, which are highly compressible either by DCT or FT. Transformation in compression defines that, most of the compressed data in transformed vector are nearly zero, only a few of the values are contained with a significant value. Thus it is based on their largest data values which then store in their index of the data vector along with their location.

Lossy decompression means applying the inverse transform function on the transformed data vector, which then able to recover the original data with a reasonable accuracy. However, there is slight difference between the original data and the one restored data, which then calibrates the lossy parameter to scale the amount of data that differs from the original data. While knowing the trade off between and compressed and decompressed data. It can be easily tuned the transform compression to achieve a desired compression ratio while sending the application desired degree of accuracy.

In the contrast, loss-less compression uses a standard entropy coding techniques, namely as Huffman coding and run-length coding. In this prospect, it keeps all the transformed data and with the help of coding techniques further suppress the stored or transmitted data. Therefore, it uses two phases of decompression, one is encoding string of data that is decoded, and the second one is inverse transform, which can be applied on the result to recover the original data. In this scenario, it can able to recover the complete original signal without any loss. Yet, the process of the method is expensive compared to the lossy compression.

## 2.6. Necessity of Data Reduction in Distributed Sensor Networks

- **Sensor:** Usually a sensor detects or generates the physical quantities that the phenomena of application interest, such as light, temperature, humidity, pressure, acoustic, solar temperature, and so on. While using an Analog-to-Digital (ADC) converter, it then converts the recorded phenomena and presents the output in a binary number that reflects to the sensed data value.

In general assumption, the energy consumption in sensing has given a less stand-point or not relevant. In fact, some of the literature and research studies have a different perspective that it is indeed relevant, yet it can be even greatest energy consumer compared to the radio or even greater than the total amount that

usually consumed in sensor node [47], this happen may be because of many different factors [48], such as power hungry transducers, A/D converters, active sensors, and long data acquisition time [8].

- **Data Transmission:** Data transmissions occur periodically over a communication channel. In WSNs, a large amount of data sets are used to be carried over a noisy wireless channels. While carrying through noisy channels, there might be packet losses, collision, congestion and inevitable errors, which causes the several re-transmissions, thus leads the system communication cost way too expensive for sensor nodes. It is inevitable for the node to handle when both communication and raw data carriage is far more than channel capacity. In order to reduce the communication or transmission cost, an effective data redundancy technique must be applied on the system.
- **Processing and Storage:** In WSNs, sink node gathers the data among sensor nodes, which then process and store the data for future evaluations. However, the data can be encoded, encrypted or reduced, it thus depending upon the application interests that what sort of information, the system needs to be gathered over the network.

## 2.7. Compression-Based Sensing in Distributed WSNs

Some of the earlier research studies and surveys have given the less proximity for compressive sensing in WSNs. They rather said compression techniques are not significantly related to the WSNs, such [8]. As explained in the earlier sections, either compression-based or prediction-based techniques, they have their own pros and cons in relation to the WSNs, and on the basis of their application-specific requirements.

Usually, the quantity of compressibility of the sensed or sampled data is unknown, because of its inaccessibility to the whole sets of data at a time and at a single point. Especially, this is not feasible in multihop sensor networks. Because each data item can be individually sensed by the nodes. The basic compressive model is not compatible to the distributed sensor networks. Because of simultaneous operations of sampling, encoding and communication in WSNs. Due to the limited computational resources of the nodes, both encoding and compressing has to be less complex in nodes. Since the sink hold with sufficient more power, then it can be adequate to perform decompression or decoding methods often in sink nodes. The complexity nature of compression/decompression system module can be sub-classified into three application classes, such as High compression ratio and easy compress/decompress complex algorithms, Fair compression/decompression resource constrained, and Resource-constrained encoders and powerful decoder.

As per the literature, WSNs fit into these sort of classifications of compression that mentioned above. Since the encoding process only matters to manage the computational resources of the nodes, and the complexity of decoding is not an issue at either

sink node or end user side, for reconstructing the received signal of the encoded data by nodes. The further sections will give a brief explanation of three main compression-based methods, such as in-network compression, Distributed Source Coding (DSC), and Compression-Based Sensing (CS). Most of the earlier research contributions are based on these three methods, and some other provide the improvements or extensions of these methods.

### **2.7.1. In-network Processing**

As mentioned in the above sections, in-network compression is with the combination of routing and encoding throughout the data forwarding process that is carrying hop-by-hop to reach the final destination as sink or BS. In-network compression is mostly built through transform compression methods. However, DCT or Fourier transform can be used and much suitable for acquiring the signals, such as temperature, humidity, light, etc. Wavelet transformation can be used for environmental scenarios or the signals that have rapid variations.

In comparison, in order to achieve the optimal compressive for a given signal in theory [49], namely DCT, Fourier, Wavelet, and Karhunen-Loeve Transform (KLT) are considered to be as better transformations. While, DCT and FT can generate finite number of matrices, but KLT does not have any practical solution for generating the entries of a matrix for the assigned signal [50]. Irrespective of compression technique that is applied for in-network processing, the data transmissions has to follow the order of compression form. Thus, the compressed data that are collecting through a spanning a tree approach, once the data received by sink, then it applies a decompression technique to retrieve the original data, for instance Shen et al [51] presented a data gathering mechanism on the basis of transform compression.

In order to achieve a best compression ratio, transformation must be used on total amount of data rather than the small portions of sensed data. In this case, each node receives the compressed data, then it has to a decompression at itself to extract the raw data and combine its own sensed data to the received data, and then the node can forward the combined data as a compressed data packet to the next hop.

The main drawback of doing this process, nodes must have a sufficient computational resources, in terms of memory and power. This sort of operation is only possible when the nodes are considered to be as powerful than the regular nodes, such as powerful CPU and a better RAM sizes. The another problem in in-network compression that the computations handling by nodes are not fully balanced across the network.

In this assumptions, the nodes that are close to the sink can perform more computation and require much RAM than the nodes that are in middle of the network tree. This imbalance in nodes leads to the in-homogeneous of the network. Moreover, this sort of inflexibility in WSN causes the disruptions of the operational tasks, due to the often configurational parameter changes. Since, homogeneous networks must be equipped same hardware platform features.

The advantages and disadvantages of in-network compression is as follows.

- **Pros:** Higher compression ratio, low bandwidth operations, and extendable to temporal domain.
- **Cons:** Complexity, imbalanced computation, and sensitive to the packet failures.

Nevertheless, apart from the packet losses and network perturbation, abnormal sensor readings by nodes may also effect the in-network compression. Especially, when the transformation choice is against to the robustness of the network. Therefore, the sort of such wavelet transform principle is more suitable choice to balance the network, when unexpected events or abnormal readings are occurred.

### 2.7.2. Distributed Source Coding

The Distributed Source Coding (DSC) is another popular method which often used in WSNs. It is based on correlation findings between any two data sources. Joint encoding can be very efficient to use when any two data sources are correlated. However, the main drawback in DSC is that it often depends upon the correlation model between the sources. In this case, DSC must be aware of its statistical model correlations between different readings. During the network operations in WSN, the correlation model in DSC should be remain constant. These requirements are hard enough to achieve in WSNs. Therefore, the sensor network gather the data among the nodes and observe their physical parameters for a longer duration to achieve an accurate correlation model. Due to the limited resource constraints of the nodes, it is usually not feasible at all to achieve such accuracy in correlation model. Without accuracy, DSC could not perform effectively, thus leads to the data collection errors.

Moreover, even if the network knows correlation model there is no proof that the given model remains unchanged during its entire operational duration of the network. It strictly depends on the phenomena variations, if the sensed data changes then the correlation model as well abrupted, and then DSC requires a new correlation model to ensue the correct data collection. Nonetheless, DSC has its own pros and cons, such as less complex in terms of coding, memory, adjustable robustness, but it is required for only accurate models, and sensitive towards the environmental changes.

### 2.7.3. Compressive Sensing (CS) Theory

CS theory presents lossy compression techniques, which differs from transform coding in measurement acquisition. Donoho [52] presents that it can be possible not to perform full transformation on the whole datasets under certain conditions. Moreover, the linearity of measurement mechanism, which can simplify the nodes hardware and software design.

Like transform compression, CS also uses a linear transformation on the sensed data. Unlike transform compression, not all of the coefficients can be calculated in the transformation. CS uses a fraction of transformed data toward sink, namely measurements.

In order to acquire the linear measurements from the network, which involves only multiplication and additions are being used to calculate in nodes and the complex calculations are very minimal. Nevertheless, this sort of operations can be performed more efficiently in nodes, when the nodes can be equipped with specific low-power hardware, in order to perform the array processing Bajwa et al [53].

The challenging task to be achieved in transform coding is that, the magnitude and location significant coefficients in transformed data has not known in advance. Thus, the transformation performs on whole data. The compressibility of the data through transformation, benefits the CS, while performing the transformation on whole data in order to reduce the compressed data. The subset of data is enough to reconstruct the signal from in order to get the original data. The CS theory has its own pros and cons, such as decentralized data processing, simple arithmetic calculations at SNs, and balanced communication and processing. The drawbacks in CS theory as complex reconstruction algorithms and measurement inefficiency in leaf nodes.

## 2.8. Data Prediction in Distributed WSNs

As explained in the above section 2.4.2, data-prediction build with the model parameters, so that the queries can be sent through the model instead of actual sensed data. Therefore, the data-prediction can be sub-classified into three, such as *Stochastic approaches*, *Time-series forecasting*, and *Algorithmic approaches*.

- **Stochastic Approaches:** Stochastic model uses a random process to exploit a characterization of the phenomena. Well, Ken et al[54] provides an example of this approach. In this general scheme, the number of models that are replicated at the source and at the sink. However, the basic model of probabilistic, i.e. probability density function (pdf), which refers to a set of attributes that are obtained after the training phase. In worst scenario, when the model is considered as not valid anymore, then the sources updates it and send the number of gathered samples directly to the sink. Therefore, the Ken approach is well adapted towards a specific phenomena for exploiting both spatial and temporal correlations. In this case, Markov chain model can be used to exploit the temporal correlations. But, it is not the same case for spatial correlations, which are somewhat more difficult to manage. In the sense, require to collect all the correlated data at one node, which then manages the model on correlated phenomena at a particular area. Thus, if the nodes are coordinated, then the communication cost will be reduced automatically. In order to achieve this, the authors provide a disjoint-clique organization solution, in terms of energy-efficiency by greedy approach. The authors of[55] presented a similar approach towards the data predictions, but uses a Kalman filter as a core model.

The extension of Ken et al work improved in [56]. It is implemented a probabilistic database view by Dynamic Probabilistic Model (DPM) exploitation, that

means it consists a snapshot of data arriving model through a user-friendly interface. This sort of hidden database view bringing to the user experience is actually implemented on the basis of model-based views [57].

- **Time-series Forecasting:** The dynamic variation of time series is given by Moving Average (MA), and the Auto-Regressive (AR) or can be termed or used as Auto-Regressive Moving Average (ARMA). These models are simple, yet effective, and can be used in many practical scenarios with better data accuracy. The other sophisticated models such as ARIMA and GARCH have also been follower ARMA, but the complexity of these models does not suit for WSNs.

The Probabilistic Adaptable Query (PAQ) system [58] uses a low-order AR model to answer the probabilistic queries. In PAQ, each sensor has a local AR model and samples the values at every given time instances. PAQ has a learning-phase criteria, where nodes store their sensor readings in a queue buffer, when the queue fills, it then applies the model parameters and send it to the sink. Although, it is limited to the parameters of the model (i.e., AR model co-efficients) for the communication between nodes and sink, which thus exclude the sensor readings. User-specified error bound is associated with each and every model. If the predicted value is found within the error bound, then the model is considered to be valid for the specified sensed data range. Else, one of the following cases are happened. First one is that the outliers can be marked by sampled data, for instance if an error reading found. In this case, outliers can be either sent to the sink or simply ignored at the source level itself. The other model can be marked as invalid, so that it has to be triggered back to the learning phase, and then re-send back to the sent. This only happens when the consecutive readings greater than the given error bound range. Furthermore to the basic principle of PAQ, distributed clustering approach is presented among the group member nodes. SAF method is an improvement of PAQ in two cases. First, They can filter the dat through smooth outliers, and it reduces the impact of outliers while enlarging the size of used data. Second, SAF presents the centralized method while choosing the optimal number of clusters and has a complexity of  $O(n \log n)$ .

In the contrast, the authors of [59] presents the time series forecasting method with an adaptive multiple model selection algorithm, where as the above approaches assumed that only a single model is being used to represent the given amount of data. The basic principle of [59] allows the system itself to choose the model update automatically rather having a-priori knowledge of the phenomena, as it cannot be available.

- **Algorithmic Approaches:** The common approach used in several other data-prediction mechanisms in WSNs, as they used the algorithmic approach in order to get the predictions, which starts from a heuristic or behavioral characterization of the physical quantities.

The first approach [60] in analogy for video compression is applied to the sensor networks. In this analogy, at a given time instant, where a sensor network can

be considered that the thought as an image, where each "pixel" of the image is a sensed data by a given node. As this statement is given the validation point that it is certainly possible to exploit the spatial correlations among the acquired samples. Furthermore, as the sensed data is gradually varied over time, then the evolution of the readings can be seen as a "sensor movie". The authors of PREMON presents a data-prediction method, which is basically on the basis of MPEG encoding. As per the monitoring process, sensor nodes at first send their initial readings to the sink, then the sink computes the model while evaluating the data correlations among the macro-blocks, which then derives a motion vector relative to each and every block. After obtaining the model, sink sends back to the sensors. For the second round, sensors make a comparison between each sample with the prediction model derived by sink. If the sensed data are close to the prediction model within its user specified error, then the nodes turn down transmitting the data towards the sink. In that case, the model considered to be as invalid and out of date. After the expiration, then the model computation can restart from the beginning.

Since PREMON uses a centralized principle. The buddy protocol [61] considers a distributed method while presenting it as an improvement of PREMON for exploiting temporal correlations among the sensed data. In this case, each node has maintained a buddy relationship with its neighboring nodes. As a result, a cluster of buddies with certain numbers are formed so that only a head node can be assigned as a representative of all the buddies. However, buddy protocol operates with two modes, such as default and PREMON. Default mode triggers the nodes to send simply their data to the cluster-head (CH). In the PREMON mode, nodes only sent the model to the CH, and data which do not fit for the predictions. Each node decides by itself whether to trigger the default or PREMON mode based on early estimation of their energy cost, since it is associated to the specific operational mode. The advantage of this mechanism, it is only valid and preforms well if the phenomena is stable, so the number of exchanged packets can be reduced. On the other not, if the sampled data is fast, then the overheads related to the PREMON. In this case, the default mode can be more energy efficient than PREMON.

## **2.9. Energy-Efficient Data Acquisition in WSNs**

As we have explained and classified several sensor subsystems in earlier section 1.4.4. As it is total contrast to the general standpoint view in sensing, as the matter of fact that the sensing not relevant from energy-consumption. But, several methods under data-acquisition techniques are illustrated that the sensing module energy consumption can be greater than the communication or even larger than the rest of the sensor node, as described in section 2.2.

Data acquisition schemes brought up that reducing communication may only be not

enough, yet energy saving methods have to use, in order to reduce the number of data samples (or data acquisition). It is further explained that aiming at energy-efficient data acquisition not only saving the energy, while reducing the energy consumption of the subsystem, and reducing the sampled data among the sensor nodes, and they also suppresses the number of non-related data communication as well. However, many data-acquisition methods have been designed, in order to achieve energy efficiency, under the circumstances of sensor energy consumption, which is considered to be negligible. Nevertheless, the main classification of the data acquisition techniques are given as, namely adaptive sampling, hierarchical sampling, and model-based active sampling.

- **Adaptive Sampling:** Adaptive sampling methods exploit the correlations among the sampled data, in order to reduce the sampled data that to be acquired from the transducer. For instance, variations in data changes gradually over time. In this prospect, temporal correlations of sampled data do not differ much over each other, however, they may be exploited in order to reduce the amount of acquisitions. A similar approach can be used, if the acquired physical phenomena has not changed quickly between the areas that are covered by the neighboring nodes. This prospect helps to reduce the both sampling rate and communication costs while taking the advantage of spatial correlations among the sensed data. It thus clearly shows that it is possible to exploit both temporal and spatial correlations altogether, in further to reduce the amount of data that supposed to be acquired.

Furthermore, adaptive sampling is more related to our propositions. Since we also used both spatio-temporal use case scenarios, in order to suppress the communication costs as well as sampling data rates.

- **Hierarchical Sampling:** A hierarchical sampling defines the trade off between accuracy and energy saving. As, this approach of the nodes are equipped with few different sort of sensors, thus each sensor characterized by its given resolution and its energy consumption. It then dynamically choose by itself that which class to be activated, in order to achieve the trade off between data accuracy and energy preservation.
- **Model-based Active Sampling:** As model-based approach is similar to the data-prediction. A model of the sensed phenomena can be triggered on the sampled data, so that the upcoming values can be forecasted well with the fine grain accuracy. In order to exploit the obtained model by model-based sampling is being used to reduce the number of data samples, as well as the amount of data that being transmitted to the sink, although this is not the main principle of this mechanism.

## **2.10. Summary**

In this chapter, we conducted the literature survey over various data-driven approaches and their certainties. The focus of this thesis mainly on spatio-temporal adaptive sampling techniques. Adaptive sampling has better proximity in WSNs, as it supports both temporal and spatial correlational scenarios at a time. According to this, in any sorts of data-driven techniques delivers a better accuracy in terms of sampled data for the signal reconstruction. Some methods only targets the signal accuracy with complex computational methods, such as compression-based sensing and data-prediction techniques, thus leads to the system complexity.

The techniques that are discussed above based on data-driven, which have been evaluated and tested either analytically or experimentally. However, there is very minimal work only proved in both analytically and experimentally based on real test-bed performances, which are mostly from indoor environments. The challenging aspects of providing the simple solutions for the low-power wireless sensor devices are remain unsolved.

The above literature survey of each and every technique in data-driven has been summarized, the advantages and disadvantages of the methods are also listed in the sections. While most of the data-driven methods are strictly application-based, which are from either time-driven or event-driven/query-driven applications. the following metrics are ensured as the most acknowledged ones, in order to achieve the system longevity in WSNs, such as energy-efficient, communication cost, message cost, accuracy in original data and delays. However, the propositions are considered to measure the following metrics in terms of energy cost, quality of data and total number of reduced transmissions.

# Chapter 3

## Single-hop WSNs: Energy-Efficient Routing Approaches

This chapter presents the background study of routing and the literature survey from various known protocols. In this chapter, increase of energy efficiency based on energy-efficient routing, thus, prolong the network lifetime. As increasing number of batteries in a system would cost the size and system inflexibility, although the second method of using a rechargeable batteries would also cost the maintenance for man power requirement. In this regard, the battery operated devices may also rely on energy harvesting methods, but, it does effect the operational cost as well as the limitation of usage in numerous application scenarios. As the adaptation of energy harvesting in WSNs restricts the applications usage, thus, led to reconsider the energy optimized mechanisms in WSNs in order for prolonging the network lifetime. The energy optimizing mechanisms not only provide the solutions for network lifetime, but also broads the importance of designing the optimized protocols in sensor nodes for better operability.

After discussing the significance of energy optimized protocols in general, and routing and data-driven in particular, as the network management for wireless sensor networks in general applications, routing study will be addressed twofold. Firstly, we discuss the literature survey of existed routing protocols in WSNs. Secondly, we adapt the functionality of hierarchical routing model based on Low-energy Adaptive Clustering Hierarchical (LEACH) protocol, and then present the improvements over LEACH structure in order to modify the mechanism and carry out the structure for the proposition needs. Third, the summary of the chapter and follow up in chapter 4.

### 3.1. Introduction

Advances in WSN technology, low power electronics and low-power radio frequency enable the low-power sensors, which can be connected to the wireless network. Small sizes and communication costs are an important goal in the sensor networks. Sensor

network nodes are equipped with sensing, computation and communication modules and limited power resources of the batteries. Wireless communications and computational modules consume significant amount of energies, sensor nodes should be capable to manage their battery power levels. The required low energy resources such as less memory capacity, low transmit power and less processing. In this process, it is necessary to use communication protocols to enhance the nodes lifetime and reduce the bandwidth.

In sensor networks, routing protocols have the higher network performances in terms of network lifetime, less transmissions and low latencies than the traditional communication methods. The primary design goals of routing WSNs is to maximize the network lifetime while providing the reliable data transmissions by utilizing different energy-efficient routing techniques. The design of routing protocols in WSNs has influenced by many energy-efficiency factors. Energy-efficiency is a basic challenge in low-power devices, and low power is a key concern for these sensors and actuators as to keep them functioning for several months to years without interruption. As the classified routing network protocols as follows from [62, 38]: topology, data-centric, hierarchical, and location-based routing. The network structure can play a prominent role into the network's operation of the routing. In our case, we consider to design the propositions in hierarchical cluster-based networks, which then follows up the data-centric topology in chapter 4. Then there is a great amount of research study being conducted on energy efficient routing protocols, and several well-known works have also been studied in the literature, such as [63, 64, 65, 66, 67, 68, 69, 70, 71].

However, processing the sensor node data through a centralized architecture demands a significant bandwidth for communication all of the sensor node information, which adds an additional communication burden on the sensors battery powered devices [72]. On the other hand, centralized systems are high latency and lack of robustness due to the internal system increased traffic. In order to reduce the communication overheads distributed systems are considered to design our propositions in WSNs. In single-hop distributed architecture, a set of cluster-head (CH) nodes are chosen based on their closest path to the base station (BS) and their highest energy levels, which can collect the information from their corresponding cluster nodes to aggregate the data before forwarding it as a single data packet towards the BS.

This chapter presents a Single-hop Clustering and Energy Efficient Protocol (SCEEP) for prolonging the network lifetime based on Cluster-Heads remaining energy and their optimal CH range levels. It is mainly considered to reduce the communication overheads and unnecessary cluster-head changes at every round. In the rest of the chapter, it organizes it as follows, section 3 focuses on the literature studies of routing protocols and their simulation scenarios, and then presents the summary of studied routing techniques. Section 4 explores the simulation study and its impact on wireless sensor networks. Section 5 presents the radio energy consumption model based on LEACH, and the SCEEP proposition and its design structures of the algorithm have also been described. Simulation results and comparisons are discussed in section 6. Final section concludes the chapter and suggestions for the future works

are given.

## 3.2. Motivation

The energy efficiency in sensor nodes is the combination of several methods, such as routing, data-driven, duty-cycling, and mobility. However, it is not completely possible to combine all the methodologies into one, due to their complex nature and system inflexibility. It is only possible to adapt the methods at some extent, like routing and data-driven or MAC and routing or routing and Mobility, vice versa, because the methods usage may vary depending upon the application that operates the tasks. In this case, the propositions mainly focuses on routing and data-driven approaches in order to achieve the energy optimization in WSNs.

This chapter focuses on the study of routing in order to extend our thesis further direction towards data-driven methodology, and then combined both the studies to modelize our propositions in chapter 4. Since, we believe routing is an primary parameter to follow up other energy-efficiency studies in WSNs. For instance, whether it comes to data-driven or duty-cycling or mobility, routing must be concatenate with any of these studies to interface the link or share the acquired data with other sensor nodes or base station. Routing is basically for maintaining the good communication links between sensor nodes, relays and the BS. Although, there extensive research study of routing has been discussed in [38] for both single-hop and multi-hop WSNs. This dissertation pursues the routing study, in particular, hierarchical clustering to adapt and design the propositions based on its drawbacks. Because, this chapter is a primary work for chapter 4 in order to study and adapt the routing strategies into the data-driven approaches. We consider the first cluster-based routing approach of LEACH [64], and then present an improvement over LEACH protocol. In particular, LEACH protocol changes its CHs every round among all the nodes irrespective of their current energy levels. However, there is a vast amount of protocols presented based on LEACH. In regard to the proposition, we consider the problem of node continuity and its remaining energy levels whereas others omitted.

## 3.3. Background and Related Works

This section presents the background study and their related works of the routing protocols. The classification of routing protocols are discussed in [62]. The topology-based control can be sub-classified as link-state and distance-vector protocols. In link-state protocol, every router node first gathers the neighbor node information, and then floods the collected information towards the network, like Open Shortest Path First protocol (OSPF) [73], and Optimized Link State Routing (OLSR) [74]. Hence, every node of the network has the global information of the topology, each node checks its next hop shortest path through a routing table by a metric. In distance-vector routing protocols [75, 76], each node exchanges and keeps the routing information with its

neighbors rather than the mapping information of the network. However, each node knows its own routes and aware of its neighboring node routes as well, but they are not aware of its entire path to a final destination node. On contrary, link-state protocols are faster convergence and less overhead, but distance-vector routing protocols are way simpler, require less storage and computational resources. Since the nodes are contained with limited resources for both storage and computational, distance-vector protocols are preferred to be a good alternative for WSNs.

The following literature explains the topology-based routing protocols with the effect of both MAC and routing protocols. Which then shows their performances in terms of energy-efficiency, reliability, and delay based on their application specific requirements.

- **CTP** [77]: Collection tree protocol (CTP) is one of the well-addressed recent protocol, which targets to be designed reliability, efficiency, robustness and hardware independence. It works mainly based on three metrics such as Route engine, data forwarder and link estimator. The authors have given the great insights while testing CTP on several real WSN testbeds to evaluate the packet delivery ratios for reliability, efficiency and robustness of the nodes between point-to-point links.

CTP considers to build Expected transmissions metric (ETX) for estimating the costs between hops based on cost of their next hop plus their links to the next hop and also used it for finding potential parent nodes. Like CTP, COOJA also used this ETX metric in collection protocol for calculating link costs between hops. CTP has also done a series of experiments on 12 different real testbeds to find the dominant causes of collection protocol performances in practice, which are link dynamics and transient loops. However, CTP may have a load-balancing problem because whatever the node's having a good quality link, then the node can be selected as a preferred parent and consumes more energy. Moreover, CTP uses a default maximum number of retransmissions is 32 times, which causes the delay and might not be very useful for controlled applications [78].

- **BCP** [79]: Backpressure collection protocol (BCP) presents the dynamic backpressure routing techniques in wireless networks. Like CTP, BCP also follows the same way of routing and forwarder decision is taken on basis of a per-packet by computing through a backpressure weight of each outgoing link, which is a function of localized queue and link-state information. Thus, the overheads of BCP algorithm are depending upon the forwarding nodes of the next hops. However, the BCP does not consider to prevent the loops of routing, which may cause to incur large delays in the network.
- **LWB** [80]: The low-power wireless bus (LWB) protocol is another technique to provide the solutions to avoid collisions between the nodes traffic, it considers fast glossy floods and time triggered operation techniques, it also follows the scheduling in accordance with the global communication schedule.

Some of the major challenges have been addressed in MAC routing protocols, which causes in the operations of node state conditions. Low power, Low Delay: Opportunistic Routing duty cycling protocol is presented the Duty-cycled protocol stack, which transmits the packets opportunistically in duty-cycled sensor networks [81].

- S. Bade et al, [82] introduces an Adaptive synchronization method is introduced to reduce overall power consumption by allowing small duty cycles while considering the overhead design factor for environmental Wireless Sensor Networks.
- **CMIMO** : Coined cooperative multi-inputs-multi-outputs (CMIMO) presents a distributed MIMO-adaptive energy efficient clustering and routing scheme, which named as coined cooperative MIMO (CMIMO). This paper considers two CHs in every cluster, namely as Master CH (MCH) and Slave CH (SCH). It has four types of transmission modes to control the transmission powers on a per-packet basis, such as SISO, MISO, SIMO and MIMO. They also made the comparison with non-adaptive clustered WSNs, [83].

Goals Minimizing total energy consumption in multi-hop networks. CMIMO only does the comparison with DCA in terms of total energy consumption. It does not prove and discuss about the throughput efficiency and communication overheads.

The data-centric routing protocols typically design to perform same role and common tasks by all the sensor nodes. In general, WSN contains with hundreds to thousands of nodes, and which is also feasible to have a global identifier to every node. This application-centric consideration has led to data-centric protocols. The basic aspect of data-centric is that if the communication is on application-specific data rather than the traditional IP global identifiers, then the routing, storage, and querying algorithms can be performed more efficiently. This data-centric methods towards routing saves additional amount of energy from the communication overhead of binding identifiers. Which then facilitates in-network processing by data aggregation and compression. However, data-centric routing is not well addressed mechanism in terms of dealing complicated queries, and only works well with static nodes. On the other hand, it is also not scalable for large sensor networks. Although the energy dissipation issue is strongly depending upon the traffic patterns, and fairness of the network resources. Two well-known data-centric routing protocols are sensor protocols for information via negotiation (SPIN) [84] and directed diffusion [85]. These two protocols use the data negotiation technique to can save the energy by suppressing the redundant data transmissions between the nodes. There has been an extensive similar research work contributed by other protocols while following these two protocols concept. In the following section, we then describe some insights and highlight the key ideas of these two protocols.

- **SPIN:** Instead of sending all the processed data, SPIN protocol uses high-level descriptors called as meta-data. In this technique, sensor node operates itself as more energy-efficiently while suppressing the redundant data transmissions. SPIN contains three sort of messages, namely as advertisement (ADV), request (REQ), and DATA. Sensor node firstly broadcasts an ADV message, which contains a descriptor of DATA message, if any neighbor interested for the DATA packet then the node sends the DATA to that requested neighbor node, and the other interested nodes also receive a copy of the DATA message. Nevertheless, SPIN solves several problems, such as redundant data passing, overlapping of sensing areas and resources blindness using meta-data negotiation. Thus, it then proves that SPIN achieves a good energy-efficiency as well as suppresses the redundant information of the network. Furthermore, the topological changes are required for localization, since each node should need to know its neighbor hops only. However, there is no guarantee in reliable delivery of the data. For example, distance-level nodes that are interested of the data mostly unable to reach the source nodes, in that case relay nodes either can not be connected or not interested to deliver the data to destination nodes from the sources, then the such data will be lost all the time.
- **Directed diffusion:** The directed diffusion mechanism uses the attribute value pairs of data to query the sensors on the demand basis by the pairs. Thus, the sink broadcasts its interest for the data to sensor nodes. When the nodes receive the sink interests, they then can do the caching for later use. Moreover, the intermediate nodes aggregate the data on the basis of data type and attribute value pairs. However, the interest entry has several gradient fields, then the receiver interest of a sensor sets up a gradient towards the sensor nodes by its receiver's interest. The each gradient adds the data rate, duration and expiration time. Based on interests and gradients, source paths are established between the sinks and nodes. Then the sensed data can be reversely propagated through the path of the interest. This protocol can be established several paths so that anyone of them is used by reinforcement. Therefore, if the path between source and sink fails then the alternate path could be chosen automatically. Hence, the key feature of on-demand query concept in directed diffusion differs from SPIN protocol. With the feature of point-to-point communication, directed diffusion protocol does not need any node addressing mechanism. Since it is energy-efficient and on-demand, which also does not maintain any global network topology. It then proves that it is energy saving mechanism with the utilization of in-network processing method. However, it might require some extra overheads in sensor nodes because of its matching process for data and queries of the nodes. Since it is query-driven data protocol, and may not be feasible for some applications like environmental monitoring.

The hierarchical routing protocols, also known as cluster-based routing. Which can also classify the node roles based on their higher residual energy levels, and the roles can be assigned as cluster-head and member nodes frequently. Each header is

a respective group head of the members, which can collect and aggregates the data from the corresponding members before forwarding it as a packet to the BS. Thus, hierarchical routing lowers the energy consumption within the cluster members while performing the data aggregation or fusion by head nodes to decrease the required number of transmitted messages. Furthermore, it helps the system to cope with additional load of the network to cover a large area of application interest without lacking the service. Nevertheless, hierarchical routing brings the overall system scalability, efficiency and network's lifetime. This prospect helps other literature studies may choose to work with the extensive cluster-head methods or channel allocation mechanisms rather than multi-hop scenarios for the routing. Although, hierarchical routing suffers with communication overhead problem due to the frequent cluster-head changes. However, earlier literature studies have demonstrated that hierarchical routing saves better energy consumption and performances compared to the flat network topologies for large-scale WSNs. Moreover, most of the hierarchical routing lacks the node mobility, network partitioning and hard to adapt it for time-critical applications due to the communication overheads and often CH evaluation procedures.

Moreover, the well-known literature protocols from various topologies that are discussed and listed in the following Table 3.1. The classification has been made based on energy  $E$  and delay  $D$  metric parameters, among these. The symbol of  $\odot$  denotes the protocols that are designed by considering the parameter indication  $\oplus$  which have also been experimentally validated in sensor nodes. The indication of  $\otimes$  explains that the protocols are designed by considering the indications of column  $\oplus$ , but they have not experimentally validated in sensor nodes. The indication of  $\ominus$  denotes the non-consideration of the parameters that is neither considered nor presented in the protocols. The symbol of  $\otimes$  also denotes that the protocol does not include the indication, but the simulation and experimental results included it. However, many protocols have done the investigation based on the strong assumptions of these metric performance indicators. The "Relay" term indicates the relay region consideration of the protocol design based on the location information. Moreover, the reliability metric performances only showed by the topological based protocols that are listed in the table, but there is no indication of this metric in hierarchical clustering protocols, yet PEGASIS only simulates and experiments the results without including the indication.

- **LEACH:** Low Energy-Adaptive Clustering Hierarchal (LEACH) [64, 86] is one of the first clustering technique protocol for different wireless sensor network applications. The basic idea of this algorithm works based on setup and steady-state phases. The idea itself of this mechanism motivated many other hierarchical routing protocols to extend or cover up the drawbacks of LEACH protocol. In LEACH is purely distributed, and the sensor nodes organize themselves into local clusters. LEACH has a randomized CH algorithm technique to change the CHs to balance their nodes energy dissipation.

In the set-up phase, every node decides by itself whether or not to become a cluster-head or master node based on a random number [01]. If the number is less than the following CH threshold function then the node becomes a CH for

Table 3.1 – A list of well-known routing protocols in WSNs.

Protocol	Class	Access Scheme	Analysis	<i>E</i>	<i>D</i>
CTP	Routing	Topology	⊙	⊕	⊖
BCP	Routing	Topology	⊙	⊙	⊕
ORW	Routing	Topology	⊙	⊕	⊖
LWB	Routing	Data-centric	⊙	⊕	⊖
SPIN	Routing	Data-centric	⊗	⊗	⊖
Directed diffusion	Routing	Data-centric	⊗	⊗	⊙
LEACH	Routing	Hierarchical	⊗	⊗	⊖
TEEN and APTEEN	Routing	Hierarchical	⊗	⊗	⊖
PEGASIS	Routing	Hierarchical	⊗	⊗	⊖
HEED	Routing	Hierarchical	⊗	⊗	⊖
LEACH-DCHS	Routing	Hierarchical	⊗	⊗	⊖
LEACH-SWDN	Routing	Hierarchical	⊗	⊗	⊖
GAF	Routing	Location-based	⊗	⊗	⊙
GEAR	Routing	Location-based	⊗	⊗	⊖
SPAN	Routing	Location-based	⊗	⊗	⊖
GeRaF	Routing	Location-based	Relay ⊗	⊗	⊗

this current round, and the threshold function  $T(n)$  is given as

$$P_i(t) = \left\{ \begin{array}{l} k \quad \text{if } n \in G \\ N - k(R \bmod \frac{N}{k}) \quad \text{otherwise} \end{array} \right\} \quad (3.1)$$

Where  $P$  is the predetermined percentage of cluster heads (e.g.,  $P = 0.05$ ),  $r$  is the current round in the network,  $G$  is a list of nodes that have not been cluster heads in the last  $1/p$  rounds. In this set-up phase cluster heads use a CSMA-MAC protocol to broadcast an advertisement message to all non-cluster-head nodes. Each node decides by itself which cluster it will belong to, and then node informs to cluster head that it will be a part of this cluster. This decision has been depended on the received signal strength of the advertisement messages. In the steady-state phase, LEACH uses TDMA slot method to do the data communications between the nodes and cluster-heads. In this scenario, cluster heads compress the collected data from all non-cluster-head nodes, and then send it as an aggregated data packet to the base station.

1. LEACH assumes that all nodes have enough power to transmit to reach BS, if needed that each node has computational power to support different MAC protocols. Therefore, it is not feasible to deploy the large region of networks.
2. Cluster-head changes occurred unnecessarily, even there is a higher

amount of remaining energy left on the nodes, and it does not count remaining energy levels of the nodes before changing the cluster-heads.

3. Cluster-head nodes obviously run out of energy very soon due to the set-up phase of every round, during the set-up phase for cluster formation or cluster-head election, it always consumes lot of energy to do the formation.
4. There is no equal size of clusters, because of uneven node distribution or extra load, some of cluster-heads have higher energy consumption, which then depletes earlier.
5. LEACH can not be applied for the time-driven applications, and presents the hot-spot problem.
6. The idea of dynamic clustering brings extra overhead (cluster-head changes, advertisement and so on.) may diminish gain in energy consumption.

LEACH has done the initiation in the area of hierarchical clustering routing protocols for Wireless Sensor Networks (WSN), which has a better performances to increase the network lifetime when compared to the traditional or direct transmission networks. It is also being motivated towards many other well known clustering protocols, which have done further improvements based on LEACH. LEACH has many features and issues to resolve while using LEACH protocol, such as evenly energy load distribution in the network, self-organizing and randomized rotation of the cluster-heads.

- **TEEN and APTEEN:** Threshold-sensitive Energy Efficient sensor Network (TEEN) protocols [87, 88]. TEEN is a hybrid protocol desined in a both combinations of hierarchical clustering and data-centric approcheds are considered to develop mainly for the time critical applications, based on LEACH with multi-level cluster-heads. Since the network architecture is based on a hierarchical clustering, which monitors to the sudden changes in the sensed attributed and responds correspondingly. It works based on hard and soft threshold levels, which can be broadcasted by cluster-heads to the nodes. In this scenario, hard threshold sensed the value when the data change is greater than the soft thresholds, a value that triggers nodes to transmit. These two protocols reduce the number of transmissions and also increases accuracy of the system. Both outperforms well than LEACH, but due to the multi-level clusters and threshold mechanisms, system complexity has been increased. However, TEEN could not be used for periodic applications, where periodic reports are needed because the values of the attributes may not reach the threshold level at all. Furthermore, sometimes the distinguish between dead nodes from alive nodes is inevitable at sink node. In TEEN, CHs do the message propagation well unless if they are not in the each other's tranmission range, then the message can be lost. The adaptive threshold sensitive energy efficient sensor network protocol (APTEEN) [88] is being used same architecture and an improved version of TEEN, and then it

targets at both periodic data collection and reacting to time critical application events.

The main drawback of these two protocols, if the thresholds are not received, then the nodes never communicate and there will be no packet transmissions to the BS at all.

- **PEGASIS:** Power Efficient Gathering in Sensor Information Systems (PEGASIS) [66], PEGASIS also designed and developed based on LEACH. Unlike LEACH cluster-head threshold algorithm, PEGASIS only works single cluster-head chain construction method, during every round it elects its cluster-head among the closest nodes of the base station. It is purely multi-hop communication protocol. It has proven that it enhances network lifetime much better than LEACH. The basic principle of the PEGASIS, in order to extend the network lifetime, nodes can only communicate with their closest neighbors, and they take turns to communicate with the base station.

Somehow, PEGASIS overcomes the clustering overhead issue, it still needs dynamic topology adjustment that node needs to about energy status and location of its neighbors to route the data. Such topology adjustment occurs significant overhead especially highly utilized networks. PEGASIS also assumes, each node can directly communicate to the BS. In real scenarios, sensor nodes use multi-hop communication to contact the BS. Moreover, PEGASIS network nodes can maintain complete info of other neighbors location and energy information to route the data. An extension of PEGASIS called, Hierarchical-PEGASIS introduced with objective of decreasing delay incurred for packets during the transmissions to BS. It is also proved that it outperforms than PEGASIS.

- **HEED:** Hybrid Energy Efficient Distributed (HEED) [63], protocol has developed with four primary goals such as
  1. Prolonging network lifetime by distributing energy consumption
  2. Terminating clustering process within a constant number of iterations/steps.
  3. Minimizing control overhead.
  4. Producing well distributed cluster-heads and compact clusters.

HEED proposes to select new CHs based on  $T_{cp} < T_{no}$ , where  $T_{cp}$  is clustering process interval and  $T_{no}$  network operation interval is the time between the end of old  $T_{cp}$  interval and the start of the subsequent  $T_{cp}$  interval. The condition  $T_{cp} < T_{no}$  must be ensured to reduce the overhead. HEED also considers the probabilistic ally selected CHs based on their residual energy and an initial rate of cluster head  $C_{prob}$  among all nodes, which is only used to limit the initial cluster-head announcements, and has no direct impact on final clusters.

During the start-up execution of HEED, it uses CH probabilistic model to elect CHs as given as

Where,  $E_{res}$  = Estimated residual node energy level  $E_{max}$  = maximum node energy level (fully charged battery)

In this protocol, node joins request takes place based on intra-cluster communication cost. HEED introduced two categories to define three communication cost functions.

Cluster properties, such as cluster size, it also includes two cost functions, namely as (a) minimum degree cost, means that a node joins the CH with minimum degree to distribute cluster-head load (possibly at the expense of increased interference and reduced spatial reuse), (b) maximum degree cost, nodes join CH with maximum degree to create dense clusters.

Whether or not variable power levels are permissible for transmission within a cluster, i.e., if each node is allowed to use minimum power level to reach its CH. In this case, the authors have defined Average Minimum Reachability Power (AMRP) that means the nodes minimum required power levels to reach a CH within the cluster range.

To follow the HEED communication costs and principles, Extended HEED (EHEED) have introduced another criteria of communication cost, which is among the non-CH nodes.

- **LEACH-DCHS** [89]: LEACH Deterministic Cluster-Head Selection (DCHS), is also another improvement of LEACH protocol. This algorithm has done the improvement based on two cluster-head threshold functions are presented in [89].

LEACH-DCHS has considered to improve the network lifetime based on three measurement metrics such as First-Node-Dies (FND), Half-Nodes-Alive (**HNA**), and Last-Node-Dies (**LND**). Based on the first modification of LEACH-DCHS equation (2) proven that the network lifetime increased by 30 percentage for FND and 20 percentage for **HNA** metrics. Nevertheless, network stuck after several rounds of simulation with Eq (2). To overcome this problem, LEACH-DCHS has been developed with Eq (3), even though LEACH-DCHS is unable to keep the optimal cluster head ranges while increasing the number of rounds.

- **LEACH-SWDN** [90]: LEACH-Sliding Window Number of Nodes (LEACH-SWDN) considered the drawbacks of LEACH-DCHS, and it has proposed and proved that to keep up optimal cluster-head range until **LND** (last node dies). Unlike LEACH random numbers, it considers sliding window approach to change the cluster-heads per round. For keeping optimal cluster-head ranges all the time, it works similar like LEACH-DCHS Eq (2).
- **DEEC**: [70] Unlike LEACH homogeneous clustering networks, Distributed energy-efficient clustering algorithm (DEEC) presents a heterogeneous cluster-based network model based on LEACH hierarchical model. However, DEEC

protocol considers different sorts of nodes at various levels for collecting the data towards the BS and reorganizing the nodes during every round. The nodes contained with different platform heterogeneous nature, and indeed, have held with different radio ranges as well. Moreover, we consider DEEC protocol in order to make a comparison between the proposed protocol SCEEP, DEEC, and the other protocol of DDEEC, in terms of number of nodes alive over time. Although, the proposition is purely homogeneous network.

- **DDEEC:** [71] Developed DEEC (DDEEC) is based on DEEC, and it presents a CH load balancing method over the network. The network contains normal and advanced nodes, which were considered as CHs, latter, it used a CH election algorithm when the advanced nodes run out of their energy.

In location-based protocols, each node aware of its own and neighbor nodes positions. The message sources are informed about the position of the destination node for energy-efficient routing path evaluations. Location information of the nodes are much needed for calculating or measuring the distances between any particular two nodes so that energy estimation can be easily calibrated. In some of the location-based applications, query can be propagated to a particular group by utilizing their sensor locations to eliminate the amount of transmissions significantly. There is a small low-power GPS receiver [14], which can be used for finding their locations in the nodes. The distances of the neighboring nodes can be estimated by using their received signal strengths. The fundamental issue using location-based routing is the availability of its accurate node positioning systems of GPS cards, which is not compatible with the current technology. However, many literature studies of location-based routing even stated that the communication costs rely on geographic position of the nodes. On the other hand, [91] stated that the information of geographic position of node is not sufficient to define the communication costs. Therefore, literature describes that it is almost not feasible to model the wireless channel for indoor offices due to the fact, too many fickleness that we need to consider. Some earlier literature contributions of location-based protocols are given as minimum energy communication network (MECN) [92], geographic adaptive fidelity (GAF) [14], and geographic energy aware routing protocol (GEAR) [93]. Nevertheless, some of the protocols are mainly designed for mobile ad hoc networks, as the following protocols can be explained and can provide a detailed information of this topic.

- **GAF** [14]: GAF is a location-based energy-aware routing protocol, which is specifically designed for mobile ad hoc networks. Each node associates itself to point in the virtual grid by using GPS card. Nodes, whoever has the same point on the grid are considered to be having an equivalent cost in terms of the packet routing. The basic principle of GAF is making a collaboration between nodes, and the nodes can be played different roles in each zone. GAF works based on three states of the nodes, such as neighbor discovery in the grid, active participation in routing, and sleeping mode whenever the radio is turn off. Thus,

the routing parameters and time intervals of the states are depending upon the applications and tuned them during the routing process accordingly. Furthermore, the sleeping neighbors manage their sleeping schedules according to the routing fidelity and load balancing. This leads to increase the network lifetime by increasing the number of nodes. In mobility scenario, each node estimates its leaving time of the grid and forward this information to the neighbors. However, the time period of the node expires, then the sleeping nodes automatically wake up, and then become as one of the active member. It then keeps the network as connected by maintaining a cluster-head, which always in active mode for each region in its virtual grid. Simulation results of GAF shown that GAF performs as well as a basic ad hoc routing protocol in terms of latency, packet loss, and maximizes the network lifetime by saving energy. On this note, GAF CH's do not support neither aggregation nor fusion like hierarchical routing protocols.

- **GEAR** [93]: The basic principle of GEAR uses geographic information while spreading queries to particular regions rather all since data attributes contain geographic information as well. Especially, GEAR restricts the number of interests in directed diffusion towards the whole network, and it chooses specific regions rather than sending to all regions. In this scenario, each node learns and keeps an estimated transmission cost of reaching the destination by its neighbors. This estimated cost includes the combination of residual energy and its distance to the destination. The learned cost is an improvement of the estimated cost, which helps to discover the routing around holes in the network. Hole usually happen when there is no any closest neighbor of the node to the target region then itself. If the holes are empty there, then the estimated cost must be equal to the learned cost.

Geographic Energy Aware Routing (GEAR) [93] protocol is a recursive data dissemination protocol, disseminate queries to appropriate regions whose data include geographic attributes. It achieves energy saving by sending the interest to certain region than the whole network. On the contrary, GEAR is not scalable and does not support data fusion.

- **SPAN** [16]: SPAN protocol is a kind of similar to GAF protocol, since it triggers a certain amount of nodes in a particular area at any given time instances. SPAN elects the nodes as coordinators based on their geographic positions, and the coordinators are being used to form as a network backbone for forwarding the messages. In SPAN, node is used to become a coordinator when the two neighbors of a non-coordinator unable to reach each other either directly or by one/two coordinators. The drawback of SPAN is that it consumes more energy when the number of nodes are being increased.
- **GeRaF** [15]: GeRAF presents the better solution with the combination of both routing and CSMA/CA mechanism used as a MAC layer. It needs both the nodes and their neighbors location information. Nevertheless, during the transmission request time among the nodes that awake can be chosen a forwarding

node. Thus, routing procedure consumes more energy and increases the network latency.

### 3.4. Radio Energy Consumption Model

The same first order radio energy model has been used from one of our fellow doctoral thesis [94], as the model originally from LEACH first order radio model [64]. However, we have modified the model with additional energy parameters [94], as shown in Figure 3.1. The parameter values, where the radio dissipates  $E_{fs} = 50/nJ/bit$  to run the transmitter and receiver circuitry and  $\epsilon_{mp} = 100/pJ/bit/m^2$  for the transmit amplifier as fixed. Therefore, radio can consume certain amount of energy to reach the authorized recipients by using the power control methods. In the mean while radios can be turned off to avoid the unauthorized listeners.

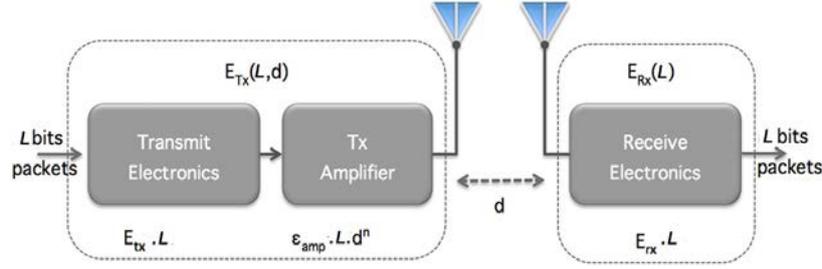


Figure 3.1 – First order radio energy consumption model.

In this model, the following equations are considered for calculating the transmitting and receiving costs of a  $L$  bits per message for node and the distance  $d$  shown below, respectively. The energy parameters of transmitter  $E_{Tx}$  and receiver  $E_{Rx}$  considerations can be defined as

$$E_{Tx}(L, d) = \begin{cases} E_{tx} \cdot L + \epsilon_{fs} \cdot L \cdot d^2, & \text{if } d \leq d_0 \\ E_{tx} \cdot L + \epsilon_{mp} \cdot L \cdot d^4, & \text{if } d > d_0 \end{cases} \quad (3.2)$$

$$E_{Tx}(L, d) = E_{fs}(L) + E_{mp}(L, d) \quad (3.3)$$

$$E_{Tx}(L, d) = E_{fs} * L + \epsilon_{mp} * L * d^2 \quad (3.4)$$

$$E_{Tx}(L, d) = E_{fs}(L) \quad (3.5)$$

$$E_{Rx}(L, d) = E_{rx} * L \quad (3.6)$$

In this radio model, the reduction of receiver cost for receiving a message is not

only the primary goal. The protocol should also consider to minimize the distances as well as the transmitter and receiver operational cost per message. The length of the packet  $S$  by CH is  $4k$  bits per round.

### 3.5. SCEEP Algorithm Designs and Specifications

This chapter presents a Single-hop Clustering Energy-Efficient Protocol (SCEEP) based on LEACH homogeneous sensor network model with an initial amount of energy among the sensor nodes. The proposition has been considered first order radio energy model [64]. This approach mainly targets to reduce the communication overheads and unnecessary cluster-head (CH) changes at every round. The CH election procedure follows two sorts of clustering principle to elect the CHs among the nodes. The first one follows LEACH algorithm of Eq .3.7 to initialize the random sensor network at time instances  $t$ . The second focuses on the probabilistic CH threshold function in order to elect the CHs among the nodes, and it can be derived as

$$P_i(t) = \begin{cases} \frac{K}{N-K(r \bmod N/K)} & C_i(t) = 1 \\ 0 & C_i(t) = 0 \end{cases} \quad (3.7)$$

Where  $K$  denotes the initial amount of optimal cluster head range,  $N$  is the total number of nodes in the network,  $r$  is the current round and,  $C_i(t) = 1$  if the node  $i$  has not been already a cluster head in the last  $N/K$  rounds at given time instances  $t$ , otherwise  $C_i(t) = 0$ . According to Eq. 3.7, nodes always have the equal chances to become a cluster head for every round, regardless of how much high or little amount of residual energy in the nodes have had. In this scenario, cluster or CH changes occur every round without considering the remaining energy levels of the nodes, which may lead to earlier nodes death and communication overheads or degrading the network lifetime. For avoiding the communication overheads and maximizing the network lifetime, we propose the second principle of clustering threshold function, which has been designed based on CHs remaining energy and fixed optimal cluster-head ranges at every round. However, the recent literature works have presented some advanced energy-efficient protocols in [69], [70] and [71], based on nodes residual energy of the network or using different sorts of nodes in the cluster-topology like a heterogeneous model, in which they achieve only at some extent level to reduce the communication and computational costs.

In this method, CHs check their energy levels and optimal CHs range by the Eq .3.10 before requesting for new cluster-heads election, The following list of features that are presented by SCEEP protocol is described as

1. Initial set-up phase of the network, we use LEACH cluster-head selection algorithm to elect the CHs during its first round.

$$E_i(r) = \sum_{i=1}^K E_{CH_i}(r)K, \quad (3.8)$$

Where  $E_i(r)$  is current remaining energy of the nodes, and  $E_{tot}$  is a total energy of the nodes.

$$p_{CH_i}(r) = \frac{E_i(r)}{E_{tot}} \quad (3.9)$$

The  $p_{CH_i}(r)$  is a probabilistic function to measure the current energy levels of current CHs,  $E_{CH_i}(r)$ .

$$E_{CH_i}(r) = \begin{cases} \hat{P} \geq \hat{P}_{Th}, & p_{CH_i}(r) > E_{Th} \\ P_i(t), & otherwise \end{cases} \quad (3.10)$$

2. After completion of the network first round (in seconds) simulations, current CHs check their energy levels based on the function of  $E_{CH_i}(r)$  to continue their intra-cluster communications with the nodes or not, by using energy threshold function  $E_{Th}$ . According to the network setup, the energy threshold value can vary based on the given initial energy levels of the nodes. If the current CH energy levels is greater than or equal to the energy threshold level, it continues its intra-cluster activities to receive the data among the cluster member nodes, otherwise CH destroys the current cluster and calls for the new cluster formation.
3. In this case, we fixed an optimal CHs threshold function  $\hat{P}_{Th}$ , if the CHs optimum level  $\hat{P}$  is lower than the optimal threshold range, then it goes back to the cluster head election mechanism for electing new cluster-heads. In this regard, we fixed the optimal threshold range at 3.
4. If current CHs are higher than or equal to the minimum required CHs range, then it continues receiving the data from their corresponding member nodes.

**Lemma 1.** Let  $E_{CH_i}(r)$  be a function which can check the current energy levels of CHs and their optimal range  $\hat{P}_{Th}$  of CHs in every round for continuous operation.

*Proof.* We remark that the functions of  $E_{CH_i}(r)$  and  $\hat{P}_{Th}$  play a key role to reduce the communication overheads. Because, the CH threshold function only restarts the setup phase when it meets the requirements of  $E_{CH_i}(r)$  and  $\hat{P}_{Th}$ . The analysis and considerations are different from the corresponding term  $E_{CH_i}(r)$  given in [94] due to the  $\hat{P}_{Th}$  ranges, in order to prevent the cluster reformations for the next round to resume the steady-state phase. But, the operational function  $E_{CH_i}(r)$  is same, and it can be simplified as. While  $C_i(r)$  defines the state of current CH round from Eq 3.7, whereas the probability of current CH is either 0 or 1.

$$\begin{aligned}
E[\#CH] &= \sum_{i=1}^K E_{CH_i}(r) * C_i(r) \\
&= \sum_{i=1}^K K \frac{E_i(r)}{E_{tot}} * 1 \\
&= \left( \frac{E_1(r)}{E_{tot}} + \dots + \frac{E_1(r)}{E_{tot}} \right) K \\
&= K
\end{aligned}$$

We simplify the above functions to check the current CHs energy level over the network.  $\square$

In the setup phase, as explained in the Algorithm 1 3.2, nodes are randomly placed within the field of  $400 m \times 400 m$ . Nodes decide by themselves to become a CH for this round or not based on CH election algorithm function of Eq. 3.7. If the node has better proximity or better amount of energy levels than others, the node can be elected as a CH for that round. However, the only deviation between LEACH CH threshold function and SCEEP algorithm is the CH election procedure repeats only when current CHs run out of energy. Otherwise, the CH continues their steady phase mode while checking their energy levels during every round until it gets depleted. Moreover, in order to check the required amount of CH range to balance the network, SCEEP uses the threshold function of  $\hat{P}_{Th}$  to check the optimal CH range. In SCEEP, the initial phase of clusters reorganization only occurs when the node passes the parameters of Eq. 3.7,  $E_{Th}$ , and  $\hat{P}_{Th}$ , respectively. The remaining setup phase functionality of the algorithm is same as LEACH.

The steady-state phase flow diagram describes the data message bit exchanges between nodes and CHs towards the BS. However, this approach did not use the complete functionality of LEACH, because it has two-tier functionality of single-hop with two different radio range communications. As the short range for the data message transactions between end devices and CHs, and the long-range communications between CHs and the BS. In this scenario, it reduces the communication overheads, as there used to be proper scheduling method for sending the data packets from CHs to the BS. However, the aggregation method at CH nodes is remain same, like LEACH protocol.

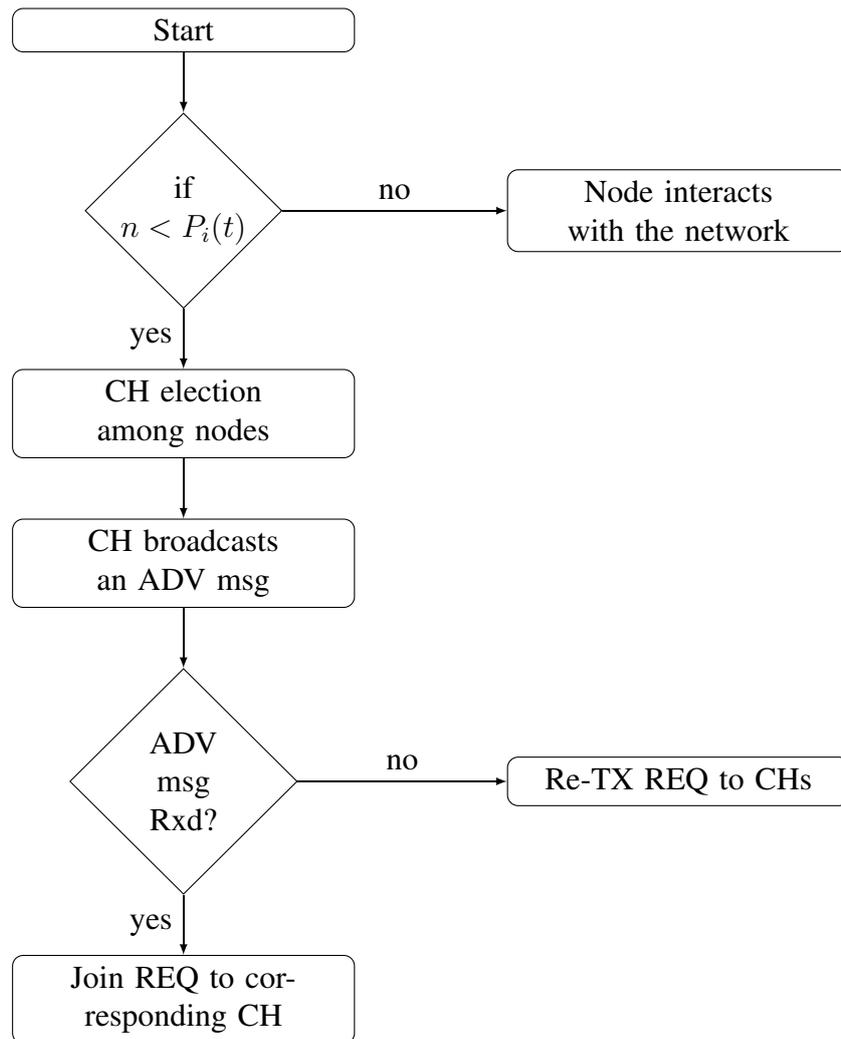


Figure 3.2 – Set-up phase of the network algorithm

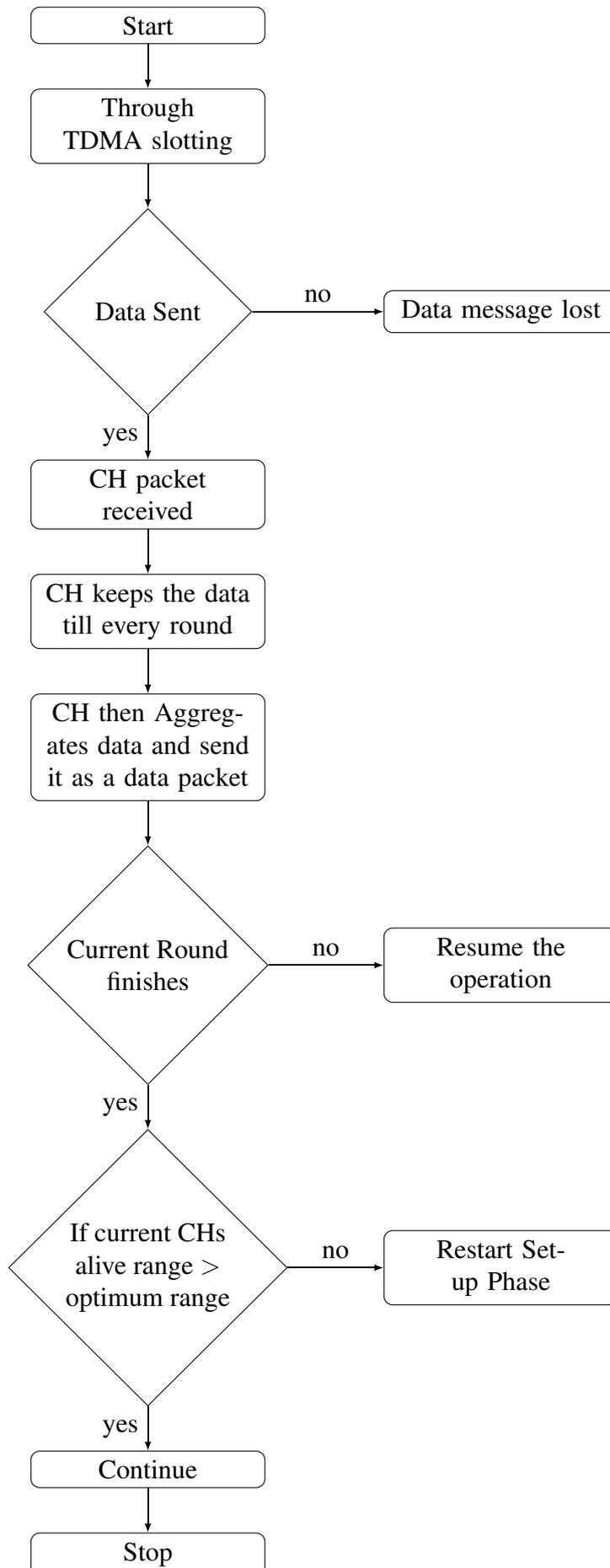


Figure 3.3 – Steady-state phase of the network algorithm

## 3.6. Simulation Results

The SCEEP has initially been implemented and integrated in the network simulator. We have ran several simulation tests to check performances of the protocol. However, as the simulator lacks the further new updates, and the drastic core changes by open source developers, thus, led the inconsistency in results. Then, the propositions are integrated and the results have been conducted through MATLAB simulations. In this simulation set-up, the amount of nodes that has been considered as 51 static wireless sensor nodes including the base station in the field of  $400\text{ m} \times 400\text{ m}$ . The total amount of network energy is  $100\text{ J}$  and the initial node energy is  $0.5\text{ J}$ . Every node can transmit  $4\text{ kb}$  bits message per round to its cluster head node.

### 1. Measurement Metrics

Several measurement metrics has been stated for evaluating the performances of proposed and existed algorithms, such as total nodes alive, network lifetime, number of cluster heads, Energy, number of data packets and Half of the Nodes Alive (**HNA**). According to the simulation results, First Node Dies (**FND**) and Last Node Dies (**LND**) metrics are not being used. In graphs 1 and 2 show that nodes are start dying from first round itself due to that, SCEEP has not been considered to plot by using **FND** metric, and the fact that **LND** metric is not necessary to measure because every last node can obviously remain to perform the clustering algorithm, as it works as a sink node in the network.

### 2. Measurement Analysis

This sub-section provides an analysis of the simulated results, which contained performance observations and comparison between proposed and existed algorithms and measurement metrics. This regards to the network lifetime, Energy and data packets.

Figure 3.4 shows the steady state region performances in the network and plot number of nodes alive over time. It then illustrates the how many nodes alive at different time periods. In this scenario, the network lifetime performance analysis is evaluated between the existed and proposed algorithms. It can clearly show that SCEEP enhanced a network lifetime 30% more compared to the DEEC [70] and DDEEC [71] protocols.

The results of SCEEP simulations are shown in Figure 3.4 and Figure 3.5. The presented metric analysis through both the figures, as per the Fig. 3.5, the packet reception rate of SCEEP 3 times more compared to the DEEC and DDEEC protocols. In Fig. 4.10 SCEEP, where the First Node Dies (**FND**) metric shows at round 172 while other protocols **FND** are at round 28 and 36 respectively. While compared with Half of the Nodes Alive (**HNA**) metric, it has clearly shown the highly varied performances, where the pointed line indicates in Fig. 4.10 at 50 nodes alive, SCEEP extends the network lifetime 500 rounds ahead than the other known protocols. From the above

analysis and demonstration, SCEEP indeed have outperforming results compared to the other protocols. In fact, DEEC and DDEEC are developed with powerful node consideration of heterogeneous network model, such as normal, advanced nodes, and super nodes. As SCEEP only homogeneous network, even though it still maximizes the network lifetime 3 times more at HNA metric compared to the other protocols.

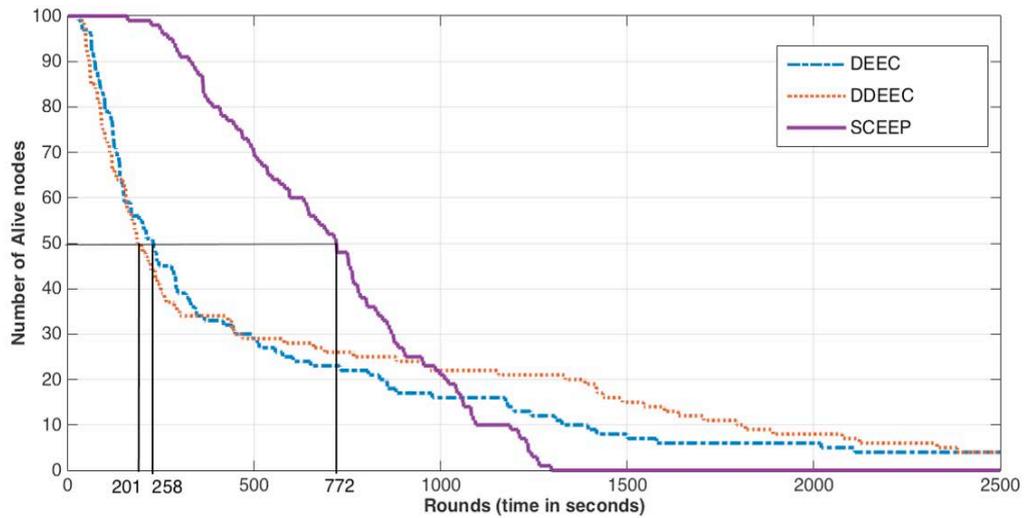


Figure 3.4 – Number of alive nodes over time.

Figure 3.5 delivers a better throughput performances while carrying a better amount of data message bits at total run time simulation. While compared to the DEEC and DDEEC protocol, SCEEP protocol delivers 3 times more data message bits that are received at the CHs. Due to their high time differences between the protocols, since DEEC and DDEEC have been used multi-tier nodes, such as advanced and super nodes, in order to increase the network longevity. In which, the nodes always send their sensed information to their corresponding CHs, which then forward their messages to neighboring CHs, in order for sending out the messages to the BS. However, sharing the same information at multi node level, cause huge time delays, but it is not the case in SCEEP protocol, which delivers the information at CHs, and then aggregates the information before sending as a packet towards the BS.

In Fig. 3.6, it has shown same performed results as Fig. 3.4. The only difference, it presents the *FND* metric over time, which shows the time duration of *FND* metric node dies at different times of the protocols. However, it is not the case in *LND* metric, as the metric describes only last node dies, when *LND* compares with DEEC and DDEEC, it shows that number of dead nodes over time in SCEEP is faster than the DEEC and DDEEC protocols, which is at 1400 rounds. Because of their different powerful network nodes consideration, whereas SCEEP built with homogeneity.

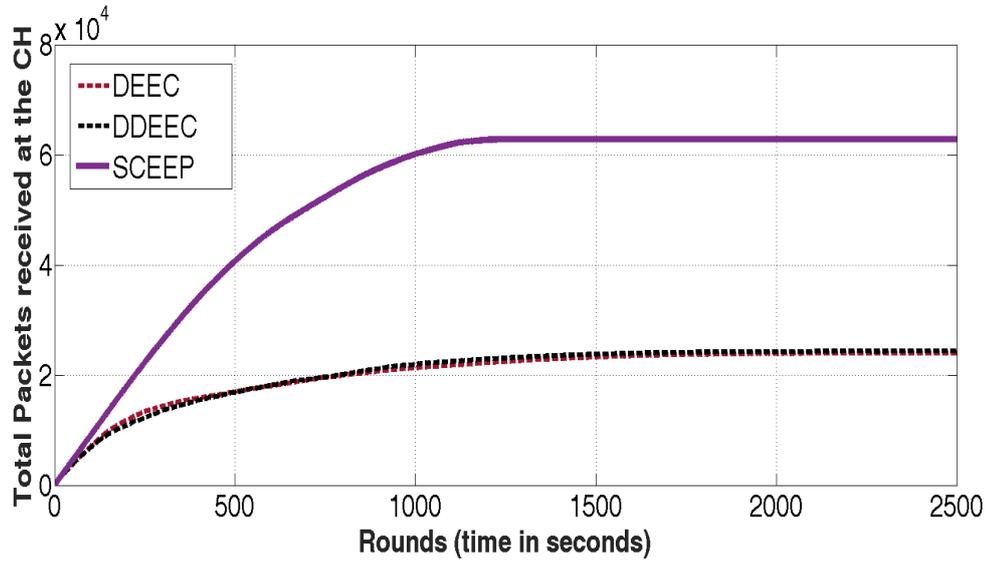


Figure 3.5 – Total number of data bits received at the CH.

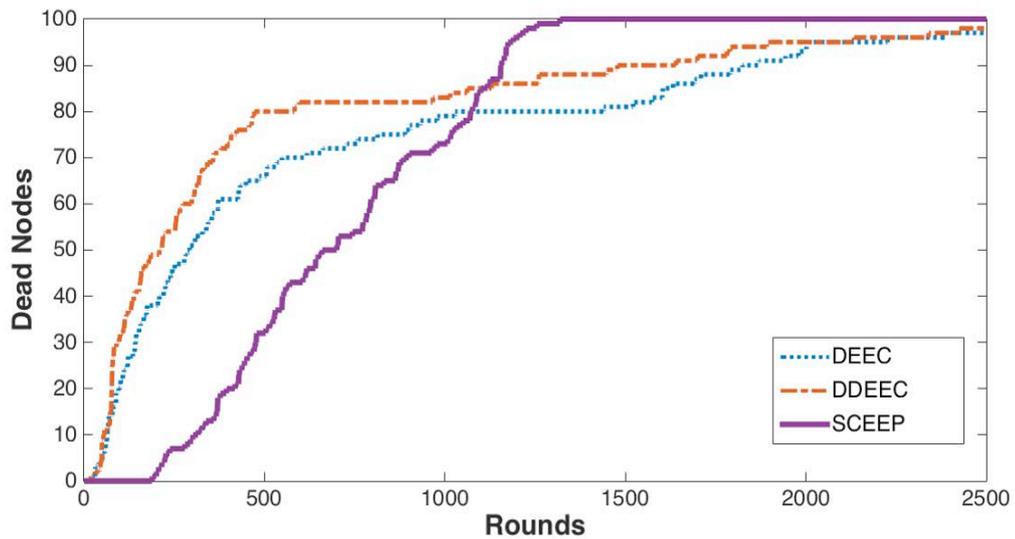


Figure 3.6 – Number of depleted nodes over time.

### 3.7. Background Study of COOJA/Contiki

Contiki is one of the leading operating system (OS) for wireless sensor networks, and it is an open source platform [95], multitasking OS for memory-efficient embedded systems and WSNs, built with a certified IPv6 stack (6LoWPAN) and Rime stack. Contiki provides the IP communication with IPv4 and IPv6, uIP is a small RFC TCP/IP stack, which makes the possibility for Contiki communications over the internet. Rime is designed for low-power radios and it is a lightweight communication stack. COOJA rime provides a wide range of communication primitives through the cross-level network stack.

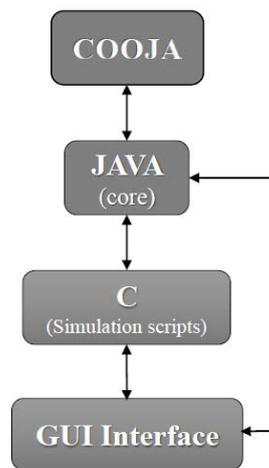


Figure 3.7 – COOJA/Contiki system overview

This section describes an overview of COOJA/Contiki simulator with variety of platforms, such as MSP430 and AVR. Contiki has a complete full implementations of IPv6, which contains with TCP, UDP, RPL and ICMP. Moreover, it further supports IPv6 on IEEE 802.15.4 by a 6LoWPAN adaption layer. It then contains with variety of duty-cycled MAC layers and supports several radio drivers for different sensor nodes and various hardware platforms of the smart objects. Contiki is an event-driven based kernel, which enables energy saving features, such as multi-threading, lightweight stackless thread construction called as photothreads [96]. Contiki serves the primitives in straight forward C programming, rich API libraries and timers. The energy profiling library of [97] measures the time spent over various components of the nodes. This prospect eases the development of sensor network applications, which can provides the accurate insights that where the energy is being spent most during the processing period of the application. Thus contiki provides many more other features in [98, 99], while providing small code footprint is on kilo bytes as flash and a low memory usage of RAM can be configured in tens of bytes.

<sup>2</sup><http://www.contiki-os.org/start.html>

### 3.7.1. Rime Protocol Stack

The rime protocol stack contains with a various set of communication primitives in both the combination of LLC and network layers, ranging from best effort local neighbor broadcast and reliable unicast, best effort network flooding and multi-hop unicast. Furthermore, some of the communication primitives may use on whatever the applications or protocols are running on top of the rime network stack. The communication primitives of rime stack in every layer are shown in figure 3.8 that how they arranged in.

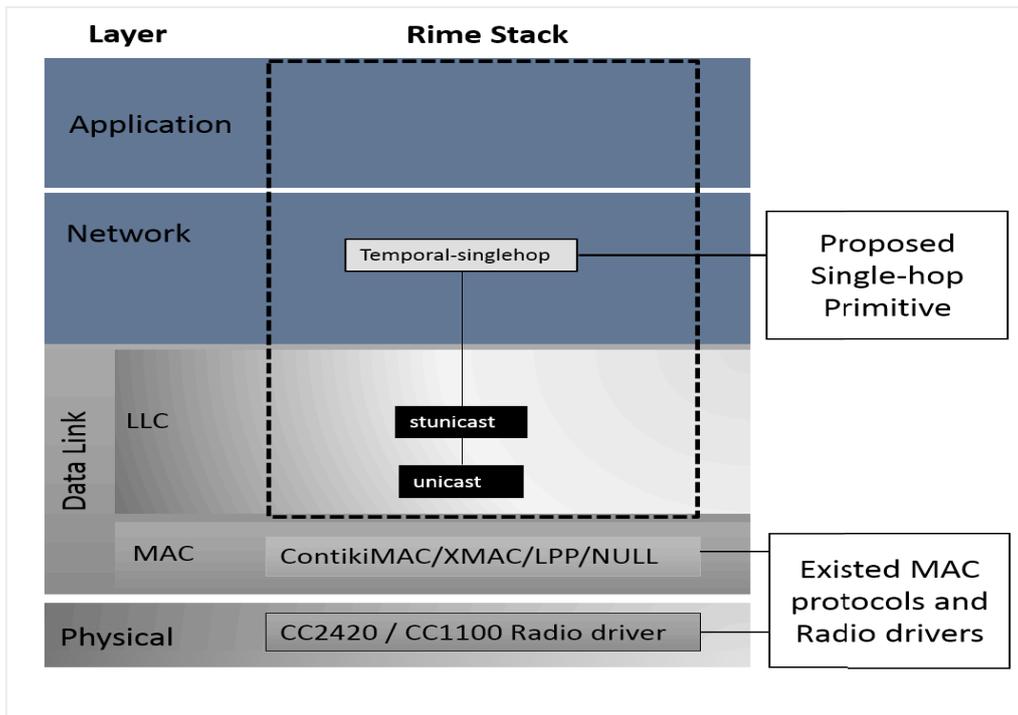


Figure 3.8 – COOJA Rime Network Protocol Stack

### 3.7.2. Cross-level Network in COOJA

The interoperability and code reuse benefits drives the contiki project to present Chameleon, which is a communication architecture for sensor networks. It consists of two parts, one as a rime network stack and another one is a set of packet transformation modules. The challenge for interoperability communication architecture is finding a universal header [100]. As part of our thesis implementations in COOJA rime stack. We further briefly explained about the rime communication stack and its pros and cons of the module utilization.

In COOJA, the rime network protocol stack provides a lightweight communication protocol primitives which are ranging from anonymous broadcast local area to reliable

network flooding protocols [101]. It presents the simpler way of layer designing solutions from the complex communication primitives based on reliable distributed programming [102]. For sensor networks, the simple layer combinations of lightweight principle has several benefits. First, layer simplicity provides the ease in implementations and testing. Second, for memory constrained devices, memory footprints of the implementations are supposed to be small. Third, applications might have the flexibility to express that whatever the communication features they need to be attached to any layer of the stack. In the contrary, heavyweight layered as TCP/IP protocol, it is not feasible to articulate such sort of fine-grained features.

The rime stack of the chameleon produce headers from the packet attributes, by utilizing the packet attributes feature, chameleon transforms it into a standard packet header format. It then also be a compatible architecture with another node, which implements the standard. However, chameleon complains that header transformation alone is not enough to copy another communication protocol.

### **3.8. Design Decisions and Specifications in COOJA**

As discussed earlier, nodes are energy-constrained devices and preserving the energy in WSNs is one of the significant challenge to be achieved. The protocol implementations and testing on the real WSNs is a cost-effective method, in terms of time and resources. With the consideration of these factors, there is manifold simulators with the replication of various real sensor node platforms and their considerations, such as COOJA (based on ContikiOS) [95], TOSSIM (TinyOS) [103], and so on. As this thesis further focuses on to develop and test its propositions in COOJA Contiki based simulator. As there is very minimal work being conducted based on COOJA developments from the WSN literature, and its more realistic simulator network, which then, drives us to design and integrate our further research investigations in COOJA.

As part of our cluster-based topology implementations in COOJA, we initially design and develop static cluster-based single-hop wireless sensor network in COOJA. The complex considerations of LEACH hierarchical method, instead, we present our own way of cluster-based routing technique in COOJA. According to our propositions, nodes are considered as normal sensor nodes (SNs) and fixed cluster-heads (CHs) as super nodes, because, there is no cluster-based topology in Rime network stack. As the rime stack built with based on cross-level network abstractions and does not hold any specific IP node addresses. Unlike IPv4 and 6LoWPAN, it only has 8-bit binary node addresses, it shifts their transmission and reception operation states based on binary bit either 0 or 1. If the node is a sender, then the rime address operated with binary number '1', otherwise '0' binary address will be assigned as receiver sensor nodes.

The earlier simulation scenarios has been conducted in both MATLAB and Ns-2, since both did not contain any real network platforms, and has shown less impact for the research investigations of recent advancements in WSNs. Moreover, both

the simulation networks are restricted to the simplistic propagation models. On the other hand, where COOJA interface designed with various radio propagation models, which gives us better opportunity to integrate and test the propositions through many propagation models, such as Unit Disk Graph Model (UDGM), Directed Graph Model (DGM), Linear Graph Model (LGM), No Radio traffic, and Multi-path Ray Model (MRM). In regard to check the performances in both the interference and non-interference environments.

### **3.9. Summary**

During the initial phase of this thesis, the existed literature protocols have been integrated and tested in popular network simulator 2 (ns-2), such as LEACH, LEACH-C, Static-Clustering, MTE and PEGASIS, which have given us the opportunity to learn their impact and performances in WSNs from such simulators like ns-2. After total integration of these known protocols, we have carried out several simulation tests in order to analyze and validate the protocol performances. Thus, it gives us different experience to reconsider the simulation tools, in order to do further the research integration based on different hands on experience of ns-2. Moreover, ns-2 is an open source simulator and changed its original behavior in terms of performing the results over time to time. As the simulator has given different performances of various protocols from period to another period. Because of drastic modifications in the core of ns-2 from various open source developers, and their source compatible issues in the simulator, thus effects the total environment of simulator, which then leads to the inconsistency in performing results. Nevertheless, there is no further active contributions in ns-2, the organizations of ns-2 shifted to ns-3 which is a new simulator built mainly with IPv4 and IPv6 abstractions other than for cross-level low power devices, and has nothing to do with ns-2.

After the various drawbacks of ns-2, the thesis has moved on to another simulator as COOJA/Contiki, designed based on contiki operating system, since COOJA is designed with multiple choices of node hardware platforms, unlike network simulators (ns-2). Furthermore, the developed codes from COOJA can be reusable directly on recommended sensor test-beds. Thus, drives us to design and develop the propositions, in order to validate the performances through the various metric considerations.

The proposed SCEEP distributed clustering protocol showed better performances while compared to the other protocols, especially in terms of extending node lifetime as well as network lifetime and the throughput over the network. SCEEP protocol also presented a preliminary use case in COOJA while integrating into it. be easily integrated in COOJA, and better performances than the other known protocols. Furthermore, it also presents a simple integration procedure into COOJA for carrying out the chapter 4 propositions, as the thesis focuses more on data reducing techniques while utilizing routing study case, in order to build a new prototype data reduction mechanism in sensor nodes.

# Chapter 4

## Single-hop WSNs: A Data Window Aggregation Techniques

The previous chapter has investigated on hierarchical cluster-based methods, which is mainly for energy optimization in order for prolonging the network lifetime. While following the chapter 1, this chapter presents a mechanism to preserve the energy in sensor nodes while suppressing their radio transmissions and, thus, prolongs the node lifetime as well as the network lifetime.

The data driven methodology of literature has been surveyed in chapter 2, and its motivation for the usage and the impact of data-driven in general, and adaptive sampling for data reduction techniques are discussed. In particular, the energy conservation method has been investigated for wireless sensor nodes, especially, while targeting the time-based applications of continuous or periodic and delay-tolerant. Adaptive spatio-temporal methodology will be presented as threefold. Firstly, we will discourse the data collection and gathering literature methods that are studied for energy saving in WSNs, especially, in relation to this chapter. Secondly, this chapter presents the novel approaches with a focus on architecture dimensioning. Finally, the design structures for the implementation of data window redundancy algorithms, in order to provide a simple and effective data pre-filtration mechanism, thus, reduces the complexity of nodes computation and provides the required measurements for signal reconstruction, and then systematic comparison possibilities.

In addition, we use a first order interpolation method at sink node, in order to achieve the original signal sparsity from the received data by sensor nodes, the interpolation method has been used for signal reconstruction in order to retrieve the original data, as demonstrated in section 4.6.5.

### 4.1. Introduction

In recent advances, WSN field of study growing as one of the most promising technology in many potential applications, such as environmental monitoring, health care,

home/office automation, structural monitoring, industrial, military surveillance, wild-life tracking, and many more. In general, WSN can be deployed and used as battery-powered devices, and as distributed sensor nodes within the network. Usually nodes are energy constrained devices with limited storage and processing capabilities. The sensor nodes periodically update their information towards the base station (BS) or sink node. If the data gathering system is in multihop network fashion among the nodes, which the topology requires an exchange of several messages. Thus reduces the nodes lifetime and depletes the batteries, and it also affects the sensor network lifetime as well. In order to achieve the challenges for periodic and continuous data collection, while maximizing the nodes battery and network lifetime. The data pre-filtration mechanism must be ensured in sensor nodes, in order for tolerating the redundancies in both time and space.

In WSNs, if sensor nodes are used to have similar application tasks to monitor the physical quantities than individual tasks, then the sensed data by nodes are highly, redundant samples in both time and space. This is also noticed in the real datasets of [104, 105] from the indoor deployments. In [106], authors constructed the models on data, and then reported the parameters instead of raw sample values.

The continuous or low-periodic applications utilize the data for two prominent cases, such as surveillance for making live or real-time decisions, and another one is offline, modeling, and analysis [34]. This chapter focuses on the work of delay-tolerant data gathering applications. There has been numerous sensor networks deployed for scientific research [34, 107, 108], which harvests the data continuously for modeling, analysis, and simulations.

This thesis specifically aims at maximizing the network lifetime by following the energy balancing resources usage to nodes. In [109], the main modules of energy consumption in nodes identified as the sensing, computational and communication. However, it has also shown that the radio is a major component which drains out the energy more in the node battery resources, as compared to the computational and sensing. However, its not the case if sensor node contains more than one sensor, as some of the literature research works presented that sensing module consumes much more energy than the communication, if it is designed with multiple sensors [12]. On contrary, according to the literature survey and [110], efficient MAC duty-cycled protocols are significant for operating the radio turn on or off states when they are strictly required [111]. Thus energy-efficiency can be achieved by exploiting the inherent both temporal and spatial redundancies (SRs) in data, such as temperature, humidity, light and pressure exhibit slight variations slowly over the period of time [112]. Moreover, this process may also balance the usage of node computational capabilities, which profitably extracts the spatio-temporal models for avoiding the unnecessary sensing. The distributed nature of data-driven approaches has been well-studied in the literature [39, 8, 41, 113, 114]. Hence, it is necessary to use an effective data reduction or simple prediction approaches, in order for tolerating the redundant data readings to ensure the reliability in time-driven and delay-tolerant applications [39, 8].

The main idea of this chapter focuses on exploiting the temporal redundancies

(TRs) in every sensor node and their spatial redundancies and correlations in cluster-head (CH) nodes. This prospect emphasizes the spatio-temporal redundancies of measured physical quantities in order to compute the given spatio-temporal models, and thus consequently reduces the overall amount of required transmissions. In this research contributions, we exploit temporal redundancies in every sensor node by using simple and inexpensive models, which consume limited computational resources of the nodes. Furthermore, data-aggregative window function (DAWF) also performs spatial exploitation exploit over the nodes in CHs, which we present as a preliminary methodological case of the CHs model behaviors over space. The proposition has been investigated based on the real measurements obtained from one of our collaborative lab as LCIS, DREAL project data sets, and the collected data from the indoor environments of thermal renovations of housing estates. DAWF mechanism demonstrates the energy efficiency in sensor nodes by showing that the overall required sensor readings can be reduced as well as their redundant data transmissions. The organization of this chapter is as follows, section 2 describes the literature survey of several energy-efficient data collection and gathering protocols. System models and design implementations of this chapter have been presented in section 3. In section 4, the pseudo codes of algorithmic approaches are discussed in both sensor nodes and CHs use cases, which were presented for spatio-temporal scenarios. Section 5 explores the proposition developments and their framework choices in COOJA. In section 6, the experimental study case results based on two real world datasets has been investigated, in order to demonstrate that the sensor nodes can generate huge amount of temporal redundancies (TRs) in very short time, and then signal reconstruction methods are proved. The temporal use case experimental results based on COOJA implementations has been carried out, and analyzed in section 7. Final section summarizes the proposition performances through the various considered metric analysis in both time and space.

## 4.2. Related Work

The taxonomy of data-driven approaches have been studied in [8]. The prior research works in this area have suggested several methods for reducing data costs as well as their energy costs in the network. The relevant work to the propositions is wang et al [114], which describes the *undirected graphical models or decomposable models* to exploit all the spatial correlations in data, and the broadcast nature of communication by using joint probability distributions. The authors have considered two sorts of data collection metrics to predict the attributes of the sensor network, such as Joint Entrophy-Based Data Collection (bit-hop metric), which measures the total cost of sending a message  $M$  bits from node A to B is given by  $M * d(A, B)$  and Suppression-Based Data Collection (num-messages metric) [54, 115]. However, this chapter mainly targets the data reduction at every independent nodes of the network to reduce the communication costs, and the window principle of this work has been approached from one of our earlier research works [24].

Especially, in [61], the authors present a method to build predictive models for exploiting the sensed data correlations by a pair of nodes. An auto-regressive model is presented in [106]. Nodes can compute a model for the sensed data until the buffer is filled up, and then it only transmits the model parameters to base station. The data reduction approach of Tan et al [116], investigates the impact of data fusion on coverage and detection delay of WSNs.

Energy efficiency in WSNs is a widely studied issue, and a taxonomy of various categories of WSNs presented in [8, 110], which are the related data-driven approaches that have discussed in this chapter.

The prior data collection works have suggested several data-aggregation methods under data-driven approaches, which are categorized as in-network aggregation, compression-based and prediction-based data aggregations. TiNA [117] used a clause condition for specifying the differed ranges, if the differed range is greater than the specified range between any two values, then the differed result can be reported, otherwise ignored. TiNA is more related to the propositions, as we also used the RV function to exploit data redundancies between every two readings of the window stored phenomena in the sensor nodes. On the other hand, Cluster-based Aggregation technique (CAG) [118] presented a clustering technique, which only reports the spatial correlations among the sensor nodes by a CH to the BS rather than the individual nodes temporal data. The authors of [119] imposed a prediction model on the various clustered area of the nodes, like CAG to build a predictive model on CH nodes instead of individual sensor nodes and let complete computational burden on the header nodes itself.

Distributed Source Coding Using Syndromes (DISCUS) [41] used a framework for distributed data compression using joint source and channel coding. This method reduces the inter-node communication cost for using a both quantized source and correlated side information from every individual node. In [113], the authors proposed a predictive temporal redundant model in the data collection, and used it for real-time error correction. Furthermore, a source correlated model is suggested in [120] under lossy wireless sensor network with multiples sinks.

Unlike the literature works, the propositions avoid the consideration of rich spatio-temporal computationally resource constrained models or prediction models, and designed a simple data redundancy algorithm based on various real-world datasets. In this considerations, normal nodes sense the environmental phenomena, and use the DAWF algorithm for exploiting the TRs in sensor nodes as well as for suppressing the redundant data transmissions. Additionally, CH nodes also have the DAWF mechanism with better computational memories, since CHs are assumed as super nodes for reducing spatial redundancies (SRs) among the nodes data messages that are received.

### 4.3. System Model Design and Specifications

As described and discussed in chapter 2, in relation to the data acquisition techniques. A data aggregative window function (DAWF) mechanisms have been proposed, in order to achieve energy optimization while exploiting the data redundancies and correlations. Thus, the DAWF can process either filtered raw data or aggregated data by the exploitation of temporal scenarios in time and spatial redundancies in space. Moreover, literature study has also showed that the sensor observations in both time and space are highly correlated [121].

Tiny AGgregation (TAG) has [122] presented a query-based aggregation service, which is based on a clause condition over EPOCH DURATION of  $i$ , and it is mainly an in-network aggregation service for TinyDB [123] and Cougar [124] on the basis of mapped and directed diffusion concepts in databases. Following the query-based engine concepts, dunkels et al presented [125, 126] a coffee file system, which enables to store a large amount of datasets in every sensor node, like a flash drive by using antelope database system. Similar to the TAG concept, TiNA protocol [117] used a clause condition in order to differentiate the variation between the consecutive readings based on *EPOCHDURATION*  $i$ , if the differed variation is greater than the temporal coherency tolerance ( $tct$ ), then the value can be reported, otherwise ignored. The following form represents the query list that can be received by the base station in query-based applications by TiNA.

```
SELECT (attributes, aggregates);
FROM sensors;
GROUP BY (attributes);
EPOCH DURATION  $i$ ;
VALUES WITHIN  $tct$ ;
```

However, this proposition has also been considered the relative variation (RV) function in order to build a temporal data acquisition model in every sensor node. Although it is purely a time-driven application model, unlike the related works. As further detailed in chapter 2, simple low computational model is an effective choice to balance the node computations in WSNs, such as DAWF propositions, yet to be effective to achieve efficiency by using the cluster-based topology.

We address two prominent cases of the existed spatio-temporal sampling techniques.

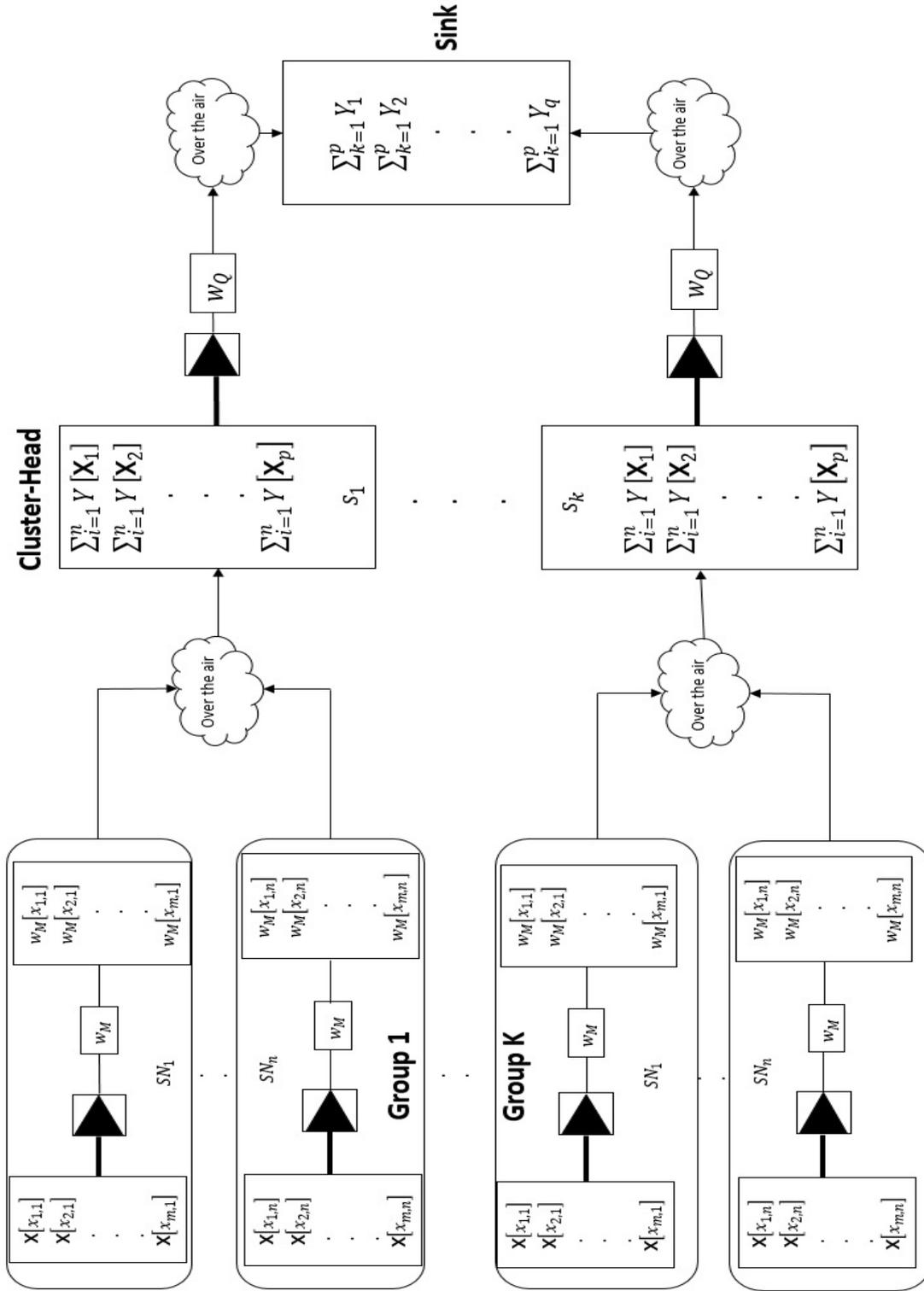
- Existed sampling techniques are considered to conduct their research based on either temporal cases or spatial correlation finding approaches. But, there is very minimal work that have done in the area of spatio-temporal sampling. Although, the existed mechanisms are purely based on spatio-temporal application dependent protocols, and they have their own set of limitations.

- We have investigated the data reduction techniques based on both CS-based and prediction-based rich computation models in WSNs, which are high energy constrained methodologies. These considerations put extra burden on the system, in order to constraint more energy resources than the predefined energy cycles, thus, leads to the system complexity, node depletion and communication overheads.

We show that simple data redundancy technique is a better solution in nodes to exploit the temporal redundancies and spatial filtration in CHs, yet effectively. We have also explored the energy efficient measurement use cases and signal recovery. The following list of design considerations that specifies the features of spatio-temporal adaptive sampling methods in single-hop WSNs.

1. A new temporal redundancy algorithm in every node enables a simple, inexpensive computation and communication costs over the network. The preliminary use case by integrating the DAWF algorithm in CH nodes has also been illustrated. However, the parameters and operational considerations are different in CH node proposed designs than the sensor nodes.
2. An improved DAWF on CH nodes for exploiting the data redundancies and correlations spatially. However, the presented method is similar to temporal approach and considered as a preliminary use case for this scenario, and further improvements offered as the future works.
3. With the help of data window concept, this proposition efficiently acquires the signal measurements among the nodes. The DAWF algorithm in every node to reduce their bandwidth before forwarding their consecutive measurements towards the cluster-heads or BS. Latter, reconstructs the original signal of each node from a received set of consequent spatio-temporal measurements by a sink node.
4. In addition, this chapter propositions are mainly categorized in two phases, namely as learning phase, and adaptive phase. In learning phase, node settles its pre-filtration activity before processing the data, and enables the windows on the amount of sensors that being held in nodes. In adaptive phase, it automatically updates and reset the window and threshold parameters whenever requires a change by the application.

Figure 4.1 shows a simple model configuration of sensor nodes (SNs), and CHs, which are connected to the sink. Every SN can generate the data and store into its own window array of  $w_M$  to access the phenomena of  $\mathbf{X}$  for that current window round  $r$ . The initial window sizes are predefined by either user or application requirements in SNs before the deployment, and performs the computation in every node, in order to exploit the temporal redundancies and forwards the required samples for signal reconstruction at sink node.



**Sensor nodes**  
Figure 4.1 – A cluster-based model diagram of Spatio-Temporal sampling through DAWF algorithm.

The model diagram describes the sensor nodes  $SN_n$  that are contained  $\mathbf{X}[x_{m,n}]$  amount of physical phenomena, where  $X$  is a physical quantities of the node, which contains  $x_{m,n}$  amount of sensors in every node, whereas  $m$  represents the range of sensors by  $n$  number of sensor nodes within their corresponding cluster section  $k$ . However, each sensor stores their sensed phenomena through the parallel to serially connected to their respective window buffers  $w_M$  towards the CHs, where  $M$  represents the nodes window range. Cluster head nodes are considered to exploit spatial redundancies among the received messages that are sent by nodes. Thus,  $Y$  defines the packet messages  $X_p$  that are sent by each cluster member nodes towards their corresponding cluster head  $s_k$  window buffers  $w_Q$ , in order to store the received messages among nodes, whereas  $Q$  is a range of the CH window, which sent latter towards the sink node.

### 4.3.1. Preliminaries

This proposition considers a cluster-based sensor network with  $n$  normal SNs, which continuously forward the unfiltered set of data attributes to the CH,  $\mathbf{X}(t) = (\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_m)$  generates the sensed physical phenomena at different time instances  $t$ , and  $s$  number of CHs as super nodes, which receive the data messages of SNs,  $Y(t) = (Y_1, Y_2, \dots, Y_p)$  during the SNs window time intervals  $r$  and their own sensor readings  $\mathbf{X}(t)$  at different time instances of  $t$ .

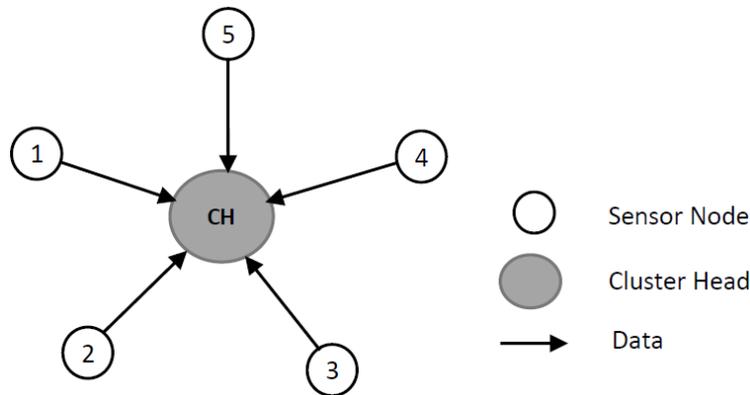


Figure 4.2 – A typical structure of a cluster-based star topology, between sensor nodes and CH.

The attribute detection of environmental phenomena  $\mathbf{X}_m$ , may be the attributes being sensed by nodes as temperature or humidity or may be the result of any application phenomena. If the sensor monitoring attributes are continuous, we consider DAWF mechanism to monitor the redundant data. This thesis primarily focus on the reduction

<sup>3</sup>In this chapter, the propositions that are DAWF and A-DAWF deviated and presented differently from the learning phase. But the core definition of the algorithm will be treated synonymously, and always termed as DAWF in the remainder of this thesis.

of instant data redundancies at individual sensor nodes rather than imposing a model on overall the network.

### 4.3.2. Learning Phase

In this section, the protocol initialization has been enabled through the learning phase. Then, the DAWF activity of sensor readings at every  $t$  time seconds. The design flow structure of Figure 4.3 enables the DAWF mechanism in every sensor node, in order to tolerate the temporal redundancies over the period of every  $t$  seconds. In this scenario, nodes generate the sensor readings at every  $t$  seconds and stores into their corresponding window queue buffer until it fills up. It then uses the temporal data redundancy function in order to pre-filtrate the data before forwarding to the CH node at every  $r$  round seconds. DAWF applies TDR function to evaluate the window data sets while using the data counter. Among the window data sets, if TDR condition is differed the range between two readings and if it is greater than the absolute threshold value, then the evaluated value is counted as non-redundant, otherwise redundant. If the consecutive readings of window length did not differ from the range of absolute threshold, then the window can do mean averaging of total window data. In that case, in spite of sending total window dataset, it only sends a mean averaged value. Therefore, the counter function also differentiates the window dataset whether the evaluated data are redundant or non-redundant. it then separates the portion of the non-redundant dataset and send it as a single packet, apart from the mean averaged portion of data.

Moreover, DAWF not only evaluates the data, but also the time differences between every two consecutive sensor readings. We have noticed that the faulty readings from several real world data sets, as there was a huge time differences between two consecutive readings rather than continuous time intervals. In order to tolerate and suppress the error readings, we present a time interval evaluation method  $\delta$ , whereas it checks the data reading time instances along with the TDR function.

### 4.3.3. Adaptive Phase

Figure 4.4 illustrates the adaptive phase of the protocol, node monitors the window data sets during every round interval of  $r$  whether the amount of variation in sampling is higher than the given threshold or not. Adaptively, nodes can auto-tune the parameters based on its varied sampling rates has introduced in this adaptive phase. The algorithm itself has the adaptive nature irrespective of the sensor phenomena, and it can also auto-tune the threshold tolerant values and window sizes or can reset the parameters based on their corresponding CHs request. In this scenario, the node window sizes may vary upto the optimal sizes of 20, as demonstrated in Figure 4.22 or the window sizes may vary even bigger, if the nodes are considered as high energy constrained devices.

In addition, we present a simple feedback acknowledgement (ACK) mechanism

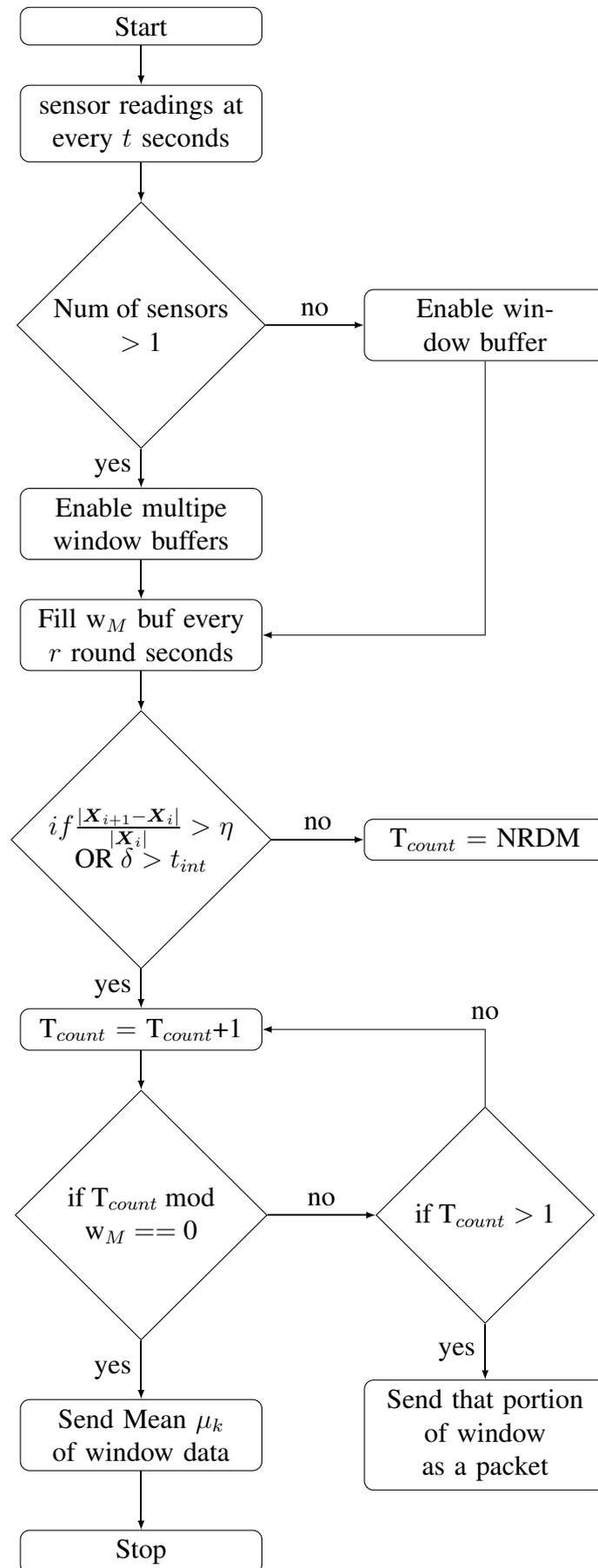


Figure 4.3 – Learning Phase in Sensor Nodes

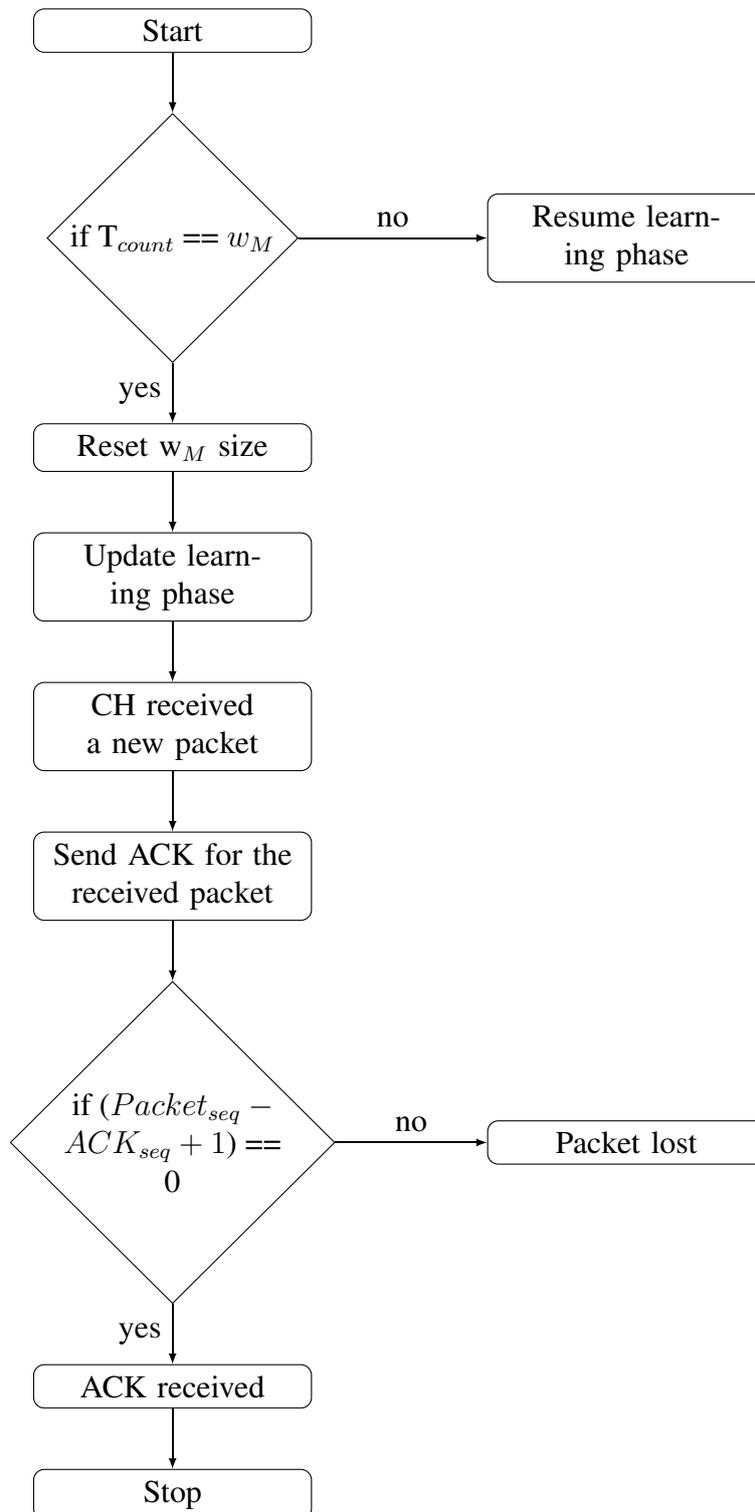


Figure 4.4 – Adaptive phase in sensor nodes

in MAC layer from one of our earlier research works [24]. In order to ensure the reliability in nodes for measuring the packet reception ratio. In this considerations, node counts the message and packet losses or no losses based on its every successful transmission feedback ACK. If sensor node does not receive any ACK for the last transmitted packet from the head node, then the packet considered to be a lost one. In this case, our mechanism calculates the difference between current transmission sequence with the last packet transmission ACK sequence number. If the difference is greater than one then the packet considered as lost, otherwise lost count equal to zero. However, this proposition does not consider the packet re-transmission principle, it only accumulates the lost count in order to reduce the communication overheads.

Moreover, the design structures are reused for the spatial scenarios of CH nodes as well. Because, the considerations are similar to temporal case other than the parameters that can be varied, and differentiated the wholesome of the work.

#### 4.4. Algorithmic Approach for Spatio-Temporal Redundancies

In this chapter, the considered network model is a single-hop cluster-based and distributed sensor nodes, which can compute and process the obtained environmental phenomena of sensors at  $t$  time instances by computing through the propositions. In general sensor nodes, computation is the second highest energy consumer after the communication, even though for computing the algorithms does not consume much energy like communication. However, the design flow structures have developed for suppressing the spatial redundant data messages over sensor nodes in CHs.

In this scenario,  $w_Q$  keeps the record of nodes received data during the observed time interval  $R$  and cleared the data on the basis of FIFO approach after every successful data window transmission to the sink node or BS.

Furthermore, we implement the DAWF mechanism in CHs for finding both SRs and TRs, since CHs receive the data messages  $Y(t)$  from member nodes as well as their own sensor readings  $X(t)$ . We assume that the minimum window  $w_Q$  size of the CH is 50, which can also be varied based on CH node computational constraints for monitoring nodes data spatially and their own readings for exploiting the TRs. In this case, if sensor nodes have the same application monitoring tasks at different periods, then there will be several SRs over nodes. Hence, for reducing the SRs among nodes, CH keeps the data records of all sensor nodes during every round time interval of  $R$  at CH windows. In order to find the SRs, the following expressions Eq. 4.1 and 4.2 must be satisfied in CH nodes.

$$SDR(t) = \begin{cases} S_{count}=0, & \text{if } \left| \frac{Y_{j+1}-Y_j}{Y_j} \right| > Th, \quad 0 \leq Th \leq 1 \\ S_{count}=1, & \text{otherwise} \end{cases} \quad (4.1)$$

$$\mu_K = \frac{\sum_{j=1}^Q Y_{j+KQ}}{Q} \quad (4.2)$$

According to the relative variation (RV) of Eq. 4.1 and 4.3, where  $S_{count}$  represents the spatial count of RV and  $T_{count}$  states the temporal count of RV. If  $S_{count}$  is greater than the threshold  $Th$  then the  $S_{count}$  will be equal to 0 otherwise count will be reported as 1. In this scenario, we set an optimum threshold value ( $Th$ ) as 0.05 based on a tradeoff between Algorithm 3 and 4 by Figure 4.22 in both temporal and spatial scenarios. The relative threshold may also vary based on various phenomena sampling rates or different application physical activities. Thus, it can be evaluated by Eq. 4.2, whereas  $\mu_K$  is the mean average value of window at  $K$  number of index times which starts from zero. The notation  $j$  is an index of the sensor nodes delivered data property at given time instances  $t$ , and  $R$  is the round time interval of  $w_Q$ . The notation  $Q$  is a window size that can be either fixed or varied based on the nodes computational resources or flash memory. During every  $w_Q$  of CH node, the window can compare its member node value  $Y_{j+1}$  with the previous member node value of  $Y_j$  through the window stored readings by Eq. 4.1. Moreover, during every window round time interval of  $R$  after the data check based on Eq. 4.1, the previous sensor readings in DAWF can be flushed itself based on First-In-First-Out (FIFO) queue method. In order to exploit the temporal redundancies in both sensor nodes and CHs, and the parameters and variable considerations of TRs have the similar features as like SR expressions. But, the threshold parameters may vary since CHs has two window constraints one window function for nodes received data and another one for CH nodes sensed physical properties, as like regular sensor nodes. In this scenario, the following constraints should be satisfied in both sensor nodes and CH for exploiting the temporal redundancies, and the given expressions can be defined as

$$TDR(t) = \begin{cases} T_{count}=0, & if \left| \frac{X_{i+1}-X_i}{X_i} \right| > Th, \quad 0 \leq Th \leq 1 \\ T_{count}=1, & otherwise \end{cases} \quad (4.3)$$

$$\mu_k = \frac{\sum_{i=1}^M X_{i+kM}}{M} \quad (4.4)$$

$$\delta = |t_{now} - t_{old}| > t_{int} \quad (4.5)$$

Where  $\delta$  presents a time evaluation method,  $t_{now}$  is a current reading time instance,  $t_{old}$  is an old reading time instance,  $t_{int}$  is a prefixed time interval of data.

According to the Equations 4.3, 4.4, and 4.5, the above functions must be satisfied in sensor nodes in order to exploit the temporal data redundancies. Similar to the spatial methodology, DAWF temporal evaluates the window data based on the expression of Eq. 4.3 and 4.5. In this scenario, both the expressions should be satisfied in nodes. Because, the evaluations not only between two readings, which makes a comparison between the current reading  $X_{i+1}$  with the old reading of  $X_i$ , but also a time difference between the intervals of current window two readings. If both the conditions are satisfied in nodes, then only the window data transmissions can be initiated to process,

otherwise the data can be reported as abnormal readings. Like special use case, temporal scenario also apply the meaning averaging  $\mu_k$  during at  $k$  number of index times by Eq 4.4, in order to carry out the redundant window data information as a mean averaged data message towards the CH. However,  $i$  is an index of sensor nodes sensed phenomena  $X$ , and  $M$  is a size of predefined length of the window index, such as 10 for the initial prefix based on Figure 4.22.

In general applications, most of the obtained environmental data are either redundant or correlated. We assume that the observed nodes data are highly redundant.

## 4.5. Design Implementations in COOJA

The second phase of wireless sensor network experiments that are performed in this thesis has been carried out from COOJA Contiki-3.0 based simulator [98]. The maximum size of the packet transmission unit is 128 bytes from the RF transceiver of CC2420 [23]. The considered cross-level network stack that has been used in this thesis as shown in Figure 4.5.

The framework of DAWF protocol has designed and implemented in COOJA/Contiki simulator. However, there is no default ADC sensors in COOJA, although, we have considered to use randomized sensory readings that will be generated by ADC physical measurement function randomly. However, the light sensor primitive is only available to integrate in COOJA Java based simulator by Contiki developers rather than all the sensors, such as temperature, humidity, etc.,. As the generated readings by sensor nodes are randomized, and data pattern is irregular, thus we consider the second principle of the learning phase proposition, which has slightly been modified according to the design considerations of this framework in COOJA. Therefore, node will send all window data as a single packet, if the window contains all non-redundant data messages (NRDM). The analysis of this investigation from COOJA simulations has been evaluated to demonstrate that even carrying a total amount of window data as single packet may not require additional energy costs, whereas it showed in section 4.6.

### 4.5.1. Design Decisions of DAWF Integration in COOJA Simulator

The propositions are carried out in both MATLAB and COOJA simulations throughout this dissertation, and the TmoteSky [127] and Zolertia Z1 node platforms have also been used in COOJA [98], in order to test the propositions in real sensor node platforms. As COOJA was built based on Contiki operating system abstractions with multiple hardware platform and node choices. In regard to the network simplicity, we choose a light-weight network platform of Rime stack in COOJA, since it is mainly designed for low-power embedded devices without any security header fields, and it has 8-bit rime node addresses and, unlike IPv4 and IPv6 protocol stacks.

The propositions that are integrated in COOJA rime network stack [101] as shown in Figure 4.5. While there is no physicality of ADC sensors in Rime network stack, we then developed the protocol in application layer, which is mainly integrated based on the temporal use case in every sensor node of the network.

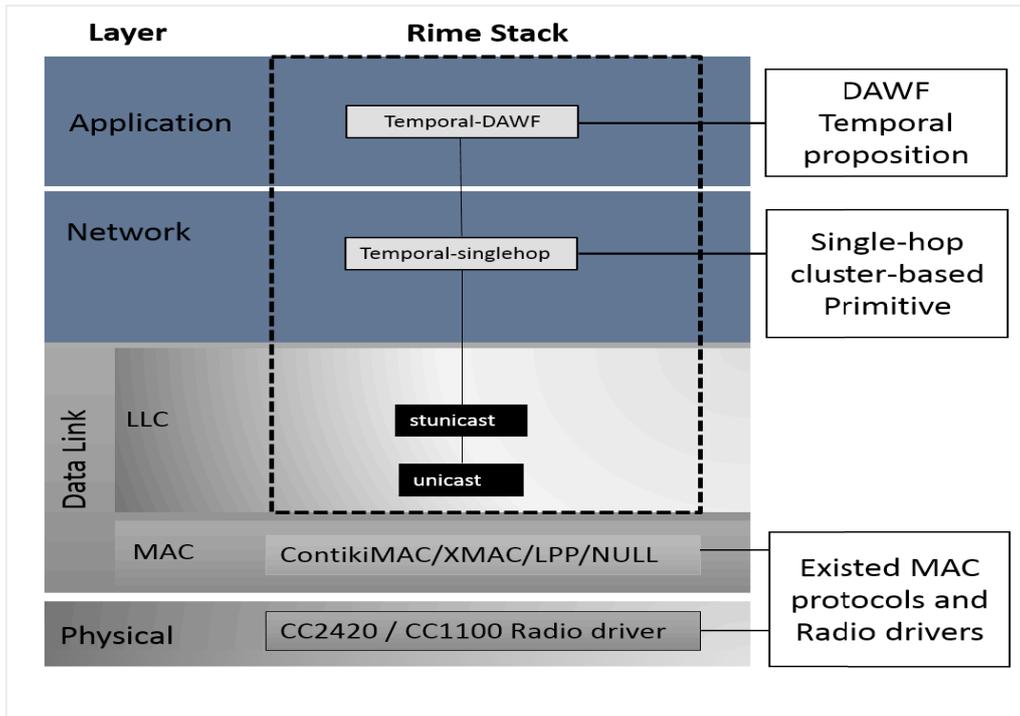


Figure 4.5 – Structure of established Rime cross-level network stack

Moreover, the default header and packet sizes of the Rime network is 48 bytes, 128 bytes, respectively. Due to that many predefined packet attributes of Rime are within the packet buffer of Rime. As we noticed that the minimum cross-level packet size of Rime was always logging through Radio as 45 bytes of data irrespective of the packet attributes that has contained. However, in the propositions, we did not consider any additional packet buffer attributes of the node, such as MAC ACK, RSSI, LQI, Battery Info, etc,.. In terms of simple header fields, we only use the source, destination's ID, and packet sequence numbers from predefined header fields, and then added the DAWF payload size of  $w_M$  as 10 bytes, which was showed in Figure 4.6.

<i>Src</i>	<i>Dst</i>	<i>Seqno</i>	<i>Payload message <math>w_M</math></i>
------------	------------	--------------	-----------------------------------------

Figure 4.6 – Rime Packet header fields with DAWF  $w_M$  payload message size.

### 4.5.2. Adaptive Compression and Prediction-Based Mechanism

The iterative Algorithm 1 starts by initializing all the variables and their parameters. The core of this algorithm, consists of a data redundancy loop that predicts the node either redundant or correlated data with the previous data values. If the current readings are different than the previous readings then only the node can forward data to the cluster-head. Otherwise, the data can be flushed itself, and the window queue buffer can be reused, iteratively.

---

**Algorithm 1** Cluster area nodes received data ( $Y_j(t)$ )

---

**procedure** : *Initialize*( $s, t, R$ )

Parameters:  $t \leftarrow$  *CH time intervals*,  $SDT(t) \leftarrow \frac{|Y_{j+1}-Y_j|}{|Y_j|} > Th$

$R \leftarrow$  *Round time interval of CH window*

**if**  $Y_j(t)$  **then**

    Store into the Window of  $\mu_K$

**if**  $\frac{|Y_{j+1}-Y_j|}{|Y_j|} > Th$  **then**

        Count spatial data redundancies and correlations

**if**  $S_{Count} \bmod w_Q == 0$  **then**

            Send mean averaging data of  $\mu_K$

**end if**

**end if**

**else**

    Count the non-redundant and non-correlated data messages

**if**  $NRDM_{count} || NCDM_{count} > 0$  **then**

        Send total window as single data packet

**end if**

    Count window amount of redundant and correlated data messages

**end if**

**end procedure**

---

The entire algorithm iteratively stored the received data of nodes in an attribute window function, computed and updated in CH nodes. In this regard, DAWF mechanism has two time intervals; one as nodes received data time intervals  $t$  and another one as window round time intervals  $R$ . While using RV function (Algorithm 1, line 4) DAWF mechanism evaluates the short window data sets of their corresponding nodes, whether they are redundant or correlated. If the RV function is greater than the threshold then it counts and forwards the non-redundant data towards the BS. If the window contains the redundant information then (Algorithm, line 8) DAWF enables the mean averaging function to send one appropriate data value rather than all window readings. Moreover this pre-filtration method also reduces the redundant data and their redundant transmissions, which can reduce the communication cost on nodes, and then maximizes the overall node lifetime compared to the predefined lifetime as well as the network lifetime.

Algorithm 2 and 3 in sensor nodes initializes the parameters for evaluating their sensed physical phenomena for exploiting TRs at different time instances  $t$ . However, the core of the algorithm iteratively works same as like Algorithm 1, except working at different time intervals.

---

**Algorithm 2** Nodes sensed phenomena ( $\mathbf{X}_i(t)$ )
 

---

```

procedure : Initialize( $n, t, r$ )
Parameters:  $t \leftarrow$  node time intervals,  $TDR(t) \leftarrow \frac{|\mathbf{X}_{i+1} - \mathbf{X}_i|}{|\mathbf{X}_i|} > Th$ 
 $r \leftarrow$  Round time interval of nodes window

  if  $\mathbf{X}_i(t)$  then
    Store into the window of  $\mu_k$ 
    if  $\frac{|\mathbf{X}_{i+1} - \mathbf{X}_i|}{|\mathbf{X}_i|} > Th$  then
      Count temporal data redundancies and correlations
      if  $T_{Count} \bmod w_M == 0$  then
        Add mean averaging data of  $\mu_k$ 
      end if
    end if
  else
    Count non-redundant data messages
    if  $NRDM_{count} > 0$  then
      Send total window as single data packet
    end if
    Count redundant data messages
  end if
end procedure

```

---

On the other hand, DAWF also carries the non-correlated window data as single packet, the authors of [128] presented that increasing a payload size (from 1 bytes to 90 bytes) to some extent does not add the additional communication costs over the nodes. Moreover, this thesis has been carried out the propositions in COOJA emulator for the real implementations, since COOJA emulator uses its own developed software's and can be uploaded directly on any COOJA recommended real nodes, for instance, we consider Tmote-sky and Z1 sensor node platforms in COOJA simulations, and then software codes can be reused and uploaded on TelosB testbeds for the real node performance evaluations.

## 4.6. Experimental Study Case and Results

In order to validate the proposition models as described in the earlier section, this chapter carried out a set of experiments by using real data sets from Intel research lab, Berkely [105] and LCIS, DREAL datasets, collected by OSAmI-FR Wired/Wireless

**Algorithm 3** Nodes sensed phenomena ( $\mathbf{X}_i(t)$ )

---

```

procedure : Initialize( $n, t, r$ )
Parameters:  $t \leftarrow$  node time intervals,  $TDR(t) \leftarrow \frac{|\mathbf{X}_{i+1} - \mathbf{X}_i|}{|\mathbf{X}_i|} > Th$ 
 $r \leftarrow$  Round time interval of nodes window

  if  $\mathbf{X}_i(t)$  then
    Store into the window of  $mu_k$ 
    if  $\frac{|\mathbf{X}_{i+1} - \mathbf{X}_i|}{|\mathbf{X}_i|} > Th$  then
      Count temporal data redundancies and correlations
      if  $T_{Count} \bmod w_M == 0$  then
        Add mean averaging data of  $\mu_k$ 
      end if
    end if
  else
    Count the redundant and non-redundant data
    Send redundant data portion as  $\mu_k$  data packet
    Send remaining fields of non-redundant data as single packet
  end if
end procedure

```

---

Sensor Network (WSN) nodes, contained and supported by DREAL, LCIS research lab from [104].

### 4.6.1. Temporal Performance Analysis

In this section, we have presented two sorts of experimental analysis. First one describes the exploitation of data redundancies over the collected data of nodes. Second one focuses on the consideration of mean prediction errors over temperature, and the transmission costs evaluation over both time and space. While some of the collected datasets dates and timings are differently formatted, thus, considered to use 11 months of data for the this study case, in order to validate the metrics of mean absolute error (MAE) and total amount of reduced transmission  $R_{TX}$  metric costs.

In this section, the approach structured in section 4.6, and the cluster-tree topologies that we assume and superimposed on the considered amount of sensor nodes, as shown in the Figures 4.7 and 4.15. Thus, allow us to investigate the outcome of the propositions, in terms of suppressing the required number of transmissions as well as preserving the energy.

### 4.6.2. Experimental Study Case Based on Intel Data sets

---

<sup>4</sup><http://db.csail.mit.edu/labdata/labdata.html>.

In this section, the simulation results among the considered sensor nodes of an area as shown in Figure 4.7 based on real-world Intel Data sets from [105]. According to this online repository, the data values are captured every 31 seconds continuously during the day and night, and the readings (temperature, relative humidity and light) are collected from the indoor network deployment of 54 sensor nodes at Intel-Berkeley Research Labs, between February 28 and April 5, 2004. The Mica2Dot sensor nodes equipped with weather boards for measuring temperature, relative humidity and light physical quantities. As the datasets are mixed of normal and abnormal readings often. Thus, for the simplification, we only considered to use one sort of physical phenomena as temperature, and one day measures of the considered sensor nodes and a CH node. In order to achieve the required amount of samples that are needed for signal reconstruction, we set the minimum window round intervals, the window sampling rate from minimum of 5 minutes to the maximum of 1.6 hours, whereas the basic sampling rate is used for acquiring the sensor readings at every 31 seconds from the datasets.

However, the gathered datasets are highly redundant data, since the collected data from both day and night within the indoor environment, thus, the temperature gradual variations is very minimal during the night times, as showed in the Figures 4.8, 4.9 and 4.10.

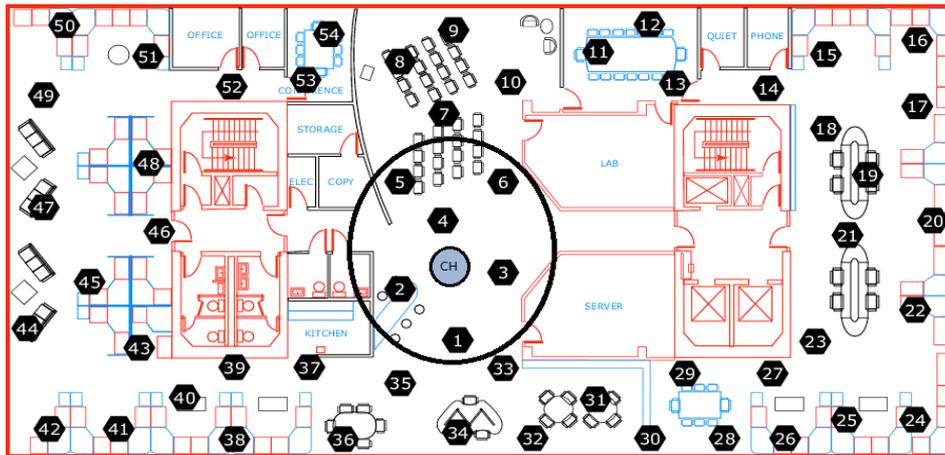


Figure 4.7 – Sensor Nodes Deployment Map of Intel Berkeley Research Lab [105]: the highlighted area has chosen for our experiments.

Figure 4.8 illustrates the pre-filtration by DAWF with a fixed window sizes of 10 at relative thresholds. The DAWF mechanism forwards non-redundant data by varying the thresholds from 0.001-to-0.1, as shown in the table 4.1. Without DAWF, 1316 sensor readings would be sent by every individual node to the CH. According to Figure 4.8, the presented graphs with minimum to maximum thresholds to find the optimum threshold parameters. We also noticed the optimum threshold values between 0.008 to 0.05, after 0.05 Th whatever the threshold value is given then the filtered data remain constant, which is obvious since RV function do filter the redundant data comparison between every two window data values at a fixed window size of 10 other than

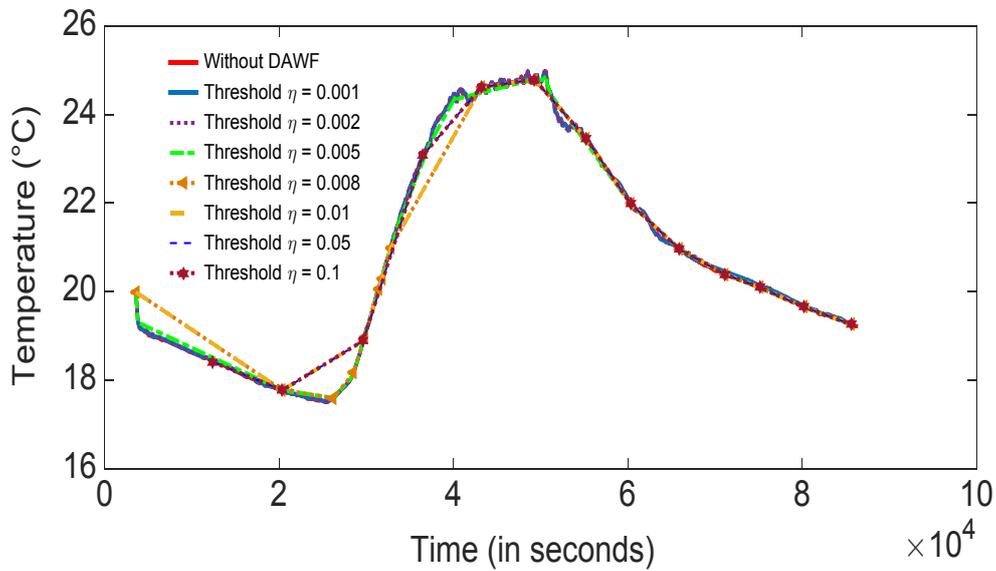


Figure 4.8 – DAWF with a fixed  $w_M$  at various thresholds of  $Th$ .

the varied window sizes, and then process the non-redundant or mean averaging data messages accordingly.

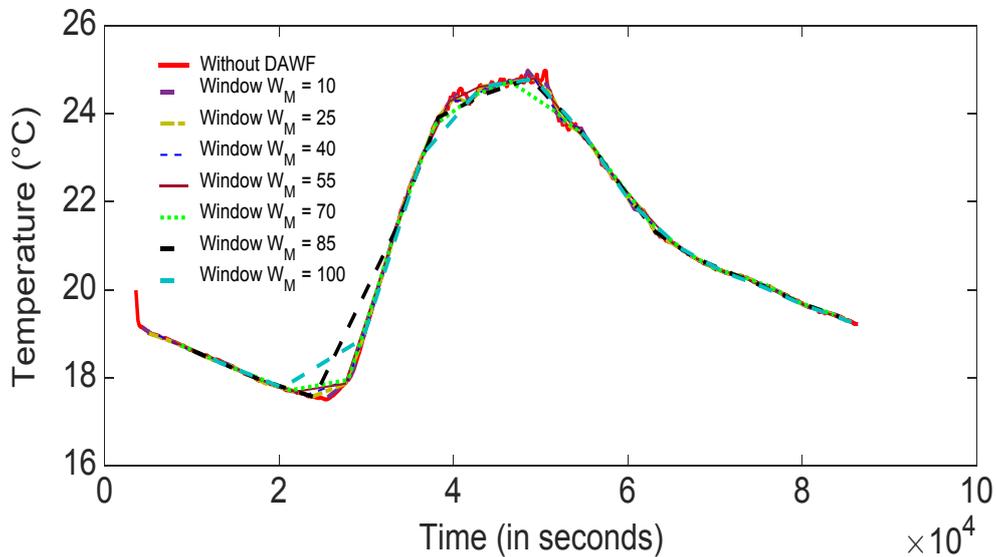


Figure 4.9 – DAWF with a fixed  $Th$  at various window sizes  $w_M$ .

On the basis of Figure 4.9 and Figure 4.10, we carried out two sorts of performance analysis. First one describes the performances with a varying window sizes at a minimal threshold of 0.001 over the original data signal, for instances  $Th$  of 0.001

Table 4.1 – DAWF performances at various thresholds and window sizes.

Threshold ( $Th$ )	Window size ( $w_M$ )	Nodes temporal data TXs (out of 1136)
0.001	10	383
0.002	25	167
0.005	40	47
0.008	55	26
0.01	70	21
0.05	85	15
0.1	100	13

filtered 383 readings, a  $Th$  of 0.005 gives 139 sensor readings, a  $Th$  of 0.01 provides 132 and with 0.1  $Th$  DAWF delivers 131 readings out of 1316. On the other hand, as tabulated in 4.1, if the window size is varying on the basis of threshold variation then the delivered data also shown huge differences. In this scenario, figure 4.9 clearly demonstrate that having an optimum window size value is indeed delivered the highest performances compared to the optimum threshold value.

In the second case, Figure 4.10 demonstrates various filtered data signals over the original signal by varied window sizes at a fixed threshold of 0.05. In Figures 4.9 and 4.10, if the threshold is fixed at maximum value then the signal bandwidth will be compressed at higher rates, because it omits the quality of data that are processed. But, when it comes to maximum window sizes at a minimum threshold, the delivered data signal is fairly compressed compared to the Figure 4.9. In this case, the user has to know that keeping the window size as longer interval than the minimal, which is only be useful for longer data evaluations than the minimal window sizes. However, in regards to the quality of data metric can only be achieved by keeping the threshold value as an optimum.

The Figure 4.11 exploits the various window performances of  $w_M$  over the different data thresholds  $Th$ , which can show that how DAWF window performances gradually increased in temporal use case scenarios. Figure 4.11 also explains the total number of window data transmissions over the varied window  $w_M$  sizes at various error thresholds  $\eta$  as showed in the table 4.1.

In this scenario, for calculating the nodes reduced transmissions for each non-redundant data, we presented a number of reduced transmissions evaluation metric, which can achieve is given by Eq.4.6. For instance, if the node follows a periodic transmission system as per the Intel data sets time interval, then it sends total day measures of 1316. In the contrast, using DAWF mechanism at a minimum fixed window interval of 10, whereas it reduces 1186 redundant messages out of 1316 sensed data, which is clearly proved by Eq. 4.6 that DAWF can suppress upto 90% of redundant transmissions compared to a periodic data transmissions. The reduced transmissions metric can be derived as

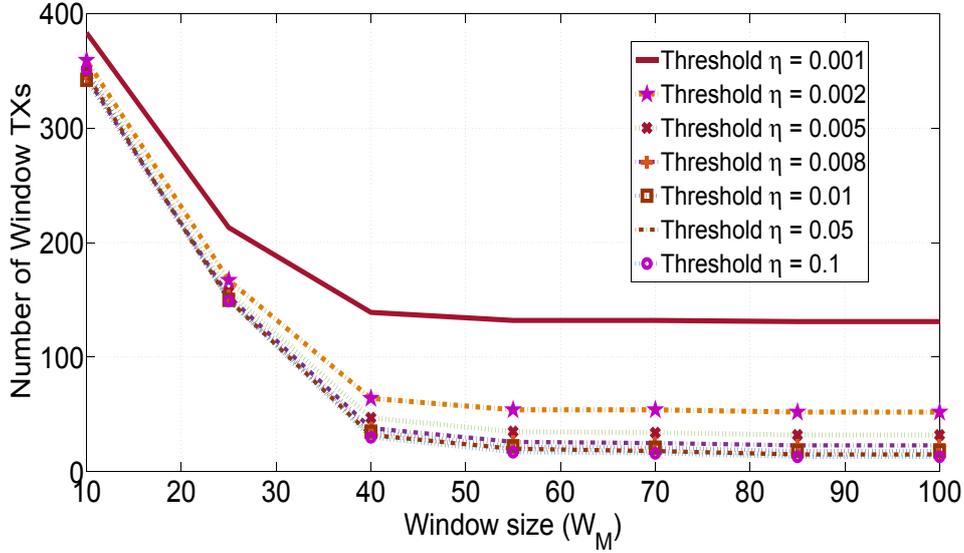


Figure 4.10 – Number of window data transmissions (TX) over the varied window sizes of time.

$$R_{TX} = \left(1 - \frac{\text{Window filtered messages}}{\text{total generated messages}}\right) \times 100 \quad (4.6)$$

Moreover, in the above results, some of the presented graphs are overlapped. Because, in nodes temporal performances, after the optimum threshold range of 0.05, whatever the threshold values thus being increased then the delivered performances are nearly same even at the larger window sizes. However, the delivered performances may vary if there is any network interference or node interference with neighboring nodes or hardware node failures/node dead. In this regard, it requires to upload the implementations on virtual test-beds or real deployments, in order to validate the real network performances. On the other hand, a simple approach is conducted based on Intel data-sets observations, we present a time interval evaluation method  $\delta$  for minimizing the faulty readings, as derived in Eq. 4.5. In this case, we consider six cluster nodes which they do the TDRs filtration, whereas sensor node transmits 131 readings to the CH out of 1316 readings compared to the data sets sampling rate. In order to measure the amount of reduced transmissions  $R_{TX}$  in every node. As, we used a  $R_{TX}$  metric of Eq. 4.6 to evaluate the performances, as shown in figure 4.11. However, we also conducted several simulation tests using different  $\eta$  and  $w_M$  parameters over the  $R_{TX}$  and MAE metrics to explore the performances by both relative thresholds and window sizes as showed in the Figures 4.11 and 4.12.

Figure 4.11 describes the total amount of reduced transmissions over varied window sizes and threshold parameters, where a minimum window of 10 reduces 66%, and a maximum window of 100 reduces more than 90% of the required transmissions. In figure 4.12, we further used the mean absolute prediction error metric of sensor

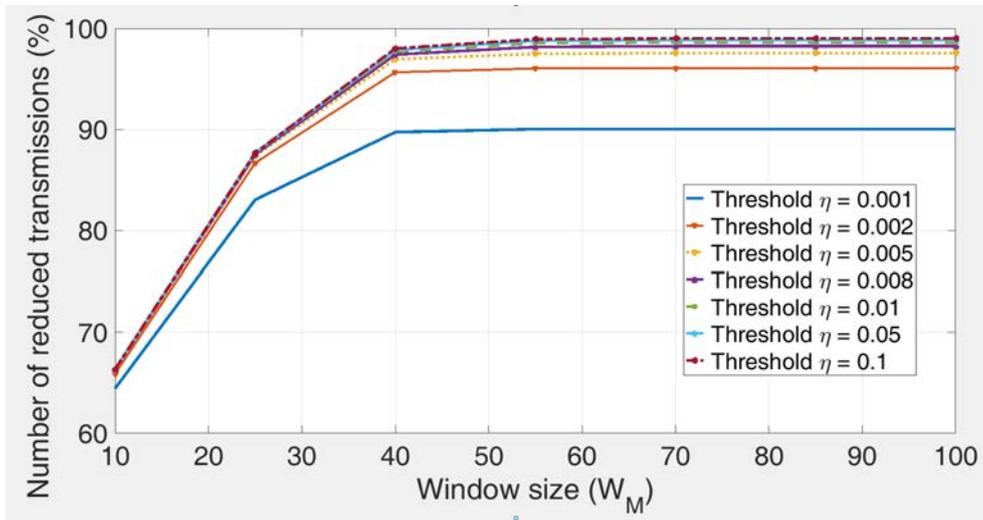


Figure 4.11 – Total number of reduced transmissions (%) over the windows.

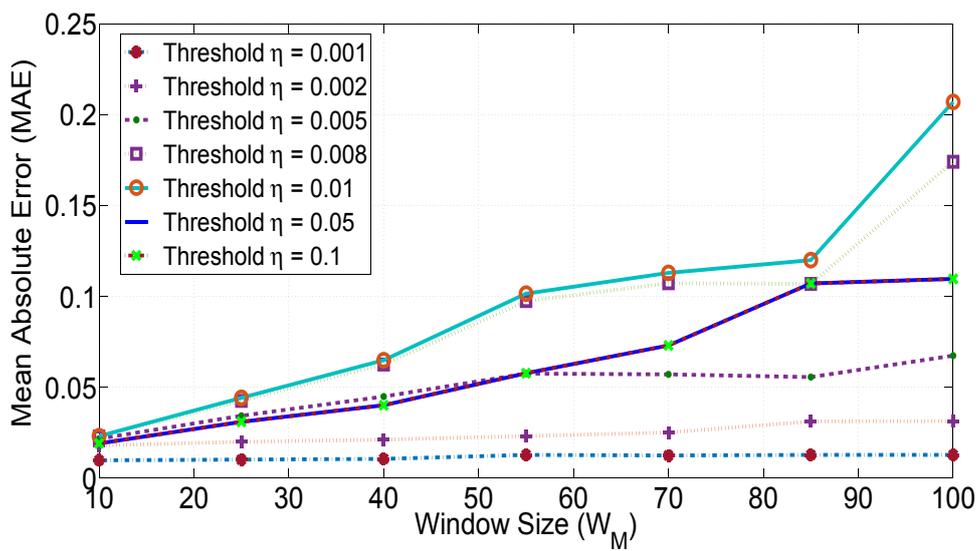


Figure 4.12 – Mean Absolute Error (MAE) cost over the window sizes.

nodes at head node, which then shows the attained error over the various window sizes for minimum to maximum prediction error of the thresholds  $\eta$  are from 0.001 to 0.1. Following the minimum to maximum thresholds of MAE is less than one percent (<1%) in all cases.

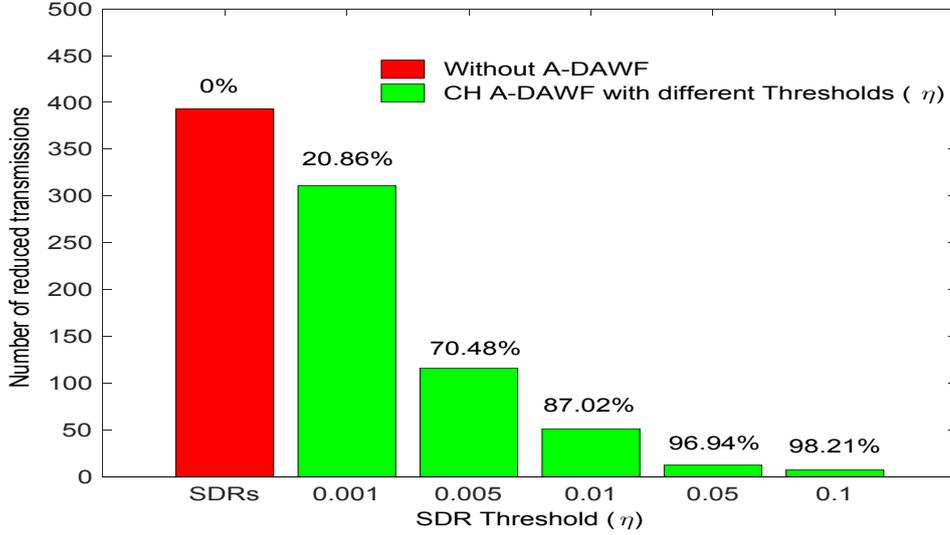


Figure 4.13 – CH A-DAWF among nodes spatial data with different thresholds.

Figure 4.13 demonstrates the spatial data redundancy exploitation on a considered area of cluster nodes to the CH at different time instances  $t$  from [105]. Figure 4.13 shows that there are still large amount of spatial redundant data among the nodes. In which, CH A-DAWF can still suppress a large amount of redundant data at a various thresholds and at a various window sizes of 10, 50 and 100. This experimental analysis explores to understand the differences and how often that the faulty readings can occur among the correlated values and their interactions between each other. Moreover, we also notice that there is no much varied performances in TDRs and SDRs after  $\eta$  of 0.05 at a fixed window intervals, thus whatever the higher values are given then the A-DAWF delivered data messages are remain constant, this is obvious since RV function only performs the redundant data comparison between every two window data values, and then process the non-redundant or mean averaging data accordingly. If the window size varies, then the window delivered performances will obviously be varied, which were showed in the above tabulated readings.

In order to calculate the energy consumption of total measurements among the datasets that are filtered by DAWF mechanism before forwarding to the CH. The total energy consumption that is spent during the DAWF transmissions is achieved by Eq. 4.7 and 4.8.

$$E_{TX} = P_{TX} * t_{TX} \quad (4.7)$$

Where  $E_{TX}$  represents the energy for transmitting,  $P_{TX}$  is a transmission power

and  $t_{TX}$  defines the time in  $TX$  mode.

$$t_{TX} = N_{TX} * TX_t \quad (4.8)$$

In Eq. 4.8,  $N_{TX}$  states number of transmissions that are occurred and  $TX_t$  derives the active transmission time.

In Table. 4.2, we present a temporal comparative analysis in terms of energy that has been spent during the datasets filtration by TAG, TiNA and DAWF algorithms. However TAG algorithm did not consider any threshold coherency tolerant value, it only compares the current reading with the previously transmitted value, if they differ to each other, then only TAG sends the data. Moreover, TAG is an aggregation service for query-based database system of TinyDB. According to Table. 4.2, where TAG transmitted total amount of data as 1316 readings, and which has consumed  $2.5662J$  of energy over total day of measures, since it did not have any  $tct$  and considered as 0. Whereas, TiNA protocol has consumed  $1.2812J$ , and DAWF  $0.2574J$ , respectively.

Protocol	Mica2Dot (in $TX$ mode)	Energy consumption for $TX_{ing}$ (per day)
TAG [122]	$65 mW$	$2.5662 J$
TiNA [117]	$65 mW$	$1.2812 J$
DAWF	$65 mW$	$0.2574 J$

Table 4.2 – Energy consumption cost at TX mode of TAG, TiNA and DAWF protocols.

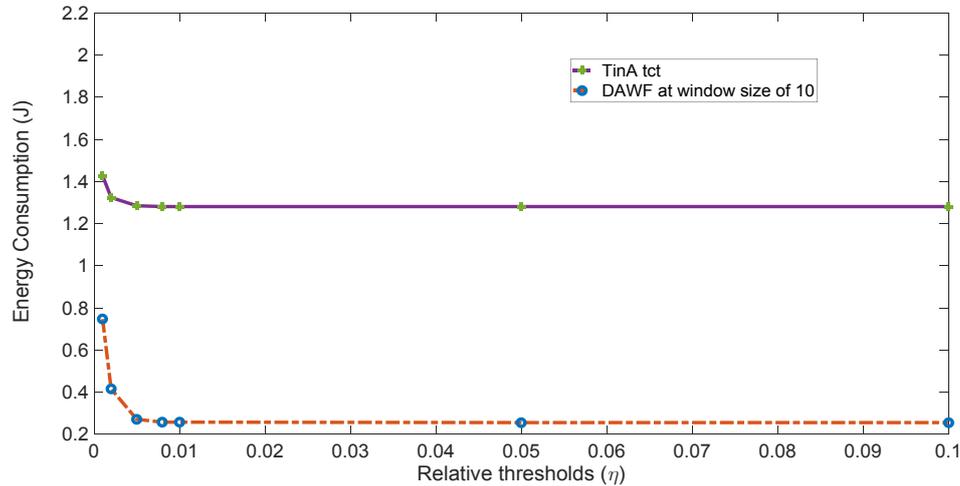


Figure 4.14 – Energy cost comparative analysis between TiNA protocol with  $tct$ , and DAWF at the window size of 10

Figure. 4.14 shows the total amount of energy that was consumed while transmitting the datasets by TiNA and DAWF protocols. While both the protocols are designed

based on Intel datasets and compared with TAG protocol. As demonstrated in Table 4.2, DAWF proposition has certainly achieved upto 90% of energy saving in terms of reducing the redundant data transmissions while compared to the TAG aggregation service, and upto 80% of energy has been saved compared to the TiNA protocol. However, the achieved performances may vary over the total energy consumption of the network. Because, during the suppression of redundant data transmissions by DAWF sender nodes, which let the receiver nodes and other listener nodes to be in either sleep or power down mode. Thus, saves a great amount of energy in both receiver and neighboring nodes. Moreover, DAWF spends more energy consumption in computation mode, although, the amount of energy that has spent for computation in order for pre-filtration evaluations is no greater than the transmission mode of energy consumption, which consumes  $25mW$  only.

### 4.6.3. Experimental Study Case Based on LCIS DREAL Data sets

In this section, the results were carried out among the test area of sensor nodes, as shown in Figure 4.15 based on real-world data sets from the indoor deployment of DREAL, OPHLM and UGINE project under LCIS research laboratory. According to the lab repository, the data values are captured at every one hour, and the readings (ambient temperature, relative humidity, solar temperature, light and  $CO_2$ ) are collected from the indoor network deployment of several sensor nodes by DREAL project, between January 1 and December 30, 2011. Moreover, the sensors appear to be placed linearly in different floors of the DREAL housing project, can be seen in the Figure 4.15. In this experimental case, we considered three sorts of physical phenomena as solar data from the solar panels, ambient temperature, and humidity. However, the datasets that are used in total a year measures of 6 sensor nodes with a CH node. In this scenario, we use different window sampling rates, in order to find the optimal ranges as discussed in the below results. Moreover the evaluated window sampling rates are varied from the minimum window interval of 5 upto the maximum window interval of 100, whereas the basic sampling rate for sensor readings has chosen at every hour from the datasets.

Figure 4.16, 4.17, and 4.18 illustrates the spatio-temporal DAWF mechanism of considered physical quantities by computing through relative thresholds  $\eta$  of 0.001 to 0.1 at a fixed window size of 10, and over different window sizes of 10 to 100 with an optimum threshold ranges from 0.01 upto 0.05 and may vary based on the data sensitivity and temporal threshold coherency ( $tct$ ), which is demonstrated in the below subsections. In this scenario, DAWF presents a comparative analysis between node threshold parameters and their window constraints. In order to validate that increasing a window size based on the nodes computational constraints would be always beneficial. In figure 4.16, it clearly shows that the node window constraints certainly plays a significant role to reduce the data bandwidth without losing any significant data

---

<sup>5</sup>DREAL, OPHLM and UGINE housing project under Essisar-INP, LCIS lab, <http://www.auvergne-rhone-alpes.developpement-durable.gouv.fr/>.

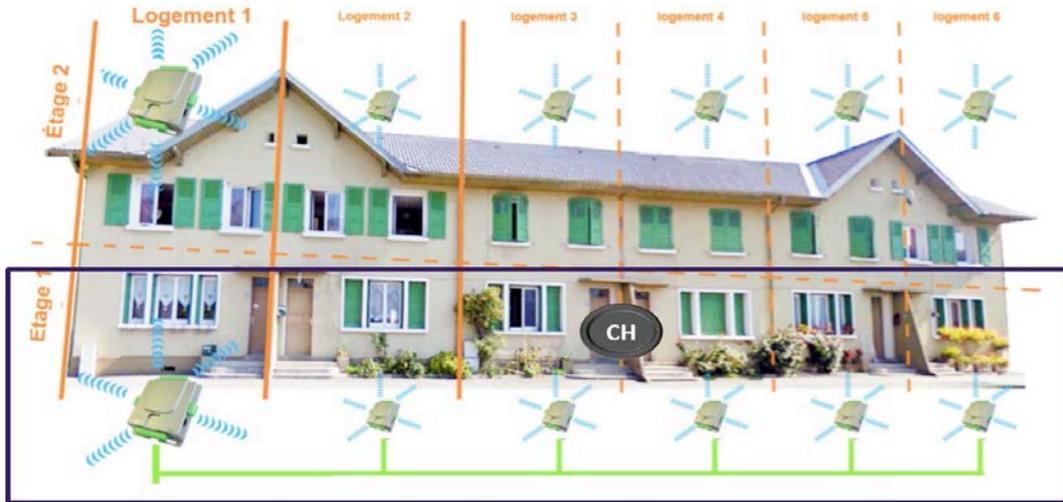


Figure 4.15 – DREAL Deployment of Indoor Wireless Sensor Network.

values. However, this is the case only for temporal exploitation in nodes. In spatial scenarios, exploiting the spatial data redundancies among the received data messages in CHs, whereas the threshold parameters are having a major role for delivering finest data performances compared to the window sizes.

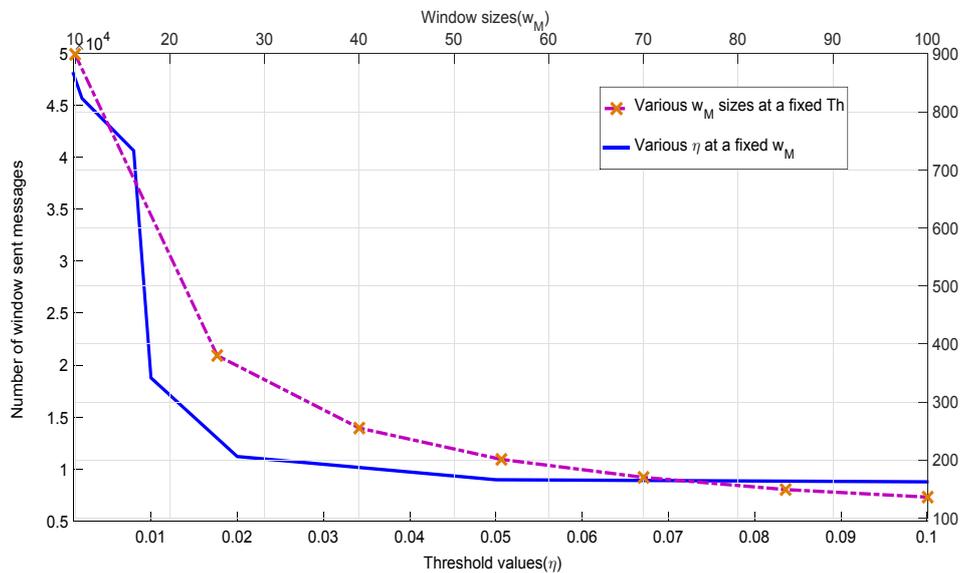


Figure 4.16 – Temperature datasets through DAWF performances with various thresholds  $Ths$  over different  $w_N$  sizes of temporal data.

In Figure 4.17, DAWF computes the humidity datasets in order to show that how often the temporal redundancies occur in indoor humidity sensors. The delivered performances are from relative thresholds at a fixed window size and over varied window sizes at a fixed threshold. As seen in Figure 4.17, various windows at an absolute

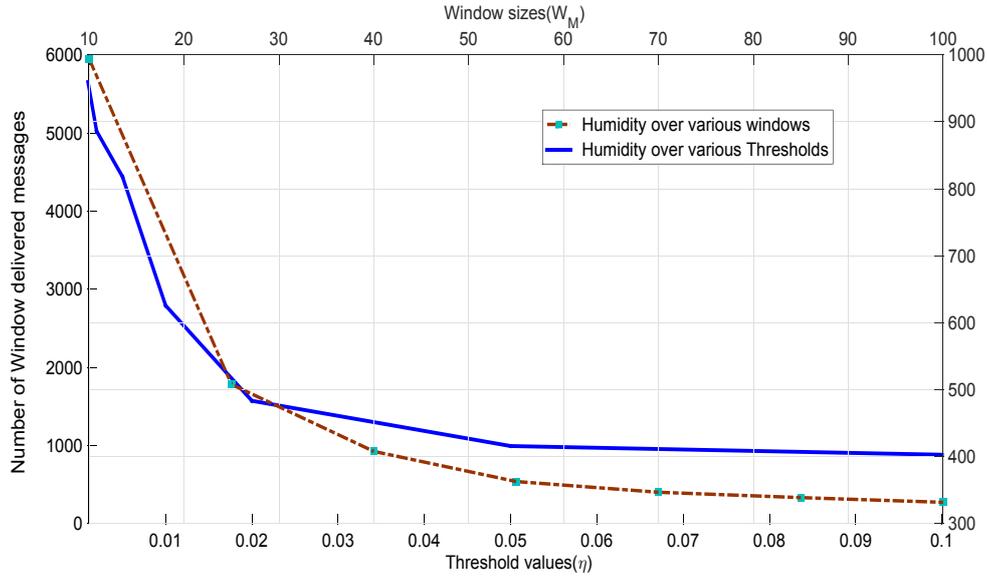


Figure 4.17 – Humidity datasets through DAWF performances with various thresholds  $Ths$  over different  $w_N$  sizes of temporal data.

threshold shows the gradually pre-filtered results and delivered better performances compared to the various thresholds at a fixed window size, whereas the delivered performances are remains constant at after 0.05 threshold value. Thus, also proved that in humidity phenomena that having better optimal window sizes can help the mechanism to perform far more better than minimal parameters.

Although, in Figure 4.18, it is not the same as in figures 4.16 and 4.17. Because, the highly varied datasets of solar phenomena in short times. In DAWF performances, the data sensitivity in various phenomena is directly proportional to the DAWF parameters. If the parameters fixed as maximum, then the performed results are high. However, we defined the data sensitivity based on two cases, one as hard data sensitivity based on outdoor sensor or highly varied indoor data sensors, and another one as soft data sensitivity, which is based on indoor regular sensors, which are only varied when there is high event triggers. Moreover, it is purely a application choice that held with the amount of sensors which are needed. As shown in Figure 4.18, the solar phenomena are highly varied data over short times, thus, led that the thresholds have higher proximity than the window sizes, as threshold parameters only decides that how much of data that needs to be filtered for this current window round. In this case, if the threshold is bigger then the loss of data also gets increased.

The following figures of 4.19, 4.20 and 4.21 differentiate the protocol performances at different datasets that how effectively the mechanism can do the pre-filtration at both highly varied phenomena as well as on gradually varied phenomena. The metric of  $R_{TX}$  demonstrate the performances at minimum to maximum DAWF parameters on various phenomena, as discussed below.

In the contrast, the presented metric of reduced number of transmissions  $R_{TX}$  is

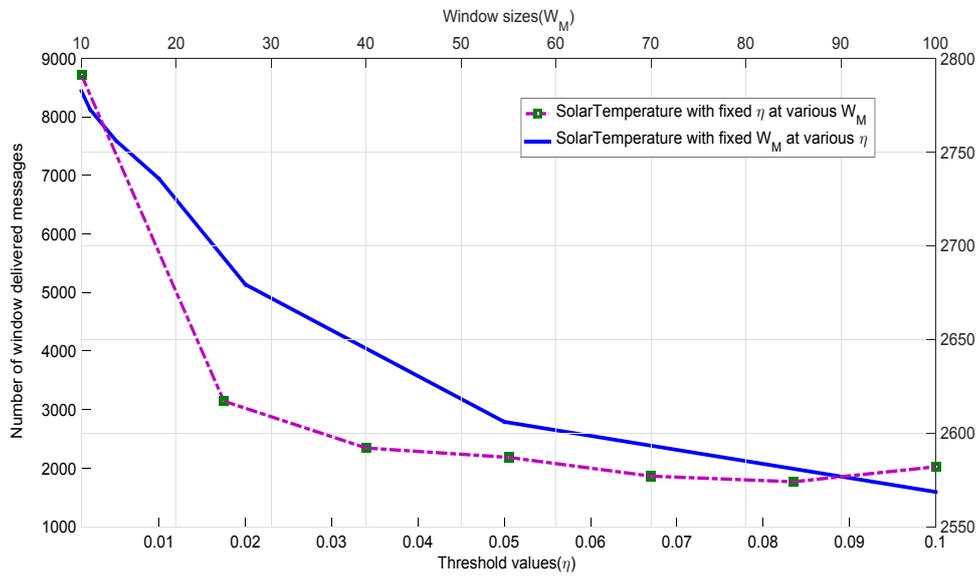


Figure 4.18 – Sensor node DAWF performances of solar temperature over various window sizes.

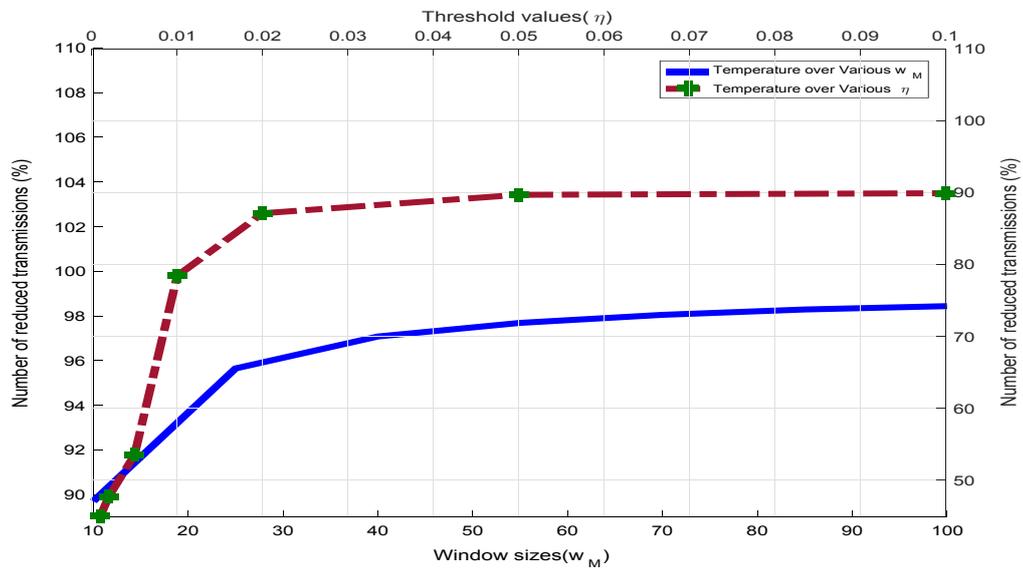


Figure 4.19 – Temperature: Sensor node temporal reduced transmission metric performances over various thresholds and the various window sizes

given by Eq. 4.6, in order to calibrate the percentage of reduced transmissions per node while exploiting the temporal data redundancies. In figures 4.19, 4.20, and figure 4.21, which also explored the different phenomena performances of nodes through the  $R_{TX}$  metric. In Figure 4.19, the  $R_{TX}$  metric delivered performances as 88.89% by a minimum window size of 10 and an optimum threshold of 0.05, and a maximum of 98.44%, whereas relative thresholds of  $R_{TX}$  metric were suppressed at 45% and 89.92%, respectively. Thus, a node can reduce the minimum of 88.62%  $R_{TX}$  in humidity by computing through the DAWF algorithm, and the maximum of 99.62% with the consideration of higher window interval as 100 and an optimum threshold parameter of 0.05. According to threshold parameters, whereas DAWF has been sup-

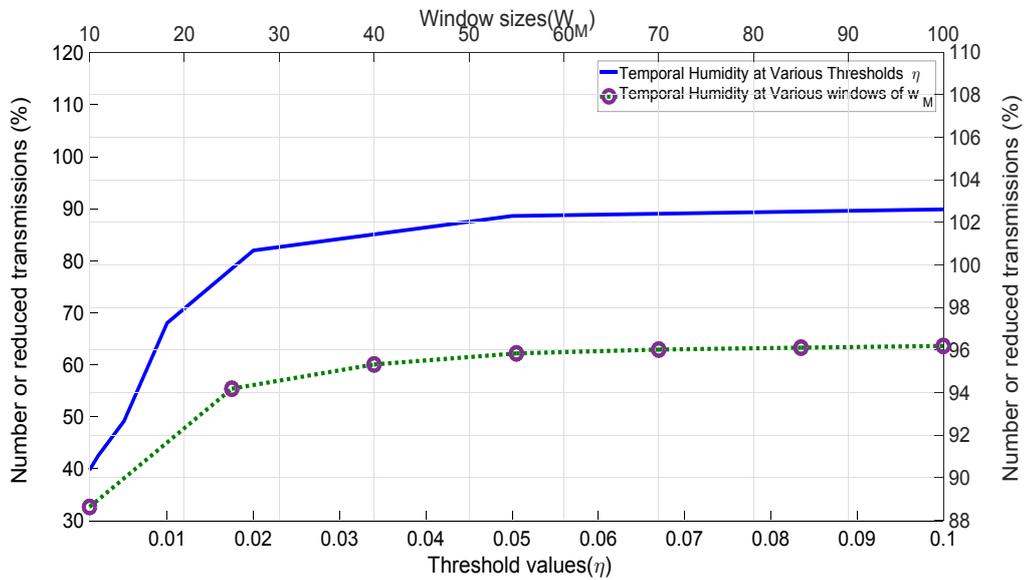


Figure 4.20 – Humidity: Sensor node temporal reduced transmission metric performances over various thresholds and the various window sizes

pressed a minimum of 39.76% redundant transmissions by a minimum window size of 10, and at the threshold of 0.001. The maximum of 89.89% redundant transmissions has been minimized by a maximum temporal window size of 100 and at an optimum threshold of 0.05. However, in Figure 4.21, DAWF algorithm of different window sizes has shown less impact on solar phenomena, as of a minimum window size of 10 reduced 63.70% and the window size of 100 has been reduced 66.19% only by an absolute threshold. On the contrary, whereas the threshold parameters on solar phenomena showed outperforming results with minimal window size of 10 by a minimum threshold of 0.001, which has minimized 3.22%, and a maximum threshold parameter of 0.1 suppressed by 79.94%. In this scenario, we noticed that the performances of different window sizes were delivered the similar results compared to the threshold parameters. Because, the node has generated less redundant datasets of solar over the given time instances, moreover, the monitored datasets are often highly varied data.

Figure 4.22 (a) and (b) show the temporal performances trade off between Algorithm 3 and Algorithm 4. As explained both the algorithms in section 1.3.2. This

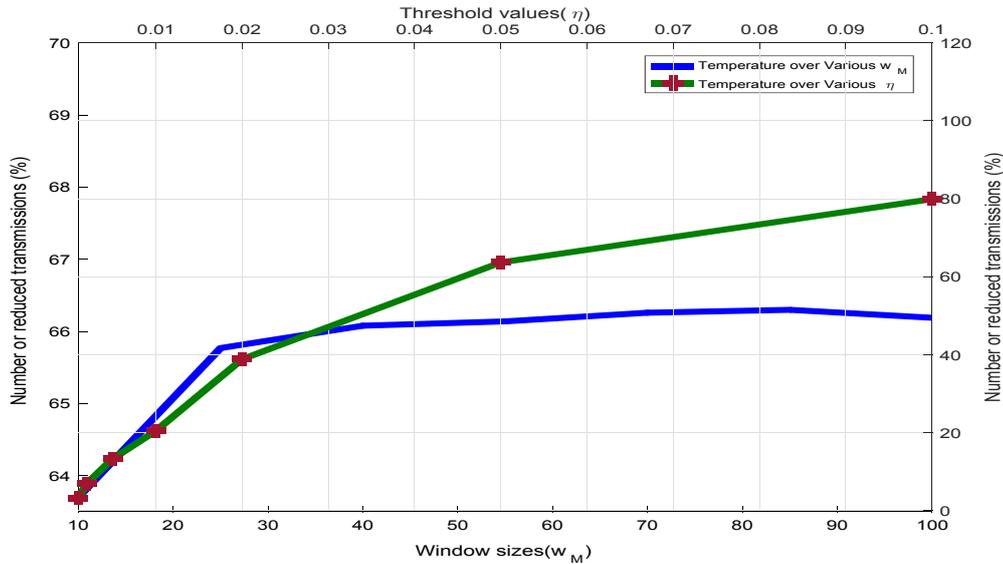


Figure 4.21 – Solar: Sensor node temporal reduced transmission metric performances over various thresholds and the various window sizes

chapter then made a comparative analysis between Figure 4.22 (a) and Figure 4.22 (b). In Figure 4.22 (a), the window counter considers to send total window dataset for that current round if the window contains any non-redundant data, as showed in the plots, thus, the window increases then the total amount of delivered data messages count has also been maximized.

In Figure 4.22 (b), it is total contrast as the window increases then the total amount of delivered measurements were decreased. Because, in every window round, the counter function subdivides the window into several, in order to check the window, whether it contains more redundant or non-redundant information. That means, the counter function divides the consecutive redundant data portion of window as a mean averaged sampled value, and the other portion of non-redundant window data will be sent as a single sub-window packet or sent as several non-redundant data messages accordingly. In this scenario, node indeed achieves the higher performances in terms of reducing the bandwidth and saving energy compared to the Algorithm 3. But, it is not the case all the time in system performances, sometimes it may lead to the communication overheads because of smaller consecutive sub-window data sets, if the main window dataset has more highly varied sensor readings than the redundant readings or any node fault tolerant occurrences.

This trade off also made an analysis, in order to find the optimal window sizes. In case of window increments from the minimum size, whereas it has showed that window sizes of 10 to 20 having the optimal points, after the optimal range of varied window sizes that are performed. Similarly, a less varied performances can be observed between both the algorithms (Figures 4.22 (a), (b)). The following Figures

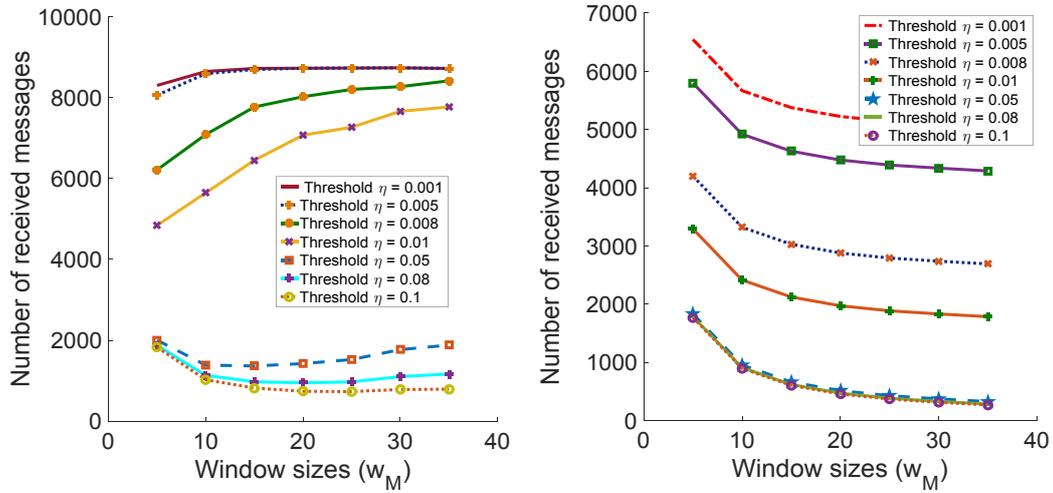


Figure 4.22 – Trade-off between DAWF first algorithm and second algorithm

describes a comparative analysis between the propositions and the related works of TiNA protocol, which is based on a TAG aggregation service from TinyDB databases. This comparative analysis indirectly related to the TAG protocol as well. However, TAG has many states of query-based aggregation engine in TinyDB databases based on EPOCH DURATION  $i$ , which is not the case in this proposition, thus, differs to compare with TAG. Moreover, In this analysis, the propositions were carried out the performances of three physical quantities, while compared to TiNA and Adaptive-DAWF (A-DAWF) protocol, in terms of relative  $tct$  and the clause condition.

Figure 4.23 and Figure 4.24 shows the ambient temperature performances of relative  $tct$  over the absolute window sizes of 5 and 10. However, TiNA protocol did not have any window principle, where it only observed data sets through the clause condition of TAG over  $tct$ . The algorithms of DAWF and A-DAWF still showed the better performances compared to TiNA protocol, while reducing the total amount of redundant data transmissions. Although, DAWF protocol shows less performances at the very minimum  $tct$  value of 0.001 compared to the TiNA. As DAWF held with the minimum absolute window sizes of 5 and 10, since it evaluates the small window dataset over that window duration of every value to each other, thus, differs the time and performances when compared to TiNA. Because the window mechanism drops the reading value when the current value similar with the new value, otherwise the value can be accounted in order to carry over that current window interval.

However, the A-DAWF window principle has been treated differently, in order to carry small portions of the window few times. As the counter function partitions the single window data as small parts of few windows, in order to separate the significant data and redundant. In Figure 4.24, where it showed the huge differences in its performances while compared to the DAWF and TiNA protocols. In temporal scenarios, as the data window varies upto the certain extent level, then the performing results also varied. While its demonstrated in both the Figures 4.24, 4.23 that the window de-

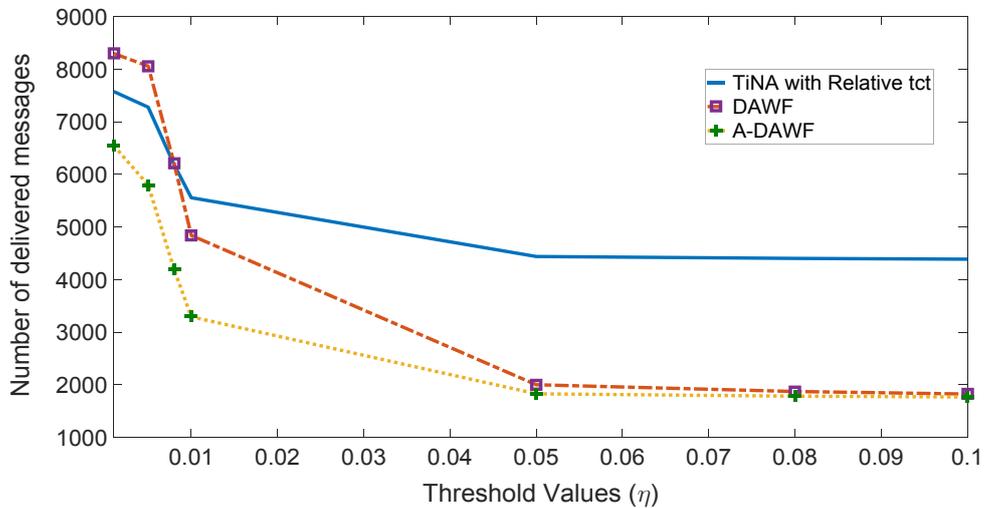


Figure 4.23 – A comparative analysis of Ambient temperature phenomena between TiNA protocol with relative  $tct$ , DAWF and A-DAWF algorithms with the window size of 5

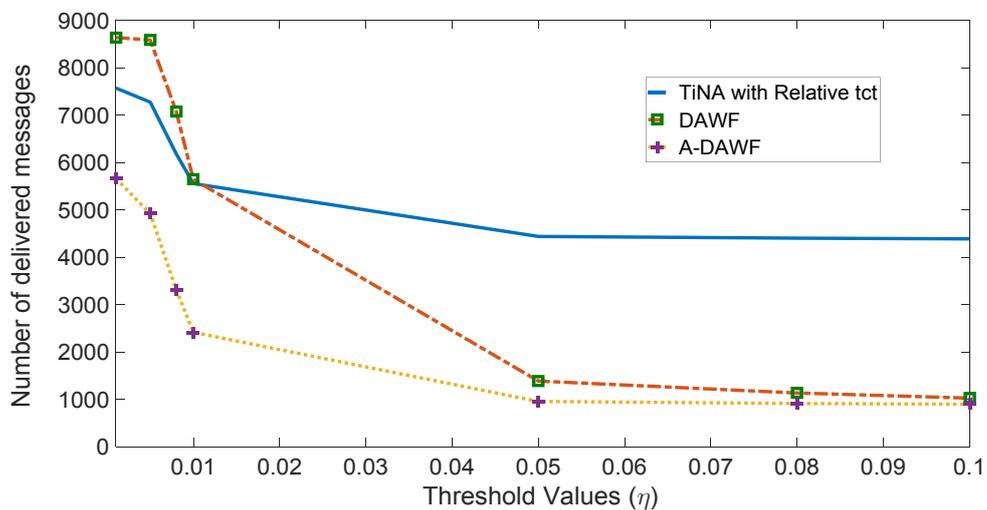


Figure 4.24 – A comparative analysis of Ambient temperature phenomena between TiNA protocol with relative  $tct$ , DAWF and A-DAWF algorithms with the window size of 10

livered performances are more than 12% compared to the Figure 4.23. However, it's not the case to all the phenomena, since the mechanism has physical data sensitivity, and works differently for both indoor and outdoor sensor environments.

For instance, Figure 4.25 and 4.25 represent the solar datasets of indoor environments from the thermal housing project. It is certainly a threshold dependency case, because of the solar sensor readings. The sensed datasets are highly varied data over the period. If the threshold tolerance range can be increased, then the delivered performances has also been maximized.

In this case, if the relative  $tct$  varies then the performing results also increased, which was shown in Figure 4.25, but it is not the case in the above figures of indoor sensor datasets. Like temperature, solar phenomena also evaluated through two absolute window sizes of 5 and 10. The trade-off between window sizes and threshold ( $tct$ ) ranges of solar phenomena,  $tct$  always showed the better performances compared to the various window sizes of 5 and 10, even at the larger window intervals, the performances are decreased, as delivered in Figures 4.25, 4.26. However, in this comparative analysis, TiNA protocol showed better performances when compared to the DAWF, but lagged than A-DAWF proposition at a window size of 10. However, DAWF and A-DAWF delivered better performances at the window size of 5 compared to TiNA, as showed in Figure 4.25. In this comparative analysis, DAWF lagged because it carries total amount of window data irrespective of  $tct$ , if current window contains greater than 0 NRDM messages.

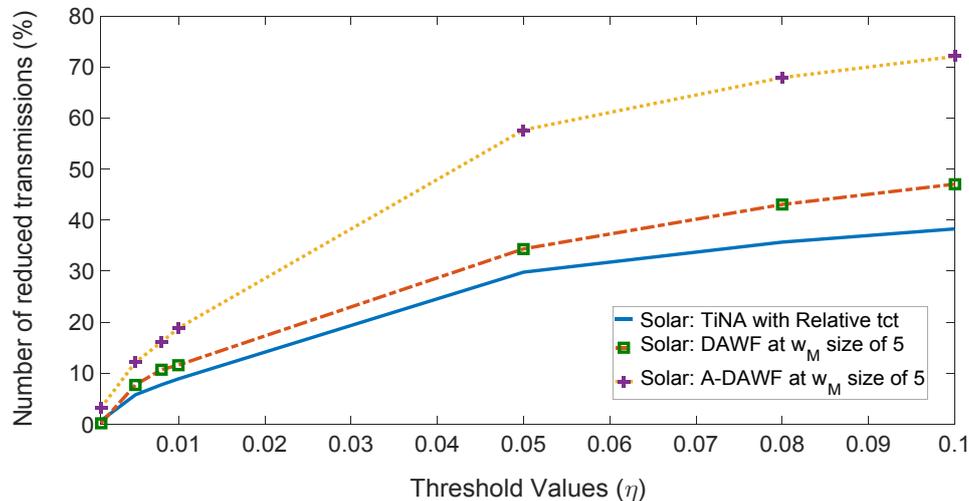


Figure 4.25 –  $R_{TX}$  comparative analysis of Solar temperature phenomena between TiNA protocol with relative  $tct$ , DAWF and A-DAWF algorithms with the window size of 5

In order to calibrate the reduced number of transmissions, this thesis introduce a  $R_{TX}$  metric, which will only present show the solar phenomena performances in both the Figures of 4.25 and 4.26, instead of the total amount of considered physical

quantities. As discussed above, if the captured data is highly varied, then the relative  $tct$  performs well over the absolute window sizes rather than varied window sizes. It is certainly common, since the algorithm works based on the both lower and higher sensitiveness of the data sets, and then they filtered accordingly on the basis of either relative  $tct$  over absolute window sizes or relative window sizes over the absolute  $tct$  values.

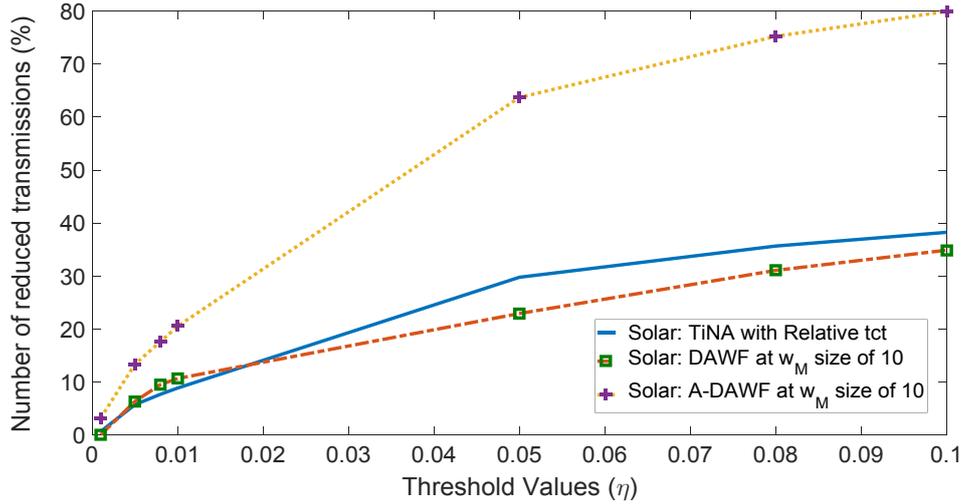


Figure 4.26 –  $R_{TX}$  comparative analysis of Solar temperature phenomena between TiNA protocol with relative  $tct$ , DAWF and A-DAWF algorithms with the window size of 10

The Figures of 4.25 and 4.26 demonstrate the filtered performances of temporal data sets at two absolute window sizes of 5 and 10 over relative  $tct$ . Although, TiNA performs better over DAWF protocol, due to the long time intervals compared to the TiNA time instances. Because DAWF carries total amount of sensed window phenomena at a time when it held with non-redundant data. However, A-DAWF showed outperforming results even at the optimal  $tct$  of 0.05 and a window size of 5, which has reduced upto 60% of the signal bandwidth.

#### 4.6.4. Spatial Performance Analysis

In order to suppress the redundant data and their redundant transmissions spatially. DAWF proposition has used the CH node to exploit spatial redundancies as well as the exploitation of data window correlations that are received by the sensor nodes. Like temporal experimental study case, spatial scenario also considers the metric of  $R_{TX}$  and  $MAE$  metric for prediction error rate among the nodes phenomena.

Figures 4.27, 4.28 and 4.29 presents the varied window performances over threshold tolerances in time. In this scenario, we show that how the window performances are gradually increased over time while tuning the parameters of  $\eta$  and

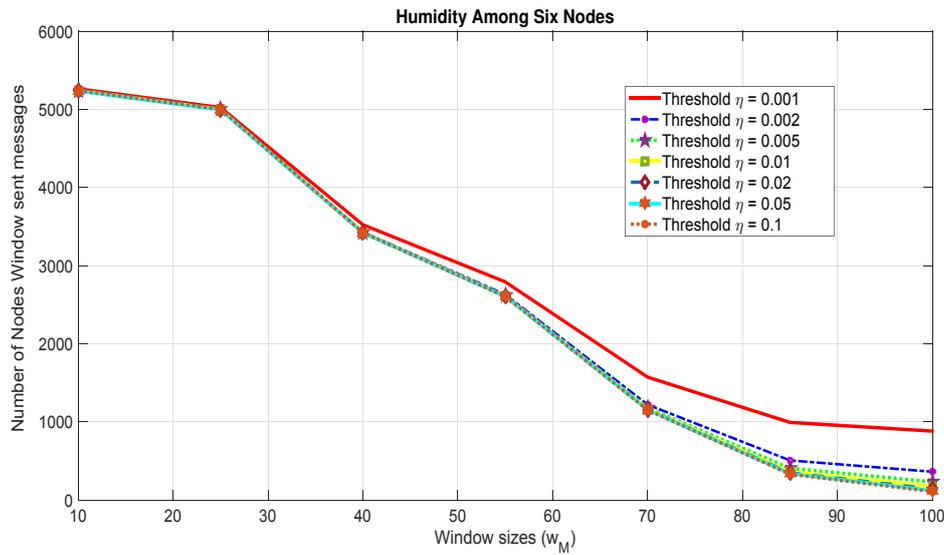


Figure 4.27 – Humidity Phenomena: CH node performances over various window sizes.

$w_M$  combinedly compared to the individual variations as shown in the Figures 4.16, 4.17, and 4.18. This comparative analysis has proved that window performances of nodes are directly proportional to the optimum thresholds of the sensor nodes and their dynamicity based on the different phenomena. However, threshold sensitivity of the phenomena depends on the application considerations that how much acquired phenomena needs to be filtered before processing the required amount of samples to the BS.

The trade-off between relative threshold performances from 4.27 and the MAE prediction error rates of 4.34 showed the performance differences, that how the minimum to maximum of relative thresholds quantify the signal over various phenomena, whereas the absolute value of 0.05 showed in Figure 4.34. The delivered performances are nearly same after the absolute value of  $0.05Th$  in both temperature and humidity. Because both the phenomena are gradually varied over the longer duration of time. But, it is not the case in solar phenomena, since the phenomena has more threshold sensitivity than the window intervals, unlike temperature and humidity.

Figures 4.30, 4.31 and 4.32 illustrate the spatial data performances over various window sizes and at relative threshold parameters. As followed by Figures 4.30 and 4.31, which are certainly showed that the threshold parameters have been delivered the outperforming results compared to the window sizes. The varied window performances in spatial scenarios, which have showed less impact compared to the threshold parameters, unlike as they performed well in temporal scenarios of the sensor nodes. In this case, whatever the window size is being maximized, most of the delivered data are showed slight variations over the varied window performances, but many remain constant, as shown in 4.31. Because of CH received data by the nodes that are not

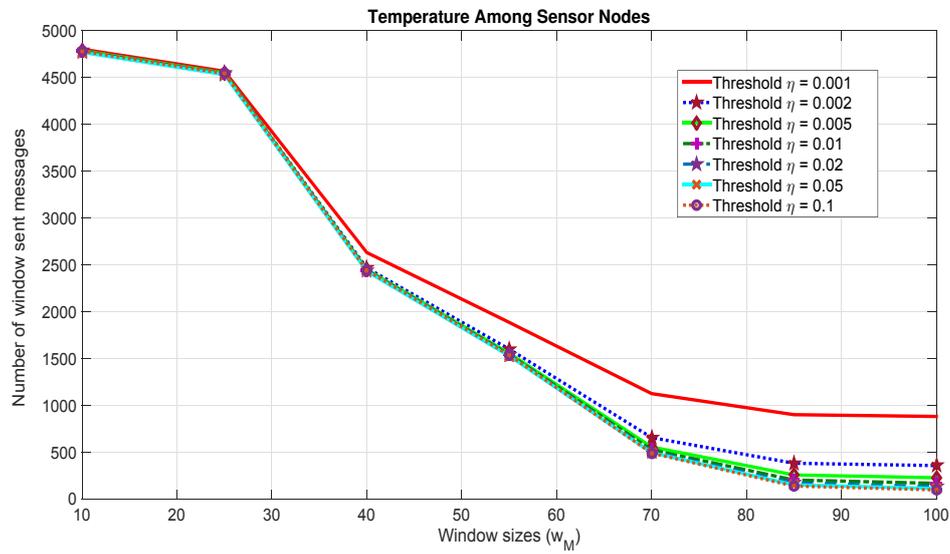


Figure 4.28 – Temperature Phenomena: CH node performances of over various the window size.

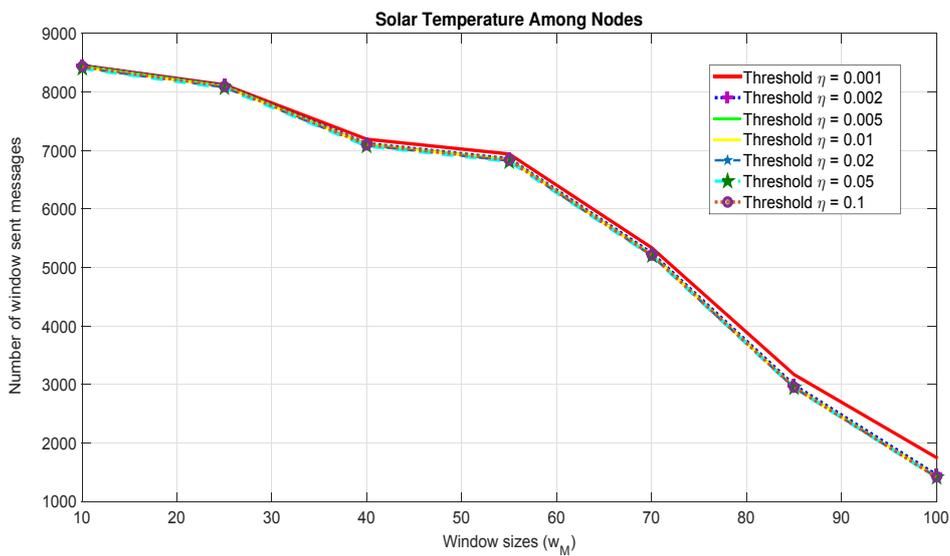


Figure 4.29 – Solar Phenomena: CH node performances of over various thresholds through the window sizes.

either highly redundant or fairly correlated. In this scenario, we presented the varied window performances with the consideration of two fixed optimum thresholds of 0.05 and 0.1. The analysis has been conducted on the considered phenomena of nodes, as shown in Figure 4.32.

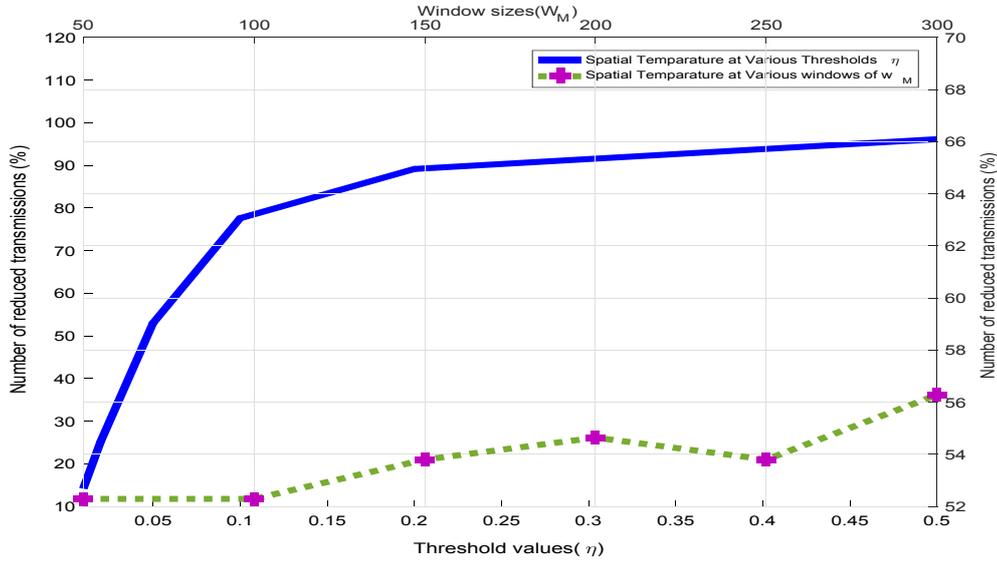


Figure 4.30 – CH DAWF  $R_{TX}$  metric performances of nodes temperature over various thresholds through the window sizes.

Figure 4.30, 4.31 and 4.32 explores the performances of spatial  $R_{TX}$  metric over received data messages of nodes. In this metric analysis is given by Eq. 4.6. As, we carried out the performances from the thresholds over window size variations. In Figure 4.30, 4.31 and 4.32, where CH suppressed the amount of 14.17%, 5.19%, and 2.90%, respectively by computing through a minimum threshold of 0.01, and the maximum threshold 0.5 performances of 96.15%, 95.25% and 78.65%. On the other hand, DAWF window sizes delivered performances are a minimum of 52.29%, 22.95% and 10.75% by an optimum threshold at 0.05, and with the maximum window constraints of 56.28%, 26.34% and 13.15% among the nodes delivered phenomena. As clearly demonstrate that the threshold parameters will always play a significant role, in order for delivering the better performances in spatial phenomena.

According to the data sets, nodes follow the periodic transmission system by every hour, then it transmits total considered nodes measure of their sensed phenomena at the given time intervals. In contrast, using DAWF mechanism at a minimum window interval of 50 through CH where it reduces 59.29%, 22.95% and 10.75% of redundant messages out of total received data, while computing through the constant threshold parameter of 0.05. Thus measured through the Eq 4.6, and shows that CH node can still suppress the unnecessary data messages among the cluster nodes.

Furthermore, temporal redundancies in CH node by various thresholds with different window sizes of  $w_M$  also performed same as sensor nodes, which were presented

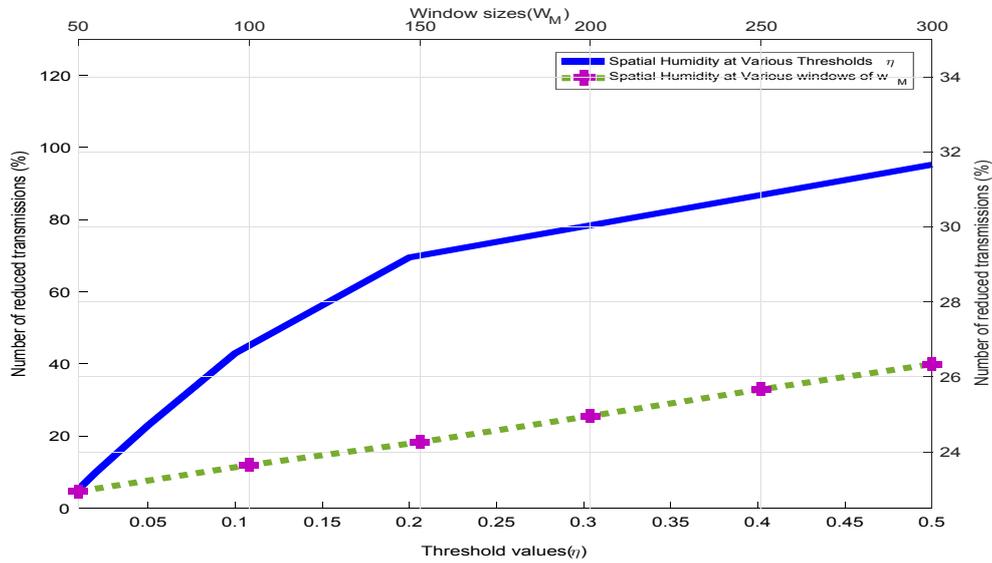


Figure 4.31 – CH DAWF reduced transmission metric performances of humidity over various thresholds through the window sizes.

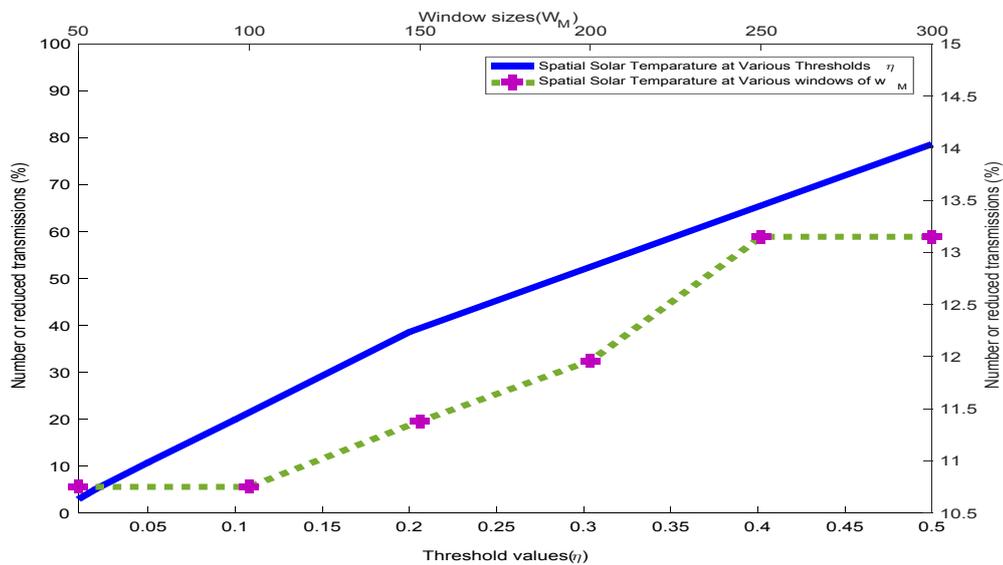


Figure 4.32 – CH DAWF reduced transmission metric performances of solar temperature over various thresholds through the window sizes.

and explained in the above section of 4.1. This is also stated that the temporal redundancies may greatly depend on the window sizes rather than threshold values. Thus, it exploits 90% of temporal redundancies in CH sensed phenomena by computing with optimal thresholds and maximum window sizes. This is obvious, because, if the window sizes gets bigger then the stored data and comparison between every two window stored readings can drastically be increased. In order to show the performance analysis of the proposed approach in terms of energy saving, as briefly illustrated in the above results. The DAWF can effectively suppress a required number of transmissions with respect to the periodic or delay-tolerant approach, whereas each node can forward its obtained data whenever it detects towards the data collector or BS.

#### 4.6.5. Signal Reconstruction

This section describes as twofold, first one presents the mean absolute error (MAE) metric over various considered phenomena. Second one focuses on the signal reconstruction procedures from the received data, and then illustrate the results and their performances. The following derivation can be achieved the MAE cost among the considered sensor nodes, and can be derived as

$$MAE = \frac{1}{N} \sum_{i=1}^N |f_i - x_i| \quad (4.9)$$

Where,  $f_i$  is a prediction values of the datasets and  $i$  is the index of prediction values over time,  $x_i$  is the original data values over the duration of  $i$ .

We measured the absolute error cost at each phenomena, in order to explore the error rate from acquired data sets. In this case, a trade off between MAE cost of various thresholds over the error cost of various window sizes, and the MAE metric cost can be calculated by Eq. 4.9. The signal reconstruction, we applied a first interpolation method to reconstruct the original signal based on the reduced sample data.

Figure 4.33 and 4.34 show the various metric performances over the considered network area. The reliability of the proposed aggregative models has been assessed by computing the transmissions ( $TX$ ) that are suppressed, and the  $MAE$  prediction error over the considered amount of sensor nodes of their physical phenomena. In this scenario, the study case of  $MAE$  metric performances demonstrated over the temperature phenomena of various sensor nodes. In Figure 4.33 and Figure 4.34, presents the performance analysis of the metric over various window sizes as well as over relative thresholds, respectively.

In Figure 4.33, the measured  $MAE$  cost in temperature has varied from 0.2356 to 1.7826 among the sensor nodes by the considered area of the network, which is the attained error of various window sizes by a fixed threshold of 0.05 for minimum to maximum prediction errors of  $MAE$ . On the other hand, Figure 4.34 presents the minimum to maximum threshold prediction errors of  $MAE$  varied from 0.0448 to 0.1720, respectively. However, the attained error shows less than 1% among all the

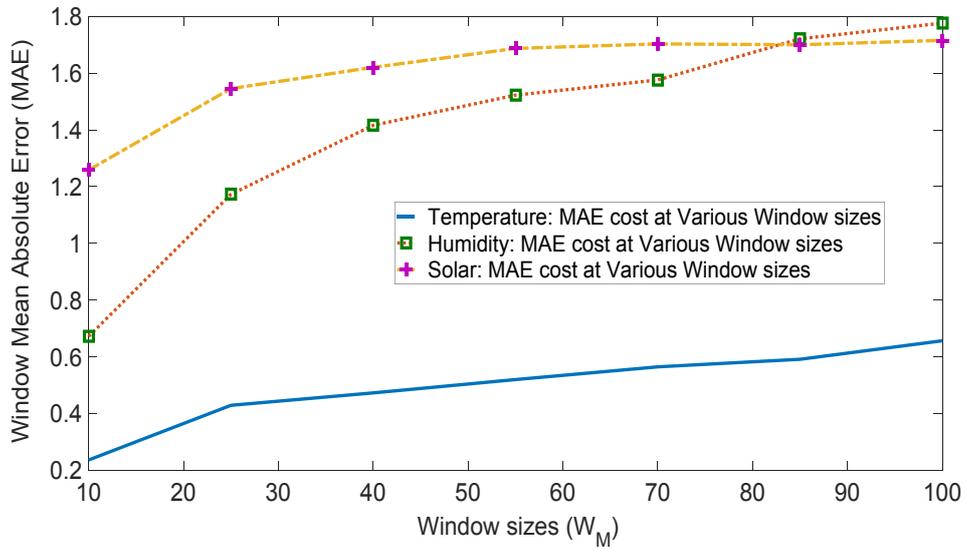


Figure 4.33 – MAE metric cost of node 1 at Interpolation over various windows and at an absolute threshold. In this scenario, we present the prediction error rates through MAE metric, while performing the interpolation of signal reconstruction for the considered amount of phenomena at various window sizes.

nodes phenomena. In this case, more the error gets increased then the more signal gets compressed, as the DAWF parameters auto resetting is strictly an application choice.

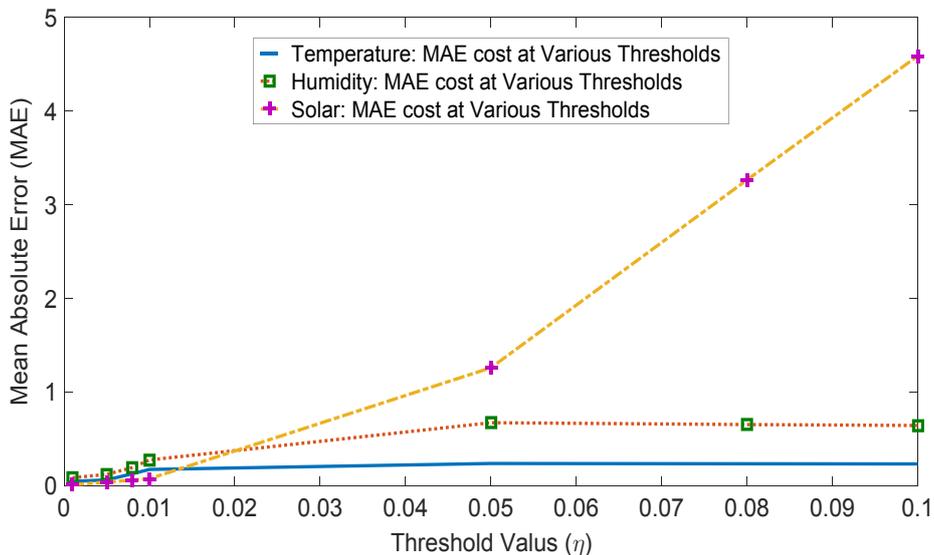


Figure 4.34 – MAE metric cost of node 1 at Interpolation over various thresholds at an absolute window size. In this scenario, we present the prediction error rates through MAE metric, while performing the interpolation of signal reconstruction for the considered amount of phenomena at various threshold values.

Figure 4.33 shows that how the performances of MAE metric can gradually vary over different phenomena at various window sizes as well as various thresholds. However, the MAE cost of various thresholds did not vary much compared to over the various window sizes, as shown in Figure 4.33. As it is not the case at solar phenomena, as showed in Figure 4.33 that DAWF error thresholds have delivered the better performances compared to the various window sizes. However, the maximum error value of various window sizes are reached by 0.6 for temperature, 1.79 for humidity, and 1.65 for solar, respectively. The minimum MAE cost of various window sizes are 0.21, 0.62, and 1.23 only, which is not much bigger compared to the threshold error rates, as shown in Figure 4.34.

In Figure 4.34, the minimum to maximum prediction error rates lesser in both temperature and humidity phenomena. In humidity datasets, the attained error rate is bit higher compared to the temperature, since the error rates are different than temperature, because of their differed data ranges. However, it is not the same scenario, when it comes to solar phenomena, because of their highly varied data sets, the relative  $tct$  plays a key role compared to the window sizes. If the relative  $tct$  varies then the amount of filtered data have also been increased over the absolute window data. In this case, the window filtration certainly rely on relative  $tct$  value, as shown in Figure 4.34.

The evaluations also showed while placing all the phenomena of MAE cost plots into one, which were demonstrated in both the Figures of 4.34 and 4.33. Figure 4.34 presents the relative  $tct$  performances of MAE cost over the three phenomena, and Figure 4.33 describes the various window size performances based on considered data sets over an absolute  $tct$ . The main objective to demonstrate that the differed performances and their rapid fluctuations between both the randomly varied and periodically varied sensor data sets.

In Figures 4.35, 4.37, we show the performances of signal reconstruction from the acquired data samples after the pre-filtration. The data sets among the three sensors, each one illustrates the original data signal and the interpolated signal from the received data of sampling window protocol. As shown in Figure 4.35, the retrieved sparse signal from the interpolation method can deliver the better results. However, we notice the signal noises during the interpolation, although, DAWF clearly showed in the below figures that it can still able to reconstruct the original signal while eliminating the signal noise. In regard to the better signal reconstructive analysis, we presented a figure with some portion of the original, and then the retrieved data signal from the original signal, as demonstrated in Figure 4.36.

However, the above figure represents the retrieved data signals that are filtered by Algorithm 3 and Algorithm 4, as shown in Figure 4.36 (a) and (b). The distinguish between these two signals pattern states that the quality of data in 4.36 (a), if only the user interested on a quality signal. Otherwise, the second one focuses on more compressed data signal, if the application challenge is only bandwidth reduction other than the signal reconstruction, then the Algorithm 4 (Figure 4.36(b)) would be a better choice to perform the effective pre-filtration.

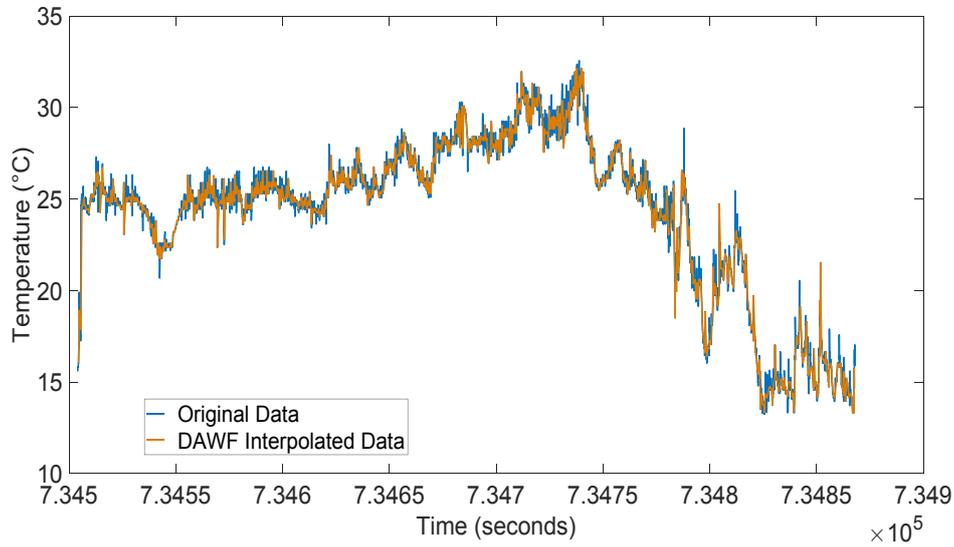


Figure 4.35 – Original and interpolated signals of ambient temperature from node 1.

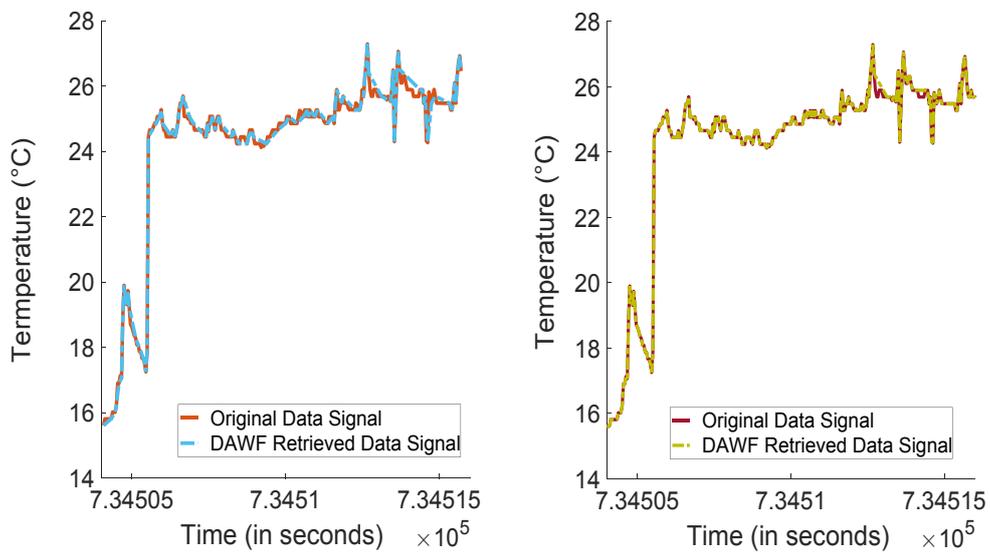


Figure 4.36 – Original and interpolated signals of ambient temperature from node 1.

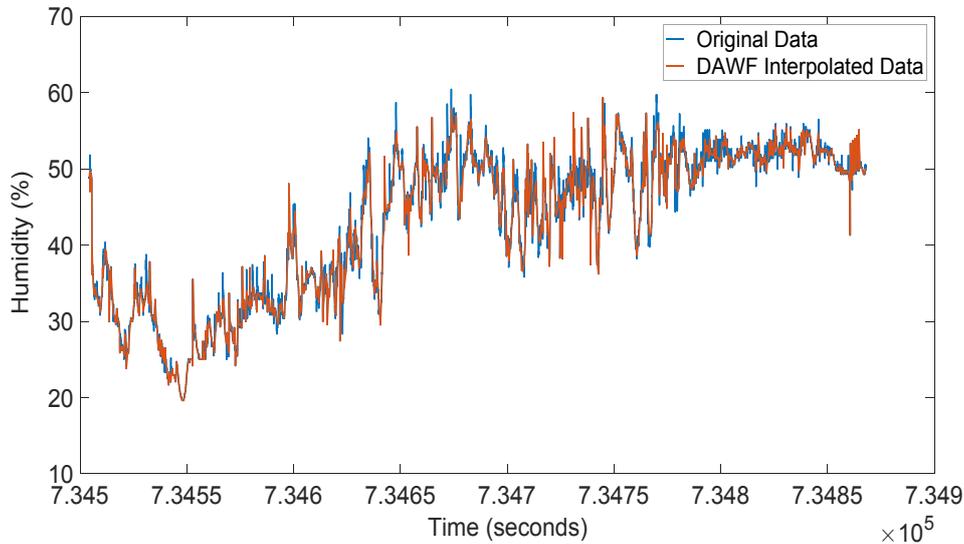


Figure 4.37 – Original and interpolated signals of humidity sensor of node 1.

## 4.7. DAWF Temporal Evaluations in COOJA Cross-level Rime Network

This section describes the DAWF experimental study case in COOJA simulator, in order to carry out the real network integration. This thesis has been considered to integrate the propositions in contiki based COOJA simulator, since the developed software in COOJA can be reused for various testbed platforms, as discussed in section 4.5.1. The propositions are integrated in Rime network stack, which is based on cross-level network for low-power embedded devices. The developments are integrated in both network and application layer. While the network layer used for network management and organization of the nodes based on single-hop primitive by chapter 3. Then, the application layer used for the DAWF design developments, although, DAWF algorithm supposed to be integrated on physical and link layer, since there is no default ADC sensors in COOJA simulator. In this case, we used the primitive of randomized light sensor at Java platform from one of the contiki sources, and then re-modified the source in application layer, in order to follow up the DAWF developments.

Throughout the thesis, for the preliminary integration and the experimental study case in COOJA, that we have developed DAWF learning phase algorithm in time. The considered network assumed as static cluster-based topology, and nodes held with default destination IDs, such as a fixed destinator node addresses during every node transmission, which we consider for our propositions as either CH or sink node.

The listed choices of module parameters and development specifications from Table 4.3, which were used for the COOJA simulations. However, the header size is prefixed in Rime platform, and the predefined packet attributes can be reused ac-

Module	Implementation
Rime	Cross-level
Node platforms	TmoteSky & Z1
CSMA, ContikiMAC	Duty-cycled
Transceiver	CC2420 driver
Sensors	Light
Clock rate	8 Hz
Payload size	2 & 10 bytes

Table 4.3 – Module parameters for the implementation of DAWF data reduction algorithm.

According to the developer considerations. While DAWF proposition primarily designed for singlehop network, which does not require any extra header packetbuf attributes, such as ACK, RSSI and LQI, and etc.,. Although, we have only used the simple header fields of source, destination node addresses and the corresponding sequence number of the packets, and the DAWF payload packet information, as showed in Figure 4.6.

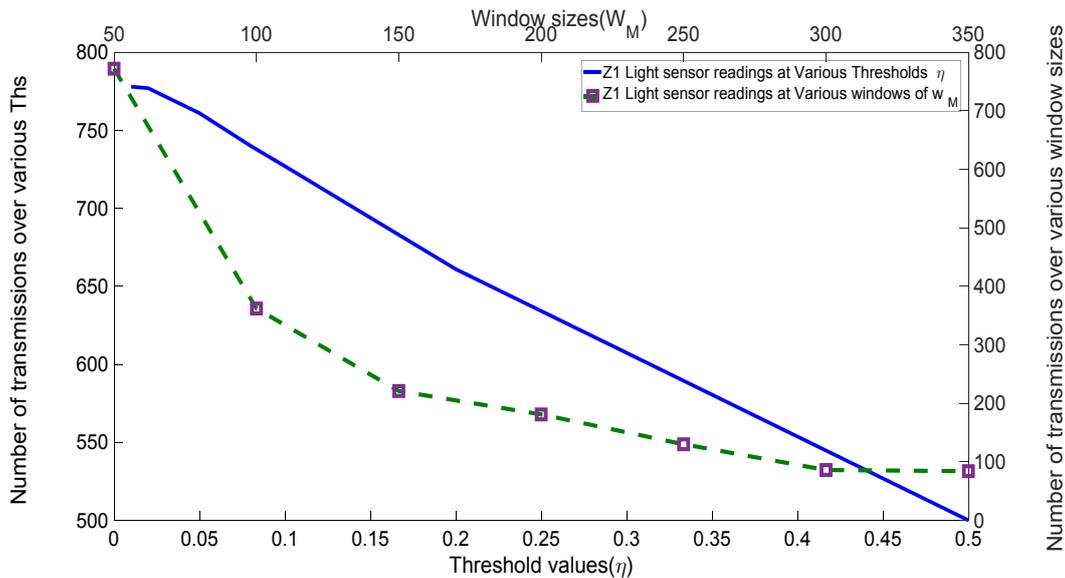


Figure 4.38 – Z1 Mote Light Sensor from COOJA.

While we used the sensory randomization method in COOJA, in order to generate the light sensor readings in both TmoteSky and Z1 mote platforms, as showed in Figure 4.38. The generated readings are randomized light sensor readings at every 2 seconds of time, which is highly varied data over short time intervals. In this case, we considered the NRDM window message principle of DAWF mechanism. As it only carries the total amount of window, if the window contains more than zero non-redundant data. However, the simulated nodes exploit few temporal redundancies only

over the short window intervals, and then carries remaining window data as single packet, since the generated sensor information is highly varied data over the time intervals. As the data window intervals do short evaluations of their corresponding windows data, and then forwards data messages to the CH. Hence, in order to exploit spatial redundancies in CH node over the received data of nodes, this proposition has used the spatial use case to exploit the spatial redundancies among the received data. Because, the considered CH held with better energy and computational resources compared to the nodes, which can also able to perform the longer evaluations over the received data.

Figure 4.38 showed that a tradeoff between varied window performances over the relative thresholds. It also clearly shows that the relative thresholds shown the better performances compared to the varied window sizes, because the generated light information is highly varied data over short time instances. Figure 4.38 also demonstrated that relative *tct* plays a key role in order to fix the sampling error rate.

In order to evaluate the performances of the proposition. We consider energy metric to measure the cost of energy at each module of the TmoteSky sensor node.

Figure 4.39 shows that how much energy it is usually consumed over simulated Sky motes of both communication and computational modules. As showed in 4.39, the least energy consumed module is LPM among all, following others. The node carries 10 bytes of DAWF payload message during every transmission, although it has consumed less energy compared to the reception mode of operation. Figure 4.39 showed that the total amount of energy that has spent at different modules of CPU, LPM, TX and RX.

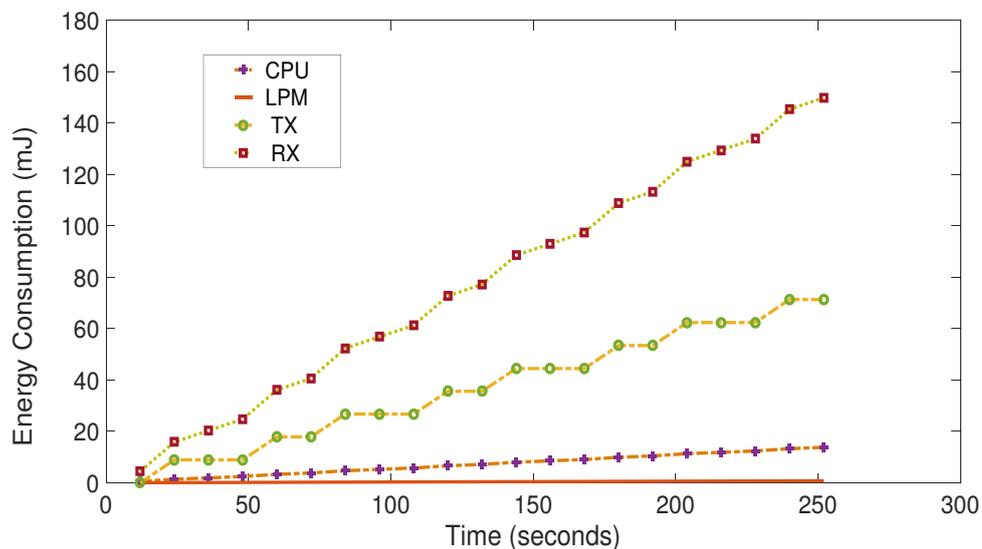


Figure 4.39 – Tmote Sky COOJA node energy consumption (milli Joules) over time. While carrying a window dataset of payload data size 10 bytes.

In Figure 4.40, TmoteSky simulations are captured at every 22 seconds of  $r$  inter-

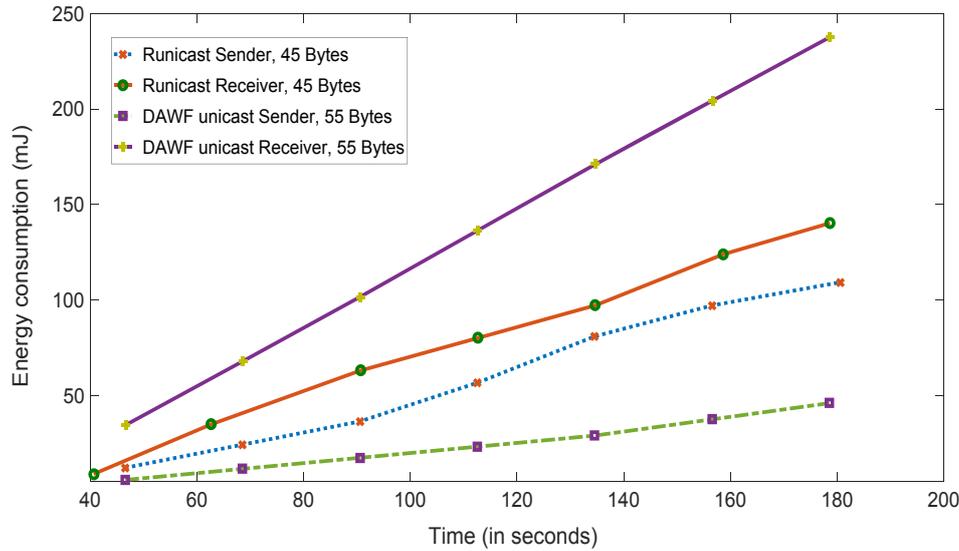


Figure 4.40 – Influence of other parameters on Radio energy consumption of TmoteSky in COOJA.

val by the DAWF sampling window, and the light sensor readings that are randomly generated by every 2 seconds. Because, the light sensor readings are highly varied data over the short time intervals. As showed in Figure 4.40, we conduct several simulation tests in order to calculate the energy cost at two different primitives. In this scenario, Rime-*Runicast* primitive carries 45 bytes of a cross-level packet, whereas DAWF temporal carried the total window of 55 bytes. In this scenario, DAWF did not use any MAC ACK retransmission method, it only does the simple unicast communication principle, unlike Runicast method. However, DAWF still showed the better performing results compared to Runicast method, whereas DAWF TX mode spent less than 50 mJ over 180 seconds of time, and Runicast consumed greater than 100 mJ of energy. Although, DAWF spends more energy in RX mode, which is greater than 230 mJ over the short duration compared to the Runicast. Because, the neighboring sensor nodes listen all the radio packet activities of sender node every time, since the packet size is large, and consumes nearly same energy like RX mode for listening the packets. As showed in Figure 4.41, whereas the other modules of both the nodes consume nearly same energy.

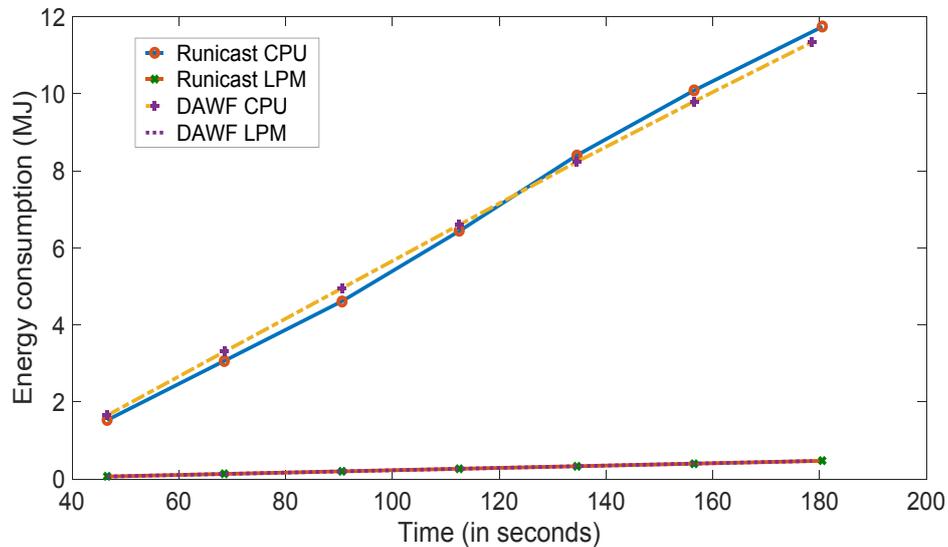


Figure 4.41 – Influence of other parameters on CPU and LPM modes energy consumption of TmoteSky in COOJA.

## 4.8. Summary

In this chapter, we have presented a spatio-temporal data reduction approach for WSNs. It provides a simple data pre-filtration solution in nodes, in order to reduce the signal bandwidth and forwards a required set of data measurements for the signal reconstruction at sink node. We first present the impact of our propositions over the considered amount of data sets by using MATLAB, later showed the trade off between threshold and window parameters that how the performances may vary over the various window sizes and relative thresholds of the algorithms. This analysis gives us an idea to hold the amount (%) of redundant information that we need to suppress through the window over the period. However, in temporal case, optimal window sizes showed better performances compared to the relative threshold variations. In addition, we also make a comparison between algorithm 1 and algorithm 2 performances. According to the algorithm 1, which sends a total window dataset for current round, if non-redundancy count is more than 1 or equal to one. But, as per algorithm 2, it sends only that the smaller portion of non-correlated window data as another new packet transmission, and the redundant portion of window data as mean averaged data packet measurement. Moreover, this chapter evaluations are based on the real world data sets of Intel research lab, and DREAL data sets of LCIS lab. This chapter not only presents a simple data reduction solution, but also demonstrated a signal reconstruction method by DAWF mechanism, which is not the case in existed literature of in-network aggregation methods, such as TiNA and TAG aggregation methods.

# Chapter 5

## Conclusions and Perspectives

The combination of routing and data-driven in WSNs can benefit greatly from the utilization of both the features, as its usage gives the hybridization of reliable data transmissions and robust communication links from routing, and the sampling, encoding, compression and prediction in both temporal and spatial data distributions. Therefore, the pre-processing and pre-filtering control techniques of the nodes have greatly been in focus of framing both the effective data reduction and routing techniques for WSNs. Although, the network timeline management of both methods is composed with individual sensor nodes, which are either centralized or distributed, and they are able to adapt the application needs accordingly. A numerous application examples and proof-of-concept studies have been presented in this area. However, they are specifically designed with either simple data aggregation along with the routing techniques or rich computational data-driven models without giving a much priority towards routing. While both the areas are too broad to target the challenges for generic applications, many existed protocols are either application-specific or non-application specific, which then have less feasibility to work based on real application scenarios. Among these, both the studies and their technical challenges must be addressed in order for prolonging the WSN network lifetime. Throughout the literature studies, this thesis also addressed that there is very minimal work that has been investigated in both the combination of simple routing and data reduction techniques. This investigation helps to reduce the communication overheads as well as prolonging the network lifetime. Moreover, the advanced or super nodes of CHs can greatly support to avoid the communication overheads in hierarchical networks. Thus also provides a solution for other methods of energy preservers, such as data reduction and duty-cycling techniques, which can be easily integrated along with the routing. On the other hand, the pre-filtration of data in nodes achieves better data compression or aggregation over the network, which has also been investigated in this thesis. The proposed data reduction method also demonstrated that it can easily achieve the original data signal, based on its processed data, as presented in the chapter 4.

The lifetime of a sensor network has been a major concern in the field of WSNs from the beginning of research in WSNs. Since the network lifetime is strongly cor-

related with the lifetime of every sensor node in the network. As prolonging the network lifetime, a vast amount of work has been conducted in WSNs, in order to maximize the nodes lifetime. The contributions based on energy-efficiency through the combination of both routing and data-driven approaches that have been studied by this research work, which are presented as twofold in this thesis. The first one has presented based on the improvements over hierarchical clustering methods, and then presents a static cluster-based approach in COOJA simulator. Second possess the research field of data reduction, which is a simple novel approach based on the hybridization of sampling and compression based. While the method exploits the data redundancies and simple correlation evaluations over the network by utilizing the data aggregative window mechanism and signal reconstruction methods.

With respect to the energy conservation in sensor nodes, a data window aggregation algorithm has been presented, which is specifically aimed for suppressing the nodes redundant data transmissions. This framework primarily designed for continuous, periodic and delay-tolerant applications, in order to filter the sensed data of various physical quantities. In general, physical phenomena recurrences are common in wireless medium due to their repeatable nature over time to time in both time and space. The utilization of pre-filtration methods in nodes avoid redundant data transmissions, while providing the significant data towards the sink node. This prospect allows the node to increase its lifetime from the predefined lifetime, as well as decreases the power of both transmitter and receiver node modules.

The limited computation and energy resources of the nodes are not compatible with the complex, rich computational models and high rate data transfers, because, the low-power sensor nodes are operated by batteries. While the improvement of energy-efficient protocols provide an extendable lifetime solution for sensor nodes that to be extended. But, the energy limitation remains due to their restrictive resources nature of the battery operated sources. Particularly, the consideration of energy-efficient protocol is an effective method that to be used, in order to extends the sensor nodes lifetime. Moreover, efficient protocol frameworks can also be useful for sensor nodes that are operated by other power sources, such as energy harvesting methods of solar, wind, temperature, etc,...

This thesis has framed as twofold, first one focuses on the single-hop routing strategies based on hierarchical networks, which then later integrated into COOJA simulator in order to follow chapter 4 methodology. In this case, we primarily considered the static cluster-based topology in COOJA on the basis of dedicated CH nodes for single-hop network, in order to avoid the system complexity in nodes. We used a simple unicast primitives in order to design and develop the propositions in COOJA. For the data reduction technique, we follow the combination of all in-network processing, compression, and prediction-based aggregation. Although, unlike fully prediction-based model, we presented a simple comparison-based relative function in window mechanism, in order to evaluate the data, and let the window pre-filtration before forwarding any messages towards the CH. This thesis provides a multi-tier solution that covers the cross-level layering of WSN sensory systems, from sampling

data upto the data transmissions, and then the signal reconstruction either at advanced powerful nodes like CHs or sink node. However, due to some unwanted distortion of the sensed data might decrease the quality of data that have been forwarded to the sink node.

The second part of this thesis shifts towards the multihop networks, as the dissertation presents a preliminary novel approach in appendix chapter, in order to integrate and evaluate the propositions into multihop scenarios that demonstrate the performances in large scale WSNs as well. However, it presents as ongoing future work from this thesis.

## 5.1. Contributions from this thesis

This thesis focuses on pre-filterized data-reduction techniques in WSNs that are based on hierarchical routing and inexpensive computational time-driven application model. The feasible solution has been provided to tolerate various sensed phenomena more efficiently that where the data collection signal is either costly or in-feasible in both the indoor and outdoor environments.

In order to support the efficient data reduction algorithm, this dissertation considered the direct involvement of hierarchical routing in single-hop WSNs. As part of the research study, this thesis pursues the fields of hierarchical, data-centric and data reduction for time-driven applications that are existed. As there is very minimal research work that has been conducted based on the combinational area of research study, such as simple data aggregation methods and routing, yet to be effective. Thus, the combination of routing and spatio-temporal data reduction approaches drives us to pursue the field of research, in order to follow both and make a simple effective data reduction algorithm in sensor nodes without losing any quality of data. Then, it also provides a solution for signal reconstruction from the amount of required data samples that is received at the sink node, which were sent after the pre-filtration operations by nodes.

This thesis carried out different sorts of data pre-filtrations in both time and space, before forwarding the final data messages or packets towards the sink node, which can benefit to reduce a huge amount of energy consumption among the node distribution. However, there are some data compression algorithms, which are mainly based on larger datasets in order to find a correlation between data. But the larger data sets require a larger delays in time domain, in order to gather the data over extended time period, thus, led to heavy computational burden on nodes. In this case, this dissertation presents a data window mechanism, the window principle has been originally adapted from one of our previous research works.

Moreover, the proposition of chapter 4 have considered two real world data sets from Intel lab and LCIS lab. The considered amount of datasets play a significant role, in order to analyze and understand the data that how often they are gradually varied over time. In this regard, chapter 4 has also presented the time evaluation method

together with TDR function, whereas it checks the time instances between current and new readings. During the data sets analysis and their investigations, we noticed that the abnormal reading time periods are either wrongly formatted and placed along with the highly varied data. The considered real datasets not only exploits the redundancies by propositions, but also gives the opportunity to demonstrate the consistency in measured performances, as detailed in chapter 4, section 4.6.

In chapter 4, the dissertation has presented two differently projected data window algorithms in various considerations, whereas Algorithm 1 stores its data into the array window buffer until the window timer up, and then evaluates the data by using RV function. In order to exploit the data redundancies and correlations during that short time window interval, which then later decides through the window aggregative function  $w_M$  that whether to carry the total amount of window at a time or the different small portion of the windows at different times on the basis of their data redundancies with other readings or non-redundant. This prospect benefits the node lifetime to decide by itself, in order to do the instant filtration before processing the data towards to their corresponding CH. In this case, the algorithm performances mainly vary based on their optimal both window and threshold parameters. In temporal scenario, if the window sizes increases upto the optimum window ranges then the total amount of delivered data that are sent by window also reduced. But, when it comes to threshold variations, the delivered performances are far lesser than the window varied performances, as demonstrated in chapter 4, section 4.6.

The DAWF Algorithm 2 presents a different sort of analysis compared to the Algorithm 1. As the window decisive algorithm is purely based on its sensed data sensitiveness. If the current window contained information that are highly redundant or greater than the given threshold value, then the window carries total amount of window as a single data packet without compressing or doing any mean averaging data message towards the CH. However, it is not the case when it comes to the spatial scenarios, in order to achieve spatio-temporal data exploitation. This dissertation further used the use case of DAWF temporal algorithms with different considerations for spatial scenarios as well. During spatial evaluations in CH nodes, where they evaluates the data among their clustered nodes. In order to evaluate whether the received data is either highly redundant or correlated over space. Moreover, we also carried out a comparative analysis with the known protocol of TiNA based on DAG protocol for TinyDB services, as described in chapter 4, section 4.6. However, there is further improvement required in DAWF for better spatial data exploitation, so far the algorithm is only able to exploit the spatial redundancies over various received data of nodes. In order for exploiting the spatial correlations among the nodes, it is necessary to improve the proposition for holding the greater evaluations before forwarding its final window data packets towards the sink node.

Nevertheless, this dissertation also demonstrated that the DAWF algorithm has a physical phenomena adaptivity nature in nodes, regardless of their sensed physical quantities, DAWF can monitor any sorts of physical phenomena. As this thesis demonstrate that simple data redundancy algorithm is sufficient to tolerant the nodes sensed

phenomena, yet effective manner. Data window algorithm also proved that it can reconstruct the original signal efficiently without losing any much of quality in data over time.

This is based on a fully distributed spatio-temporal novel data reduction approach that is specifically designed for individual sensor nodes in temporal scenarios, which aimed to keep the WSN as energy optimized as it can be. The use of pre-filtration data techniques for exploiting the physical quantities allowed to reduce the communication costs over nodes, and thus maximizes the sensor nodes lifetime, by adaptively tuning their data rate. The use of simple pre-filtration techniques rather than complex computationally expensive models was the basic idea to design a window technique, which adapts the dynamicity regardless of any particularly monitored data. The experiments carried out on real data sets showed better performances, in particular when the threshold set to 5%, the overall amount of energy saving in CH node among the received nodes data phenomena are 52.2%, 22.9% and 10.75%, respectively. Thus, reducing the temporal redundancies in nodes, which is upto 90% of redundant data transmissions and saves upto 89.98% of the energy while compared to the TAG and TiNA protocols. Furthermore, this approach requires to make a comparison between a much lower sampling rate than what chosen for the original configuration, in terms of measuring the quality of data.

In overall conclusion, the work that has been presented in this thesis, is contributed to node level improvements, which reduce the signal bandwidth as well as to suppress the redundant data transmissions of WSNs in communication transactions for the applications of periodic and delay-tolerant. The presented methodologies and implementations support the prolonged nodes operational lifetime.

## 5.2. Perspectives

The on-going work still need to make further improvements, in order to make the approach to have complete adaptivity, by which a node can tune the threshold parameters or the window constraints based on monitored data attributes. In particular, during the data patterns that are irregularly incremented, but the correlation between near by data values. In this case, therefore non-correlated data counter should be effective in DAWF to list the all non-correlated data as single packet. Further experiments are also being examined, in order to present the comparative analysis with different metric considerations to measure the system performances, especially in terms of quality of data, latency, message costs in CH nodes if the window constraints are high, and energy cost in overall the network.

The further investigations required to be extended based on this thesis, in order to develop complete implementations of DAWF cluster-based topology in COOJA. While there is no default ADC sensors in COOJA, which is difficult to measure the performances through randomly varied readings of light sensor. In order to demonstrate the complete framework of DAWF in COOJA platforms. Although, DAWF has

been partially developed and evaluated the learning phase performances from COOJA, as discussed in chapter 4 and section 4.7. The proposed design structures of a simple feedback mechanism needs to be integrated in COOJA. In order to evaluate the nodes reliability in terms of number of successful transmissions that are sent. The idea of enabling a simple feedback ACK mechanism in receiver nodes overcomes the communication overheads, as well as traditional ACK re-transmission method. Because, the receiver node sends an ACK during the reception of every successful data message. In the case of reliable data transmissions in DAWF algorithm, which sends a filtered window data packet every time, in order to know the total loss or succession of packet transmissions over the time. The feedback ACK mechanism might be useful to estimate the packet reception rate in nodes.

Furthermore, the direction of this thesis has been considered to extend the propositions in multihop networks as well. As part of the preliminary proposition, we have presented a framework of multi-tier multihop heterogeneous wireless sensor network in the appendix. In the homogeneous networks, nodes have the similar features and operational activities that are alike to each other, which may have less impact to show case the performances of various filtered data sets by nodes in a broad way. Although, the considerations of DAWF in a multihop network is a major challenge to achieve in WSNs, since the heterogeneous sensor nodes might hold with different sensors from one to each other, and have less common tasks. Especially, in order for making a comparison and between each other when they are in a same cluster group. Moreover, it will be a hard choice to integrate with the existed well-known multihop protocols, because of two different framework considerations and their more incompatible nature with each other. Thus, we further consider to test and evaluate the performances in this proposed multihop network considerations of our own.

# BIBLIOGRAPHY

- [1] Anish Arora, Rajiv Ramnath, Emre Ertin, Prasun Sinha, Sandip Bapat, Vinayak Naik, Vinod Kulathumani, Hongwei Zhang, Hui Cao, Mukundan Sridharan, and others. Exscal: Elements of an extreme scale wireless sensor network. In *Embedded and Real-Time Computing Systems and Applications, 2005. Proceedings. 11th IEEE International Conference on*, pages 102–108. IEEE, 2005.
- [2] Owen Holland, John Woods, Renzo De Nardi, and Adrian Clarck. Beyond swarm intelligence: the ultraswarm. 2005.
- [3] Mounir Achir and Laurent Ouvry. Power consumption prediction in wireless sensor networks. In *Proc. 16th ITC Specialist Seminar on Performance Evaluation of Wireless and Mobile Systems*. Citeseer, 2004.
- [4] Yonggang Jerry Zhao, Ramesh Govindan, and Deborah Estrin. Residual energy scan for monitoring sensor networks. In *Wireless Communications and Networking Conference, 2002. WCNC2002. 2002 IEEE*, volume 1, pages 356–362. IEEE, 2002.
- [5] Torsha Banerjee, Bin Xie, and Dharma P. Agrawal. Fault tolerant multiple event detection in a wireless sensor network. *Journal of Parallel and Distributed Computing*, 68(9):1222–1234, September 2008.
- [6] Xue Wang, Jun-Jie Ma, Liang Ding, and Dao-Wei Bi. Robust forecasting for energy efficiency of wireless multimedia sensor networks. *Sensors*, 7(11):2779–2807, 2007.
- [7] Azad Ali, Abdelmajid Khelil, Faisal Karim Shaikh, and Neeraj Suri. MPM: Map based predictive monitoring for wireless sensor networks. In *Autonomic Computing and Communications Systems*, pages 79–95. Springer, 2010.
- [8] Giuseppe Anastasi, Marco Conti, Mario Di Francesco, and Andrea Passarella. Energy conservation in wireless sensor networks: A survey. *Ad hoc networks*, 7(3):537–568, 2009.
- [9] Kemal Akkaya and Mohamed Younis. A survey on routing protocols for wireless sensor networks. *Ad hoc networks*, 3(3):325–349, 2005.

- [10] Theofanis P. Lambrou and Christos G. Panayiotou. A Survey on Routing Techniques Supporting Mobility in Sensor Networks. pages 78–85. IEEE, 2009.
- [11] G. Pottie and W. Kaiser. Wireless integrated network sensors (WINS): Principles and practice. *CACM00*, 2000.
- [12] Sebastian Bader. *On the lifetime and usability of environmental monitoring wireless sensor networks*. PhD thesis, Department of Electronics Design, Mid Sweden University, Sundsvall, 2013. OCLC: 940771739.
- [13] Geoffrey Werner-Allen, Konrad Lorincz, Mario Ruiz, Omar Marcillo, Jeff Johnson, Jonathan Lees, and Matt Welsh. Deploying a wireless sensor network on an active volcano. *Internet Computing, IEEE*, 10(2):18–25, 2006.
- [14] International Conference on Mobile Computing and Networking, ACM SIGMOBILE, Association for Computing Machinery, Special Interest Group on Data Communications, ACM-Sigmetrics, and IEEE Communications Society, editors. *Mobile computing and networking 2001: proceedings of the Seventh Annual International Conference on Mobile Computing and Networking, July 16-21, 2001, Rome, Italy*. Association for Computing Machinery, New York, N.Y., 2001.
- [15] Michele Zorzi and Ramesh R. Rao. Geographic random forwarding (GeRaF) for ad hoc and sensor networks: energy and latency performance. *Mobile Computing, IEEE Transactions on*, 2(4):349–365, 2003.
- [16] Benjie Chen, Kyle Jamieson, Hari Balakrishnan, and Robert Morris. Span: An energy-efficient coordination algorithm for topology maintenance in ad hoc wireless networks. *Wireless networks*, 8(5):481–494, 2002.
- [17] Deborah Estrin Alberto Cerpa. ASCENT: Adaptive Self-Configuring sEnsor Networks Topologies.
- [18] Curt Schurgers, Vlasios Tsiatsis, Saurabh Ganeriwal, and Mani Srivastava. Optimizing sensor networks in the energy-latency-density design space. *Mobile Computing, IEEE Transactions on*, 1(1):70–80, 2002.
- [19] Curt Schurgers, Vlasios Tsiatsis, and Mani B. Srivastava. STEM: Topology management for energy efficient sensor networks. In *Aerospace Conference Proceedings, 2002. IEEE*, volume 3, pages 3–1099. IEEE, 2002.
- [20] W. Ye, J. Heidemann, and D. Estrin. Medium Access Control With Coordinated Adaptive Sleeping for Wireless Sensor Networks. *IEEE/ACM Transactions on Networking*, 12(3):493–506, June 2004.
- [21] Tijs Van Dam and Koen Langendoen. An adaptive energy-efficient MAC protocol for wireless sensor networks. In *Proceedings of the 1st international conference on Embedded networked sensor systems*, pages 171–180. ACM, 2003.

- [22] Mario Neugebauer, Joern Ploennigs, and Klaus Kabitzsch. Evaluation of Energy costs for Single Hop vs. Multi Hop with respect to topology parameters. In *IEEE International Workshop on Factory Communication Systems*, 2006.
- [23] C. A. CC2420 Preliminary Datasheet (rev 1.2). Technical, rep, 2004.
- [24] Somasekhar Kandukuri, Vicknesan Ayadurai, Johan Siden, and Mikael Prytz. *Power Control Mechanisms on WARP Boards*. Master Thesis Report, Mid Sweden University, Sundsvall, Sweden, January 2013.
- [25] Isabel Dietrich and Falko Dressler. On the lifetime of wireless sensor networks. *ACM Transactions on Sensor Networks*, 5(1):1–39, February 2009.
- [26] Jason L. Hill and David E. Culler. Mica: A wireless platform for deeply embedded networks. *Micro, IEEE*, 22(6):12–24, 2002.
- [27] Sikka P, Corke P., Valencia L., and Wark T. Fleck - A Platform for Real-World Outdoor Sensor Networks. In *Proceedings of the 3rd International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP)*, pages 709–714, 2007.
- [28] Henri Dubois-Ferrere, Laurent Fabre, Roger Meier, and Pierre Metrailler. TinyNode: a comprehensive platform for wireless sensor network applications. In *Proceedings of the 5th international conference on Information processing in sensor networks*, pages 358–365. ACM, 2006.
- [29] Prabal Dutta, Jay Taneja, Jaein Jeong, Xiaofan Jiang, and David Culler. A building block approach to sensor network systems. In *Proceedings of the 6th ACM conference on Embedded network sensor systems*, pages 267–280. ACM, 2008.
- [30] Lynch C and O’Reilly F. PIC-based TinyOS implementation. In *Proceedings of the Second European Workshop on Wireless Sensor Networks*. IEEE, 2005.
- [31] JeongGil Ko, Qiang Wang, Thomas Schmid, Wanja Hofer, Prabal Dutta, and Andreas Terzis. Egs: A cortex m3-based mote platform. In *Sensor Mesh and Ad Hoc Communications and Networks (SECON), 2010 7th Annual IEEE Communications Society Conference on*, pages 1–3. IEEE, 2010.
- [32] Tim Wark, Peter Corke, Jim Liu, and Darren Moore. Design and evaluation of an image analysis platform for low-power, low-bandwidth camera networks. In *Workshop on Applications, Systems, and Algorithms for Image Sensing*, volume 2008, 2008.
- [33] Alan Mainwaring, David Culler, Joseph Polastre, Robert Szewczyk, and John Anderson. Wireless sensor networks for habitat monitoring. In *Proceedings of the 1st ACM international workshop on Wireless sensor networks and applications*, pages 88–97. ACM, 2002.

- [34] Gilman Tolle, Joseph Polastre, Robert Szewczyk, David Culler, Neil Turner, Kevin Tu, Stephen Burgess, Todd Dawson, Phil Buonadonna, David Gay, and others. A macroscope in the redwoods. In *Proceedings of the 3rd international conference on Embedded networked sensor systems*, pages 51–63. ACM, 2005.
- [35] Holger Karl and Andreas Willig. *Protocols and architectures for wireless sensor networks*. Wiley, Chichester, reprint. with corr edition, 2007. OCLC: 253959574.
- [36] The sensor network museum, 2012.
- [37] Amit Sinha and Anantha Chandrakasan. Dynamic power management in wireless sensor networks. *IEEE Design & Test of Computers*, 18(2):62–74, 2001.
- [38] Jamal N Al-Karaki and Ahmed E Kamal. Routing techniques in wireless sensor networks: a survey. *IEEE*, (IEEE Wireless Communications):6 – 28, December 2004.
- [39] Elena Fasolo, Michele Rossi, Jrg Widmer, and Michele Zorzi. In-network aggregation techniques for wireless sensor networks: a survey. *Wireless Communications, IEEE*, 14(2):70–87, 2007.
- [40] Hseyin zgr Tan and Ibrahim Krpeolu. Power efficient data gathering and aggregation in wireless sensor networks. *ACM Sigmod Record*, 32(4):66–71, 2003.
- [41] S.S. Pradhan and K. Ramchandran. Distributed source coding using syndromes (DISCUS): design and construction. *IEEE Transactions on Information Theory*, 49(3):626–643, 2003.
- [42] Min Wu and Chang Wen Chen. *Multiple Bit Stream Image Transmission over Wireless Sensor Networks*. Chapter 13 in book *Sensor Network Operations*. IEEE & Wiley Interscience, 2006.
- [43] Cauligi S. Raghavendra, Krishna M. Sivalingam, and Taieb Znati, editors. *Wireless sensor networks*. Springer, New York, NY, 2006. OCLC: 255376743.
- [44] Zixiang Xiong, Angelos D. Liveris, and Samuel Cheng. Distributed source coding for sensor networks. *IEEE Signal Processing Magazine*, 21(5):80–94, 2004.
- [45] David S Taubman, Marcellin , and Michael W. *JPEG 2000: Image Compression Fundamentals, Standards and Practice*. Kluwer Academic Publishers, 2001.
- [46] Amara Graps. An introduction to wavelets. pages 2:50–61, June 1995.

- [47] Cesare Alippi, Giuseppe Anastasi, Cristian Galperti, Francesca Mancini, and Manuel Roveri. Adaptive sampling for energy conservation in wireless sensor networks for snow monitoring applications. In *2007 IEEE International Conference on Mobile Adhoc and Sensor Systems*, pages 1–6. IEEE, 2007.
- [48] Vijay Raghunathan, Saurabh Ganeriwal, and Mani Srivastava. Emerging techniques for long lived wireless sensor networks. *IEEE Communications Magazine*, 44(4):108–114, 2006.
- [49] Michael Gastpar, Pier Luigi Dragotti, and Martin Vetterli. The Distributed Karhunen&ndash;Lo&egrave;ve Transform. *IEEE Transactions on Information Theory*, 52(12):5177–5196, December 2006.
- [50] M. F. Duarte, G. Shen, A. Ortega, and R. G. Baraniuk. Signal compression in wireless sensor networks. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 370(1958):118–135, January 2012.
- [51] Godwin Shen and Antonio Ortega. Transform-based distributed data gathering. *IEEE Transactions on Signal Processing*, 58(7):3802–3815, 2010.
- [52] Donoho D. L. Compressed Sensing. *IEEE Transactions on Information Theory*, 52:1289–1306, April 2006.
- [53] Waheed Bajwa, Jarvis Haupt, Akbar Sayeed, and Robert Nowak. Compressive wireless sensing. In *Proceedings of the 5th international conference on Information processing in sensor networks*, pages 134–142. ACM, 2006.
- [54] David Chu, Amol Deshpande, Joseph M. Hellerstein, and Wei Hong. Approximate data collection in sensor networks using probabilistic models. In *Data Engineering, 2006. ICDE'06. Proceedings of the 22nd International Conference on*, pages 48–48. IEEE, 2006.
- [55] A. Jain, E. Y. Chang, and Y.-F. Wang. Adaptive Stream Resource Management Using Kalman Filters. *Proc. of the ACM International conference on Management of Data (SIGMOD2004)*,, pages 11–22,, June 2004.
- [56] Bhargav Kanagal and Amol Deshpande. Online filtering, smoothing and probabilistic modeling of streaming data. In *2008 IEEE 24th International Conference on Data Engineering*, pages 1160–1169. IEEE, 2008.
- [57] Amol Deshpande and Samuel Madden. MauveDB: supporting model-based user views in database systems. In *Proceedings of the 2006 ACM SIGMOD international conference on Management of data*, pages 73–84. ACM, 2006.
- [58] Daniela Tulone and Samuel Madden. An energy-efficient querying framework in sensor networks for detecting node similarities. In *Proceedings of the 9th ACM international symposium on Modeling analysis and simulation of wireless and mobile systems*, pages 191–300. ACM, 2006.

- [59] Yann-Al Le Borgne, Silvia Santini, and Gianluca Bontempi. Adaptive model selection for time series prediction in wireless sensor networks. *Signal Processing*, 87(12):3010–3020, December 2007.
- [60] P. Brighten Godfrey and David Ratajczak. Naps: scalable, robust topology management in wireless ad hoc networks. In *Information Processing in Sensor Networks, 2004. IPSN 2004. Third International Symposium on*, pages 443–451. IEEE, 2004.
- [61] Samir Goel, Andrea Passarella, and Tomasz Imielinski. Using buddies to live longer in a boring world [sensor network protocol]. In *Pervasive Computing and Communications Workshops, 2006. PerCom Workshops 2006. Fourth Annual IEEE International Conference on*, pages 5–pp. IEEE, 2006.
- [62] Pangun Park. *Modeling, analysis and design of wireless sensor network protocols*. School of Electrical Engineering, KTH Royal Institute of Technology, Stockholm, 2011.
- [63] Ossama Younis and Sonia Fahmy. HEED: a hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks. *Mobile Computing, IEEE Transactions on*, 3(4):366–379, 2004.
- [64] Wendi Rabiner Heinzelman, Anantha Chandrakasan, and Hari Balakrishnan. Energy-efficient communication protocol for wireless microsensor networks. In *System sciences, 2000. Proceedings of the 33rd annual Hawaii international conference on*, pages 10–pp. IEEE, 2000.
- [65] Wendell H Fleming, William M McEneaney, George Yin, and Qing Zhang. *Stochastic analysis, control, optimization, and applications: a volume in honor of W.H. Fleming*. Birkhuser, Boston, 1999.
- [66] Stephanie Lindsey and Cauligi S. Raghavendra. PEGASIS: Power-efficient gathering in sensor information systems. In *Aerospace conference proceedings, 2002. IEEE*, volume 3, pages 3–1125. IEEE, 2002.
- [67] V. Loscri, G. Morabito, and S. Marano. A two-levels hierarchy for low-energy adaptive clustering hierarchy (TL-LEACH). In *IEEE Vehicular Technology Conference*, volume 62, page 1809. IEEE; 1999, 2005.
- [68] Jalil Jabari Lotf, Mehran Hosseinzadeh, and Rasim M. Alguliev. Hierarchical routing in wireless sensor networks: a survey. In *Computer Engineering and Technology (ICCET), 2010 2nd International Conference on*, volume 3, pages V3–650. IEEE, 2010.
- [69] Rwan Ibrahim, Quang-Dung Ho, and Tho Le-Ngoc. An Energy-Efficient and Load-Balancing Cluster-Based Routing Algorithm for CSMA-Based Wireless Sensor Networks. In *Vehicular Technology Conference (VTC Spring), 2013 IEEE 77th*, pages 1–5. IEEE, 2013.

- [70] Li Qing, Qingxin Zhu, and Mingwen Wang. Design of a distributed energy-efficient clustering algorithm for heterogeneous wireless sensor networks. *Computer Communications*, 29(12):2230–2237, August 2006.
- [71] Brahim Elbhiri, Rachid Saadane, Sanaa El Fkihi, and Driss Aboutajdine. Developed Distributed Energy-Efficient Clustering (DDEEC) for heterogeneous wireless sensor networks. In *I/V Communications and Mobile Network (ISVC), 2010 5th International Symposium on*, pages 1–4. IEEE, 2010.
- [72] Chee-Yee Chong, Feng Zhao, Shozo Mori, and Sri Kumar. Distributed tracking in wireless ad hoc sensor networks. Technical report, DTIC Document, 2003.
- [73] Deepinder Sidhu, Tayang Fu, Shukri Abdallah, Raj Nair, and Rob Coltun. Open shortest path first (OSPF) routing protocol simulation. *ACM SIGCOMM Computer Communication Review*, 23(4):53–62, 1993.
- [74] Clausen T and Jacquet P. Optimized Link State Routing Protocol (OLSR). Technical, IETF, 2003.
- [75] David A. Maltz. David B. Johnson Department of Computer Science Rice University David A. Maltz, Josh Broch Carnegie-Mellon University.
- [76] Perkins C, Belding-Royer E, and Das S. Ad hoc On-demand Distance Vector Routing (AODV) Protocol. Technical, IETF, July 2003.
- [77] Omprakash Gnawali, Rodrigo Fonseca, Kyle Jamieson, David Moss, and Philip Levis. Collection tree protocol. In *Proceedings of the 7th ACM Conference on Embedded Networked Sensor Systems*, pages 1–14. ACM, 2009.
- [78] Joo P. Hespanha, Payam Naghshtabrizi, and Yonggang Xu. A Survey of Recent Results in Networked Control Systems. *Proceedings of the IEEE*, 95(1):138–162, January 2007.
- [79] Scott Moeller, Avinash Sridharan, Bhaskar Krishnamachari, and Omprakash Gnawali. Routing without routes: the backpressure collection protocol. In *Proceedings of the 9th ACM/IEEE International Conference on Information Processing in Sensor Networks*, pages 279–290. ACM, 2010.
- [80] Federico Ferrari, Marco Zimmerling, Luca Mottola, and Lothar Thiele. Low-power wireless bus. In *Proceedings of the 10th ACM Conference on Embedded Network Sensor Systems*, pages 1–14. ACM, 2012.
- [81] Olaf Landsiedel, Euhanna Ghadimi, Simon Duquennoy, and Mikael Johansson. Low power, low delay: opportunistic routing meets duty cycling. In *Information Processing in Sensor Networks (IPSN), 2012 ACM/IEEE 11th International Conference on*, pages 185–196. IEEE, 2012.

- [82] Sebastian Bader and Bengt Oelmann. Adaptive synchronization for duty-cycling in environmental wireless sensor networks. In *Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP), 2009 5th International Conference on*, pages 49–54. IEEE, 2009.
- [83] Mohammad Z. Siam, Marwan Krunz, and Ossama Younis. Energy-efficient clustering/routing for cooperative MIMO operation in sensor networks. In *INFOCOM 2009, IEEE*, pages 621–629. IEEE, 2009.
- [84] Wendi Rabiner Heinzelman, Joanna Kulik, and Hari Balakrishnan. Adaptive protocols for information dissemination in wireless sensor networks. In *Proceedings of the 5th annual ACM/IEEE international conference on Mobile computing and networking*, pages 174–185. ACM, 1999.
- [85] C. Intanagonwiwat, R. Govindan, D. Estrin, J. Heidemann, and F. Silva. Directed diffusion for wireless sensor networking. *IEEE/ACM Transactions on Networking*, 11(1):2–16, February 2003.
- [86] W.B. Heinzelman, A.P. Chandrakasan, and H. Balakrishnan. An application-specific protocol architecture for wireless microsensor networks. *IEEE Transactions on Wireless Communications*, 1(4):660–670, October 2002.
- [87] Arati Manjeshwar and Dharma P. Agrawal. TEEN: a routing protocol for enhanced efficiency in wireless sensor networks. In *Parallel and Distributed Processing Symposium, International*, volume 3, pages 30189a–30189a. IEEE Computer Society, 2001.
- [88] Arati Manjeshwar and Dharma P. Agrawal. APTEEN: A hybrid protocol for efficient routing and comprehensive information retrieval in wireless sensor networks. In *Parallel and Distributed Processing Symposium, International*, volume 2, pages 0195b–0195b. IEEE Computer Society, 2002.
- [89] M. J. Handy, Marc Haase, and Dirk Timmermann. Low energy adaptive clustering hierarchy with deterministic cluster-head selection. In *Mobile and Wireless Communications Network, 2002. 4th International Workshop on*, pages 368–372. IEEE, 2002.
- [90] Aimin Wang, Dailiang Yang, and Dayang Sun. A clustering algorithm based on energy information and cluster heads expectation for wireless sensor networks. *Computers & Electrical Engineering*, 38(3):662–671, May 2012.
- [91] IEEE. In *IEEE 802.15 task group 4e: Wireless Medium Access Control and Physical Layer Specifications for Low-Rate Wireless Personal Area Networks*. IEEE, 2010.
- [92] Volkan Rodoplu and Teresa H. Meng. Minimum energy mobile wireless networks. *Selected Areas in Communications, IEEE Journal on*, 17(8):1333–1344, 1999.

- [93] Yan Yu, Ramesh Govindan, and Deborah Estrin. Geographical and energy aware routing: A recursive data dissemination protocol for wireless sensor networks. Technical report, Technical report ucla/csd-tr-01-0023, UCLA Computer Science Department, 2001.
- [94] A. Rabdriatsiferana Rivo Sitraka. *Energy Optimization Protocols Communication of Wireless Sensor Networks*. PhD thesis, Universite de La Reunion, May 2015.
- [95] Contiki, Available at <http://www.contiki-os.org/start.html>, August 2015.
- [96] Adam Dunkels, Oliver Schmidt, Thiemo Voigt, and Muneeb Ali. Protothreads: simplifying event-driven programming of memory-constrained embedded systems. In *Proceedings of the 4th international conference on Embedded networked sensor systems*, pages 29–42. Acm, 2006.
- [97] Adam Dunkels, Fredrik Osterlind, Nicolas Tsiftes, and Zhitao He. Software-based on-line energy estimation for sensor nodes. In *Proceedings of the 4th workshop on Embedded networked sensors*, pages 28–32. ACM, 2007.
- [98] Adam Dunkels, Bjrn Grnvall, and Thiemo Voigt. Contiki-a lightweight and flexible operating system for tiny networked sensors. In *Local Computer Networks, 2004. 29th Annual IEEE International Conference on*, pages 455–462. IEEE, 2004.
- [99] Fredrik Osterlind, Adam Dunkels, Joakim Eriksson, Niclas Finne, and Thiemo Voigt. Cross-level sensor network simulation with cooja. In *Local Computer Networks, Proceedings 2006 31st IEEE Conference on*, pages 641–648. IEEE, 2006.
- [100] Cheng Tien Ee, Rodrigo Fonseca, Sukun Kim, Daekyeong Moon, Arsalan Tavakoli, David Culler, Scott Shenker, and Ion Stoica. A modular network layer for sensorsets. In *Proceedings of the 7th symposium on Operating systems design and implementation*, pages 249–262. USENIX Association, 2006.
- [101] International Conference on Embedded Networked Sensor Systems and Association for Computing Machinery, editors. *SenSys'06: proceedings of the Fourth International Conference on Embedded Networked Sensor Systems: Oct. 31 - Nov. 3, 2006, Boulder, Colorado, USA*. Association for Computing Machinery, New York, NY, 2006.
- [102] Rachid Guerraoui and Luis Rodrigues. *Introduction to reliable distributed programming*. Springer, 2006.
- [103] Philip Levis, Sam Madden, Joseph Polastre, Robert Szewczyk, Kamin Whitehouse, Alec Woo, David Gay, Jason Hill, Matt Welsh, Eric Brewer, and others. TinyOS: An operating system for sensor networks. In *Ambient intelligence*, pages 115–148. Springer, 2005.

- [104] DREAL, Thermal renovation of social housing estates laboratory,, 2011.
- [105] Samuel Madden. Intel Lab Data,, 2004.
- [106] Daniela Tulone and Samuel Madden. PAQ: Time series forecasting for approximate query answering in sensor networks. In *Wireless Sensor Networks*, pages 21–37. Springer, 2006.
- [107] Ian F. Akyildiz, Dario Pompili, and Tommaso Melodia. Underwater acoustic sensor networks: research challenges. *Ad Hoc Networks*, 3(3):257–279, May 2005.
- [108] Geoffrey Werner-Allen, Jeff Johnson, Mario Ruiz, Jonathan Lees, and Matt Welsh. Monitoring volcanic eruptions with a wireless sensor network. In *Wireless Sensor Networks, 2005. Proceedings of the Second European Workshop on*, pages 108–120. IEEE, 2005.
- [109] Ian F. Akyildiz, Weilian Su, Yogesh Sankarasubramaniam, and Erdal Cayirci. Wireless sensor networks: a survey. *Computer networks*, 38(4):393–422, 2002.
- [110] Jin Cui and Fabrice Valois. Data aggregation in wireless sensor networks: Compressing or Forecasting? Research hal-00861598, University of Lyon, Inria, INSA Lyon, CITI, France, September 2013.
- [111] Wei Ye, John Heidemann, and Deborah Estrin. An energy-efficient MAC protocol for wireless sensor networks. In *INFOCOM 2002. Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, volume 3, pages 1567–1576. IEEE, 2002.
- [112] E.-O. Blass, J. Horneber, and M. Zitterbart. Analyzing Data Prediction in Wireless Sensor Networks. In *IEEE Vehicular Technology Conference, 2008. VTC Spring 2008*, pages 86–87, May 2008.
- [113] S. Mukhopadhyay, C. Schurgers, D. Panigrahi, and S. Dey. Model-Based Techniques for Data Reliability in Wireless Sensor Networks. *IEEE Transactions on Mobile Computing*, 8(4):528–543, April 2009.
- [114] Lidan Wang and Amol Deshpande. Predictive modeling-based data collection in wireless sensor networks. In *Wireless Sensor Networks*, pages 34–51. Springer, 2008.
- [115] Adam Silberstein, Rebecca Braynard, and Jun Yang. Constraint chaining: on energy-efficient continuous monitoring in sensor networks. In *Proceedings of the 2006 ACM SIGMOD international conference on Management of data*, pages 157–168. ACM, 2006.
- [116] Rui Tan, Guoliang Xing, Xue Liu, Jianguo Yao, and Zhaohui Yuan. Adaptive calibration for fusion-based wireless sensor networks. In *INFOCOM, 2010 Proceedings IEEE*, pages 1–9. IEEE, 2010.

- [117] Mohamed A. Sharaf, Jonathan Beaver, Alexandros Labrinidis, and Panos K. Chrysanthis. TiNA: a scheme for temporal coherency-aware in-network aggregation. In *Proceedings of the 3rd ACM international workshop on Data engineering for wireless and mobile access*, pages 69–76. ACM, 2003.
- [118] SunHee Yoon and Cyrus Shahabi. Exploiting spatial correlation towards an energy efficient clustered aggregation technique (cag)[wireless sensor network applications]. In *Communications, 2005. ICC 2005. 2005 IEEE International Conference on*, volume 5, pages 3307–3313. IEEE, 2005.
- [119] Alessandra De Paola, Giuseppe Lo Re, Fabrizio Milazzo, and Marco Ortolani. Adaptable data models for scalable ambient intelligence scenarios. In *Information Networking (ICOIN), 2011 International Conference on*, pages 80–85. IEEE, 2011.
- [120] Mahnaz Roshanei and Mehrdad Valipour. Energy-Efficient Data Gathering over Wireless Sensor Networks: Correlated Sources and Lossy Channels. pages 449–451. IEEE, May 2009.
- [121] Mehmet C. Vuran, zgr B. Akan, and Ian F. Akyildiz. Spatio-temporal correlation: theory and applications for wireless sensor networks. *Computer Networks*, 45(3):245–259, June 2004.
- [122] Samuel Madden, Michael J. Franklin, Joseph M. Hellerstein, and Wei Hong. TAG: A tiny aggregation service for ad-hoc sensor networks. *ACM SIGOPS Operating Systems Review*, 36(SI):131–146, 2002.
- [123] Samuel R. Madden, Michael J. Franklin, Joseph M. Hellerstein, and Wei Hong. TinyDB: an acquisitional query processing system for sensor networks. *ACM Transactions on database systems (TODS)*, 30(1):122–173, 2005.
- [124] Samuel Madden, Michael J. Franklin, Joseph M. Hellerstein, and Wei Hong. The design of an acquisitional query processor for sensor networks. In *Proceedings of the 2003 ACM SIGMOD international conference on Management of data*, pages 491–502. ACM, 2003.
- [125] Nicolas Tsiftes, Adam Dunkels, Zhitao He, and Thiemo Voigt. Enabling large-scale storage in sensor networks with the coffee file system. In *Proceedings of the 2009 International Conference on Information Processing in Sensor Networks*, pages 349–360. IEEE Computer Society, 2009.
- [126] Nicolas Tsiftes and Adam Dunkels. A database in every sensor. In *Proceedings of the 9th ACM Conference on Embedded Networked Sensor Systems*, pages 316–332. ACM, 2011.
- [127] Joseph Polastre, Robert Szewczyk, and David Culler. Telos: enabling ultra-low power wireless research. In *Information Processing in Sensor Networks, 2005. IPSN 2005. Fourth International Symposium on*, pages 364–369. IEEE, 2005.

- [128] Christian Haas and Joachim Wilke. Energy evaluations in wireless sensor networks: a reality check. In *Proceedings of the 14th ACM international conference on Modeling, analysis and simulation of wireless and mobile systems*, pages 27–30. ACM, 2011.
- [129] Adam Dunkels. The contikimac radio duty cycling protocol. 2011.

# Appendix A

## Multi-hop WSNs: A Multi-tier Heterogeneous Cluster-based Approach

This dissertation study an overview and the various examples of cluster-based aspect of multi-hop WSNs. In this thesis, we consider the task of designing a data redundancy algorithm for heterogeneous environments for an energy-efficient multi-hop WSN. The assumption in the communication overhead problem is that the network is operating by multi-tier multi-hop heterogeneous network within the data reduction approach, which is designed for in-network processing. In this case, we first describe the significance of data reduction in cluster-based networks that are being conducted, and then present the proposition of multi-tier multi-hop distributed sensor network framework.

This chapter shows the proposition of multi-tier multi-hop distributed wireless sensor network, which can be used in periodic and delay-tolerant applications. Multiple cluster-based hierarchical features are considered in order for self-controlling and evaluating of the node for reliable WSN and data reduction techniques. In this scenario, we first present an analytical framework of the proposition, which has been addressed to estimate the packet loss probability, energy cost and delay of the cross-level network. In this design, we considered CSMA ContikiMAC protocol, in order for balancing the asynchronous system of the network. The framework is designed based on a HEED protocol. As the multi-hop network architecture example has been illustrated in 1.4, chapter 1.

### A.1. Hybrid Energy-Efficient and Throughput Protocol for Multi-Hop WSNs

We propose a Hybrid Energy Efficient and Throughput (HEET) routing protocol for reliable wireless sensor networks. The literature study of well-known existed proto-

cols, in chapter 1, have their own pros and cons. Some of them proven analytically and in simulations that they performed well. In practice, many routing protocols have failed to shown any impact even at lower data rates, which is also discussed and proved why? In CTP protocol, and also addressed the causes, because of this link dynamics and transient loops. Keeping the issues in mind based on CTP discussions and experiences, we are proposing a new hybrid energy efficient and throughput routing protocol for reliable wireless sensor networks. The following challenges are primary objective of this proposition, in order to achieve energy efficiency, reducing the communication overheads, short-range interference and throughput.

1. Cluster-head and cluster-manager radio ranges: short for inter cluster (Point-to-Point) communication and long range for BS communication via neighboring CHs. CHs and CMs are higher energy level nodes than normal sensor nodes.
2. Contention based MAC protocol CSMA RTS/CTS, ACKs, and Transmission Timeouts.
3. Broadcast messages at CHs and CMs for discovering member nodes RL nodes, and RL broadcasts for non-cluster group nodes.
4. Interesting to see the performance results at various network deployments, such as Random, Manual, and Linear, etc.
5. Data aggregation principles at CH, CM and RL nodes.
6. This proposition has been addressed lively status of the nodes and their failures, in order for collecting the energy levels of all the cluster nodes during every round by CM node, because energy consumption per transmission we already known by the given radio data-sheets. Here, we only interest to collect the info to estimate or know their further lifespan or depletion period of the nodes per round.
7. Collecting member node energy levels and their RSSI and LQI along with data packets can create extra overheads at CHs. For this again we need to keep some additional storage medium at CHs to keep all the node energy levels, RSSI and LQIs, which makes additional computational, energy and communication costs

Cluster-head selection mechanism main parameter as higher energy levels. In this network, we have five type of nodes in every cluster group, namely as CH, CM, RL, member and non-cluster member nodes.

Pros:

1. No cluster formation and Cluster-heads are fixed, CH changes occur when the original CH depletes, then only the back-up node of CM can be as a CH to lead that cluster.

2. There is only one time setup-phase of the clusters.
3. No communication overheads.
4. Less traffic or interference.
5. No external interference due to the different radio range mediums

In this network setup, we have considered ZigBee 802.15.4 CSMA MAC protocol. There is several power saving synchronous and asynchronous MAC protocols are available, such as S-MAC and T-MAC, asynchronous as B-MAC, WiseMAC, X-MAC, BoX-MAC, and ContikiMAC protocols. Among all, every MAC protocols have their original set of parameters in order for following the application needs. While comparing to synchronous MAC protocols, asynchronous protocols save a lot of energy as well as the delays. The authors of X-MAC protocol stated that the shorter period of wake up preambles while still maintaining low duty cycles, lead to reduce the latencies, energy consumption and higher throughput. In this proposition, it assumed to consider one of the existing asynchronous MAC protocols, in order to design and develop the proposed framework. The proposition considers IEEE 802.15.4 CSMA ZigBee specification for maintaining the scalable networks. In general, time synchronized mechanisms are not scalable, and also have higher packet latency. Nevertheless, if we consider the proposed methods for only risk monitoring application scenarios. Then the system should be considered to design in two cases either long duration event detection or continuous event detection systems. Because, in general applications, if we look at outdoor application scenarios event occurrences are not often unless the application is being used in the urban areas. In this case, motorization of the nodes are kept in either sleep state or turnoff state for long duration instead of an active mode.

The following timeline figures explain the messages exchange between leaders and slave nodes. These are all designed with non-beacon enabled mode. There is no prefixed Beacon Intervals (BI) or Superframe Durations(SD).

The Fig. A.1 explains the different sort of messages exchange between Cluster-Head (CH) node, Route-leader nodes, and other member nodes whoever within the cluster transmission range. It works based CSMA RTS/CTS or Clear Channel Assessment (CCA) mechanism. In this scenario, time specifications are defined by IEEE 802.15.4 and CC2420 radio transceiver data sheets. Furthermore, every sender can do the re-transmissions when there is no feedback ACKs from receivers until the maximum re-transmissions timeout.

This is the Fig. A.2 communication process between inter-cluster node as RL looking for to discover unsuccessful nodes who are not part of any cluster group or who has low level of connectivity to join in any cluster. This method helps us to discover all the nodes in a network instead of missing some, which are not being used or undiscovered. The main purpose to consider this RL nodes to discover non-cluster nodes as well to reduce the transmissions instead of letting every unnecessary signal transmissions between every point-to-point nodes, which can be seen at Fig. 5. In this

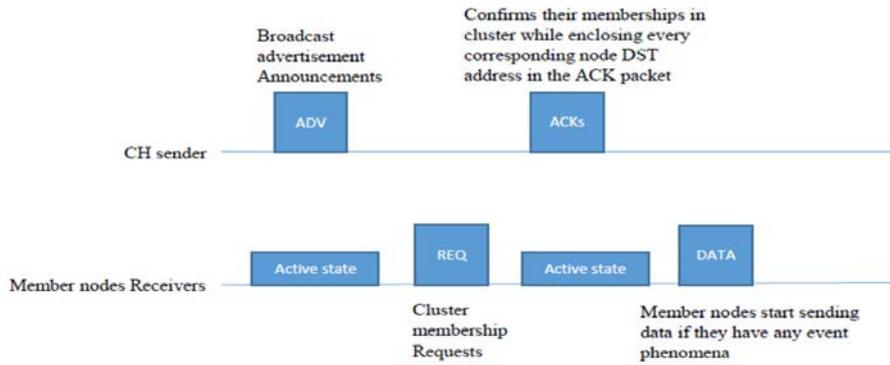


Figure A.1 – Cluster-Head and Member nodes communication.

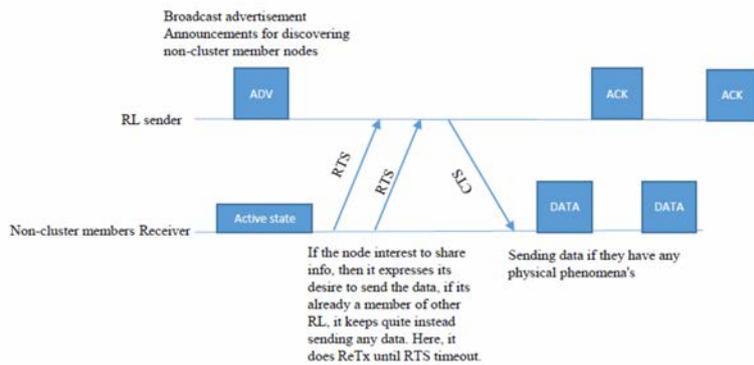


Figure A.2 – Route-leaders, member nodes and non-member nodes Communication.

case, the RL node has packet buffer to store incoming data from non-cluster nodes. Fig. A.3 the packet transmission between RL and CH nodes. The RL node would not

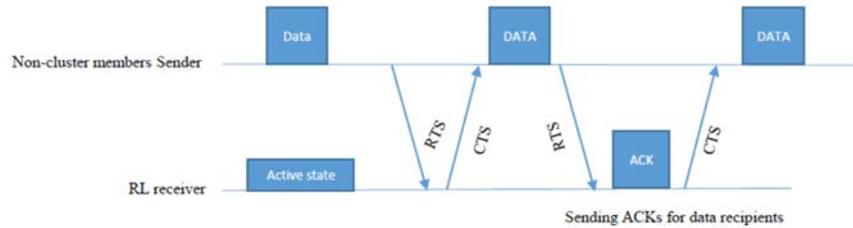


Figure A.3 – Network nodes and Route-leaders communication.

process any regular transmissions like other cluster member nodes. It does send the data packet whenever the packet buffer length is above the threshold, which means, in this scenario we fix buffer length, upper and lower packet size thresholds. Because, here we use packet buffer at every RL node to keep the data from non-cluster nodes instead of sending every event occurrences at different nodes at different times. If it does like other regular nodes, it does loose lot of energy to forward every single event occurrences of other nodes and itself.

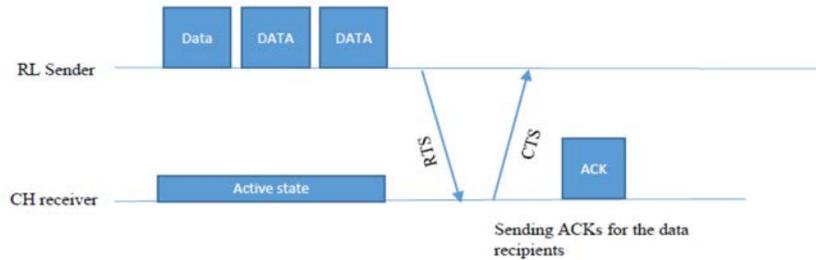


Figure A.4 – Data transmissions between RL and CH nodes

### A.1.1. Algorithm 1

Cluster-Head selection algorithm on a sensor node (SN). Parameters:

1. Battery and RSSI measurements: total energy of all nodes and signal strength measurements for point-to-point communications for maintaining better connectivity between hop-to-hop.
2. Total energy of super nodes of E(CH) and Cluster Managers (CM), E(CM) and total energy of Route Leader (RL) nodes E(RL). Total energy of member nodes E(MN)

Total energy of member nodes is:

$$E_{CH} = SN_n E_{CH}(1 + 3\alpha) \quad (A.1)$$

Similarly, CH, CM and RL nodes are supplied with higher energy considerations, whereas both CH and CM nodes have 3 and RL nodes has 2 factors more energy than member nodes E(SN) alpha.

Total energy of CH and CM nodes can be calculated as

$$E_{CM} = SN_n E_{CM}(1 + 3\alpha) \quad (A.2)$$

Total energy of RL nodes is given as

$$E_{RL} = SN_n E_{RL}(1 + 2\alpha) \quad (A.3)$$

Average energy of CH super nodes is given as

$$E_{CH}(r) = \frac{1}{CH} \sum_{i=1}^{CH} E_i(r) \quad (A.4)$$

Average energy of Cluster Manager Nodes is given as

$$E_{CM}(r) = \frac{1}{CM} \sum_{i=1}^{CM} E_i(r) \quad (A.5)$$

Average energy of Route Leader nodes given as

$$E_{RL}(r) = \frac{1}{RL} \sum_{i=1}^{RL} E_i(r) \quad (A.6)$$

Average energy of member nodes given as

$$E_{MN}(r) = \frac{1}{MN} \sum_{i=1}^{MN} E_i(r) \quad (A.7)$$

### A.1.2. Algorithm 2

For example: parameters

1. Packet buffer size: 80 bytes
2. Minimum event occurrence data length: 20 bytes
3. Just consider 2 or 3 nodes in the RL node

---

```

if  $E_{CH} \leq E_{MN}$  then
    CM node will become a CH
    if CM node is cluster-manager then
        broadcast an ADV message for collecting all node E levels
        if RL node has discovered Non-Cluster nodes then
            Start Collecting data from them
        end if
    end if
end if
    
```

---

4. Threshold length: 50 bytes
5. Upper – Threshold > 50bytes
6. Lower – Threshold < 50bytes

---

```

if packet-buffer size is greater than or equal to Threshold-value then
    Start data-aggregation
    Send packet to the CH
    Send RTS to non-cluster nodes
end if
    
```

---

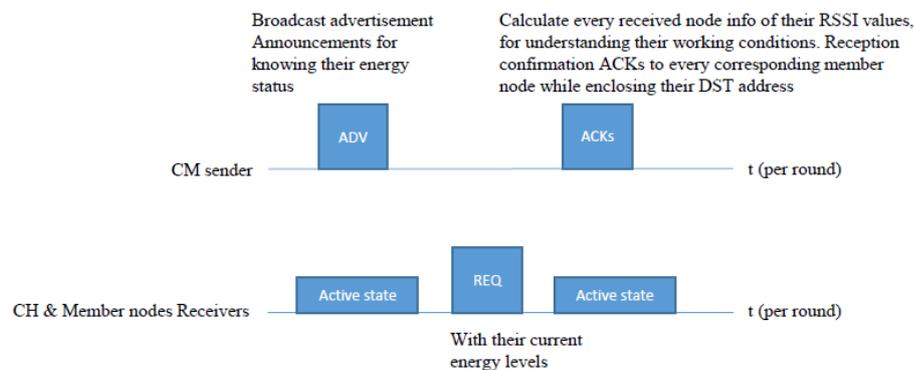


Figure A.5 – Broadcast and unicast message exchanges between CM and all cluster nodes.

Fig. A.5 describes the communication process between CM and all cluster nodes. Cluster-manager node broadcasts a message request for knowing current energy levels information of all cluster nodes including cluster-head and non-cluster nodes through RL nodes at every round. This prospect helps CM node to understand the connectivity levels and node remaining energy levels of the cluster. If there is any route failures or node failures due to their low level of energy or node depletion, it requests other

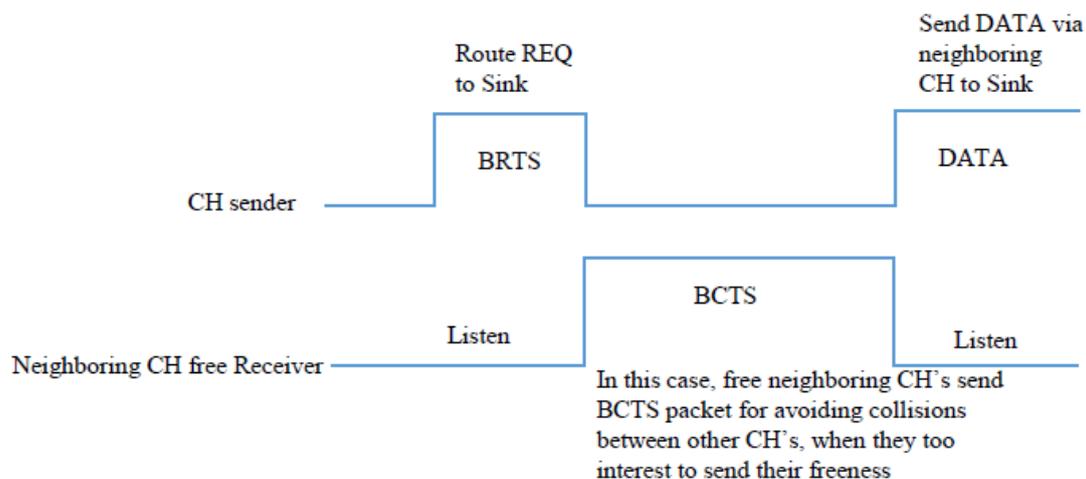


Figure A.6 – Data exchange between Current CH via other neighboring CH nodes to sink.

nodes to connect with that broken route nodes or it sends a request itself to share the data. The above Fig. A.6 explores the aggregated data packets forwarding through other CH nodes to sink node. Table. A.1 states the protocol stack that we are going to use layer by layer for lower layers. In CSMA-CA, we have four time interval states to work between sender and receiver, which are Back-off, Collision, Success and Sensing. For this time notations, we have been followed from IEEE 802.15.4 radio data sheet specifications. Table. A.2 describes the simulation parameters of this work; I

Layer	Protocols
Network layer	Rime
MAC	CSMA/CA
Radio Duty Cycling (RDC)	NullRDC or ContikiMAC or X-MAC
Physical Link	IEEE 802.15.4
Radio	CC2420 chip

Table A.1 – A list of Protocol Stack Considerations.

have considered the parameters based on COOJA [99] simulator, we can see at below given tabulation. In COOJA, it has several path-loss models and have been developed based on the considerations of distance losses and radio traffics. In COOJA, we have the power tracker application to know the power consumption values per transmission based on Rtimer function. In this scenario, we cant see the direct power consumption values in dBm or mW, we just receive only radio TX/RX times on timeline per packet transmission and reception, based on these values, we can apply the above power consumption formula while multiplying current and voltage values per transmission and reception given by hardware data sheets. The above tables shown real energy consumption values at different TX/RX and signal states, which were stated

Simulation Parameters	
Node type	T-mote Sky
Total packet size in rime stack	127 bytes
Radio medium	Several disk graph models in COOJA, as UDGM (distance loss & constant loss), DGRM, MRM, and No Radio traffic
Bit rate	250 kbps
MAC layer queue size	30 packets
Node carrier sensing range	100 meters
Network Deployment	Manual
Radio types (we considered to use)	Short TX range (for clusters) and long TX range (for CHs to Sink)
Power consumption per transmission (Contiki/COOJA)	$\text{Power (pW)} = (\text{txend} - \text{txstart}) * 20\text{mA} * 3\text{V} / 32768$ (Voltage/Current values have given by radio datasheet)

Table A.2 – A list of Simulation Parameters.

Activity	Energy (uJ)
Wake-up, no signal detected	12
False positive wake-up	100
Broadcast reception	178
Unicast reception	222
Broadcast transmission	1790
Non-synchronized unicast transmission	1090
Synchronized unicast transmission	120
Unicast transmission to awake receiver	96

Table A.3 – ContikiMAC Energy rates.

by ContikiMAC/COOJA [129]. Furthermore, I have added few more parameters and protocols in the below comparison tables for making comparison in radio, power control and MAC protocols.

## A.2. Discussion

As discussed in the chapter 2. Energy efficient in either single-hop or multi-hop depends upon the radio range of their devices and the network over coverage area. Usually, in many applications are being used two kind of network areas, such as network is away from the field of nodes or network locates within the field of nodes. It also depends on the application requirements. As per best of our knowledge, single-hop networks very reliable when it comes for general application domains, such as smart-homes, smart-cities and health-care systems, Industrial and many more. Because the coverage area of the nodes does not much require to be connected with network. In remote area systems, nodes have longer distances and networks too away from the field, in such cases multi-hop communication systems are preferable to be communicated. In this scenario, choice of network model either homogeneous or heterogeneous decides the type of communication. For example, Fig.A.7 shows the transmission differences between single-hop and multi-hop networks.

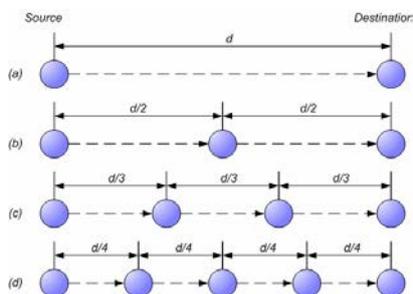


Figure A.7 – Transmission Distances for Single-hop and Multi-hops.

## A.3. Summary

The proposed algorithm static cluster-head algorithm work in progress for the implementation and results. The measurement results will be presented in the near future. The following proposals present the further new works of this thesis.

1. there is several possible extensions in this multi-hop algorithm to integrate or design a simple duty-cycling MAC CSMA (RTS/CTS) protocol. We also plan to integrate this with several well-known duty-cycling MAC protocols to see the energy-efficiency and throughput performances.

2. We have considered a multi-hop data-driven energy-efficient approach as nodes forward the received data of them through relays to the base station. In this case, relays compare their data with every current received data of neighbors, if the both data are similar or same, they forward single data instead two with two hops IDs to the another relay or to the parent node or base station.
3. For another proposal, we are proposing a new power control algorithm on transmitter side to manage power levels based on the data losses or node fault-tolerances or route-failures .



# Appendix B

## DAWF Experimental Study Case in COOJA Interface

This section describes the proposed integration's in COOJA, as discussed in chapter 3 and chapter 4 regarding COOJA, Contiki based simulator and its advantages that the propositions are being carried out in COOJA. As presented in the Figures B.1, B.2, B.3, and B.4, we have developed and tested the learning phase proposition of the algorithm in both the platforms of TmoteSky nodes and Zoltera Z1 motes based on a light sensor readings. As stated that the window mechanism successfully carried out the window payloads over at the fixed time intervals, such as every 30 seconds. Because, the sensor detection interval at every 2 seconds, which the nodes are sensed their phenomena at every 2 seconds of time, and then stored the sensed phenomena into their corresponding data windows.

However, there is slight delay differences in time over both the TmoteSky and Z1 platforms, because of their different hardware choices that are designed or the COOJA functional behavior of GUI. Although, we still need to carry out further analysis, in order to integrate the total propositions and prove their performances while comparing to the performed proposition results from MATLAB.

### B.1. DAWF Propositions Based on Real-World Datasets

In this section, Figures of B.5 and B.6 show a snippet of real world datasets from various labs, such as Intel datasets from Intel Berkley research lab and DREAL datasets from LCIS research lab. As both the datasets are given various phenomena that are being captured by every 30 seconds and 1 hour, respectively. Moreover, the considered datasets have been used in this thesis for making an experimental study case, and then let the propositions on datasets, in order to exploit both the temporal and spatial redundancies and correlations, as briefly explained and presented in chapter 4 based on the performance and comparative analysis.

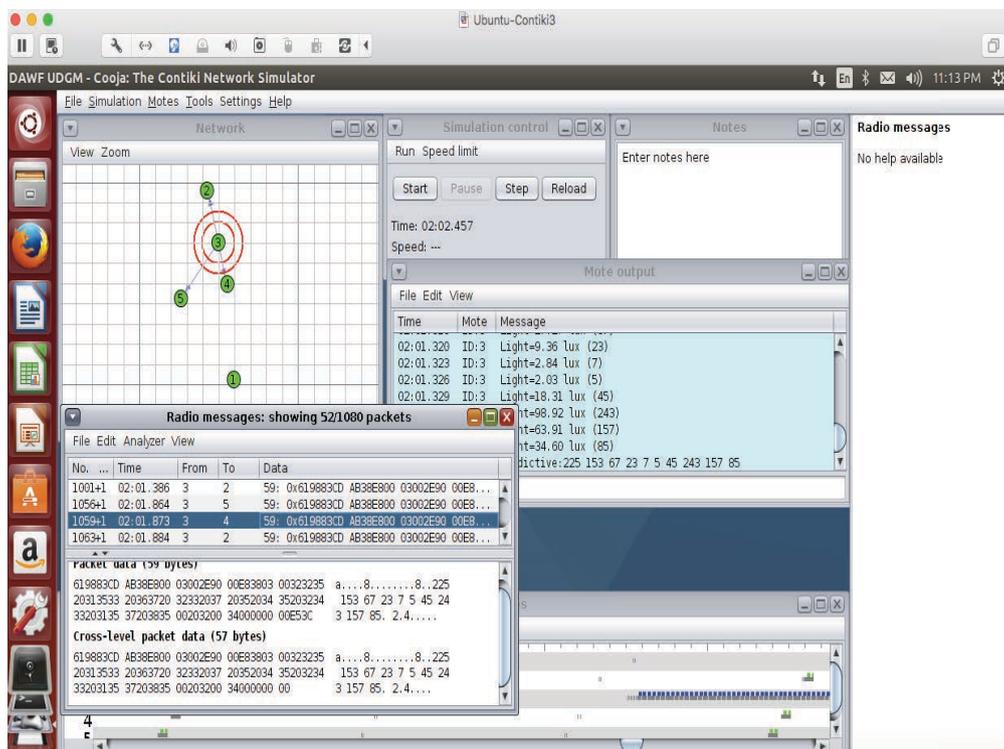


Figure B.1 – DAWF Sky Platform Unit Disk Graph Model Graphical Interface

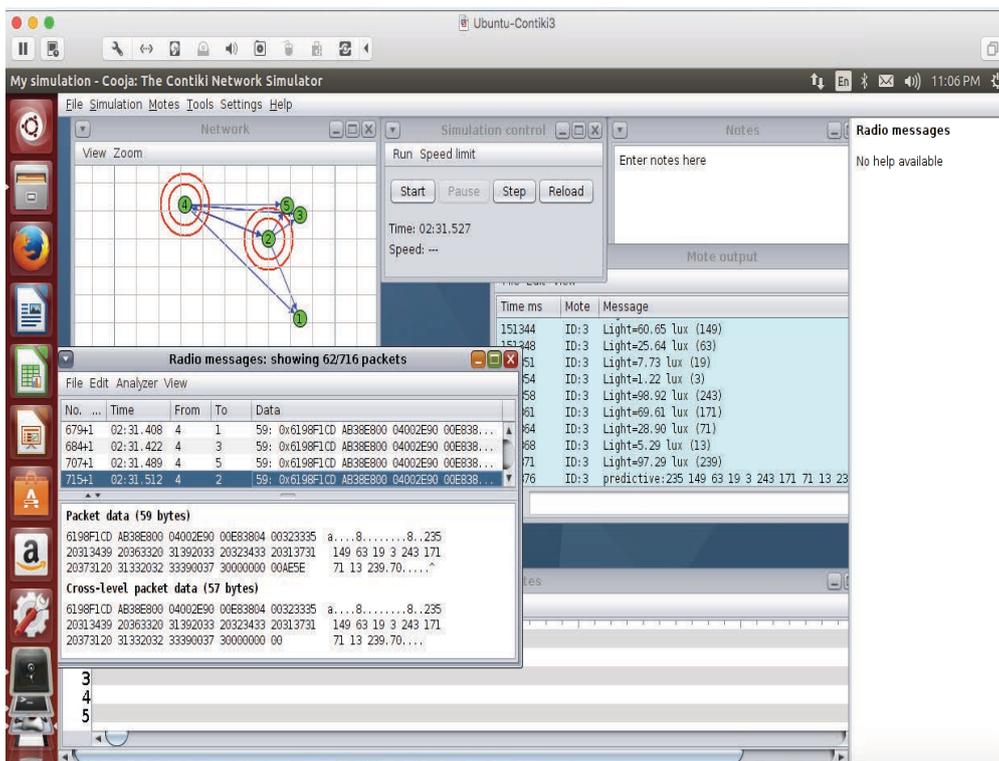


Figure B.2 – DAWF Sky Platform Multi-Path Ray Model Graphical Interface.

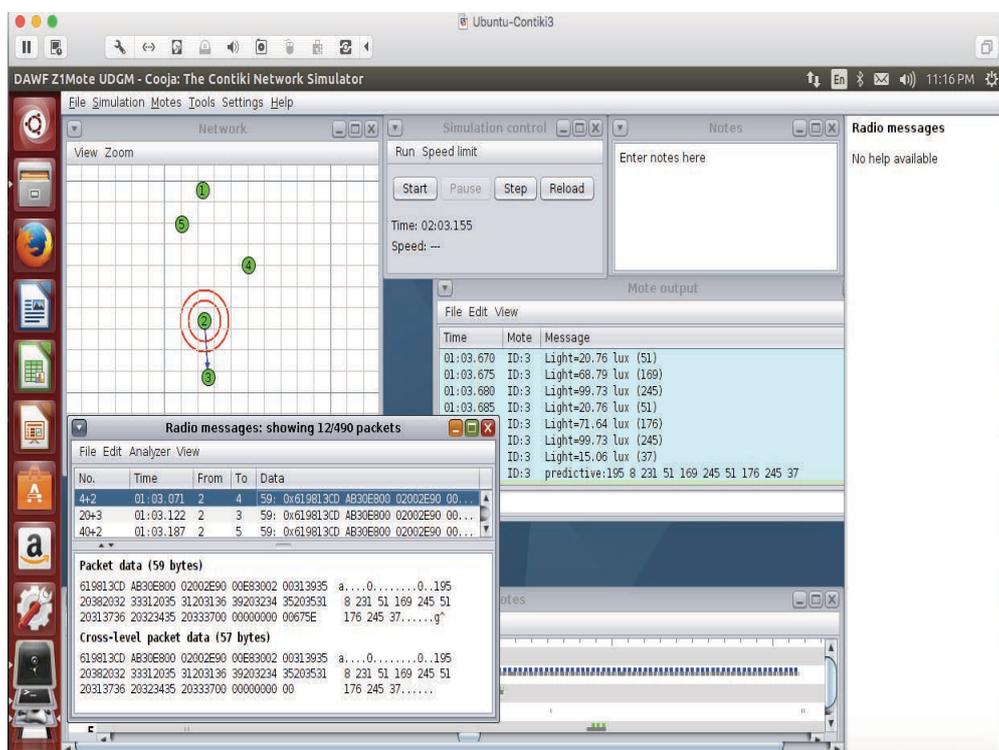


Figure B.3 – DAWF Z1 Platform Unit Disk Graph Model Graphical Interface.

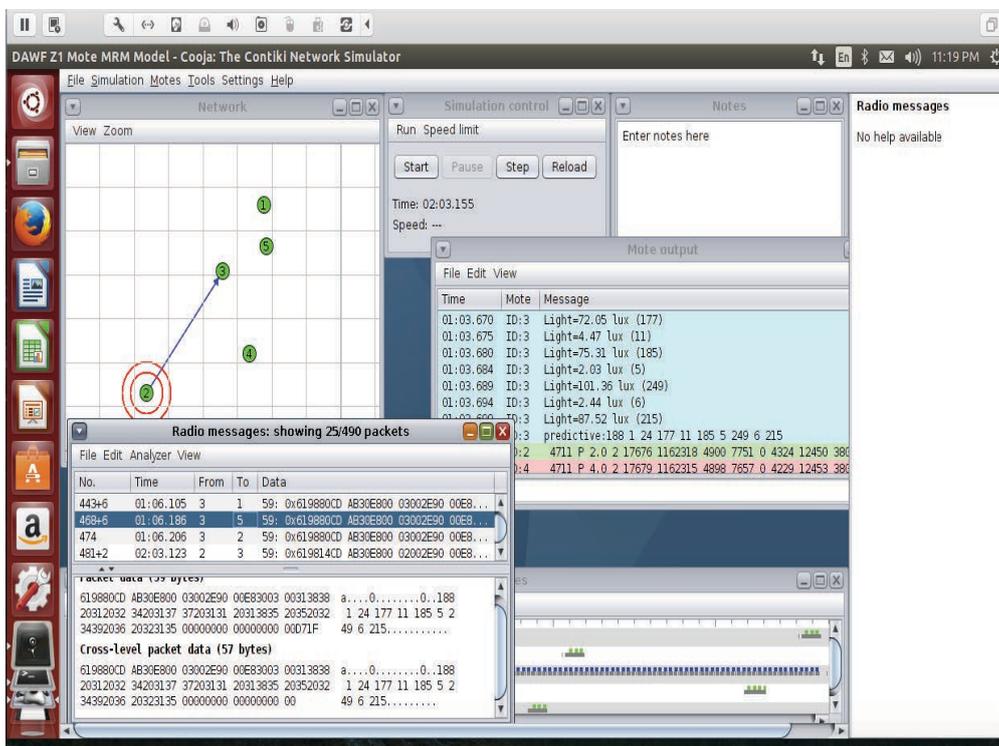


Figure B.4 – DAWF Z1 Platform Multi-Path Ray Model Graphical Interface.

	A	B	C	D	E	F	G	H
404	2/28/04	10:49.0	1106	1	23.1146	31.3139	507.84	2.7244
405	2/28/04	12:48.2	1110	1	23.193	31.1374	507.84	2.7244
406	2/28/04	15:48.0	1116	1	23.2616	31.1374	507.84	2.73696
407	2/28/04	16:46.9	1118	1	23.3008	31.1021	507.84	2.7244
408	2/28/04	18:17.4	1121	1	23.3596	31.0314	507.84	2.73696
409	2/28/04	18:47.3	1122	1	23.3302	31.0667	507.84	2.7244
410	2/28/04	21:17.4	1127	1	23.34	31.1727	507.84	2.7244
411	2/28/04	23:17.5	1131	1	23.4576	30.9608	507.84	2.73696
412	2/28/04	23:47.3	1132	1	23.487	30.8901	507.84	2.73696
413	2/28/04	24:17.6	1133	1	23.5262	30.8547	507.84	2.73696
414	2/28/04	24:48.2	1134	1	23.5164	30.8901	507.84	2.73696
415	2/28/04	25:47.2	1136	1	23.6046	30.7486	507.84	2.7244
416	2/28/04	26:17.8	1137	1	23.6536	30.6779	507.84	2.73696
417	2/28/04	26:47.7	1138	1	23.683	30.6779	507.84	2.74963
418	2/28/04	27:18.7	1139	1	23.6732	30.6779	507.84	2.73696
419	2/28/04	27:47.0	1140	1	23.683	30.6779	507.84	2.73696
420	2/28/04	29:17.4	1143	1	23.732	30.6071	507.84	2.7244
421	2/28/04	29:47.5	1144	1	23.732	30.6071	507.84	2.73696
422	2/28/04	30:17.3	1145	1	23.7418	30.5717	507.84	2.73696
423	2/28/04	32:17.2	1149	1	23.8104	30.5363	507.84	2.73696
424	2/28/04	35:18.2	1155	1	23.8496	30.5363	456.32	2.73696

Figure B.5 – A Snippet of Intel Research Lab Indoor Datasets.

	A	B	C	D	E	F
1	Date	Temperature échangeur solaire vers haut ballon (T3) - Moyenne - (°C)	Temperature échangeur solaire vers bas ballon (T2) - Moyenne - (°C)	Temperature retour échangeur solaire (T1) - Moyenne - (°C)	Temperature Eau Froide entrée échangeur ECS (T4) - Moyenne - (°C)	Temperature Eau Chaude sortie échangeur ECS (T5) - Moyenne - (°C)
2	1/1/11 0:00	16.70	16.85	16.11	13.66	50.
3	1/1/11 1:00	16.75	16.88	16.16	10.74	51.
4	1/1/11 2:00	16.81	16.93	16.19	12.48	51.
5	1/1/11 3:00	16.86	17.00	16.27	14.33	53.
6	1/1/11 4:00	16.88	17.01	16.29	16.22	51.
7	1/1/11 5:00	16.93	17.09	16.32	16.21	50.
8	1/1/11 6:00	16.96	17.09	16.37	13.15	50.
9	1/1/11 7:00	17.01	17.14	16.39	12.80	52.
10	1/1/11 8:00	17.01	17.17	16.45	14.41	50.
11	1/1/11 9:00	17.06	17.19	16.47	13.73	50.
12	1/1/11 10:00	17.09	17.22	16.50	9.63	51.
13	1/1/11 11:00	17.14	17.29	16.55	10.75	50.
14	1/1/11 12:00	20.60	20.46	20.48	9.93	50.
15	1/1/11 13:00	23.25	35.47	40.93	10.66	53.
16	1/1/11 14:00	31.56	38.90	39.34	10.27	50.
17	1/1/11 15:00	55.10	45.69	48.68	11.63	52.
18	1/1/11 16:00	48.79	48.18	50.08	11.27	51.

Figure B.6 – A Snippet of DREAL Datasets from LCIS Lab.



POLE RECHERCHE  
Ecoles Doctorales

### LETTRE D'ENGAGEMENT DE NON-PLAGIAT

Je, soussigné(e) **Somasekhar Reddy Kandukuri**,  
en ma qualité de doctorant(e) de l'Université de La Réunion, déclare être conscient(e) que le plagiat est un acte délictueux passible de sanctions disciplinaires. Aussi, dans le respect de la propriété intellectuelle et du droit d'auteur, je m'engage à systématiquement citer mes sources, quelle qu'en soit la forme (textes, images, audiovisuel, internet), dans le cadre de la rédaction de ma thèse et de toute autre production scientifique, sachant que l'établissement est susceptible de soumettre le texte de ma thèse à un logiciel anti-plagiat.

Fait à **Saint Denis** le : **22/11/2016**

Signature :

**VALIDATION**

**Extrait du Règlement intérieur de l'Université de La Réunion**  
(validé par le Conseil d'Administration en date du 11 décembre 2014)

**Article 9. Protection de la propriété intellectuelle – Faux et usage de faux, contrefaçon, plagiat**

L'utilisation des ressources informatiques de l'Université implique le respect de ses droits de propriété intellectuelle ainsi que ceux de ses partenaires et plus généralement, de tous tiers titulaires de tels droits.

En conséquence, chaque utilisateur doit :

- utiliser les logiciels dans les conditions de licences souscrites ;
- ne pas reproduire, copier, diffuser, modifier ou utiliser des logiciels, bases de données, pages Web, textes, images, photographies ou autres créations protégées par le droit d'auteur ou un droit privatif, sans avoir obtenu préalablement l'autorisation des titulaires de ces droits.

**La contrefaçon et le faux**

Conformément aux dispositions du code de la propriété intellectuelle, toute représentation ou reproduction intégrale ou partielle d'une œuvre de l'esprit faite sans le consentement de son auteur est illicite et constitue un délit pénal.

L'article 444-1 du code pénal dispose : « Constitue un faux toute altération frauduleuse de la vérité, de nature à causer un préjudice et accomplie par quelque moyen que ce soit, dans un écrit ou tout autre support d'expression de la pensée qui a pour objet ou qui peut avoir pour effet d'établir la preuve d'un droit ou d'un fait ayant des conséquences juridiques ».

L'article L335\_3 du code de la propriété intellectuelle précise que : « Est également un délit de contrefaçon toute reproduction, représentation ou diffusion, par quelque moyen que ce soit, d'une œuvre de l'esprit en violation des droits de l'auteur, tels qu'ils sont définis et réglementés par la loi. Est également un délit de contrefaçon la violation de l'un des droits de l'auteur d'un logiciel (...) ».

**Le plagiat** est constitué par la copie, totale ou partielle d'un travail réalisé par autrui, lorsque la source empruntée n'est pas citée, quel que soit le moyen utilisé. Le plagiat constitue une violation du droit d'auteur (au sens des articles L 335-2 et L 335-3 du code de la propriété intellectuelle). Il peut être assimilé à un délit de contrefaçon. C'est aussi une faute disciplinaire, susceptible d'entraîner une sanction.

Les sources et les références utilisées dans le cadre des travaux (préparations, devoirs, mémoires, thèses, rapports de stage...) doivent être clairement citées. Des citations intégrales peuvent figurer dans les documents rendus, si elles sont assorties de leur référence (nom d'auteur, publication, date, éditeur...) et identifiées comme telles par des guillemets ou des italiques.

Les délits de contrefaçon, de plagiat et d'usage de faux peuvent donner lieu à une sanction disciplinaire indépendante de la mise en œuvre de poursuites pénales.