ABSTRACT DATA EXTRACTION AND REFORMATION FOR IOT

BY

LIM XIN HUI

A

REPORT

SUBMITTED TO

Universiti Tunku Abdul Rahman

in partial fulfillment of the requirements

for the degree of

BACHELOR OF COMPUTER SCIENCE (HONOURS)

Faculty of Information and Communication Technology

(Kampar Campus)

JANUARY 2022

UNIVERSITI TUNKU ABDUL RAHMAN

REPORT STATUS DECLARATION FORM

Title: ______Abstract Data Extraction and Reformation for IoT_____________

Academic Session: ___JANUARY 2022____

I ________________________LIM XIN HUI____________________________

(CAPITAL LETTER)

declare that I allow this Final Year Project Report to be kept in Universiti Tunku Abdul Rahman Library subject to the regulations as follows:

1. The dissertation is a property of the Library.
2. The Library is allowed to make copies of this dissertation for academic purposes.

Verified by,

(Author's signature)

(Supervisor's signature)

Address:
_31, Lorong Idaman 1/2,_______
_Taman Idaman,______________
14100 Simpang Ampat, Penang.

Goh Hock Guan
Supervisor's name

Date: ____16/04/2022____

Date: 21/4/2022

FACULTY/INSTITUTE* OF INFORMATION AND COMMUNICATION TECHNOLOGY

UNIVERSITI TUNKU ABDUL RAHMAN

Date: ____16/04/2022____

SUBMISSION OF FINAL YEAR PROJECT /DISSERTATION/THESIS

It is hereby certified that ______***Lim Xin Hui***_________ (ID No: __***18ACB01058***____) has completed this final year project/ dissertation/ thesis* entitled "__Abstract Data Extraction and Reformation for IoT _" under the supervision of ______Ts Dr Goh Hock Guan_______ (Supervisor) from the Department of __Computer and Communication Technology_, Faculty/Institute* of __Information and Communication Technology__.

I understand that University will upload softcopy of my final year project / dissertation/ thesis* in pdf format into UTAR Institutional Repository, which may be made accessible to UTAR community and public.

Yours truly,

____

(LIM XIN HUI)

DECLARATION OF ORIGINALITY

I declare that this report entitled "**ABSTRACT DATA EXTRACTION AND REFORMATION FOR IOT**" is my own work except as cited in the references. The report has not been accepted for any degree and is not being submitted concurrently in candidature for any degree or other award.

Signature : ___________________________

Name : ______LIM XIN HUI_______

Date : _______16/04/2022_______

ACKNOWLEDGEMENTS

I would like to express my sincere thanks and appreciation to my supervisor, Ts Dr Goh Hock Guan who has given me this bright opportunity to engage in a Data Abstraction and Reformation (DAR) project. It is my first step to establish a career in the Internet of Things (IoT) field. A million thanks to you.

To a very special person in my life, Chang Yu Huei, for her patience, unconditional support, and love, and for standing by my side during hard times. Finally, I must say thanks to my parents and my family for their love, support, and continuous encouragement throughout the course.

ABSTRACT

In a Wireless Sensor Network (WSN), sensor nodes are able to communicate with each other and relay the sensed data to a base station. In general, sensor nodes have a limited energy supply and each packet transmission consumes a certain amount of energy from a sensor node. Therefore, if each sensor node can strategically abstract only the critical data from the sensed data to report, a substantial amount of packet transmission and energy consumption can be reduced, thus prolonging the lifetime of WSN. Data abstraction is a term used to describe such a scheme. On the other hand, to prevent data loss, the abstract data reported to the base station must contain sufficient information in such a way that the full set of data can be reconstructed as accurately as possible at the base station using a data reformation scheme. This paper studies the zeroth-, first-, and second-order DAR algorithms, as well as their applications in WSN, which are investigated in this work. Through the study, it was found that the number of packet transmissions can be greatly reduced with a proper selection of DAR algorithms, and yet the data can be reformed at the base station with acceptable accuracy.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

IoT	Internet of Things
DAR	Data Abstraction and Reformation algorithms
WSN	Wireless Sensor Network
CH	Cluster Head
RMSE	Root Mean Square Error
K-RLE	K-Run Length Encoding
EDCD	Effiecient Data Collection and Dissemination

Chapter 1
Introduction

1.1 Problem Statement and Motivation

The problem domain in this project is the energy limitation of the sensor node. A sensor node is a microelectronic device powered by batteries that the power supply is constrained. As a result, the battery life determines the sensor node's lifespan. Moreover, energy consumption is the biggest challenge faced by the sensor node. The sensor node is deployed to transmit data, communicate with adjacent sensor nodes, detect events and collect specified data. According to this, energy consumption can be categorized into three categories such as sensing, communication and data processing [1]. In other words, the energy consumed can be differentiated into sensing energy, computing energy and communication energy [2]. In the sensing phase, each of the sensor nodes needs to detect the specified data within a time range and thus convert the sensed data into digital signals. Furthermore, each sensor node spends the majority of its energy conversing and transmitting data to others during the communication phase. Next, in the data processing phase, the sensor nodes need to do internal computations. The sensor nodes will then unavoidably consume energy.

The problem of energy consumption is essential to be concerned in order to have a better improvement in the field of the Internet of Things (IoT). Moreover, to reduce energy usage, the data abstraction and reformation (DAR) methods are proposed in this study. This is because the sensor node is relying on the battery to operate its functionalities. The sensor node is unable to measure data if it does not have enough power. If the sensor node does not collect data, no data may be examined or observed. Besides, the sensor node lacks sufficient energy to interact with other sensor nodes and is thus unable to send data and perform internal computations. The data will not be properly processed. As a result, the motivation of the project suggested in this paper is to tackle the energy limitation problem by lowering energy usage and extending the sensor nodes' lifetime.

1.2 Project Objectives

The project aims to use DAR algorithms to lower the energy consumption of sensor nodes. The data reformation algorithm must be used in conjunction with the data abstraction algorithm. This is because the data abstraction algorithm is used to filter the significant data and reduce the number of transmissions of the sensor nodes. In addition, data reformation is used to reform the unknown data in order to improve data accuracy. According to this, energy conservation can be achieved. The sensor nodes are able to extend the battery lifetime.

Furthermore, using the data abstraction algorithm, the research attempts to reduce the amount of sensor node transmissions. In this paper, a threshold value is determined. Then, the data abstraction algorithm is used to abstract significant data while the data value is without the range. For example, the threshold value is determined as 25.0. If the value observed by the sensor node is equal to or greater than the threshold value, the data will be abstracted and sent to the sink node. In contrast, if the sensed data is less than the threshold value, it is not transmitted to the sink node. Besides, the data abstraction can be differentiated into three orders such as zeroth-order, first-order and second-order. The present perceptual data will be reported if it is equal to or greater than the threshold of zeroth-order data abstraction. Next, the previous sensed data and current sensed data will be reported if the difference between prior sensed data and current sensed data is equal to or larger than the threshold in first-order data abstraction. Moreover, when the difference between current sensed data, previous sensed data, and previously reported data is equal to or greater than the threshold value, the current sensed data, previous sensed data, and second previous sensed data are reported in the second-order data abstraction. According to this, the sensor nodes only need to communicate data when the threshold is surpassed. It is possible to minimise the number of transmissions between sensor nodes.

Next, the project aims to increase data accuracy by using the data reformation algorithm. As the sink node is receiving the abstracted data from other sensor nodes, the data reformation algorithm is needed to reform the non received data for further analysis. Furthermore, if the data can be reformed properly, it is effective to predict future value as the threshold. The data reformation can be categorized into three order such as zeroth-order, first-order and second-order. The unknown data is reformed using the prior received data value in zeroth-order data reformation. Next, the unknown value of data is reformed linearly by linking the previously collected value to the next sensed value of data in first-order data reformation. Moreover, the unknown value of data is reformed by combining two previously collected values to the next sensed value of data in the second-order data reformation. Based on this, the unknown can be reformed in order to improve data accuracy.

1.3 Project Scope and Direction

This project develops two algorithms for wireless sensor networks (WSN) which are data abstraction algorithm and data reformation algorithm (DAR). The DAR strategy is used to help wireless sensor nodes consume less energy. In WSN, each sensor node may need to sense the data and submit it to the destination every second. The data collected is massive. However, most of the data are common values. The sensor nodes frequently relay similar data to the sink node, indicating that the amount of energy consumed has increased. As a result, the data abstraction method is used to abstract meaningful data by comparing the difference between currently observed and previously felt data with a threshold value. The sensor node does not provide data to the sink node if the difference value is within the range. If this is not the case, the data will be delivered to the sink node. In addition, at the sink node, the data reformation technique is utilised to rebuild the data. Some vital data may not be provided to the sink node because the data has been abstracted. In order to analyse the acquired data, the sink node must implement the data reformation method to reconstruct the data using the previously received data set.

1.4 Project Contributions

To limit the quantity of data transmissions, the data abstraction algorithm can extract key values in multiple orders with varied thresholds. For example, there are three sorts of thresholds such as threshold value in zeroth-order, threshold of rate in first-order and threshold of acceleration in second-order. According to this, the significant data can be efficiently abstracted and thus the changes in different orders with different thresholds can be observed and analysed. Besides, the number of data transmissions can be compared between different orders and study which orders are able to get the least number of data transmissions.

By having this data reformation algorithm, it is able to solve the problem of data loss and enhance data accuracy. To recreate unknown data, the data reformation algorithm can be used to the sink node. Instead of assuming the unknown value of data is the same as the previously reported value of data, this data reformation technique uses the set of previously collected data to reconstruct the non reported data. The data reformation algorithm proposed in this project has three orders such as zeroth-order data reformation, first-order data reformation and second-order data reformation. Regarding this, the unknown value of data can be rebuilt and hence determine which orders are extremely close to the actual perceived value.

1.5 Report Organization

This report is organised into 5 chapters: Chapter 1 Introduction, Chapter 2 Literature Review, Chapter 3 System Design, Chapter 4 System Implementation and Testing and Chapter 5 Conclusion. The first chapter is the introduction of this project which includes problem statement and motivation, project objectives, project scope and direction, project contributions and report organisation. The second chapter is the literature review carried out on several existing energy conservation techniques to evaluate the strengths and weaknesses of each schemes. The third chapter is discussing the overall system design of this project. The fourth chapter is regarding the details on how to implement the experiment to evaluate the DAR performance and benchmarking. Lastly, the fifth chapter reports the conclusion and recommended future work.

Chapter 2
Literature Review

In order to make sure the sensor nodes can perform well, different solutions are proposed and studied by numerous researchers in incompatible fields. The power management approaches can be differentiated into three categories such as energy conservation, energy harvesting and energy transferring [3]. However, for this paper, only energy conservation will be focused on and studied. There are various solutions are proposed in a homogeneous manner.

Energy conservation approaches are concerned with minimizing a sensor node's power usage by managing the sensor node's components. A sensor node involves four major components such as sensing device, processing computing unit, communications unit and power supply unit. The sensing component is used to collect data, then the collected data is transmitted to the processing component. The processing components are responsible to manage the sensor node and perform the internal computational and aggregation of the data. The communication component is used to connect with other sensor nodes in the network. The power components such as batteries or low-power capacitors are used to provide energy to the sensor node. The solutions include in energy conservation are data reduction [4], data aggregation [5] and routing protocol [6].

2.1 Fog Computing Architecture-Based Data Reduction Scheme for WSN

According to Deng et al. [4], the authors proposed a data compression scheme that is using a synchronous prediction model based on fog computing architecture. The strategy aims to minimise network congestion while also lowering energy usage. Sensor nodes, CH nodes, fog nodes, and cloud centers make up the network architecture in this method, which is divided into three layers: sensor, fog computing and cloud computing.

Firstly, the sensor nodes collect data and compare it to the parameter predicted by the fog nodes at the sensor layer. The data collected is redundant if the prediction error is smaller than the threshold value. Otherwise, the data is transmitted to the CH node.

Next, the data will be processed and delivered to the fog nodes via the CH nodes. Thus, the CH nodes are frequently chosen via polling in clusters. After that, the fog nodes role is to condense the amount of data collected by analysing the time correlation of observed data at the fog computing layer using the synchronous prediction mode, an autoregressive prediction (AR) model. Therefore, if the fog nodes wait a long time without receiving data, a new parameter is constantly anticipated and give to all sensor nodes for the next time slot. The fog nodes, on the other hand, receive the data and record the error in the AQ error queue. The Accumulated Error Queue AQ is used to analyse the probability of successful prediction. According to this, the prediction model needs to be readjusted if the likelihood of success forecast is within the threshold. Besides, the fog nodes employ the geographical correlation of the data to further filter and analyse the received data before delivering the data to the cloud center. The data received at the cloud computing layer is stored and visualized in the cloud centers.

With the AR prediction model, the data can be compressed and reduced energy consumption and the number of data transmissions. However, the fog nodes and sensor devices are frequently communicate with each other due to the synchronous prediction model. As a result, the transmission energy which may or may not be lowered efficiently. Moreover, the latency is increasing to transmit the data via three layers with several data filtering processes. Lastly, the unreported data is thought to be identical to the previous data. In my point of view, the data abstraction algorithm is more effective in abstracting out important data without waiting for the parameters predicted by the prediction model to be compared. In addition, the data reformation algorithm can help further enhance the prediction model.

2.2 A data aggregation transfer protocol based on clustering and data prediction in wireless sensor network

According to Meng, Zhang and Zou [5], the scholars proposed a method called DACP. In this paper, the scholars hope to reduce the quantity of data transmissions and enhance the efficiency of data aggregation. The DACP is a clustering and data prediction-based data aggregation transfer protocol. The DACP consists of four stages such as DACP_Send, DACP_CH, DACP_Prediction and DACP_DA. Every sensor node sends a message to the sink node at the DACP_Send stage, which includes an ID, current remaining energy, position, and sensed data. The sink node selects a suitable set of CH nodes and broadcasts the message consists of CH node information and CH node ID. Besides, the network's sensor nodes will be split into numerous clusters. A CH node is present in each cluster. Next, the sensor nodes will then elect the CH nodes in the DACP_CH step, based on the message broadcast by the sink node. The sink node must receive information from the designated CH nodes. Furthermore, the sink node uses an AR algorithm to anticipate the next data value and sends the parameter to all sensor nodes in the DACP_Prediction stage. Each CH node combines data from the sensor nodes inside its cluster field in the final step of DACP_DA. Then, compare the sink node's accepted data to the aggregated data. The CH nodes do not provide data to the sink node if the comparison result is inside the range. Otherwise, CH nodes send the data to the sink node. The sink node saves the data from the CH nodes and updates the historical data in order to forecast the next future value.

The strength of this approach is used to decrease the number of transmissions in order to reduce energy consumption. This is because the energy used for data transmission is greater than the energy used for data receiving. Therefore, This method can minimise total energy use while also extending the energy lifespan. Some data, however, does not make it to the sink node. As a result, the sink node then assumes the unknown data in the same way that the previously received data is predicted. This may cause data loss and reduce data accuracy.

For this paper, the method proposed is able to select the CH node with multiple roles instead of each node to perform multiple functions. As the selection CH node is done in each cycle, this is able to balance the energy required by each node. However, in order to increase the data accuracy, the prediction mechanism in this paper needs to be improved further. According to this, the data reformation algorithm is better than the prediction model to reconstruct the unknown data. Moreover, the predefined threshold, the equation proposed by Goh et al. [7] shows that the threshold can be defined by using the equation of $th_n = a + (n-1)s$. As the threshold can be defined with a proper algorithm, it might increase the data accuracy.

2.3 Efficient routing protocol in IoT using modified genetic algorithm and its comparison with existing protocols

The researches proposed the Maximum Enhanced Genetic Algorithm (MEGA), which combines the Local Search mechanism with the Sleep-Wake up mechanism [6]. The researchers presented this strategy with the goal of lowering sensor node energy consumption, extending sensor node service life, and improving network performance. Through multiple cycles in the network, the Local Search method is utilised to identify the quickest path to the destination in order to send the data. The sensor node's energy is conserved via the Sleep-Wake up process, which selects a few active nodes at random. The data only can be transmitted through these active nodes. When the active nodes use up its power, then the other sleeping nodes will be woken up and continue the work. According to the MEGA, the method generates a set of active nodes as the first generation. Then, use the Local Search mechanism to find the shortest path. After that, from the first generation of nodes, the MEGA select the next generation of nodes. Furthermore, the Local Search technique discover and identify the shortest path that is available to transmit the data once more. Moreover, the Sleep-Wake up mechanism check the availability of the nodes for the next iteration. Last but not least, end the current segment by sending the data to the sink node.

The strength of the MEGA is based on the total number of successful packets transmitted, delay time and average energy consumption. From the results given, the number of successful packets transmitted is the highest. Although the MEGA is using Sleep-Wake up mechanism and some of the nodes is sleeping, the data can still be transmitted successfully to the destination. Besides, the data is transmitted with less delay. Because the route is the shortest, it cuts down on the time it takes for sensor nodes to transfer data to the sink node. Overall, MEGA is able to reduce energy consumption. Regarding this, the Sleep-Wake up mechanism do well to improve the energy-saving by setting some nodes to sleep. Nevertheless, the Local Search mechanism and the Sleep-Wake up mechanism must be conducted back and forth, the operation is complicated. In addition, the energy is consumed during the time of searching for the shortest path and active nodes.

2.4 Table of analysis

Table 2.1 Table of analysis.

	2.1 Fog	2.2 DACP	2.3 MEGA
Advantage	• reduce energy consumption • Minimise network congestion • reduce number of data transmissions	• reduce the number of data transmission • improve data aggregation	• successful to transmit data • less delay time
Disadvantage	• frequently communicate between the fog nodes and sensor devices to synchronize the prediction error • latency may increase	• The unknown data is assumed to be the same as previously received data • low data accuracy	• The execution of MEGA is complex
Comment	• DAR is more effective compare to synchronous prediction model	• Compared with the prediction mechanism in DACP, the first-order data reformation algorithm is more accurate	• Compared with MEGA, the deployment and execution of the DAR algorithm is very simple

2.5 Concluding Remarks

Finally, unlike [4], DAR algorithms do not require threshold synchronization and can still extract critical data in order to operate efficiently. In comparison to [5,] DAR methods can use the data reformation algorithm to reform non-received data with better data accuracy. In addition, when compared to [6,] DAR algorithms are simple to execute. In sum, DAR algorithms are more effective to utilise, more accurate to reconstruct non-received data values and easier to implement than the approaches of [4-6].

Chapter 3
System Model

3.1 System Model

3.1.1 Research Methodology

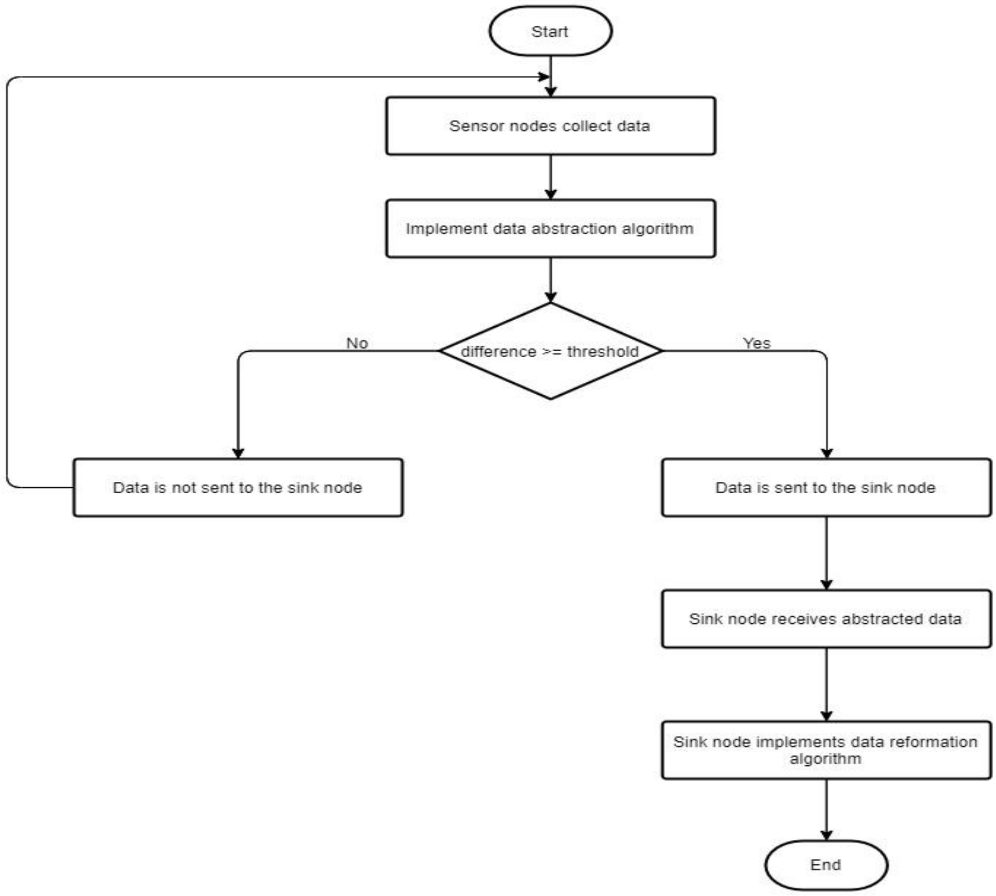


Figure 3.1 Flowchart of DAR between the sensor nodes and sink node.

This project develops two algorithms for WSN which are data abstraction algorithm and data reformation algorithm. Wireless sensor nodes employ DAR algorithms to lower their energy usage. Assume that the sensors, cluster head nodes and sink node have been deployed to form a WSN. Every second, each sensor node in a WSN must perceive the data and transmit it to the sink node through the cluster head node. The data collected is massive. However, most of the data are common value. Most of the time, the sensor nodes send the common data to the cluster head node. After that, cluster head node aggregates the data and delivers to the sink node which will increase the energy consumption during transmission. Therefore, by comparing the difference between currently detected data and previously sensed data with a threshold value, the data abstraction method is utilised to abstract relevant data. The cluster head node does not transmit data received from the sensor nodes to the sink node if the difference in sensed data is within the range. The data is sent to the sink node if this is not the case. In addition, the data in the sink node is reconstructed using the data reformation algorithm. Some vital data may not be provided to the sink node because the data has been abstracted. As a result, in order to analyse the gathered data, the sink node must implement the data reformation method to reconstruct the data.

3.1.2 Tools to use

The technologies involved are computer for hardware and Python for software.

Hardware

The hardware involved is a computer. The computer has 8.00GB of RAM. The system type is 64-bit operating system (x64-based processor). Moreover, the processor is Intel Core i7-8550U. The computer is used to execute the Python software.

Software

The software involved is Python. Python software provides Matlab library, which can be used to draw graphs for further observation and analysation. In addition, data abstraction algorithms and data reorganization algorithms can be coded in Python with fewer lines of code, which is more efficient and effective. As a result, the structure of the code will be executed to abstract the data and reform the data in order to plot the graph in zeroth-order, first-order and second-order. With the graph, the zeroth-order, first-order and second-order of data abstraction and data reformation can all be seen and evaluated.

3.2 Algorithms Design

3.2.1 Data Abstraction

Data abstraction is a filtering procedure that checks each piece of sensed data according to a set of criteria to determine whether it should be transmitted to the base station. Take a look at Figure 3.2, where data is only given at times 1, 4, and 9. For $i, j \geq 0$, let y_i represent a serial number of the i^{th} sensor data, z_j represent a serial number of the j^{th} reported data and t_j represent the timestamp of the z_j reported data. If z_j is assigned by the reported y_i, then $t_j = i$, where $j \leq i$. Thus, the data are reported $t_1 = 1$, $t_2 = 4$, and $t_3 = 9$. Therefore, $z_1 = 1$, $z_2 = 4$ and $z_3 = 9$.

Data	●	○	○	●	○	○	○	○	●	○
i	1	2	3	4	5	6	7	8	9	10
j	1			2					3	
t_j	1			4					9	

● Data sensed and reported
○ Data sensed but not reported

Figure 3.2 An example of data abstraction.

Table 3.1 Data abstraction algorithms.

Order of Data Abstraction	Algorithms	
Zeroth	$y_i \geq th_{val}$	(1)
First	$\lvert y_i - z_j \rvert \geq th_{rat}$	(2)
Second	$\lvert y_i - 2y_{i-1} + z_j \rvert \geq th_{acc}$	(3)

In essence, the data abstraction can be described in a variety of ways. The zeroth-, first-, and second-orders of data abstraction are investigated in this study. In zeroth-order data abstraction, the sensor data is compared to a pre-determined threshold value, which denotes as th_{value}. Let y_i represent the sensor data at time i, where $i \geq 1$. In this scheme, y_i is reported to the base station if the data of y_i is larger than or equal to th_{value}, as in Equation (1). Note that when $i = 1$, y_1 has to be sent to the base station as the first reference data, z_1, for the completion of the zeroth-order reformation scheme, which is to be discussed in Section 3.2.2. Typically, there is only one value is reported at a time in one packet.

In first-order data abstraction, the change of sensor data is monitored and compared to a pre-fixed threshold of rate, th_{rate}. Let z_{t_j} denote the last reported data, where t_j is the last reporting time and $j \geq 1$. In this scheme, for any $i > t_j$, if the difference between the latest sensed data y_i and the last reported data z_{t_j}, denoted as Δy_i is larger than or equal to th_{rate} as shown in Equation (2), then y_i and y_{i-1} are reported to the base station. In other words, there are two sensed data to be reported in one packet. When $i = 1$, y_1 must be sent to the base station as the first reference data in order to complete the scheme for first-order reformation, and thus $t_1 = 1$. Subsequently, if at time $i > t_j$ data y_{i-1} and y_i are reported, then t_{j+1} and t_{j+2} are set to be $t_{j+1} = i - 1$ and $t_{j+2} = i$, and t_{j+2} will then become the final report time for subsequent data evaluation.

In second-order data abstraction, the acceleration of the data is measured and compared to a pre-fixed threshold of acceleration, denoted as th_{accel}. In this scheme, for any $i > t_j$, if the difference between the current collected data and the last received data, i.e., y_i, y_{i-1}, and z_{t_j}, denoted as $\Delta^2 y_i$ is larger than or equal to th_{accel} as shown in Equation (3), then y_i, y_{i-1}, and y_{i-2} have to be transmitted to the base station. When $i = 1$ and $i = 2$, y_1 and y_2 must be sent to the base station as the first two reference data in order to complete the scheme for second-order reformation, and thus $t_1 = 1$ and $t_2 = 2$. Moreover, if at time $i > t_j$, data (i.e., y_i, y_{i-1} and y_{i-2}) are reported, then $t_{j+1} = i - 2$, $t_{j+2} = i - 1$, and $t_{j+3} = i$, and t_{j+3} will become the final report time for subsequent data evaluation. Typically, there are three data can be transmitted at a time in one packet.

3.2.2 Data Reformation

Data reformation is a type of data interpolation that uses received data to reform non-received data. Data reformation schemes can be developed in different orders, similar to data abstraction, this work only looks at the zeroth, first, and second orders of data reformation.

For $j \geq 1$ and $t_j \leq i < t_{j+1}$, let $\hat{y}_i$ represent the reformed data of y_i at the base station after $z_{t_{j+1}}$ is received, where $\hat{y}_{t_j} = z_{t_j} = y_{t_j}$ and $\hat{y}_{t_{j+1}} = z_{t_{j+1}} = y_{t_{j+1}}$. However, for any $i \notin \{t_j \mid j \geq 1\}$, $\hat{y}_i$ is unknown, and thus its value needs to be reconstructed.

The zeroth-order data reformation tries to reconstruct non-received data using piece-wise constant data; the first-order data reformation uses piece-wise linear lines to recover that data; and the second-order data reformation uses piece-wise parabolic curves to reform the data. These schemes are shown in Table 3.2.

Table 3.2 Data reformation schemes.

Order of data reformation	Algorithms	
Zeroth	$\hat{y}_i = z_j$, for $t_j \leq i \leq t_{j+1}$	(4)
First	$\hat{y}_i = \frac{(i - t_{j+1})}{(t_j - t_{j+1})} z_j + \frac{(i - t_j)}{(t_{j+1} - t_j)} z_{j+1}$, for $t_j \leq i \leq t_{j+1}$	(5)
Second	$\hat{y}_i = \frac{(i - t_j)(i - t_{j+1})}{(t_{j-1} - t_j)(t_{j-1} - t_{j+1})} z_{j-1} + \frac{(i - t_{j-1})(i - t_{j+1})}{(t_j - t_{j-1})(t_j - t_{j+1})} z_j + \frac{(i - t_{j-1})(i - t_j)}{(t_{j+1} - t_{j-1})(t_{j+1} - t_j)} z_{j+1}$, for $t_j \leq i \leq t_{j+1}$	(6)

3.3 Verification Plan

To evaluate the suggested systems' performance, two parameters are introduced which are the number of packet transmissions and the Root Mean Square Error (RMSE). The total number of packets sent to the base station is defined as the number of packet transfers. A packet does not have to contain only one piece of data. For example, a packet comprises two most recent sensor data in the first-order data abstraction, while a packet contains three most recent sensor data in the second-order data abstraction. As demonstrated in Equation (7), RMSE is utilised to calculate the difference between actual sensor data and reconstructed data.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \quad (7)$$

The total number of detected data is n. Theoretically, the value of *RMSE* will be affected by the number of data transmitted. In zeroth-order data abstraction, the number of packet transmissions decreases when th_{rate} increases. In first-order data abstraction, there is a chance that the number of packet transmissions will increase when th_{rate} as well if there is any $\Delta y_i = th_{rate}$ and $\Delta y_i < th_{rate+1}$. Furthermore, in second-order data abstraction, there's a chance that when th_{accel} grows, the number of packet transmissions grows as well when th_{accel} is increasing when there is any $\Delta^2 y_i = th_{accel}$ and $\Delta^2 y_i < th_{accel+1}$. However, lower *RMSE* is not always associated with a higher number of packet transmissions. As a result, a lower th_{rate} and th_{accel} value does not always imply a lower *RMSE*. In 1A1R, *RMSE* is always less than or equal to th_{rate}, $RMSE \leq th_{rate}$. As a result, the value of th_{rate} in the implementation of 1A1R can be specifically set to the application requirement's accuracy of the reconstructed data in terms of *RMSE*. Aside from that, *RMSE* is equal to 0 if all sensed data is reported to the base station, as there is no inaccuracy in the "reformed" data. In other words, if there are 600 data points, precisely 600 of them are reported. This indicates that the base station has no unreceived data. As a result, the *RMSE* is 0 because the data received at the base station and the data perceived by the sensors are identical.

3.4 Timeline

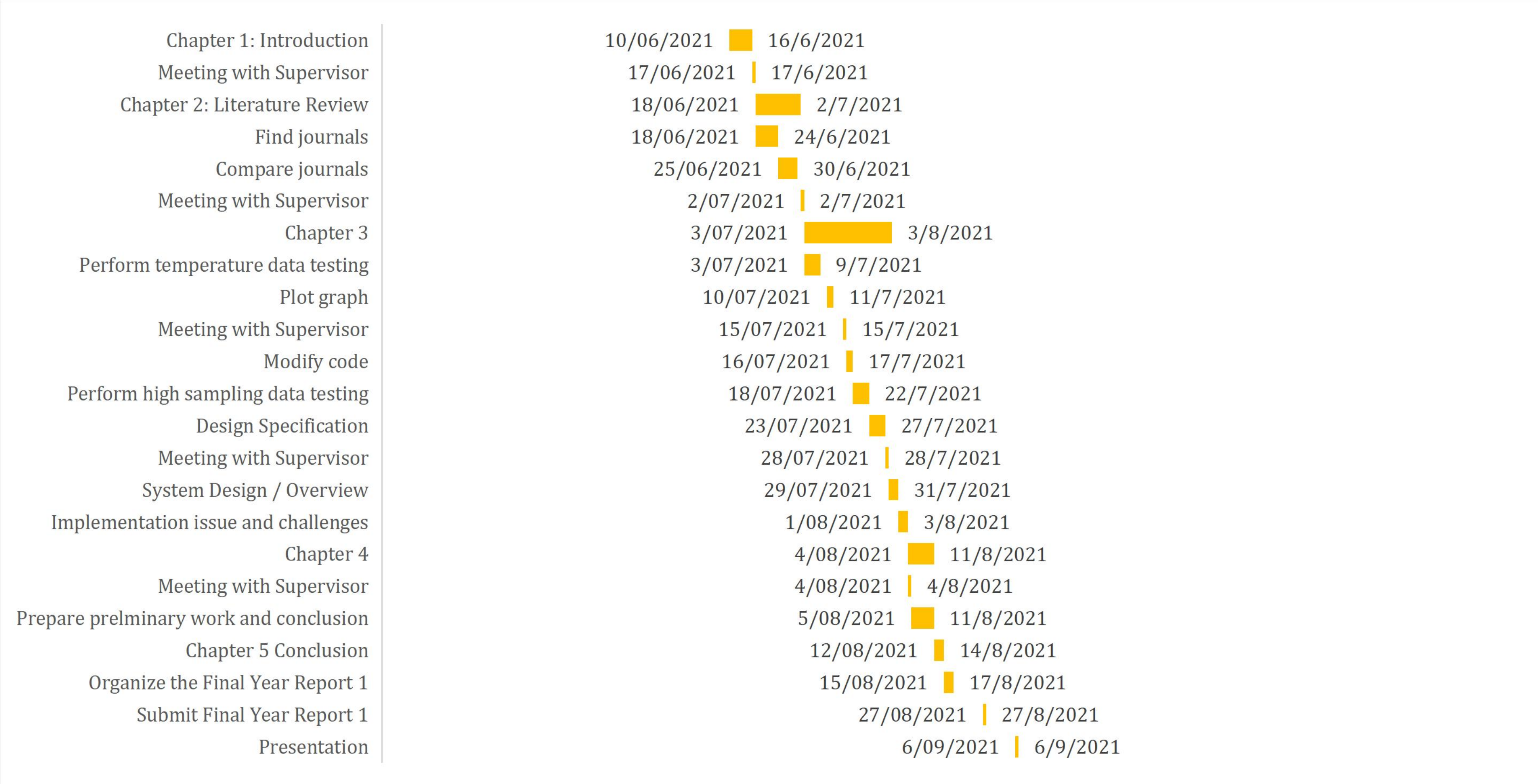


Figure 3.3 FYP 1 Timeline.

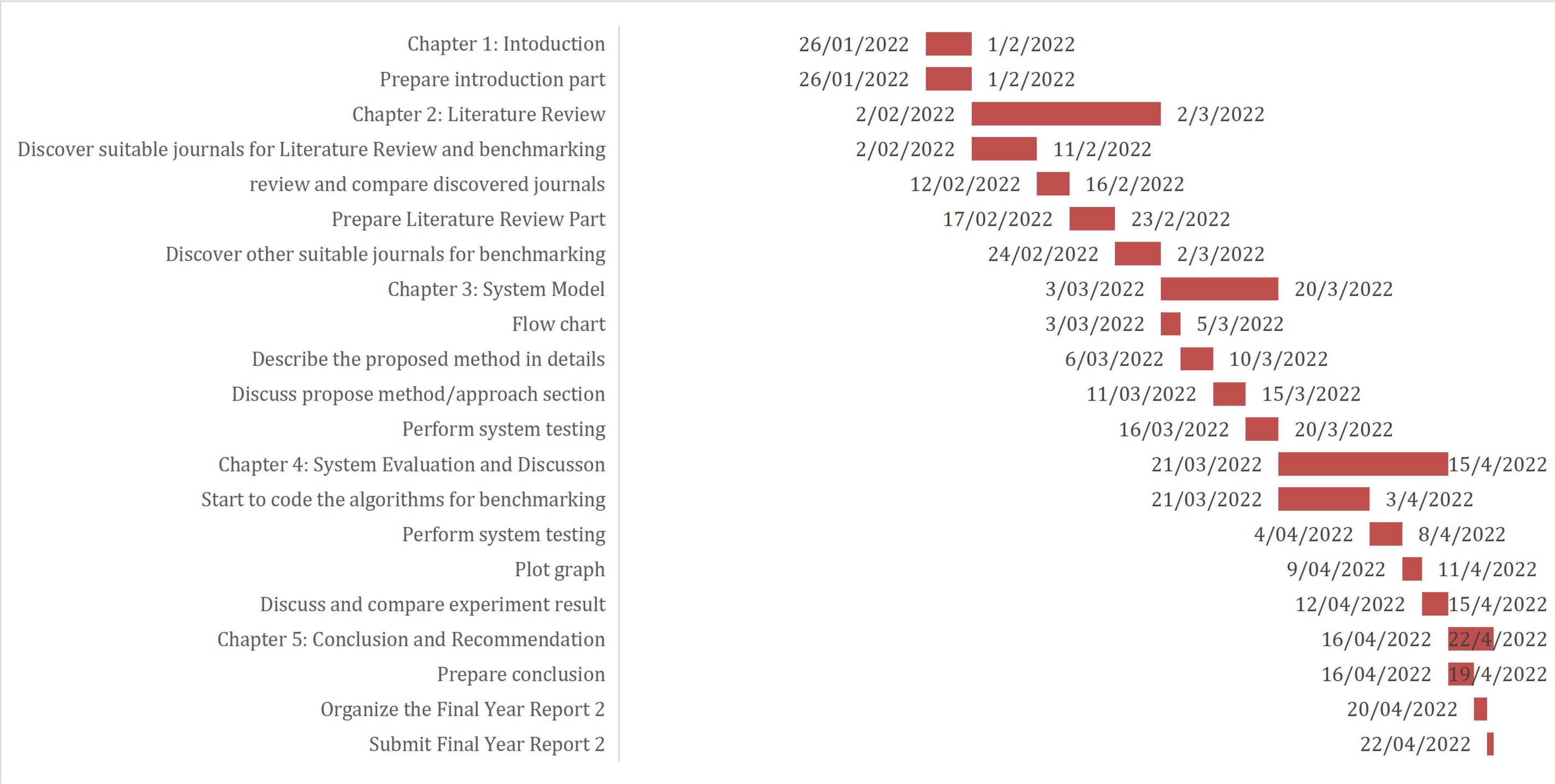


Figure 3.4 FYP 2 Timeline.

3.5 Concluding Remarks

In conclusion, data abstraction algorithm will reduce the number of data transmissions. As a result, not all of the data is transmitted and the acquired data accuracy in the sink node or base station is compromised. Data reformation algorithm is a component of DAR scheme that reconstructs non-received data to ensure that the accuracy of the information is not affected by data abstraction.

CHAPTER 4

System Evaluation and Discussion

Different data sampling rates are used to evaluate DAR algorithms in two sorts of applications. One is low-changing-rate application and the other is high-changing-rate application.

4.1 Algorithms Evaluation

4.1.1 Application/attribute with low changing rate

In this part, the project demonstrates how DAR works for temperature monitoring, where the temperature attribute has a modest changing rate in general. A temperature sensor is deployed to collect temperature values and 600 sensor samples are presented in the experiment. There are three different types of temperature data patterns: burst, incremented and decremented and random.

In addition, assuming each sensor node sends temperature data at the same time, the total number of data sent is 600 (100%). However, in order to save energy, the number of data communicated can be substantially decreased, because data transmission, as part of the communication operation, consumes the most energy of all the actions performed by a sensor node. The minimal number of packet transmissions in zeroth- and first-order data abstraction is 2 whereas the minimum number of packet transmissions in second-order data abstraction is 3.

In addition, in order to have worth information at the base station, the reformation schemes need to be applied. In temperature monitoring, there is merely a temperature value. Therefore, the sensor data can be directly compared with the reformed data. Note that $RMSE$ is generally increasing when th is increasing if the minimum and maximum values are reported before the sign of the slope changes. The following are the outcomes of the experiment.

4.1.1.1 Burst temperature data

The sensor is put to the test at a temperature setting that resulted in burst temperature data, which means that the temperature value was rapidly increased. Figure 4.1 depicts the experiment's outcome.

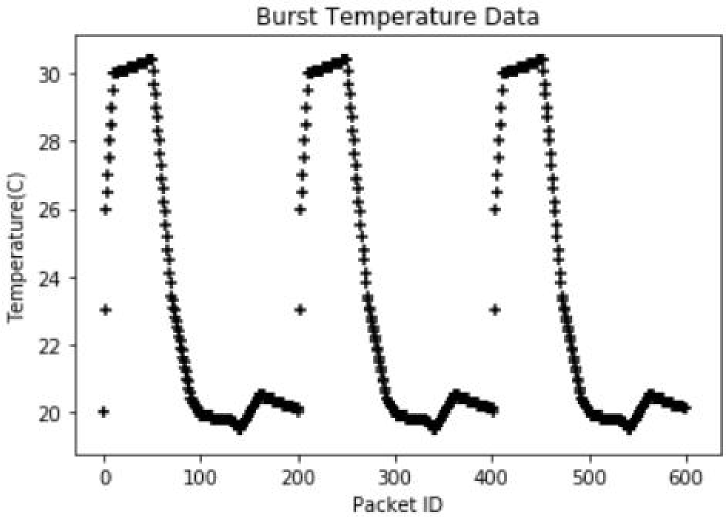


Figure 4.1 Graph result of burst temperature data.

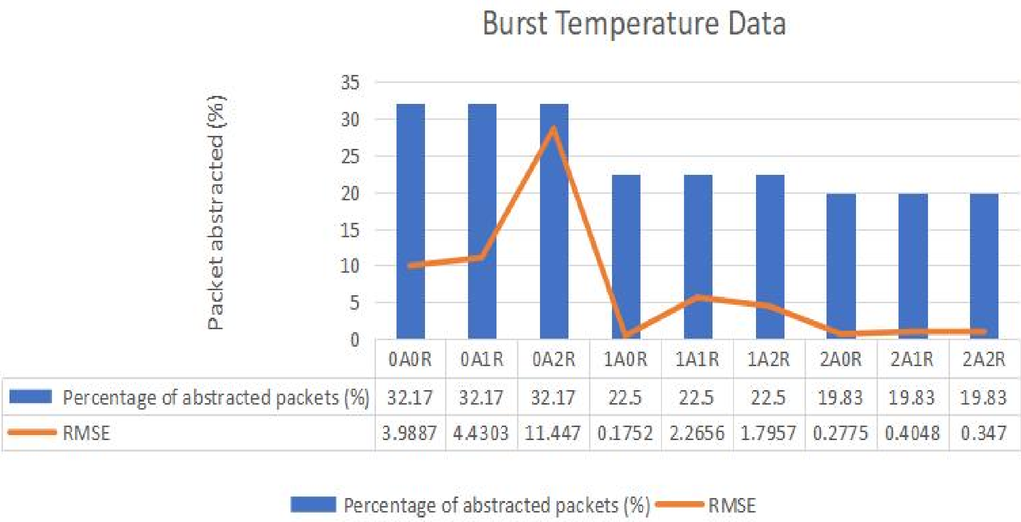


Figure 4.2 Graph results of DAR algorithms of number of packets abstracted (%) v.s. Root Mean Square Error in burst temperature data.

Figure 4.2 also shows that the first-order data abstraction and zeroth-order data reformation may reform the burst temperature data with $RMSE = 0.175$ by sending 22.5 percent of total data. Second-order data abstraction and zeroth-order data reformation are able to reform the burst temperature data with $RMSE \leq 0.8$ by just delivering 19.83 percent of total data.

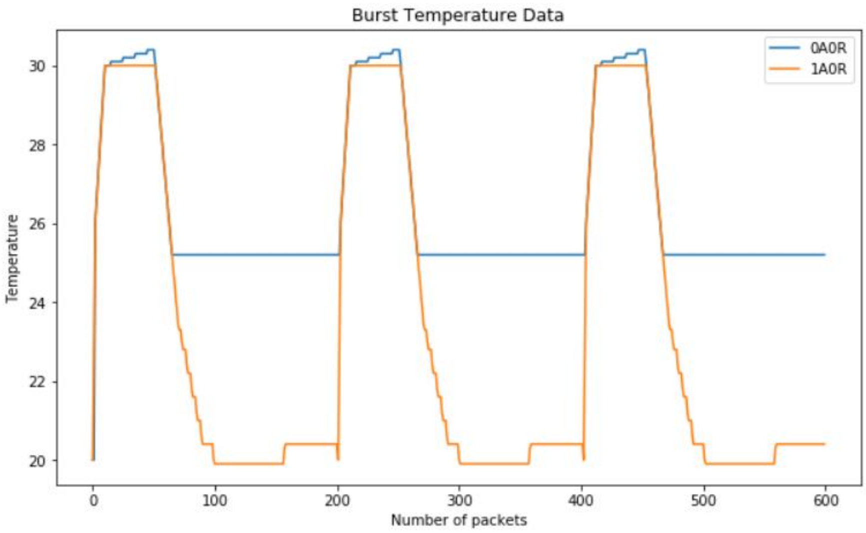


Figure 4.3 Comparison of the greatest value of RMSE and the worst value of RMSE in burst temperature data.

4.1.1.2 Incremented and decremented temperature data

The sensor was tested with a configuration that resulted in incremented and decremented temperature data, meaning the temperature value was steadily increased and then gradually decreased. Figure 4.4 depicts the experiment's outcome.

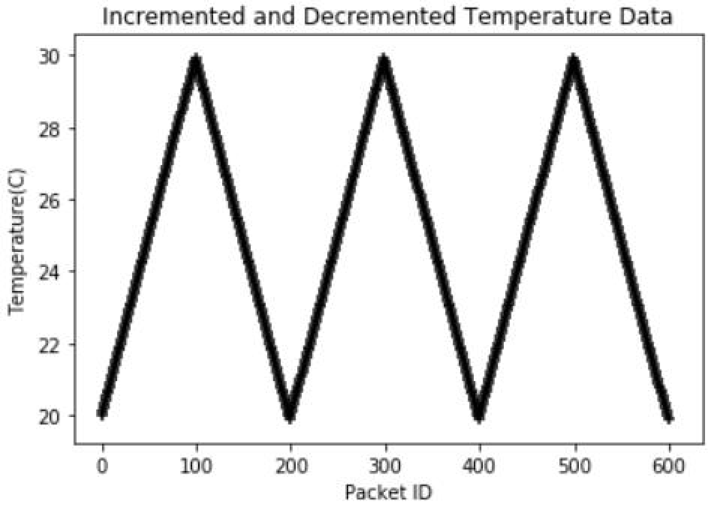


Figure 4.4 Graph result of incremented and decremented temperature data.

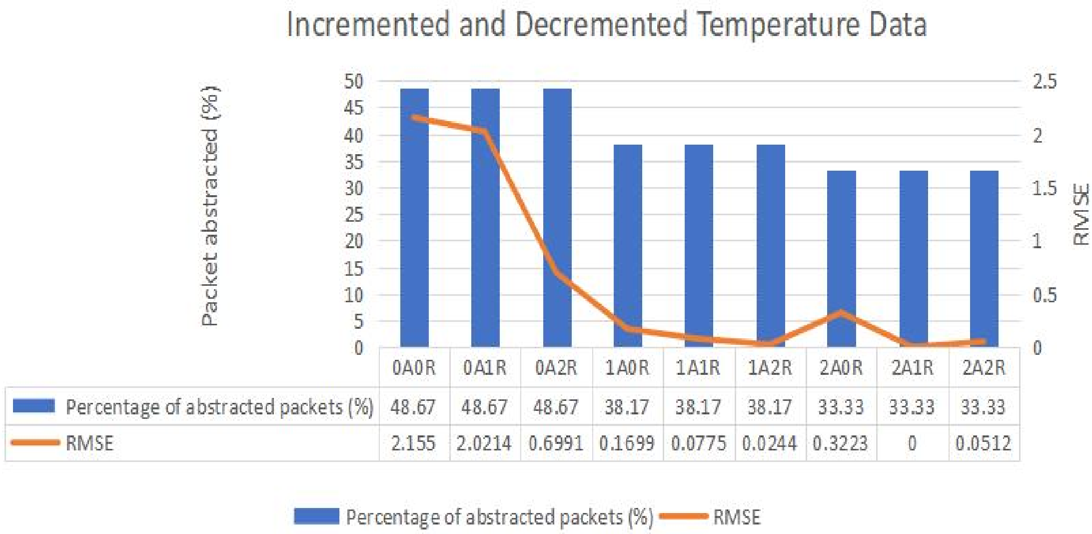


Figure 4.5 Graph results of DAR algorithms of number of data transmitted (%) v.s. Root Mean Square Error in incremented and decremented temperature data.

Figure 4.5 shows that the second-order data abstraction and first-order data reformation can reform the incremented and decremented temperature data with $RMSE = 0$ by just delivering 33.33 percent of total data. Furthermore, the data can be reformatted with $RMSE = 0.024$ using first order abstraction and second order reformation.

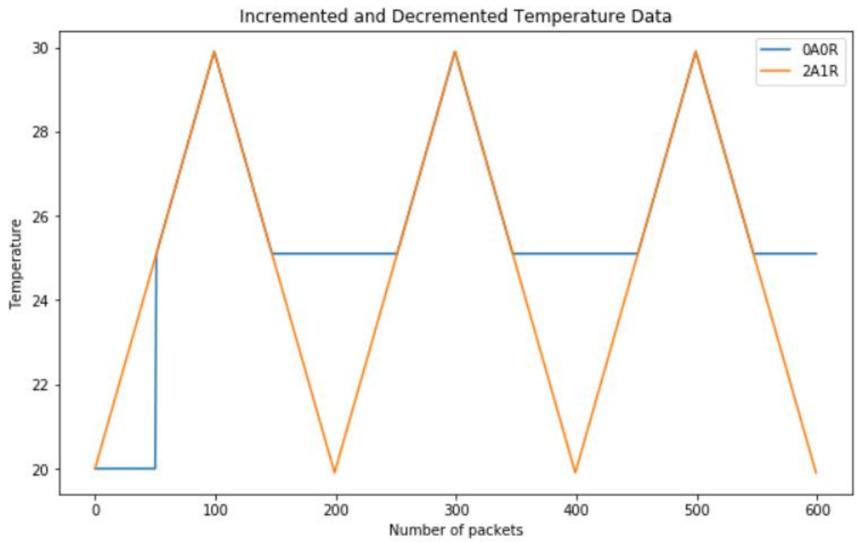


Figure 4.6 Comparison of the greatest value of RMSE and the worst value of RMSE in incremented and decremented temperature data.

4.1.1.3 Random temperature data

The sensor was put through its paces with a setup that generated random temperature readings from an object. Figure 4.7 depicts the experiment's outcome.

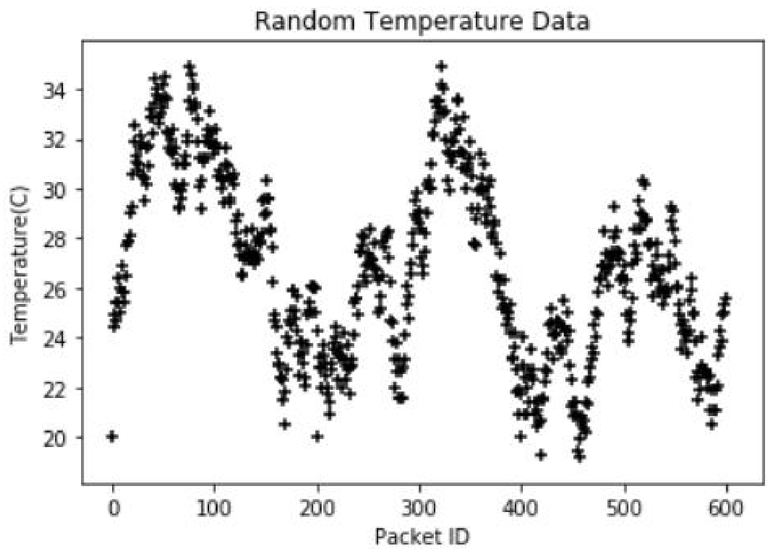


Figure 4.7 Graph result of random temperature data.

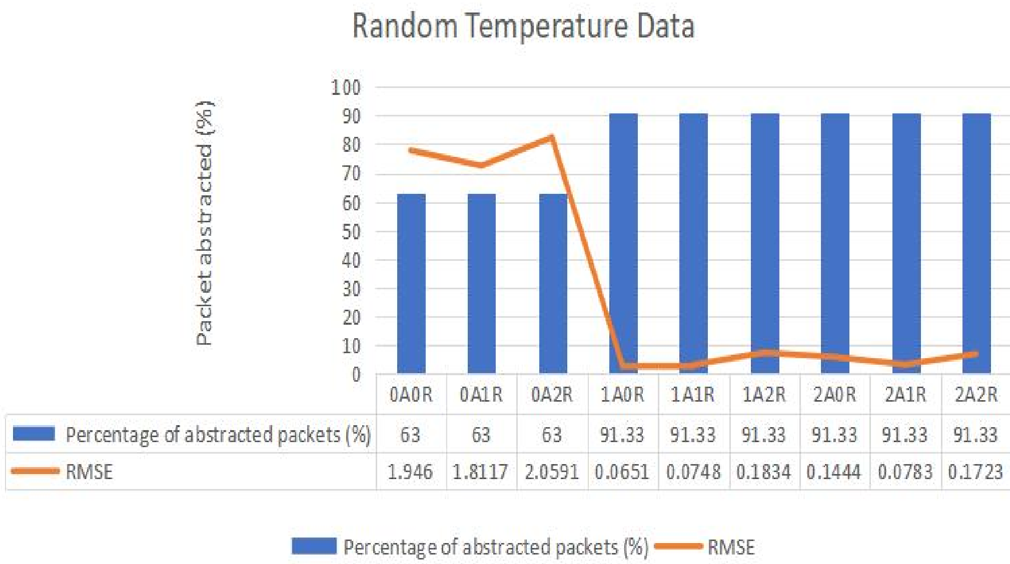


Figure 4.8 Graph results of DAR algorithms of number of data transmitted (%) v.s. Root Mean Square Error in random temperature data.

Figure 4.8 shows that the first-order data abstraction and zeroth-order data reformation can reform random temperature data with $RMSE = 0.065$ by sending 91.33 percent of total data. Furthermore, first-order data abstraction and first-order data reformation, as well as second-order data abstraction and first-order data reformation, can reform random temperature data with $RMSE \leq 0.08$.

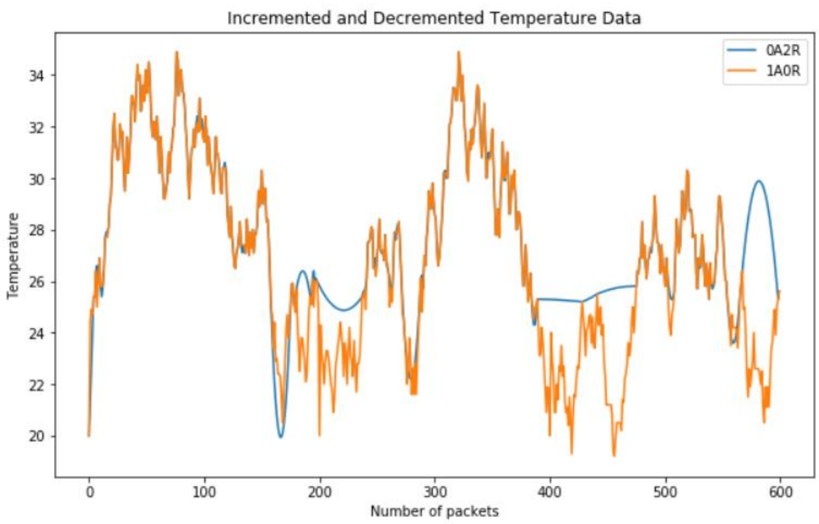
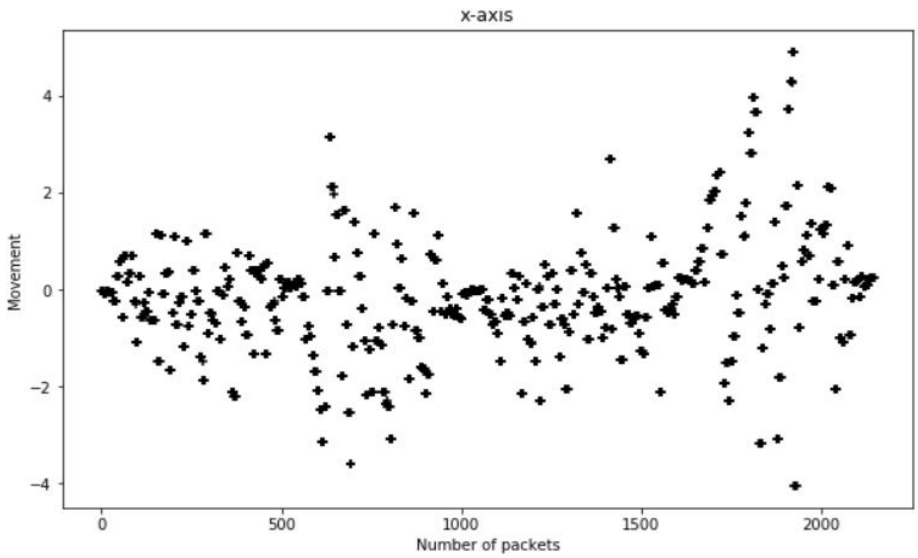


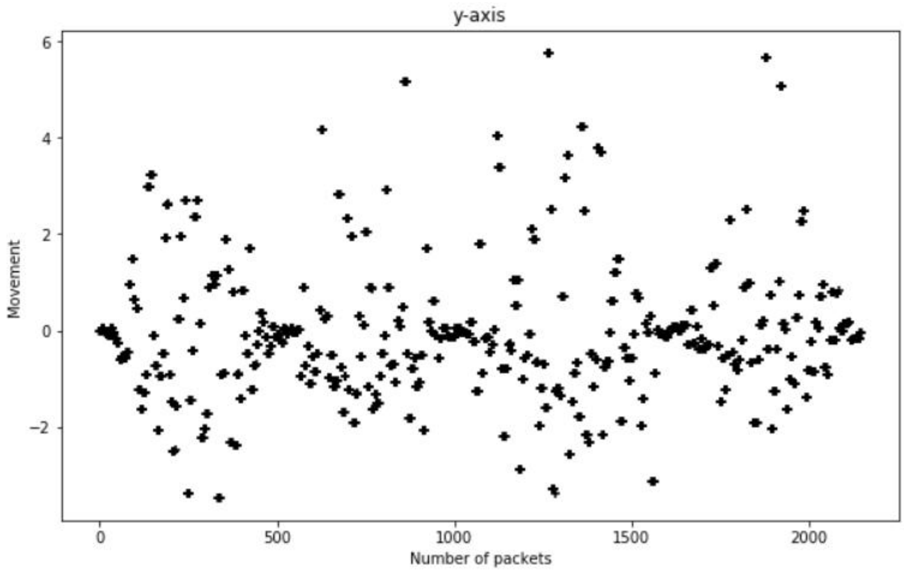
Figure 4.9 Comparison of the greatest value of RMSE and the worst value of RMSE in random temperature data.

4.1.2 Application/attribute with abrupt changing rate

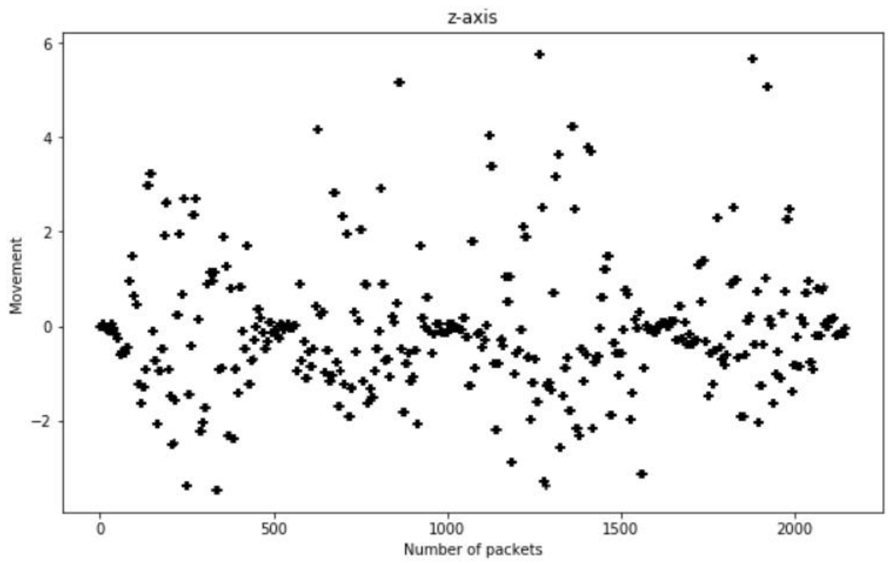
This section demonstrates how DAR algorithms function for movement tracking, where movement has a somewhat abrupt change rate in general. In the experiment, a sensor node with an accelerometer sensor was used to collect data and 2148 of the sensor samples are displayed. The accelerometer captures the movement of an object in three dimensions, namely the x-axis (left and right), y-axis (up and down), and z-axis (forward and backwards). The object was a human walking forward, right, backward, and left, and the accelerometer utilised is the LSM303D Accelerometer [8]. The following are the outcomes of the experiment.



(a)(i)



(a)(ii)



(a)(iii)

Figure 4.10 (a) Graph results of walking accelerometer data x – axis, y – axis, and z axis.

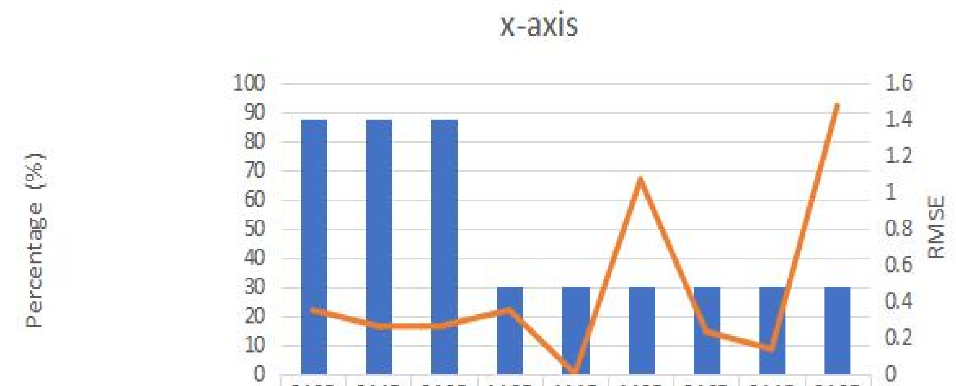
4.1.2.1 Walking accelerometer data

The sensor node is tested under different walking movement designated as shown in Table 4.1.

Table 4.1 Packet ID in walking movements.

Packet ID	Movement
1 – 519	Walking straight
520 – 1004	Turn right
1005 – 1603	Walking backward
1604 – 2149	Turn left

The sensor node sends seated accelerometer data one at a time, with a total of 2148 data sent (100 percent). However, in order to save energy, the amount of data sent might be significantly reduced. The minimal number of packet transmissions in zeroth- and first-order data abstraction is 2, while the lowest number of packet transmissions in second-order data abstraction is 3.

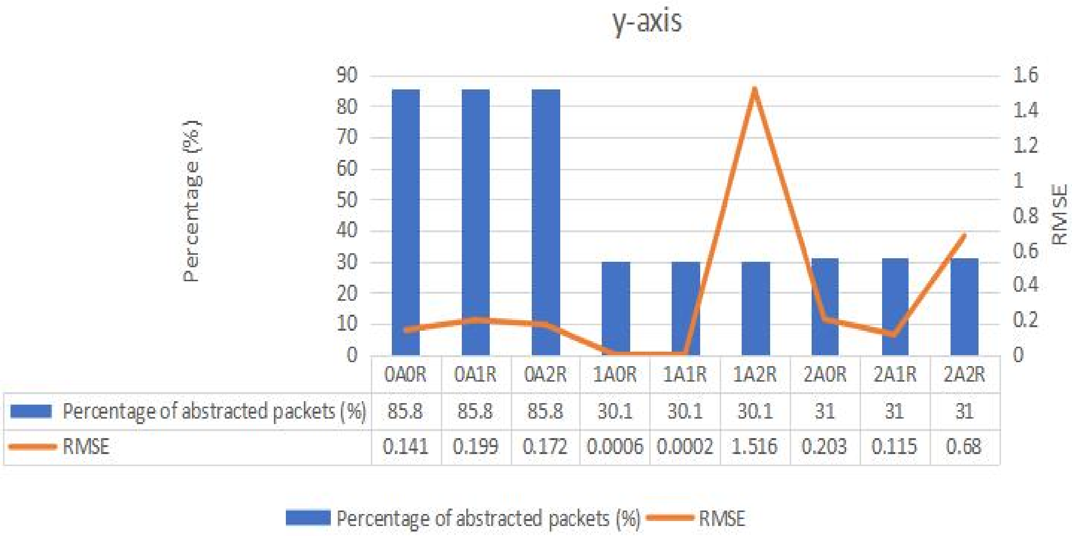


	0A0R	0A1R	0A2R	1A0R	1A1R	1A2R	2A0R	2A1R	2A2R
Percentage of abstracted packets (%)	87.4	87.4	87.4	30.3	30.3	30.3	30.3	30.3	30.3
RMSE	0.349	0.26	0.265	0.349	0.0003	1.066	0.231	0.135	1.47

(a) (i)

Figure 4.11 (a) Graph results of DAR algorithms of number of data abstracted (%) v.s. Root Mean Square Error in walking accelerometer data.

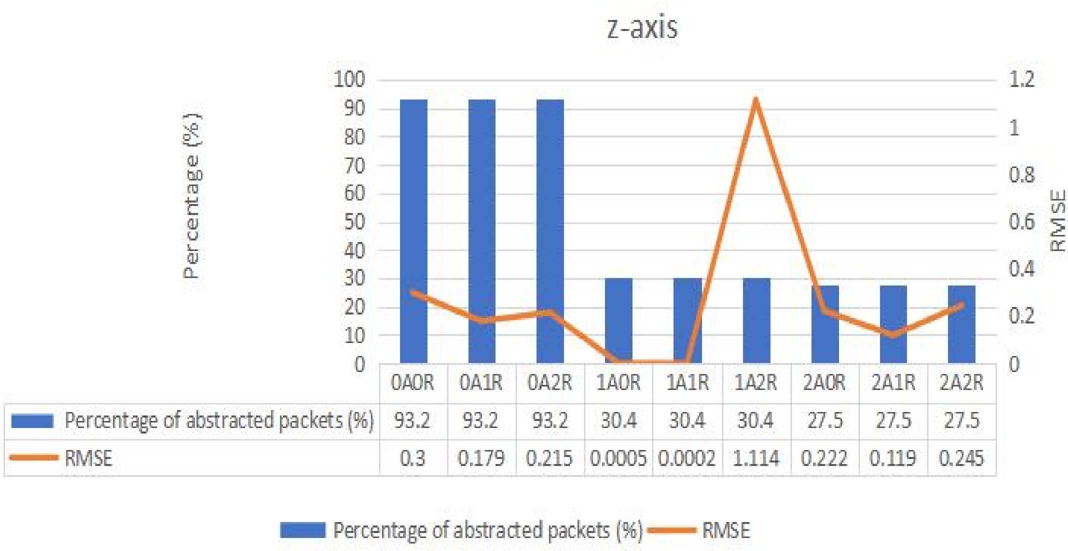
Figure 4.11 (a)(i) shows that by transmitting 30.3 percent of total data, first-order data abstraction and first-order data reformation were able to reform the data in the x-axis with $RMSE = 0.0003$ around 0. Second-order data abstraction and zeroth-order data reformation; and second-order abstraction and first-order reformation are able to reform the data with $RMSE \leq 0.3$ by delivering 30.3 percent of total data.



(a)(ii)

Figure 4.11 (a) Graph results of DAR algorithms of number of data abstracted (%) v.s. Root Mean Square Error in walking accelerometer data.

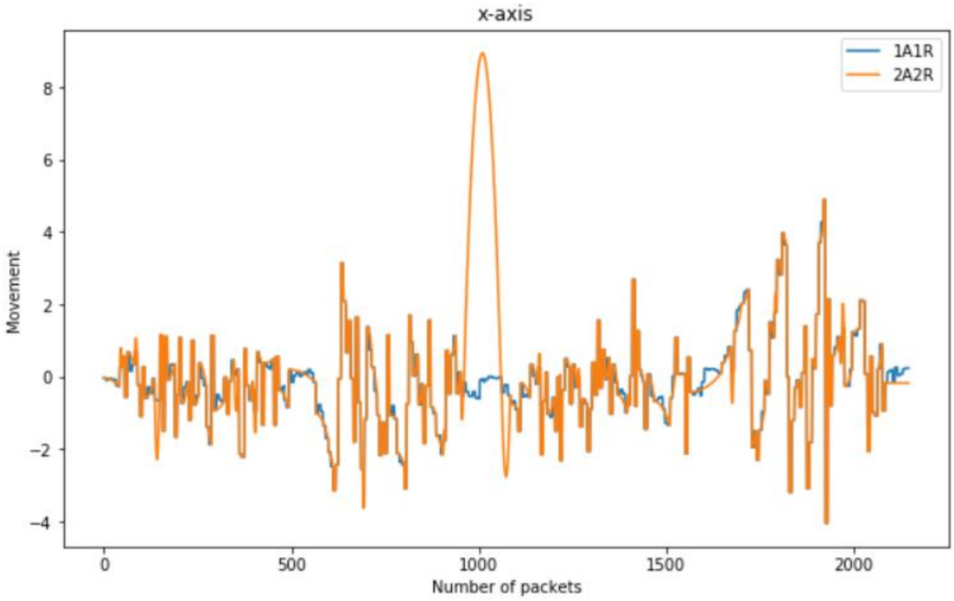
Figure 4.11 (a)(ii) shows that the first-order data abstraction and first-order data reformation were able to reform the data in the y-axis with $RMSE = 0.0002$ by delivering 30.1 percent of total data. Furthermore, first-order data abstraction and zeroth-order data reformation can reform data in the y-axis with an RMSE that is closer to 1A1R's result. The second-order data abstraction and zeroth-order data reformation, as well as the second-order data abstraction and first-order data reformation, were able to reform the data with $RMSE \leq 0.2$.



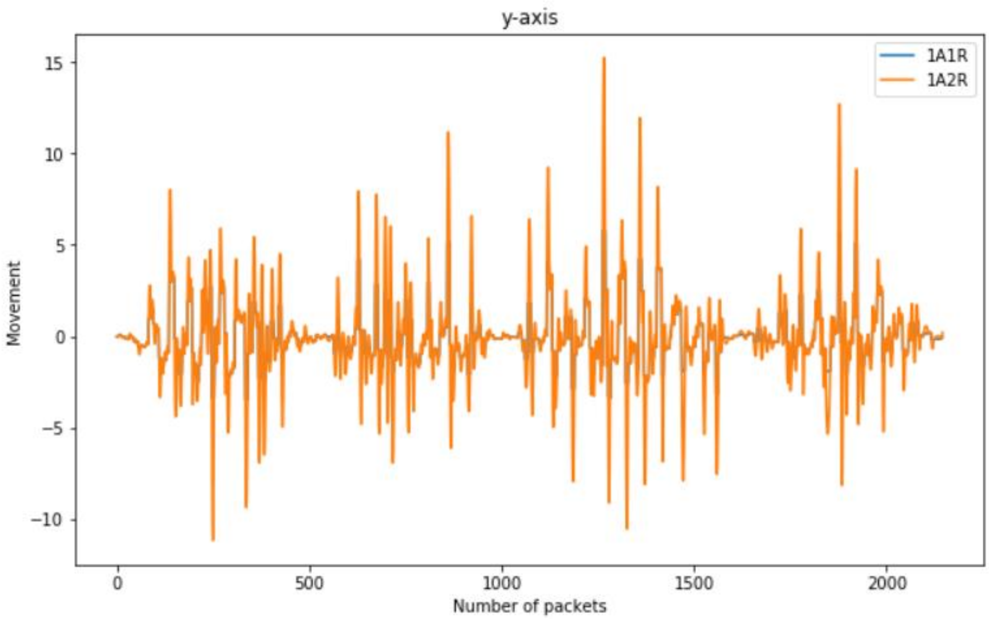
(a)(iii)

Figure 4.11 (a) Graph results of DAR algorithms of number of data abstracted (%) v.s. Root Mean Square Error in walking accelerometer data.

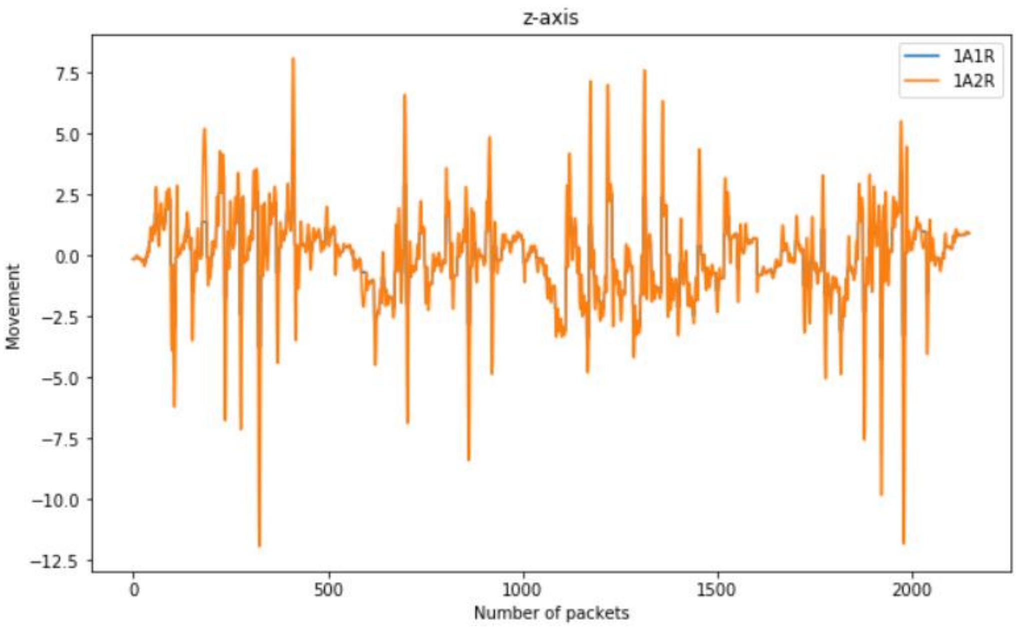
It can be shown in Figure 4.11 (a)(iii) that by sending 30.4 percent of total data, first-order data abstraction and first-order data reformation were able to reform the data in the z-axis with $RMSE = 0.0002$. Furthermore, first-order data abstraction and zeroth-order data reformation can reform data in the z-axis with $RMSE = 0.0005$ that is closer to 1A1R's result. Second-order data abstraction and zeroth-order data reformation; second-order data abstraction and first-order data reformation; and second-order data abstraction and second-order data reformation are able to reform the data in the z-axis with $RMSE \leq 0.25$ by delivering 27.5 percent of total data.


x-axis
1A1R
2A2R
8
6
4
2
0
−2
−4
Movement
0
500
1000
1500
2000
Number of packets

(a)(i)


y-axis
1A1R
1A2R
15
10
5
0
−5
−10
Movement
0
500
1000
1500
2000
Number of packets

(a)(ii)



(a)(iii)

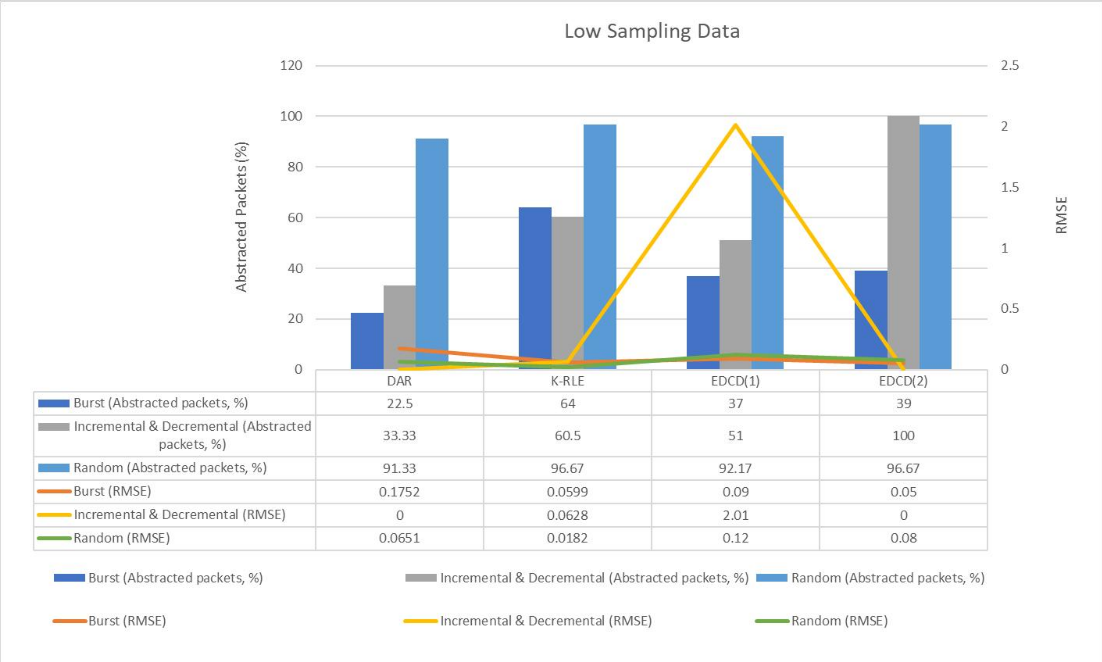
Figure 4.12 (a) Graph results of DAR algorithms of packet transmissions v.s. Root Mean Square Error in walking accelerometer data.

4.2 Performance Metrics and Benchmarking

To verify the effectiveness of our DAR algorithm, the performance of the DAR algorithm is compared to K-RLE [9] and EDCD [10] algorithms. The algorithms will be tested based on two different datasets which are low sampling data with 600 data values and high sampling data with 2147 data values. The low sampling data can be categorized into three categories such as burst, incremental and random. Besides, the high sampling data is the movement of the object in three dimensions which are x-axis, y-axis and z-axis captured by accelerometer.

Furthermore, with low sampling data and high sampling data, the thresholds determined by each algorithms are different. In order to get a better comparison, the EDCD algorithm is tested with two alternative thresholds, EDCD(1) and EDCD(2), in both low and high sampling data. EDCD(1) has a greater threshold, while EDCD(2) has a lower threshold. This is due to the fact that the EDCD algorithm's result is quite small, which may result in only one data value being abstracted. As a result of this condition, it is impossible to reform unknown data and establish a performance comparison. Besides, in order to ensure the experiment's fairness, the EDCD algorithm applied the DAR method's first-order reformation to reform the unknown data.

In three categories of low sampling data, the threshold is set to 0.1 in DAR and K-RLE algorithms. The threshold for the EDCD(1) algorithm is 0.001 and the threshold for the EDCD(2) algorithm is 0.0001. Moreover, in high sampling data, the threshold is set to 0.01 in DAR, K-RLE and EDCD(1) algorithms. The threshold for the EDCD(2) algorithm is 0.0001.



	DAR	K-RLE	EDCD(1)	EDCD(2)
Burst (Abstracted packets, %)	22.5	64	37	39
Incremental & Decremental (Abstracted packets, %)	33.33	60.5	51	100
Random (Abstracted packets, %)	91.33	96.67	92.17	96.67
Burst (RMSE)	0.1752	0.0599	0.09	0.05
Incremental & Decremental (RMSE)	0	0.0628	2.01	0
Random (RMSE)	0.0651	0.0182	0.12	0.08

Figure 4.13 Comparison of the greatest result between DAR, K-RLE and EDCD in low sampling data.

Figure 4.13 shows that DAR transmits the least percentage of total data and able to reform the data in high accuracy with the lower RMSE. In burst, DAR in 1A0R abstracted 22.5% of total data that is lesser than EDCD(1) and EDCD(2). The K-RLE abstracted the highest percentage of total data with 64%. However, DAR can reform the data in low accuracy with the RMSE = 0.1752 compared to K-RLE and EDCD. The DAR method in 2A1R then abstracted just 33.33% of total data in incremental and decremental, which is lesser than K-RLE and EDCD(1), and was able to reform the data with RMSE = 0. Since the EDCD(2) abstracted 100% of the whole data, therefore the RMSE = 0. According to random datasets, the DAR in 1A0R still abstracted the least percentage of total data that is 91.33%. The K-RLE and EDCD(2) abstracted the highest percentage of total data with 96.67%. DAR is able to reform the data with RMSE = 0.0651 that is lesser than EDCD. In comparison to K-RLE and EDCD, the DAR algorithm is able to reduce the number of transmissions. Furthermore, the DAR can reform the unknown data with low RMSE in incremental and decremental datasets as well as random datasets.

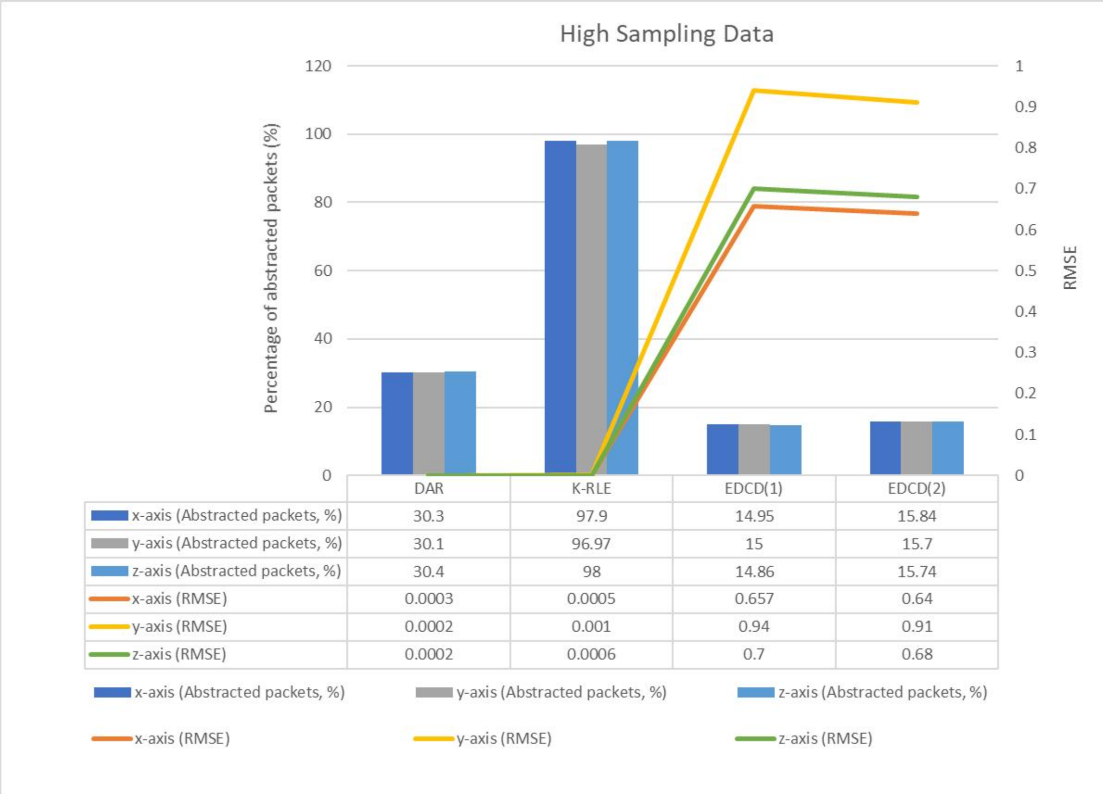


Figure 4.14 Comparison of the greatest result between DAR, K-RLE and EDCD in high sampling data.

Figure 4.14 shows that DAR in 1A1R transmitted the least percentage of total data and able to reform the data in high accuracy with the lower RMSE. In x-axis, DAR abstracted 30.3% of total data that is lesser than K-RLE. The EDCD(1) abstracted 14.95% and EDCD(2) abstracted 15.84% of total data. However, DAR reformed the data in high accuracy with the RMSE = 0.0003. The DAR method then abstracted just 30.1% of total data in y-axis and was able to reform the data with RMSE = 0.0002. Since the K-RLE abstracted 96.97% of the total data, but the RMSE = 0.001 that is higher than DAR. According to z-axis, the DAR abstracted the 30.4% of total data. The K-RLE abstracted the highest percentage of total data with 98%. DAR is able to reform the data with RMSE = 0.0002 that is lesser than K-RLE and EDCD. In comparison to K-RLE and EDCD, the DAR is able to reduce the number of transmissions and reform the unknown data with high accuracy in high sampling data.

4.3 Project Challenges

Setting a sufficient threshold to abstract the significant data is the most difficult part of this endeavour. According to DAR algorithm, when th_{rate} increases, the number of packet transmissions decreases in zeroth-order data abstraction. In addition, when th_{rate} increases, the number of packet transmissions decreases in first-order data abstraction. However, there is a chance that the number of packet transmissions will increase when th_{rate} as well if there is any $\Delta y_i = th_{rate}$ and $\Delta y_i < th_{rate+1}$. Furthermore, when th_{accel} is increased in second-order data abstraction, the number of packet transmissions generally decreases. However, there's a chance that when th_{accel} grows, the number of packet transmissions grows as well when th_{accel} is increasing when there is any $\Delta^2 y_i = th_{accel}$ and $\Delta^2 y_i < th_{accel+1}$.

The mean value of the data is used as the threshold in zeroth-order abstraction to identify the best appropriate threshold. If the abstracted data represents less than half of the entire data, the threshold will be increased or decreased and the test will be repeated. Furthermore, for the first-order abstraction and the second-order abstraction, the difference between the currently collected value of data and the previously obtained value of data is around 0.5. As a result, start by setting the threshold to 0.5. The threshold is then increased or decreased by 0.1 increments until an appropriate threshold is discovered.

Besides, in order to have worth information at the base station, the reformation schemes need to be applied. In temperature monitoring, there is merely a temperature value. Therefore, the sensor data can be directly compared with the reformed data. Note that $RMSE$ is generally increasing when th is increasing. However, $RMSE$ is decreasing when th is increasing if the minimum and maximum values are reported before the sign of the slope changes.

Aside from that, employing the EDCD algorithm to set a sufficient threshold to abstract the significant data was also one of the hurdles overcome. This is due to the fact that the EDCD algorithm's result is quite small, which may result in only one data value being abstracted. As a result, if the threshold number is set too high, only a little amount of data will be abstracted. Otherwise, if the threshold is set too low, the EDCD algorithm may be fine-tuned to provide the optimum results. According to this, the EDCD algorithm will be tested using 2 thresholds which are 0.001 and 0.0001 in low sampling data and thus 0.01 and 0.0001 in high sampling data. Note that RMSE is increasing when the threshold value is increasing. Nevertheless, the RMSE is decreasing when the threshold is decreasing.

4.4 Objectives Evaluation

From the objectives in section 1.2, which are:

1. The DAR algorithms are able to lower the energy consumption of sensor nodes by lowering energy usage and extending the sensor nodes' lifetime. The objective was achieved by using data abstraction algorithm to filter the significant data and reduce the number of transmissions. The lower the communication frequency, the lower the energy consumption. This objective was achieved and discussed in Chapter 4.

2. The data abstraction algorithm can abstract the significant data in order to limit the quantity of data transmissions. This objective was achieved by using zeroth-order, first-order, and second-order data abstraction techniques, which are addressed in Chapter 4.

3. The data reformation algorithm can reconstruct missing data values into a complete full set of data without sacrificing data accuracy. This objective was achieved by using zeroth-order, first-order, and second-order data reformation techniques, which are addressed in Chapter 4.

4.5 Concluding Remarks

In conclusion, experiments were carried out by DAR algorithms. Two studies were carried out on each of the application categories, namely application/attribute with low sampling rate and application/attribute with abrupt sampling rate. In general, each of the sensed data will be reported to the base station, and the acquired data accuracy is quite dependable. As a result, Data Abstraction and Reformation (DAR) schemes are presented as a way to reduce the number of data transmission while maintaining the accuracy of the data especially in 1A0R for low sampling data and 1A1R for high sampling data.

Furthermore, comparison to K-RLE and EDCD, the DAR in 1A1R is able to reduce the number of transmissions. Furthermore, the DAR can reform the unknown data with low RMSE in incremental and decremental datasets as well as random datasets. Besides, DAR abstracted approximately 30% of data with the RMSE close to 0, indicating that the data accuracy is quite high.

CHAPTER 5
Conclusion and Recommendation

5.1 Conclusion

In conclusion, the goal of the project suggested in this paper is to tackle the energy limitation problem by lowering energy usage and extending the sensor nodes' lifetime in order to have a better improvement in IoT field. Besides, the problem of energy limitation is able to be resolved by using DAR techniques. With the use of DAR techniques, the amount of data and packets broadcast is successfully minimised and the accuracy of the reformation is within the acceptable range, $0 \leq RMSE \leq 1$. As a result of this research, it has been established that DAR algorithms are suitable for use in both low and abruptly changing data rate applications. The increment of th_{value}, th_{rate} and th_{accel} are generally decreasing the number of data and packet transmissions and increasing the *RMSE*'s value. Overall, first-order data abstraction and zeroth-order data reformation, 1A0R, has better performance in low changing data rates and first-order data abstraction and first-order data reformation, 1A1R, is the best DAR schemes in abruptly changing data rates that is able to reduce the number of data transmission while maintaining the accuracy of the data with *RMSE* close to zero. Therefore, DAR may consistently deliver gratifying outcomes for low and abruptly changing data rates while saving significant energy and improving data reformation accuracy. Furthermore, the first-order data abstraction and first-order data reformation, 1A1R, can be done directly based on the requested *RMSE* value. While second-order data abstraction and first-order data reformation, 2A1R, both lower the number of transmissions, they also necessitate more intricate calculations.

5.2 Future Work

For the future work, it is recommended to improve the data accuracy of the reconstructed data values by adjusting the data reformation algorithm or examining a new algorithm to pair with the data reformation algorithm. In addition, set up a physical test to evaluate the energy-saving capabilities, as well as gather more real-world evidence to support the efficiency and effectiveness of DAR algorithms.

References

[1] I. F. Akyildiz, Weilian Su, Y. Sankarasubramaniam and E. Cayirci, "A survey on sensor networks," in IEEE Communications Magazine, vol. 40, no. 8, pp. 102-114, Aug. 2002, doi: 10.1109/MCOM.2002.1024422.

[2] John, T. Ogbiti, H. C. Ukwuoma, S. Danjuma and M. Ibrahim. "Energy consumption in wireless sensor network," Computer Engineering and Intelligent Systems, Kaduna, Nigeria, 2016, vol. 7, no. 8.

[3] Engmann, F., Katsriku, F. A., Abdulai, J.-D., Adu-Manu, K. S., & Banaseka, F. K. "Prolonging the Lifetime of Wireless Sensor Networks: A Review of Current Techniques", Wireless Communications and Mobile Computing, volume 2018, pp. 1–23, (August 2018).

[4] H. Deng, Z. Guo, R. Lin and H. Zou, "Fog Computing Architecture-Based Data Reduction Scheme for WSN," 2019 1st International Conference on Industrial Artificial Intelligence (IAI), 2019, pp. 1-6, doi: 10.1109/ICIAI.2019.8850817.

[5] L. Meng, H. Zhang and Y. Zou, "A Data Aggregation Transfer Protocol Based on Clustering and Data Prediction in Wireless Sensor Networks," 2011 7th International Conference on Wireless Communications, Networking and Mobile Computing, 2011, pp. 1-5, doi: 10.1109/wicom.2011.6040376.

[6] A. S. Hampiholi and B. P. Vijaya Kumar, "Efficient routing protocol in IoT using modified Genetic algorithm and its comparison with existing protocols," 2018 3rd International Conference on Circuits, Control, Communication and Computing (I4C), 2018, pp. 1-5, doi: 10.1109/CIMCA.2018.8739759.

[7] Goh, H.G., Liew, S.Y., Kwong, K.H., Michie, C., Andonovic, I. (2012). Abstract Reporting and Reformation Schemes for Wireless Sensor Networks. In: Sénac, P., Ott, M., Seneviratne, A. (eds) Wireless Communications and

Applications. ICWCA 2011. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol 72. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-29157-9_7

[8] Pololu.com, (November, 2013). Ultra-compact high-performance eCompass module: 3D accelerometer and 3D magnetometer. [online] Available at: < https://www.pololu.com/file/0J703/LSM303D.pdf> [Accessed 31 August 2021].

[9] M. Roukhami, Y. Lahbib and A. Mami, "A new efficient energy implementation of K-RLE algorithm for WSN," 2016 7th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications (SETIT), 2016, pp. 66-70, doi: 10.1109/SETIT.2016.7939843.

[10] N. A. M. Alduais, J. Abdullah, A. Jamil and L. Audah, "An efficient data collection and dissemination for IOT based WSN," 2016 IEEE 7th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), 2016, pp. 1-6, doi: 10.1109/IEMCON.2016.7746084.

APPENDIX A

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: Y3S3	Study week no.: 2
Student Name & ID: Lim Xin Hui, 18ACB01058	
Supervisor: Ts Dr Goh Hock Guan	
Project Title: ABSTRACT DATA EXTRACTION AND REFORMATION FOR IoT	

1. WORK DONE
[Please write the details of the work done in the last fortnight.] None
2. WORK TO BE DONE
i. Search for 2 - 3 new journals/conference papers published within 5 years.
3. PROBLEMS ENCOUNTERED
i. There are fewer journals where reliable references and authors can be found.
4. SELF EVALUATION OF THE PROGRESS
i. Study many different methods proposed by other researchers to reduce energy consumption.

Supervisor's signature

Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: Y3S3	Study week no.: 4
Student Name & ID: Lim Xin Hui, 18ACB01058	
Supervisor: Ts Dr Goh Hock Guan	
Project Title: ABSTRACT DATA EXTRACTION AND REFORMATION FOR IoT	

1. WORK DONE

[Please write the details of the work done in the last fortnight.]

i. Found 3 papers and sent to supervisor to check if the found papers can be used for benchmarking. At last, K-RLE and TBDRT papers are used for benchmarking.

2. WORK TO BE DONE

i. Start to code the algorithms proposed by the new papers that I found.
ii. Test both low and high sampled data values using the algorithm proposed in the new paper.

3. PROBLEMS ENCOUNTERED

i. Unfamiliar with the algorithms of the new papers.

4. SELF EVALUATION OF THE PROGRESS

ii. Study different methods proposed by other researchers to abstract significant data in order to reduce energy consumption.

Supervisor's signature

Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: Y3S3	Study week no.: 6
Student Name & ID: Lim Xin Hui, 18ACB01058	
Supervisor: Ts Dr Goh Hock Guan	
Project Title: ABSTRACT DATA EXTRACTION AND REFORMATION FOR IoT	

1. WORK DONE

[Please write the details of the work done in the last fortnight.]

i. Tested high sampling data values by using K-RLE algorithm.
ii. Plotted the greatest and the worst RMSE graph for high sampling data.
iii. Coded K-RLE algorithm to abstract high sampling data.

2. WORK TO BE DONE

i. Need to find one new journal/conference paper. Because the logic of the TBDRT algorithm I found before is the same as the first-level data abstraction algorithm in my own paper.
ii. Start to code the algorithms proposed by the new paper.
iii. Test both low and high sampled data values using the algorithm proposed in the new paper.

3. PROBLEMS ENCOUNTERED

i. There are fewer journals where reliable references and authors can be found.

4. SELF EVALUATION OF THE PROGRESS

i. Learn about K-RLE algorithm and familiar with it.

Supervisor's signature

Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: Y3S3	**Study week no.: 8**
Student Name & ID: Lim Xin Hui, 18ACB01058	
Supervisor: Ts Dr Goh Hock Guan	
Project Title: ABSTRACT DATA EXTRACTION AND REFORMATION FOR IoT	

1. WORK DONE

[Please write the details of the work done in the last fortnight.]

i. Found new paper by using EDCD algorithm to abstract significant data.
ii. Tested high sampling data values by using EDCD algorithm.
iii. Plotted the greatest and the worst RMSE graph for high sampling data.
iv. Coded EDCD algorithm to abstract high sampling data.

2. WORK TO BE DONE

i. Prepare FYP2 report.

3. PROBLEMS ENCOUNTERED

None

4. SELF EVALUATION OF THE PROGRESS

i. After testing the high sampling data, first order abstraction and first order reformation is the best model to get the lowest RMSE. Besides, first order abstraction and zeroth order reformation got the highest RMSE that is the worst model.
ii. Learn about EDCD algorithm and familiar with it.

Supervisor's signature

Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: Y3S3	Study week no.: 10
Student Name & ID: Lim Xin Hui, 18ACB01058	
Supervisor: Ts Dr Goh Hock Guan	
Project Title: ABSTRACT DATA EXTRACTION AND REFORMATION FOR IoT	

1. WORK DONE

[Please write the details of the work done in the last fortnight.]

i. Completed FYP2 report Chapter 1, 2, 3.

2. WORK TO BE DONE

i. Prepare FYP2 report Chapter 4, 5.

3. PROBLEMS ENCOUNTERED

None

4. SELF EVALUATION OF THE PROGRESS

i. Have a better understanding of the project while doing the report.

Supervisor's signature

Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: Y3S3	Study week no.: 12
Student Name & ID: Lim Xin Hui, 18ACB01058	
Supervisor: Ts Dr Goh Hock Guan	
Project Title: ABSTRACT DATA EXTRACTION AND REFORMATION FOR IoT	

1. WORK DONE

[Please write the details of the work done in the last fortnight.]

i. Completed the report.

2. WORK TO BE DONE

i. Submit the report.
ii. Prepare the presentation.

3. PROBLEMS ENCOUNTERED

None

4. SELF EVALUATION OF THE PROGRESS

i. Acquire knowledge about DAR algorithms and make better improvements in the field of Internet of Things (IoT).

Supervisor's signature

Student's signature

POSTER

ABSTRACT DATA EXTRACTION AND REFORMATION FOR IOT

DAR Algorithms

AUTHORS
Lim Xin Hui

SUPERVISOR
Ts Dr
Goh Hock Guan

INTRODUCTION

The deployment of Wireless Sensor Network (WSN) for the environment or movement monitoring has tremendously increased in recent years. In general, a WSN monitoring system requires a large number of wireless sensor nodes such as ordinary sensor nodes, cluster head nodes and sink nodes to be deployed over a field that is to be monitored.

OBJECTIVES

- Lower the energy consumption of sensor nodes
- Reduce the amount of sensor node transmissions
- Increase data accuracy

PROBLEM STATEMENT

Energy limitation of the sensor node powered by batteries

METHODOLOGY

Data Abstraction and Reformation (DAR)
There are three sorts of orders:
Zeroth, First and Second orders.

RESULTS/FINDINGS

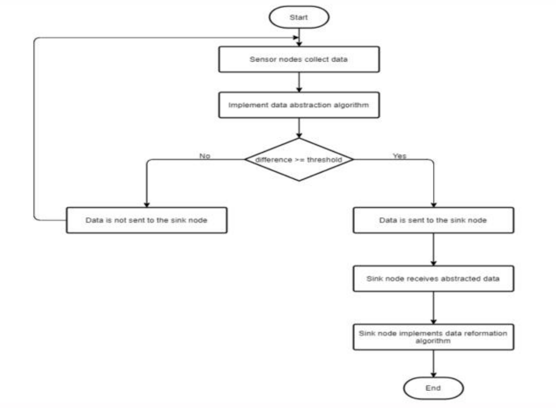
Low Temperature Sampling Data:

- Burst : 1A0R the best
- Increment & Decrement: 2A1R the best
- Random: 1A0R the best

High Sampling Data:

- 1A1R model is the greatest
- 1A0R model is the worst

PROCEDURE

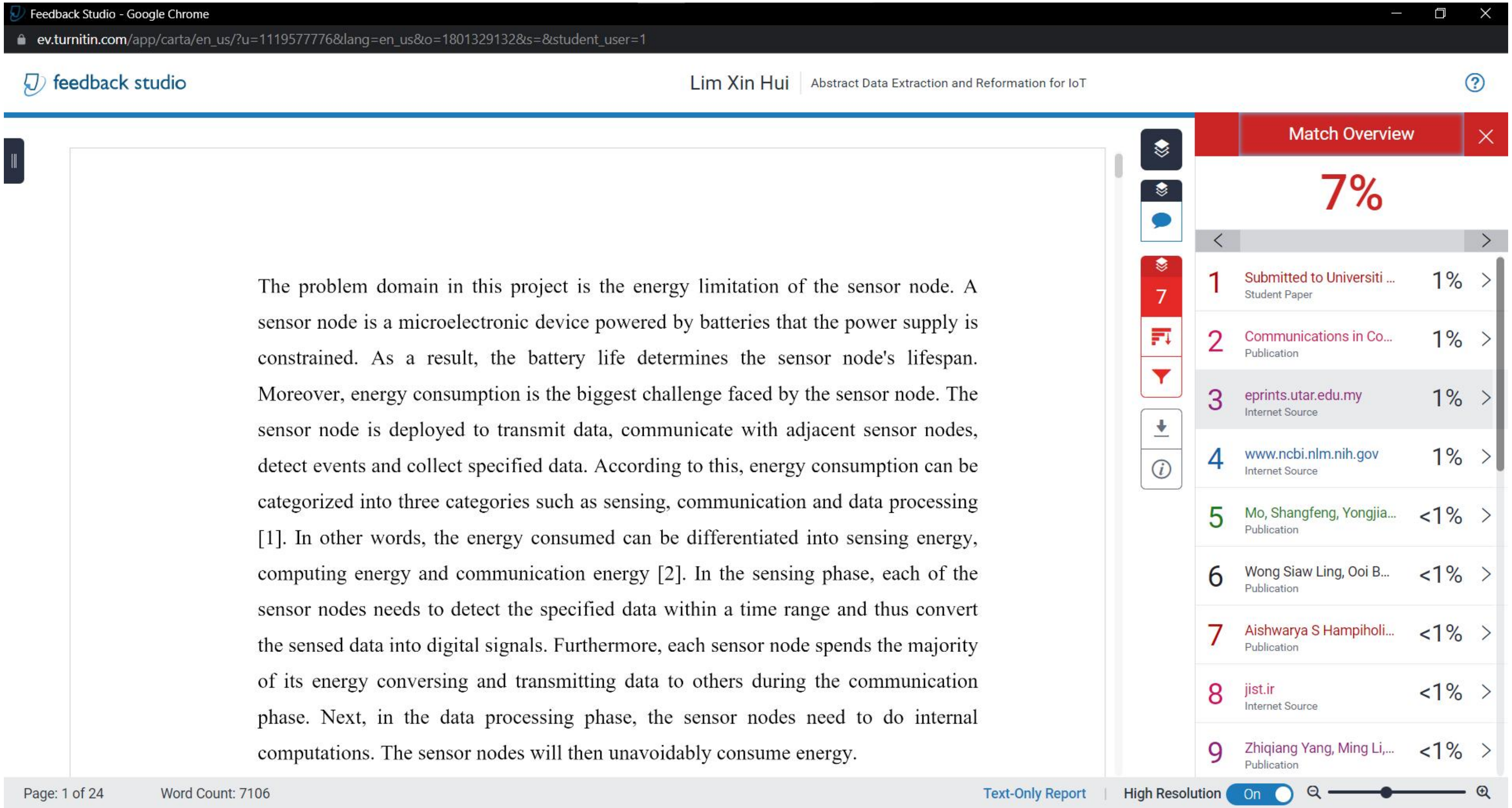


CONCLUSION

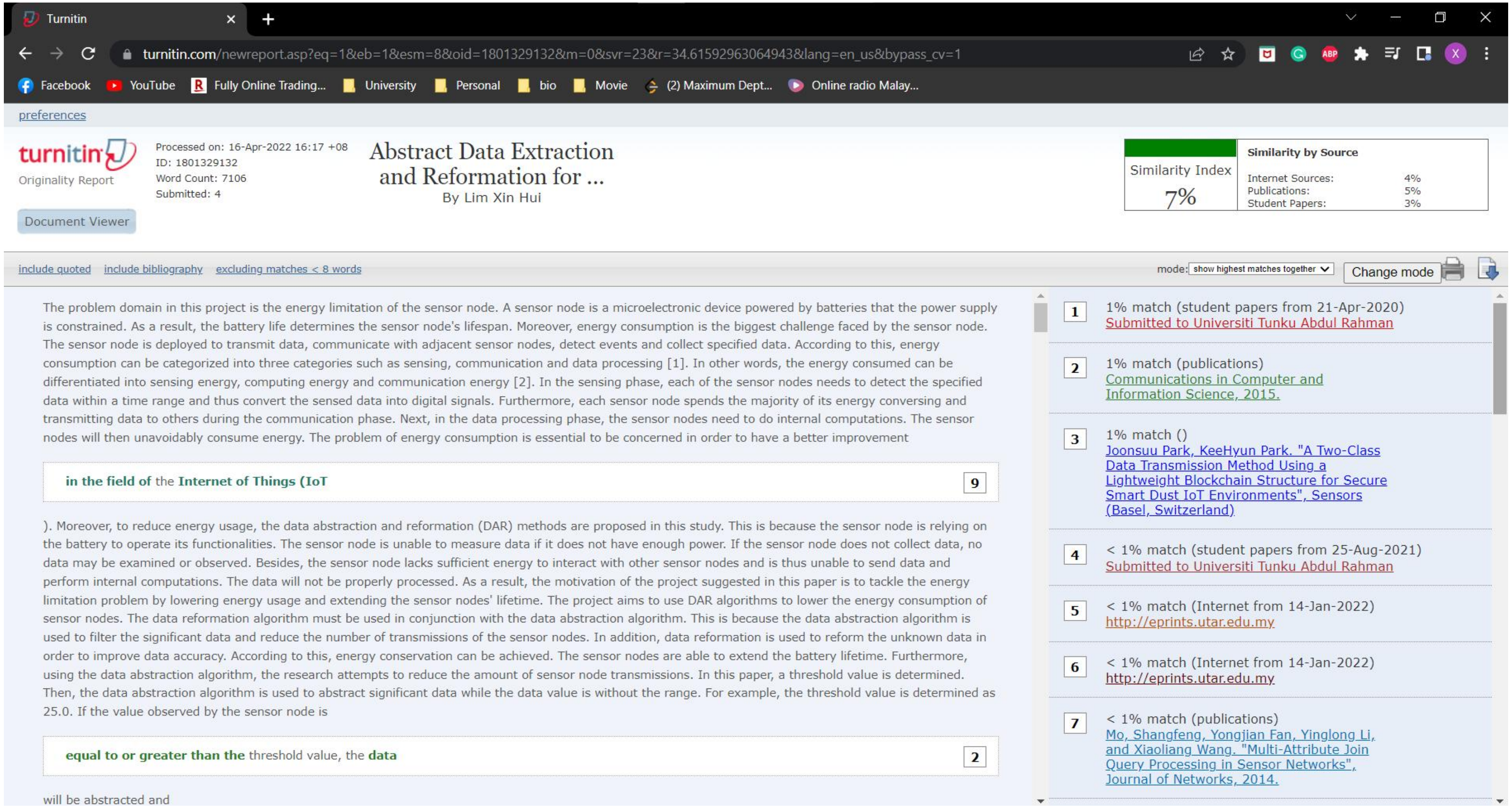
Through the study, it was found that the number of packet transmissions can be greatly reduced with a proper selection of DAR algorithms, and yet the data can be reformed at the base station with acceptable accuracy.

APPENDIX B

PLAGIARISM CHECK RESULT



APPENDIX B


Turnitin

turnitin.com/newreport.asp?eq=1&eb=1&esm=8&oid=1801329132&m=0&svr=23&r=34.61592963064943&lang=en_us&bypass_cv=1

Facebook YouTube Fully Online Trading... University Personal bio Movie (2) Maximum Dept... Online radio Malay...

preferences

turnitin Originality Report

Processed on: 16-Apr-2022 16:17 +08
ID: 1801329132
Word Count: 7106
Submitted: 4

Abstract Data Extraction and Reformation for ...
By Lim Xin Hui

Document Viewer

Similarity Index	Similarity by Source	
7%	Internet Sources:	4%
	Publications:	5%
	Student Papers:	3%

include quoted include bibliography excluding matches < 8 words

mode: show highest matches together Change mode

The problem domain in this project is the energy limitation of the sensor node. A sensor node is a microelectronic device powered by batteries that the power supply is constrained. As a result, the battery life determines the sensor node's lifespan. Moreover, energy consumption is the biggest challenge faced by the sensor node. The sensor node is deployed to transmit data, communicate with adjacent sensor nodes, detect events and collect specified data. According to this, energy consumption can be categorized into three categories such as sensing, communication and data processing [1]. In other words, the energy consumed can be differentiated into sensing energy, computing energy and communication energy [2]. In the sensing phase, each of the sensor nodes needs to detect the specified data within a time range and thus convert the sensed data into digital signals. Furthermore, each sensor node spends the majority of its energy conversing and transmitting data to others during the communication phase. Next, in the data processing phase, the sensor nodes need to do internal computations. The sensor nodes will then unavoidably consume energy. The problem of energy consumption is essential to be concerned in order to have a better improvement

in the field of the Internet of Things (IoT 9

). Moreover, to reduce energy usage, the data abstraction and reformation (DAR) methods are proposed in this study. This is because the sensor node is relying on the battery to operate its functionalities. The sensor node is unable to measure data if it does not have enough power. If the sensor node does not collect data, no data may be examined or observed. Besides, the sensor node lacks sufficient energy to interact with other sensor nodes and is thus unable to send data and perform internal computations. The data will not be properly processed. As a result, the motivation of the project suggested in this paper is to tackle the energy limitation problem by lowering energy usage and extending the sensor nodes' lifetime. The project aims to use DAR algorithms to lower the energy consumption of sensor nodes. The data reformation algorithm must be used in conjunction with the data abstraction algorithm. This is because the data abstraction algorithm is used to filter the significant data and reduce the number of transmissions of the sensor nodes. In addition, data reformation is used to reform the unknown data in order to improve data accuracy. According to this, energy conservation can be achieved. The sensor nodes are able to extend the battery lifetime. Furthermore, using the data abstraction algorithm, the research attempts to reduce the amount of sensor node transmissions. In this paper, a threshold value is determined. Then, the data abstraction algorithm is used to abstract significant data while the data value is without the range. For example, the threshold value is determined as 25.0. If the value observed by the sensor node is

equal to or greater than the threshold value, the data 2

will be abstracted and

1 1% match (student papers from 21-Apr-2020)
Submitted to Universiti Tunku Abdul Rahman

2 1% match (publications)
Communications in Computer and Information Science, 2015.

3 1% match ()
Joonsuu Park, KeeHyun Park. "A Two-Class Data Transmission Method Using a Lightweight Blockchain Structure for Secure Smart Dust IoT Environments", Sensors (Basel, Switzerland)

4 < 1% match (student papers from 25-Aug-2021)
Submitted to Universiti Tunku Abdul Rahman

5 < 1% match (Internet from 14-Jan-2022)
http://eprints.utar.edu.my

6 < 1% match (Internet from 14-Jan-2022)
http://eprints.utar.edu.my

7 < 1% match (publications)
Mo, Shangfeng, Yongjian Fan, Yinglong Li, and Xiaoliang Wang. "Multi-Attribute Join Query Processing in Sensor Networks", Journal of Networks, 2014.

APPENDIX B

Universiti Tunku Abdul Rahman			
Form Title : Supervisor's Comments on Originality Report Generated by Turnitin for Submission of Final Year Project Report (for Undergraduate Programmes)			
Form Number: FM-IAD-005	Rev No.: 0	Effective Date: 01/10/2013	Page No.: 1of 1

FACULTY OF INFORMATION AND COMMUNICATION TECHNOLOGY

Full Name(s) of Candidate(s)	Lim Xin Hui
ID Number(s)	1801058
Programme / Course	CS
Title of Final Year Project	Abstract Data Extraction and Reformation for IoT

Similarity	**Supervisor's Comments (Compulsory if parameters of originality exceeds the limits approved by UTAR)**
Overall similarity index: 7 **%** **Similarity by source** Internet Sources: 4 % Publications: 5 % Student Papers: 3 %	
Number of individual sources listed of more than 3% similarity: 0	
Parameters of originality required and limits approved by UTAR are as Follows: **(i) Overall similarity index is 20% and below, and** **(ii) Matching of individual sources listed must be less than 3% each, and** **(iii) Matching texts in continuous block must not exceed 8 words** ***Note: Parameters (i) – (ii) shall exclude quotes, bibliography and text matches which are less than 8 words.***	

Note Supervisor/Candidate(s) is/are required to provide softcopy of full set of the originality report to Faculty/Institute

Based on the above results, I hereby declare that I am satisfied with the originality of the Final Year Project Report submitted by my student(s) as named above.

Signature of Supervisor	Signature of Co-Supervisor
Name: Goh Hock Guan	Name:
Date: 21/4/2022	Date:

UNIVERSITI TUNKU ABDUL RAHMAN

FACULTY OF INFORMATION & COMMUNICATION TECHNOLOGY (KAMPAR CAMPUS)

CHECKLIST FOR FYP2 THESIS SUBMISSION

Student Id	18ACB01058
Student Name	Lim Xin Hui
Supervisor Name	Ts Dr Goh Hock Guan

TICK (√)	**DOCUMENT ITEMS** Your report must include all the items below. Put a tick on the left column after you have checked your report with respect to the corresponding item.
-	Front Plastic Cover (for hardcopy)
√	Title Page
√	Signed Report Status Declaration Form
√	Signed FYP Thesis Submission Form
√	Signed form of the Declaration of Originality
√	Acknowledgement
√	Abstract
√	Table of Contents
√	List of Figures (if applicable)
√	List of Tables (if applicable)
-	List of Symbols (if applicable)
√	List of Abbreviations (if applicable)
√	Chapters / Content
√	Bibliography (or References)
√	All references in bibliography are cited in the thesis, especially in the chapter of literature review
√	Appendices (if applicable)
√	Weekly Log
√	Poster
√	Signed Turnitin Report (Plagiarism Check Result - Form Number: FM-IAD-005)
√	I agree 5 marks will be deducted due to incorrect format, declare wrongly the ticked of these items, and/or any dispute happening for these items in this report.

*Include this form (checklist) in the thesis (Bind together as the last page)

I, the author, have checked and confirmed all the items listed in the table are included in my report.

____

(Signature of Student)
Date: 16/04/2022

APPENDIX C