University of Cape Town

EEE3097S

Engineering Design: Electrical and Computer Engineering

Final Report

Report Authors:

Thomas Clegg, CLGTHO001

Kudzayi Samakande, SMKKUD001

Content

Admin Documents

Division of Work

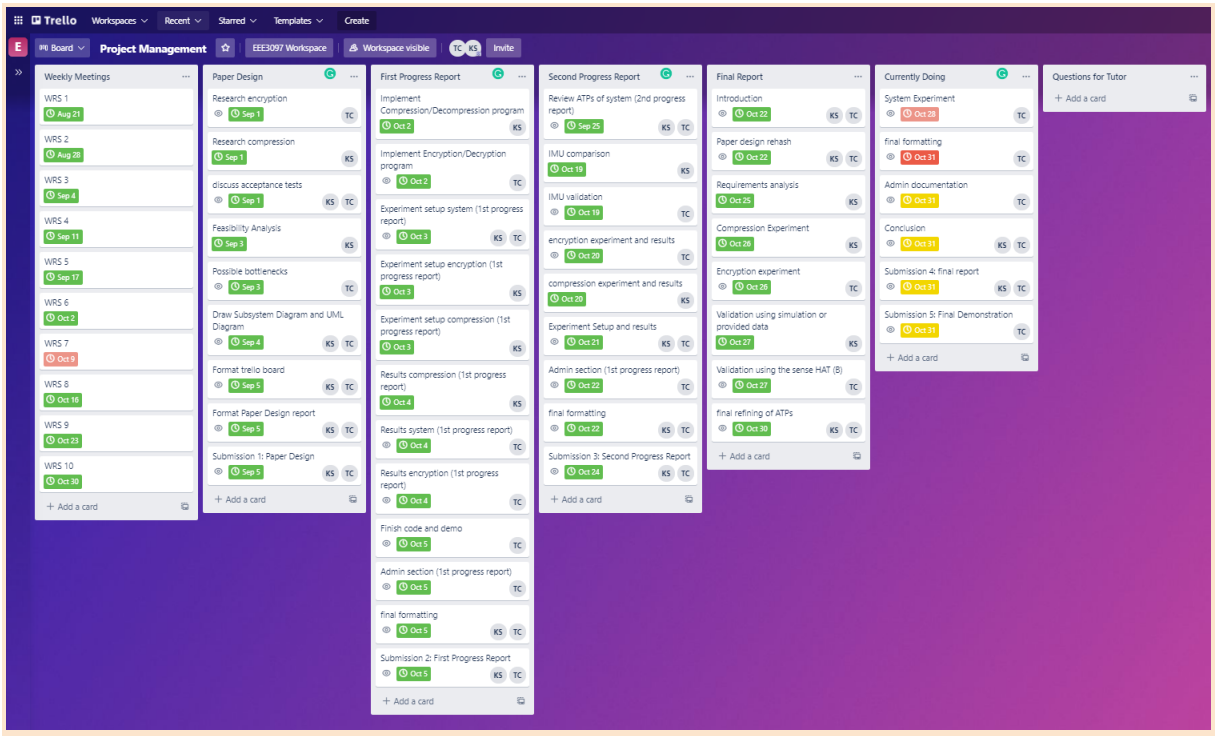
Task		Member	Page Numbers
Project management, division of work, formatting		Thomas	5-6
Introduction		Kudzayi	7
Requirements analysis		Kudzayi	9
Paper design	Requirements analysis	Both	11
	Compression comparison	Kudzayi	10-11
	Encryption comparison	Thomas	11
	Feasibility analysis	Kudzayi	12
	Possible bottlenecks	Thomas	12
	UML diagrams	Both	12-13
	ATPs	Both	13-14
Validation using sensor data	Choice, justification and analysis of data	Kudzayi	15
	Compression subsystem	Kudzayi	19
	Encryption subsystem	Thomas	21
	System	Both	23
Validation using Sense HAT (B)	Feature comparison and justification	Kudzayi	26
	Sensor validation	Thomas	27
	Compression subsystem	Kudzayi	27
	Encryption subsystem	Thomas	28
	System	Thomas	32
ATPs and future plan		Both	36
Conclusion		Kudzayi	39
Demonstration video		Thomas	----------

Github Repository

https://github.com/Tom-Clegg/EEE3097S-2021-Group6

Project Management and Timeline

We used Trello to manage our project and timeline. Below shows our Trello project management board which we used to track our tasks, who's assigned to them and when they are due.



Our timeline worked mostly to plan. Some tasks were delayed a day or two due to workload from other courses we take, but all reports were handed in on time. We missed one WRS

Introduction

There is an expedition that endeavours to study winter conditions in the Southern Ocean and Antarctica's sea ice. This is prompted by the unexpectedly extreme changes to Antarctic sea ice during 2016. Scientists intend to explore how storms affect sea ice. The proposed solution is ice-tethered buoys equipped with several sensors e.g an IMU and is installed on ice pancakes designed to communicate with scientists via satellites.

Testing

Testing is very crucial towards design implementation. We are going to test out algorithms rigorously to make sure they meet all the specifications. A robust IP is needed for this application because it can be costly to rectify mistakes after deployment. Firstly we are going to validate the system by simulating the IMU and then using a different real IMU.

Simulation

For this part we have a choice to use either simulated data or old data from a boat cruise (different from the SHARC Buoy). Simulation based testing is essential in testing our system, we can test extreme conditions which may be difficult to do under real testing. The IMU can be subjected to extreme pressures which if we do with the real IMU we can damage it and it becomes costly.

Requirements Analysis

The task at hand is to design an ARM based digital IP using the Raspberry-Pi to encrypt and compress and subsequently decompress and decrypt the IMU data.

Problem Identification

Sections below give context of the problem and constraints the designers must address and adhere to.

Power Supply

A reliable power supply is needed for a remotely deployed device to extend its functionality. Advantages of robust power supply are increased processing capabilities and duration. It's impossible to power the remote buoy using outside sources hence the need of a portable power source. Batteries are a suitable solution, preferably rechargeable cells[1].

Cost

A compromise must be made between high speed devices and low cost devices. For example, modems of transmission that have higher bandwidth are expensive[1].

Remote Communication

The SHARC buoy uses an iridium satellite network with global coverage. The selected modems for this project offer limited bandwidth.

Specifications and ATPs

System Software RST as a whole are divided into 3 categories namely general constrain, functional requirements and performance requirements

1. General Constrain

Requirements	Specifications	ATP
Perform minimum computations to save power	Use at most 75% of the pi cores	Pi connected to a 1500mAh battery must run for 20hrs

2. Functional requirements

Requirements	Specification	ATP
The system shall digitally sample the movement and orientation of the SHARC BUOY. A proper sensor is	Use ICM-20649 IMU to capture information and an ARM based microcontroller	Subject the system to movement, shaking, orientation and magnetic field and check if the

needed to collect acceleration and orientation status of the BUOY periodically	(raspberry Pi)	measured data is changing accordingly
Retain the lowest 25% Fourier coefficients of the sampled data	Algorithm to perform perform frequency analysis and discard the the upper 95% Fourier coefficients of the sampled data	Compare the sampled data and the output of the system in the frequency domain and check if the output data has the 25% Fourier coefficients
The system must pass compressed and encrypted data to the transmission module	Use compression libraries eg Zlib or GZip Use encryption libraries	Check if the file size to be transmitted is considerably lower than the file size of the sampled data Inspect if the encrypted file if difficult to read/decipher
The transmitted data should be recovered easily	Implement decompression and encryption algorithms	Compare the decrypted and decompressed file with the original file. There must be 0 differences

3. Performance Requirements

Requirements	Specifications	ATP
System`s data processing speed must keep up with the sampling rate to avoid backlogs and data losses	Compression rate plus encryption rate must be greater than the sampling rate	Processing speed in(Kb/s) = <Sampling rate(Kb/s)

The respective sub-system Requirements, Specifications and ATPs of compression and encryption block are stipulated in the Paper design Section below.

Paper Design

Requirements Analysis

The design of the subsystem in this project surrounds the collection, compression and encryption of gyroscopic accelerometer data from an IMU to be used in the SHARC buoys that will be installed on pancake ice in the Southern Ocean around Antarctica

The design criteria list three requirements for the project: the ICM-20649 IMU is to be used, at least 25% of the Fourier coefficients of the data along with data being encrypted, reducing the amount of processing done in the Raspberry Pi processor. This project will not work with the actual IMU sensor, therefore simulations and raw data provided will be used. Compression and encryption modules will be made to satisfy the second requirement and throughout the design process choices will be made to limit the processing power needed by tasks carried out on the Raspberry Pi to satisfy the third requirement

Comparison of Compression Algorithms

Compression refers to the process of reducing the data size without losing information. Compressed data is cheaper to store and transmit. In order to recover the original data from compressed data a decompressor is used. This is termed lossless data compression and decompression.

The Deflate algorithm has two stages i.e. LZ77 and Huffman encoding as shown in figure 1.[2]

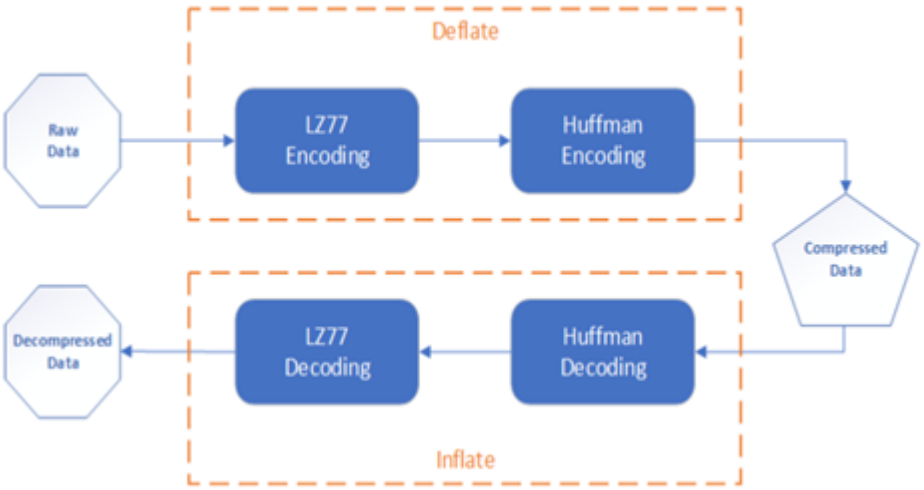


Figure 1. Breakdown of the deflate technique into two algorithms.

LZ77

Replace the repeated occurrences of the data stream with references. To find matches the algorithm keeps track of recently read data (search buffer) and input data (look ahead buffer) in a sliding window.

HUFFMAN ENCODING

The algorithm cleverly transforms fixed binary representation of data symbols (e.g. ASCII) to variable length codes based on frequency of appearance. The rule of thumb is that shorter codes are assigned to more frequently occurring data symbols. Each code is uniquely decoded and this is made possible by making each code a prefix code.

Here is a comparison of different libraries and tools that implement deflate algorithms [4].

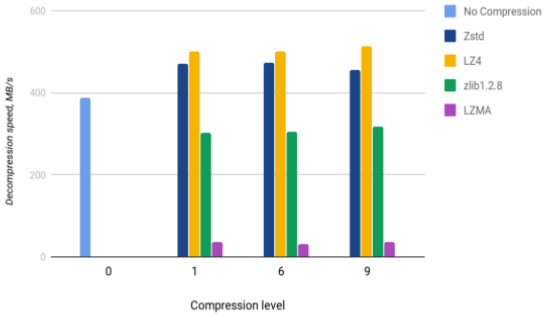


Figure 2:Diagram demonstrating the compression speed

We have chosen to use the GZip library which implements the two techniques discussed above. The GZip compression tool has the best trade-off between speed and compression ratio. GZip has a wide range of support in many applications[4].

Comparison of Encryption Algorithms

Current techniques of encryption are symmetric and asymmetric. Symmetric encryption requires that the sender and receiver have access to the same key, meaning the recipient needs to have access to the key before the message is received. Symmetric encryption algorithms include Data Encryption Standard (DES), Triple Data Encryption Standard (3DES/Triple DES), Advanced Encryption Standard (AES), Blowfish, Twofish, International Data Encryption Algorithm (IDEA), etc [5]. Asymmetric encryption on the other hand uses two keys, one public key and one private key that are mathematically linked to each other. Asymmetric encryption algorithms include Rivest Shamir Adleman (RSA), Elliptical Curve Cryptography (ECC), Diffie-Hellman exchange method, Digital Signature Standard (DSS), Digital Signature Algorithm (DSA), etc [5].

We have chosen to use symmetric encryption because it is faster than asymmetric and requires less computational power [6]. There are downfalls with symmetric encryption such as using a single secret key and potentially having to transmit this key to the receiving decryption entity, however given that we want to optimise speed and lower computational power required, symmetric encryption is best suited for our use case.

Of the available symmetric encryption methods, we have chosen to use AES. AES is quickly becoming the industry standard, it is widely used and trusted, and it is considered invulnerable to all attacks except for brute force (albeit this would take many years to break). Other forms of symmetric encryption are either outdated (such as DES), not as robust (such as Triple DES or IDEA) or not as widely adopted (such as Blowfish and Twofish), therefore making AES the best option to integrate in our subsystem [5, 6].

Feasibility Analysis

1. Technical Feasibility: The current hardware and software supports the implementation of the digital IP using the Raspberry-Pi to encrypt and compress the IMU data. Existing compression and encryption libraries will be used. The team has the technical skills for the project development

2. Schedule Feasibility: Deadlines will be met timely because we will be recycling implemented algorithms for compression and encryption.

3. Economic Feasibility: The project is not economically feasible. Funds are not available to purchase the expensive sensors(IMU).

4. Operational Feasibility: All the requirements have been turned into testable specifications which are attainable. Hence the project has a greater chance of satisfying the user requirements.

Therefore, the project can go ahead regardless of the unavailability of the sensor. As described below this challenge will be overcome by using simulations.

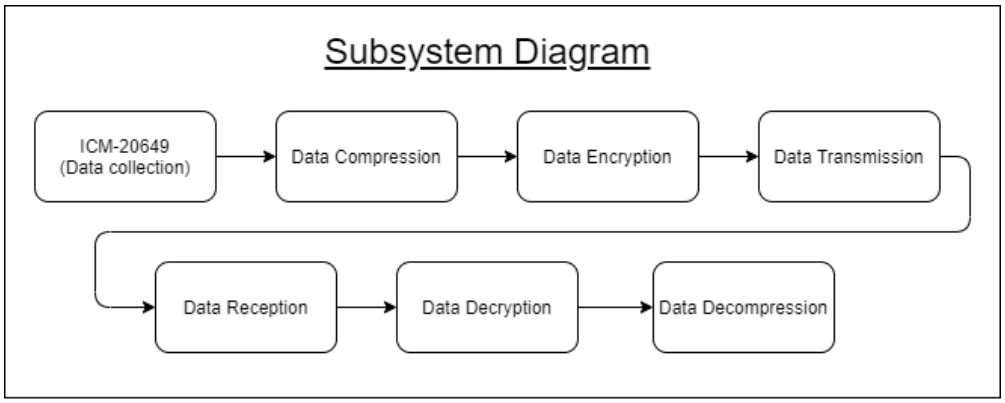
Possible Bottlenecks

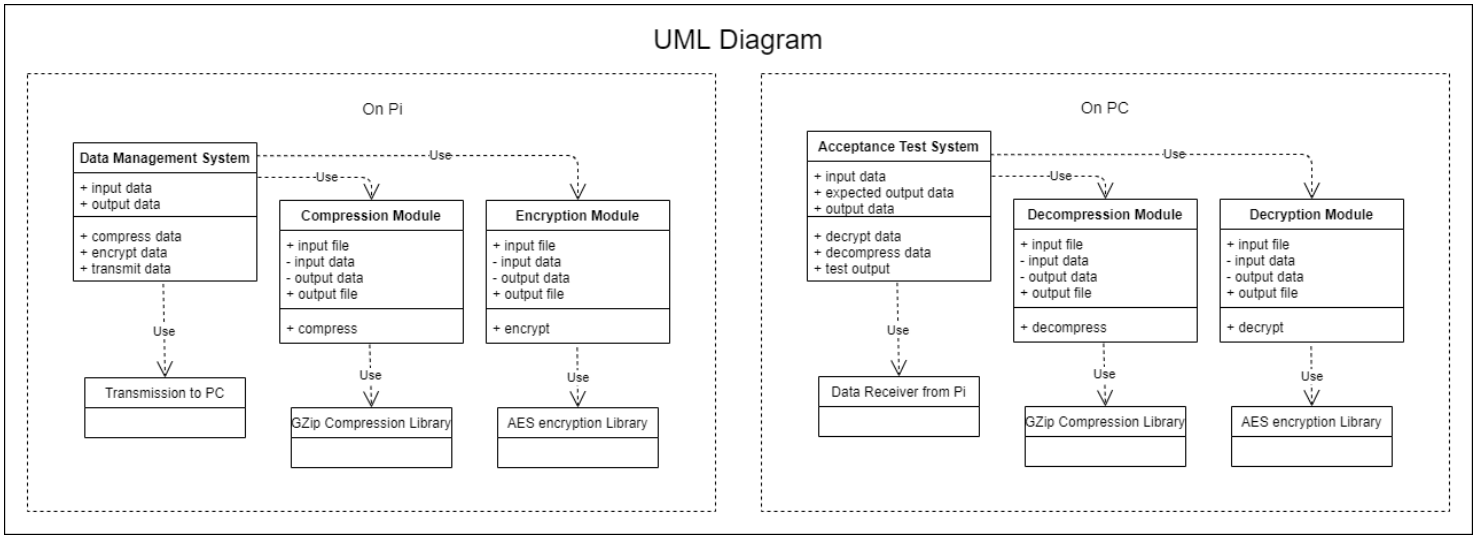
This project is still in the design phase and implementation has not begun. Faults could arise in our use of compression and encryption as well as transmitting the data between the different sub-subsystems. These bottlenecks will be dealt with if these problems occur, this may mean we need to continually update our design and methods used.

Another bottleneck may be in simulating the IMU dataset. The simulations will need to represent the actual output of the sensor as close as possible. If the simulation is not representative of the real world use, this subsystem will not be valid. To combat this, raw data from a similar sensor will be used to test the validity of our design alongside the simulation.

Subsystem Design

Our subsystem is made up of the following sub-subsystems; Data Collection, Data Compression, Data Encryption, Data Transmission, Data Reception, Data Decryption, Data Decompression.





- Data Collection

This unit is the first stage of the subsystem. Since we will not be working with the ICM-20649 sensor, the data collection phase will happen through datasets of simulations and raw data provided. In the simulations, the signal data will need to correspond as close to the sensor data as possible. The sampling rate will be the same as the actual unit and it will have to be realistic.
Additionally, this stage involves filtering of the data to extract at least 25% of the Fourier coefficients of the data.

- Data Compression

RSTs

Requirements	**Specifications**	**Acceptance Test Procedure**
Fast compression	Compression speed >>sampling rate	Verify by testing several file sizes
Great compression ratio	compression ratio > 1.5	Compare compressed file size and compare to the original file size.
Recover the original file	Lossless Compression and Decompression	The decompressed file must be identical to the original file.

The output from this subsystem is a gzip compliant file.

- Data Encryption

Once the data has been compressed correctly it will then be encrypted in this module. This will be done using the AES encryption library for the Raspberry Pi.

Requirements	Specifications	Acceptance Test Procedure
Use encryption library	AES encryption library for Raspberry Pi	Decrypt the transmitted data and compare it to raw input
Must be able to receive output from compression as input and must be able to send output after encryption to pc	Use python scripts to handle data being sent between the sub-subsystems	Data files need to be transmitted properly without fault.

- Data Transmission

The final system would require that the dataset is sent via satellite link. We will only be working with the subsystem of capturing, compressing and encrypting the data. Therefore, this Data Transmission unit will send the final compressed and encrypted dataset from the Raspberry Pi to a pc from which our acceptance tests can be run to validate the functionality of the subsystem.

On the other side of transmission the data is decrypted and decompressed to the original format.

Intra-subsystem Interaction

The modules of our subsystem will communicate and pass data files through python scripts. The Encryption module needs to be designed to take in gzip complaint files and the output from the decryption module will be a gzip compliant file which will then be decompressed by the gzip decompressor.

Validation Using Simulated or Old Data

Choice of Data

We have decided to use IMU raw data from another cruise different from SHARC BUOY. However the data sample has readings from other sensors such as temperature, humidity, pressure, magnetometer, etc. The IMU that we will eventually use will out X-, Y-, and Z-axis accelerometer and X-, Y-, and Z-axis angular rate sensors (gyroscopes)

Justification for using Raw data

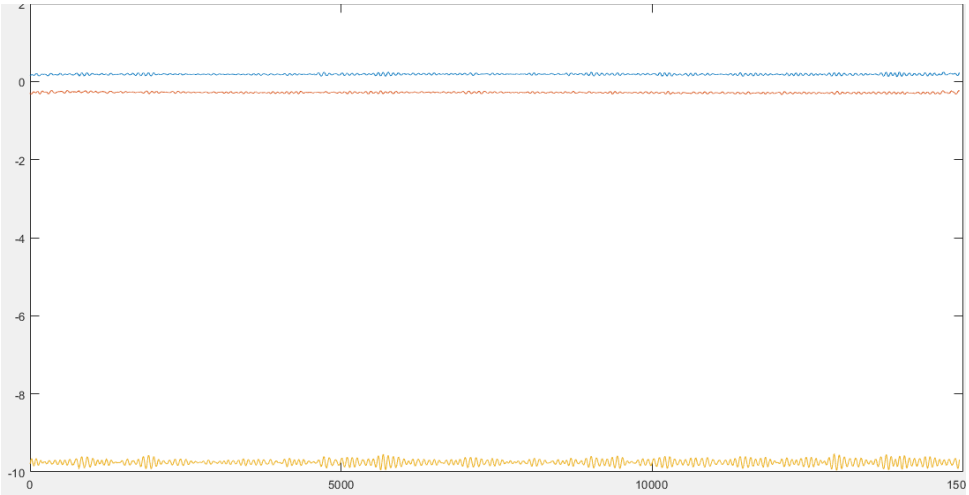
1. Data set is very large, this allow us to test our system performance under extreme conditions
2. Raw data represent the actual sampling rate of the IMU we are going to use.

3. More importantly the raw data is in the same format as the data that will be produced by the actual IMU.
4. Using old data saves time. It's simple and easy to implement, no worries about writing simulation algorithms that match the IMU sampling rate. This allows us to meet our deadlines.

Analysis of Data

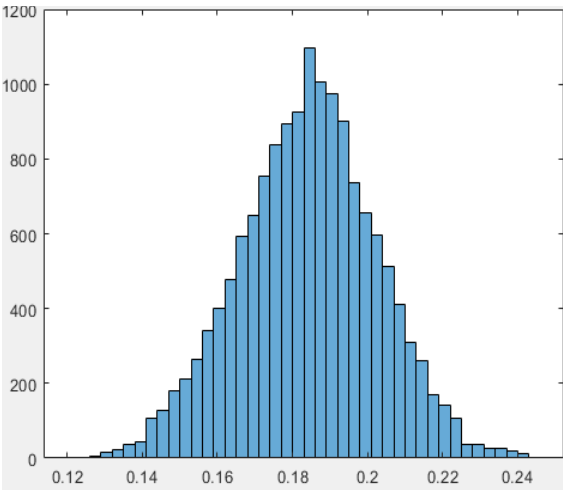
For analysis i`m using the file “2018-09-19-03_57_11_VN100.csv”

AccXYZ time Analysis

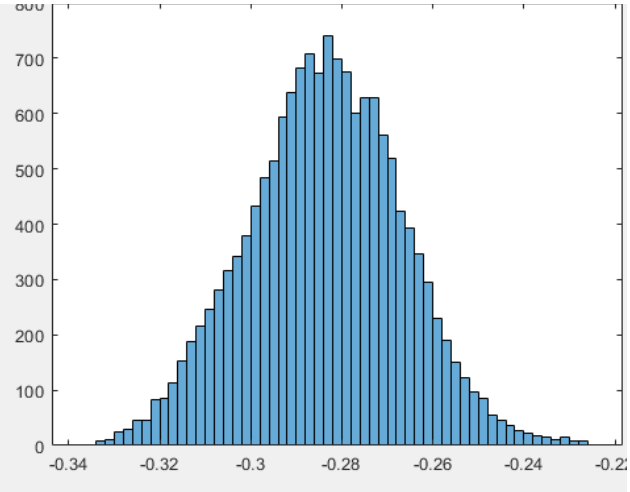


Blue=X, Red=Y, Orange=Z

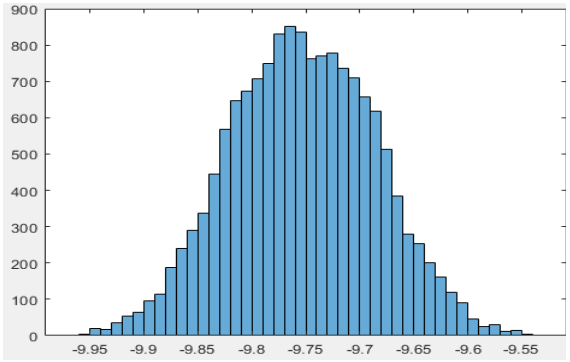
In the X and Y direction acceleration is close to 0



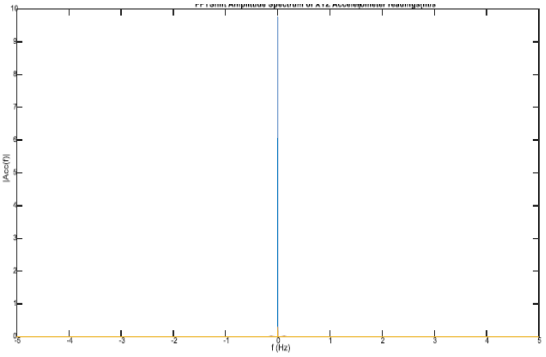
Histogram for AccX



Histogram for AccY



Histogram for AccY

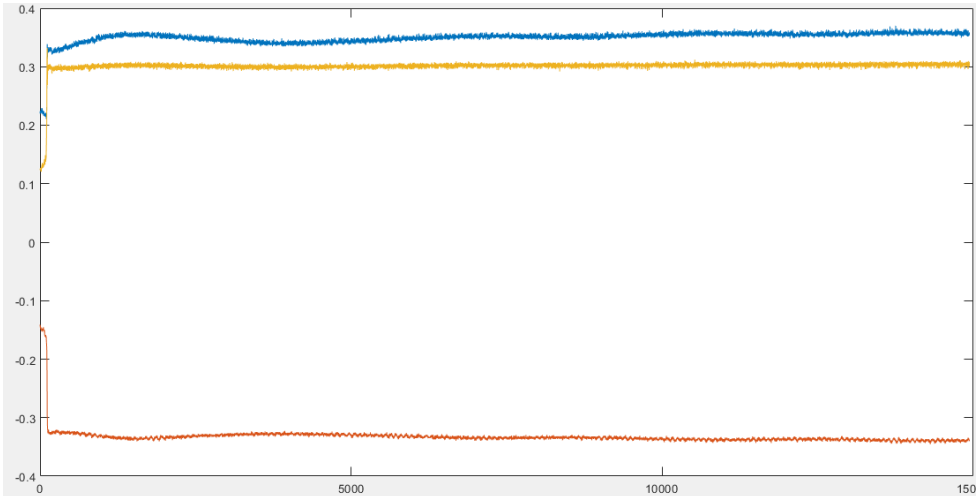


Frequency Spectrum of Acceleration

Variable	Mean	Mode	Std
AccX	0.1841	0.1476	0.0183
AccY	-0.2837	-0.3060	0.0170
AccZ	-9.7528	-9.7897	0.0674

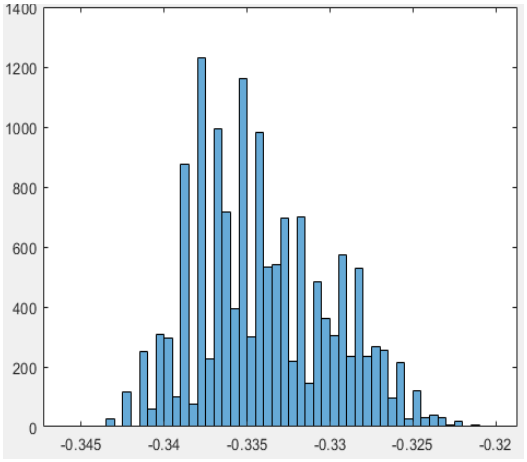
All three histograms show that the data is normally distributed about the mean.
Acceleration in the Z direction is interesting because it is equal to the gravitational acceleration of 9.8m/s^2. The assumption here is that Z represents the vertical motion, and the boat is somewhere in a free fall.
The variables closely follow the mean, small standard deviation
The frequency spectrum shows that the dominant component has a frequency of 0 which means acceleration values are fairly constant within the mean and this is confirmed by small standard deviation.

MagXYZ time Analysis

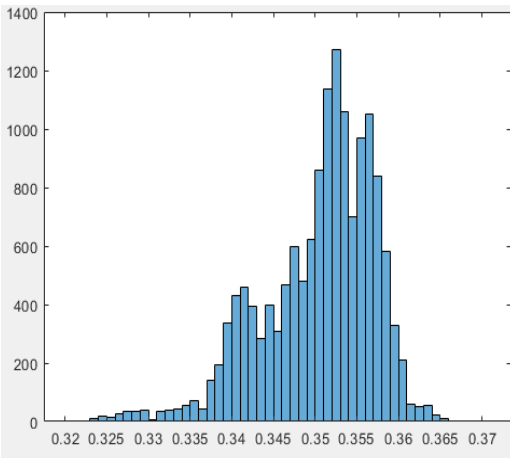


Blue=Y, Red=X, Orange=Z

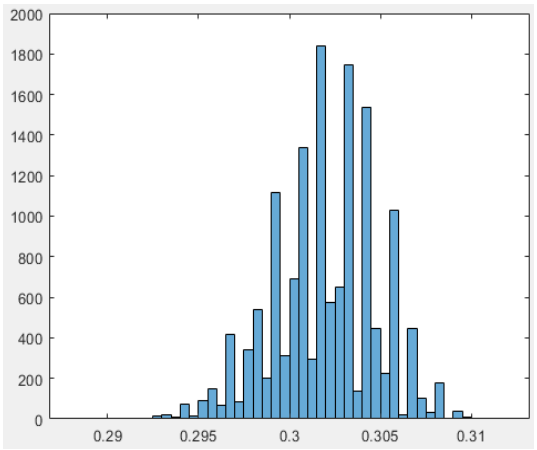
For the Histogram, I removed the first 150 elements because they are outliers, they will hinder proper data analysis.
The outliers might be a result of the boat changing direction during the initial time recording. Constant magnetic field thereafter shows that the boat is moving in a straight line



Histogram for MagX



Histogram for MagY

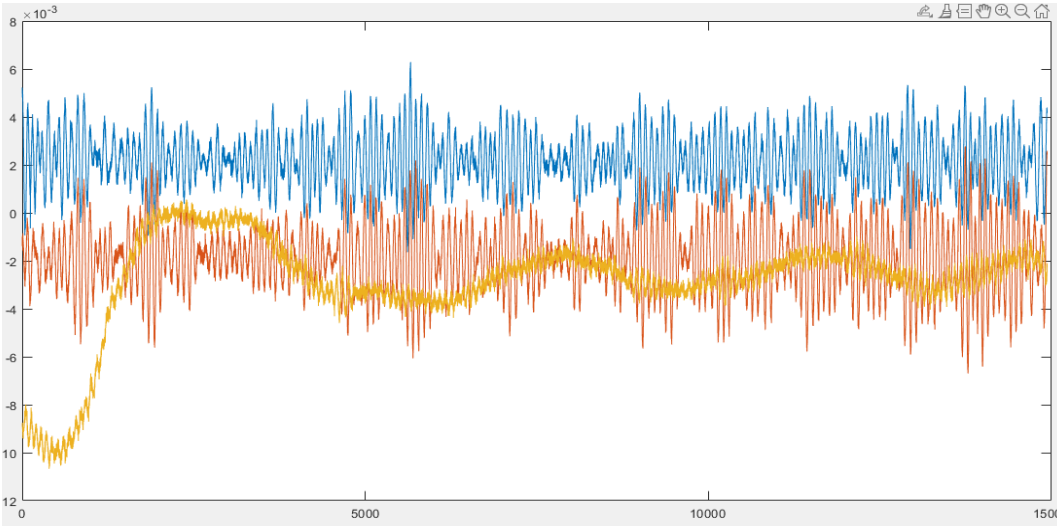


Histogram for MagZ

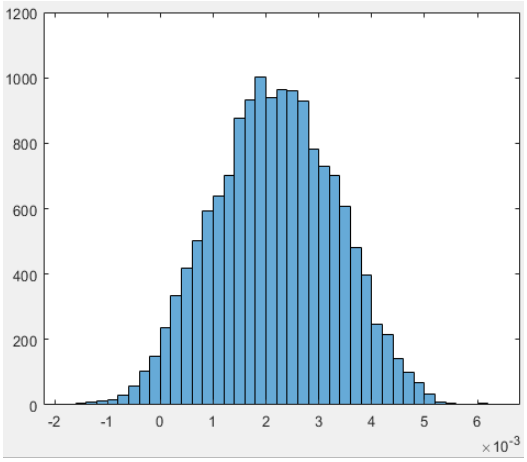
Variable	Mean	Mode	Std
MagX	-0.3339	-0.3390	0.0160
MagY	0.3504	0.3502	0.0067
MagZ	0.3020	0.3044	0.0029

MagX is positively skewed
The MagY is negatively skewed.
MagZ follows a gaussian distribution , its occurrence is random, no skewness.

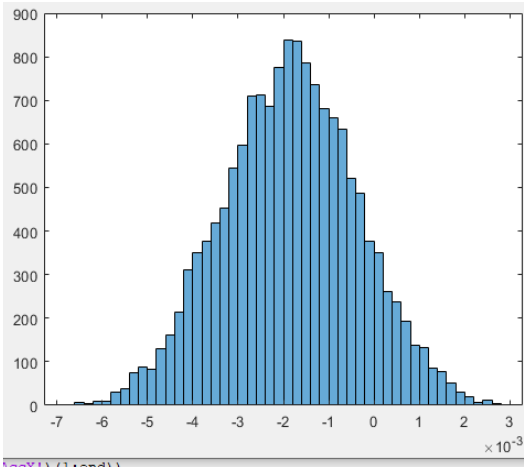
GyroXYZ time Analysis



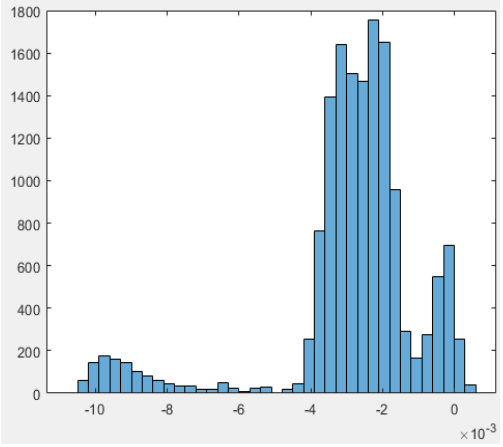
Blue=X, Red=Y, Orange=Z



Histogram for GyroX



Histogram for GyroY



Histogram for GyroZ

Variable	Mean	Mode	Std
GyroX	0.0022	0.0011	0.0012

GyroY	-0.0019	-0.0028	0.0015
GyroZ	-0.0029	-0.0037	0.0021

GyroX and GyroY are normally distributed
The occurrence of GyroZ is not random.

Validation of Data in the system

Experiment

Compression Block

The objective of the experiment is to test the compression ratios if they meet the specifications. The experiment seeks to check if the decompressed files are the same as the original file. The experiment will check the best compression level/algorithm to archive speeds required to avoid backlog and keep up with the transmission speeds.

Method

The zlib library we are using has 9 different compression levels
Choose a file large enough to represent the required throughput/bandwidth.
For each level run the compression and record time it takes to complete compression
Calculate speed by diving original file size with time
Calculate compression ratio by dividing original file size with compressed file size
The above steps are done by running the method *compress(fineName)*
Run the decompression algorithm on the compressed file, use the method *decompress(fileName)*
Decompress the compressed file and check for differences with the original file by running *comparison(file1, file2)*
There is no need to test decompression speeds because this process is not done with the IP on the buoy.
Repeat the procedure with different data samples.

Results

Table#:2018-09-19-03_57_11_VN100.csv

Level	File Size (Kb)	Compressed File Size(Kb)	Compression Time(s)	Speed(Kb/s)	Compression Ratio	Differences registered
3	8859	5133	0.54	16402	1.726	0
5	8859	4962	0.9	9843	1.785	0
7	8859	4923	1.3	6814	1.799	0

9	8859	4913	2.0	4430	1.803	0

Table# : 2018-09-19-09_49_31_VN100.csv

Level	File Size (Kb)	Compressed File Size(Kb)	Compression Time(s)	Speed(Kb/s)	Compression Ratio	Differences registered
3	8919	5126	0.46	19319	1.74	0
5	8919	4943	0.72	12423	1.80	0
7	8919	4896	1.09	8176	1.82	0
9	8919	4885	1.76	5072	1.83	0

Table #: combined two file ie 2018-09-19-03_57_11_VN100.csv and 2018-09-19-09_49_31_VN100.csv

Level	File Size (Kb)	Compressed File Size(Kb)	Compression Time(s)	Speed(Kb/s)	Compression Ratio	Differences registered
3	18678	10270	0.77	24306	1.82	0
5	18678	9895	1.26	14874	1.89	0
7	18678	9800	1.8	10370	1.906	0
9	18678	9770	2.79	6695	1.911	0

Input and output of the compression block

The input is a csv file
The output is also csv file

Input and output of the decompression block

The input is a csv file
The output is also csv file

Calculating the Sampling rate

The sample period(T) = 0.1s
Fs = 1/0.1 = 10Hz
Average file size = 8901 Kb contains 14935 samples
Each sample size = 8901/1435 = 0.6Kb
Sampling rate = 0.6Kb/s

Comments

Compression ratio increases with increasing compression level. This in turn increases time to complete compression(speed decreases) because the algorithms are using larger windows to find matches to compress the file even smaller.
With increasing file size the compression speed and ratios increase for each level this is attributed to the fact that larger files have greater chances of matches, there are many matches therefore the compression ratios are larger.

Compression Blocks ATPs

ATP	Met
Compression Speed >= Sampling Speed(0.6Kb/s)	✓
Compression ratio > 1.5	✓
Decompressed file identical to the original file	✓

Encryption Block

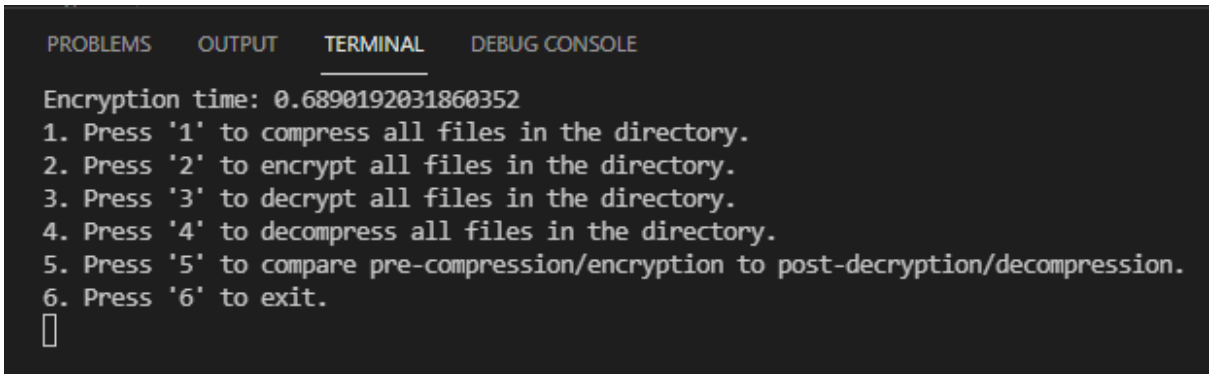
We used AES encryption from the pycrypto library and an algorithm adapted from M.A. Zia [7]. The encryptor class contains all the methods necessary to encrypt and decrypt files. This program recursively runs through each file in the same directory (except for the files containing the python code). When the encrypt_all_files method is called, a timer is run until every file has been encrypted. A second timer is used similarly for the decrypt_all_files method

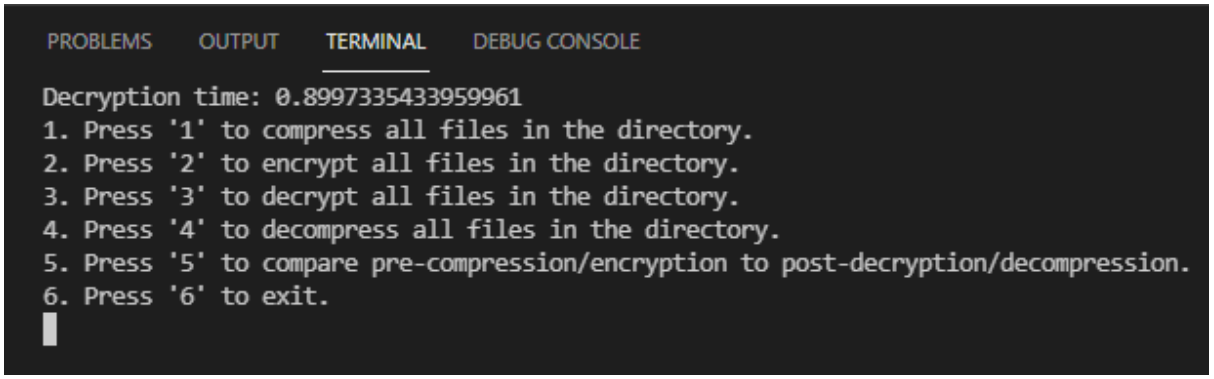
Method

To test the encryption block, the program will encrypt and decrypt all 9 data sets in the same directory. This is first to check that the program can encrypt multiple files while still managing them correctly. Secondly it is used to test the speeds of the encryption and decryption and finally a comparison check to see if the files match. This is done by reading all the files to a string variable (pre_text) before the user interacts with the program. After they are done decrypting then the compare function can be called to re-read the files to a new string variable (post_text). These two strings are then compared and if they are not exactly the same, a message will prompt the user “File is different after decryption”. If the strings are the same then it will print out “File is the same after decryption”

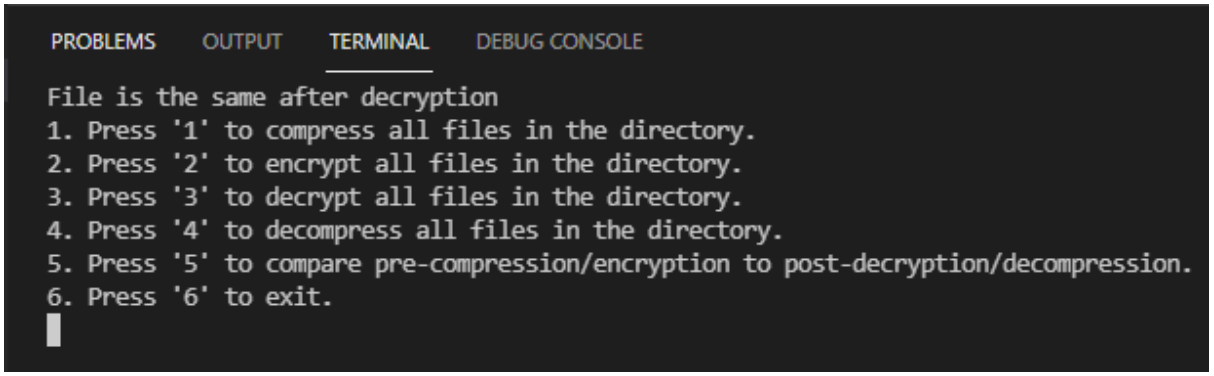
Results

Through running trials multiple times on all 9 data sets contained in the same directory, it was found that both encryption and decryption run between 0.5-2.0 seconds.





Comparing file to see if it matches after decryption:



Input and Output of the encryption block

The inputs are all the files of any type in the same directory as the program
The outputs are similarly named files except they will have a .enc extension appended to the end and the original files are deleted.
When decrypting, .enc files are deleted and the files are all saved again with their original names.

Comments

The encryption block functions as required, with no data loss between encrypting and decrypting. All files were managed correctly.

Encryption block ATPs

Encryption speed < 10 seconds for entire dataset	ATP met

System

Our system is designed to run linearly through compressing, encrypting, decrypting, decompressing and finally a comparison of the data before and after. To do this, three classes are used; one to handle compression and all its methods, one to handle encryption and all its methods, and one to compare a file's data after the first four stages. Below shows the terminal message when running the program:

```
1. Press '1' to compress all files in the directory.
2. Press '2' to encrypt all files in the directory.
3. Press '3' to decrypt all files in the directory.
4. Press '4' to decompress all files in the directory.
5. Press '5' to compare pre-compression/encryption to post-decryption/decompression.
6. Press '6' to exit.
```

It is designed such that each number is pressed sequentially, although this does allow for one to try only compressing and decompressing (by entering 1 then 4) or only encrypting and decrypting (by entering 2 then 3). Comparisons can still be done when only compressing/decompressing or only encrypting/decrypting. After each process, the time it took is printed to the terminal.

When the program is executed, it will act on every other file in the same directory except for command '5'. In the current version of our program, this only compares the data of the first data set instead of every file in the directory. This is done by first reading and storing the file content before the user is able to input. When the user calls '5', it will read and store the content of the file again in a separate variable. These two are then compared to see if the file is able to return to its original state after decrypting and decompressing since one of our ATPs was that the system should have no data loss.

Method

Since the system runs sequentially from user input, the total time is calculated from the summation of times of each process.

System results

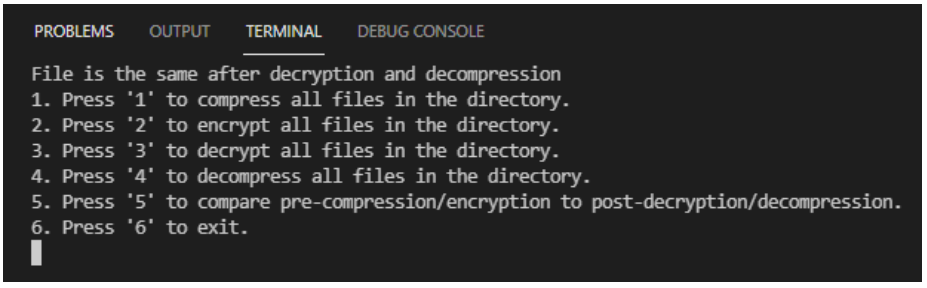
Table#: Results from running encryption followed by compression

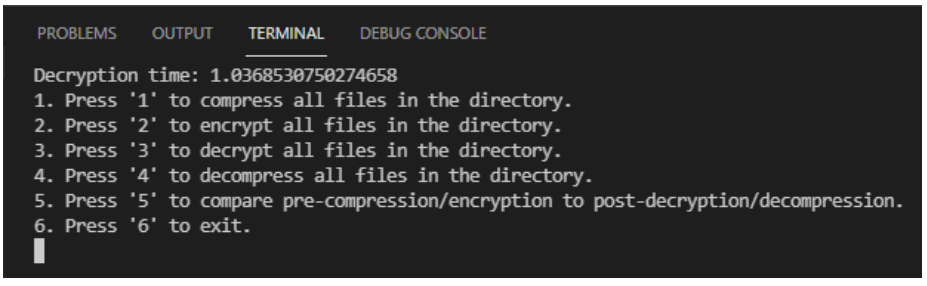
File Size(Kb)	Encryption Time(s)	Compression Time(s)	Output File Size(Kb)	Total Time(s)	Processing Speed(KB/s)	Differences Registered
8859	1.1	1.3	21	2.4	3692	0
8856	1.1	1.3	40	2.4	3692	0
8871	1.1	1.3	61	2.4	3696	0
18678	1.6	2.0	79	3.6	5188	0

Table#: Results from running compression followed by encryption

File Size(Kb)	Encryption Time(s)	Compression Time(s)	Output File Size(Kb)	Total Time(s)	Processing Speed(KB/s)	Differences Registered
8859	0.8	1.3	4662.7	2.1	4219	0
8856	0.8	1.3	4663	2.1	4217	0
8871	0.81	1.3	4667	2.11	4204	0
18678	0.9	2.0	9830	2.9	6441	0

Comparing the file after decryption and decompression





Comments

Encryption followed by compression has the worst performance. However the end files are smaller compared to the files if we did compression and then encryption after. The smaller file is attributed to the fact that encryption might have increased chances of many matches which makes compressed files smaller.

The compression speed for the st sequence is the worst because both algorithms are operating on the file of roughly the same size. If compression starts 1st then the encryption algorithm is working on a much smaller file which results in a speed up.

As expected, processing time increases for larger file sizes however the processing speed remains somewhat constant. Lastly the decrypted and decompressed file is compared against the original file, no differences were registered.

System ATPs

Our ATPs needed to be redesigned from the last submission. Below depicts a table of our ATP and whether they were met or not

Compression speed < 10 seconds for entire dataset	ATP met
Compression ratio >= 1.5	ATP met
Decompression speed < 10 seconds for entire dataset	ATP met
Encryption speed < 10 seconds for entire dataset	ATP met
One block should not operate more than 5 times slower than the other so that bottleneck does not occur	ATP met

Validation Using ICM20948 Sense HAT (B) IMU

Hardware based validation of our software system is very important and needful to address the limitations of simulation. Hardware based validation will expose the system to real world constraints such as power and processor speed constraints. Validating the system using the actual pi instead of a laptop to run algorithms gives accurate speed measurements of compression and encryption. The Sense HAT(B) is going to give an insight of how the data is actually stored.
With hardware validation we can test and configure other things which are assumed during simulation such as the I2C communication between the pi and the HAT.

Feature comparison of the ICM-20649 and Sense HAT (B)

ICM-20649	**Sense HAT (B)**
3-Axis accelerometer, FSR of ±4g, ±8g, ±16g, and ±30g	3-axis accelerometer, FSR of ±2/4/8/16 g
3-Axis gyroscope, FSR of ±500 dps, ±100 dps, ±2000 dps, and ±4000 dps	3-axis gyroscope,FSR of ±250/500/1000/2000 dps
N/A	3-axis magnetometer
N/A	Barometer, Temperature and Humidity
On-Chip 16-bit ADCs	Resolution: 12-bits
Host interface: 7 MHz SPI or 400 kHz I2C	I2C

Justification for using the sense HAT (B)

Sense HAT(B) is a board with a collection of sensors i.e. ICM20948, SHTC3, LPS22HB, TCS34725 and other processing chips. For this project we will make use of ICM20948 which closely resembles the ICM-20649 IMU to be used in the actual buoy.

- Pressure, temperature, humidity and magnetic readings are going to be ignored.
- Will use a force sensitive resistor(FSR) of ±16g for the accelerometer since it is common to both IMUs and to better insure that critical data is not lost at the point of high impact
- Will use a force sensitive resistor(FSR) of ±2000dps for the accelerometer since it is common to both IMUs and to better insure that critical data is not lost at the point of high speed rotation
- The sense HAT ((B) has lower ADC resolution, meaning the digital values obtained are not as accurate as what would have been observed when using ICM-20649. However, for testing purposes it can be ignored.

Validating the sensors on the IMU

The sense HAT was validated by running operational tests using the 'ICM20948.py' file provided by Waveshare [7]. These tests include the switch-on sequence (to test that all sensors record and used as a baseline for comparing the changes in the other tests), gyroscopic motion, acceleration test, roll pitch, and yaw and magnetic field test. For each test, other recordings are not considered; we only want to compare outputs for the specific test running.

Test Procedures

1. Switch on sequence
    - Method: Run the ICM20948.py file while the sensor remains motionless
    - Expectations: program should run and start displaying readings
2. Gyroscopic Motion
    - Method: Lift the HAT vertically upwards and downwards in varying motions, Slide the HAT across a table in different directions
    - Expectations: Gyroscope readings should change
3. Rotation Test
    - Method: rotate the HAT along its three axes
    - Expectations: Roll, Pitch and Yaw readings should change.
4. Acceleration Test
    - Method: Shake the sensor haphazardly
    - Expectations: Acceleration readings should increase
5. Magnetic field test
    - Method: Bring a strong magnet closer to the HAT
    - Expectations: Increase in magnetic readings

Test Results

Test:		Switch on Sequence	Gyroscope	Rotation	Acceleration	Magnetic
Gyroscope	X	1	-137	~	~	~
	Y	-1	226	~	~	~
	Z	0	-832	~	~	~
Rotation	Roll	-0.60	~	32.42	~	~
	Pitch	0.96	~	17.13	~	~
	Yaw	-33.43	~	-67.61	~	~
Acceleration	X	-248	~	~	-9116	~
	Y	-190	~	~	-32768	~
	Z	16518	~	~	7876	~
Magnetic	X	-122	~	~	~	1175
	Y	-82	~	~	~	-2749

	Z	99	~	~	~	-2398
Test Passed		✓	✓	✓	✓	✓

Validating the IMU in the system

Experiment Setup

Compression Block

The objective of the experiment is to test the compression ratios if they meet the specifications. The experiment seeks to check if the decompressed files are the same as the original file. The experiment will check the best compression level/algorithm to archive speeds required to avoid backlog and keep up with the transmission speeds.

Method

The zlib library we are using has 9 different compression levels
The files are generated by the IMU, Save the samples to a text file.
Pass the file to the compression module
For each level run the compression and record time it takes to complete compression
Calculate speed by dividing original file size with time
Calculate compression ratio by dividing original file size with compressed file size
The above steps are completed by running the method *compress(fileName)*
Run the decompression algorithm on the compressed file, use the method *decompress(fileName)*
Now check for differences between the the original and decompressed file by running *comparison(file1, file2)*
There is no need to test decompression speeds because this process is not done with the IP on the buoy.
Increase the sampling time and generate another text file of samples, repeat the experiment.

Results

Table#:data_1000.txt

Level	File Size (Kb)	Compressed File Size(Kb)	Compression Time(s)	Speed(Kb/s)	Compression Ratio	Differences registered
3	91	56	0.54	14578	1.626	0
5	91	53	0.9	7210	1.695	0
7	91	53	1.3	6807	1.719	0
9	91	51	2.0	3994	1.780	0

The file contains 1000 samples

Table# : data_1500.txt

Level	File Size (Kb)	Compressed File Size(Kb)	Compression Time(s)	Speed(Kb/s)	Compression Ratio	Differences registered
3	141	87.5	0.009	15071	1.611	0
5	141	85.4	0.012	11458	1.652	0
7	141	82.4	0.02	7176	1.711	0
9	141	79.3	0.03	4078	1.779	0

The file contains 1500 samples.

Table# : data_2000.txt

Level	File Size (Kb)	Compressed File Size(Kb)	Compression Time(s)	Speed(Kb/s)	Compression Ratio	Differences registered
3	180	97.3	0.010	18000	1.85	0
5	180	94	0.013	13846	1.91	0
7	180	91.8	0.024	7500	1.96	0
9	180	89.6	0.033	5454	2.01	0

The file contains 2000 samples.

Input and output of the compression block

The input is a txt file
The output is also txt file

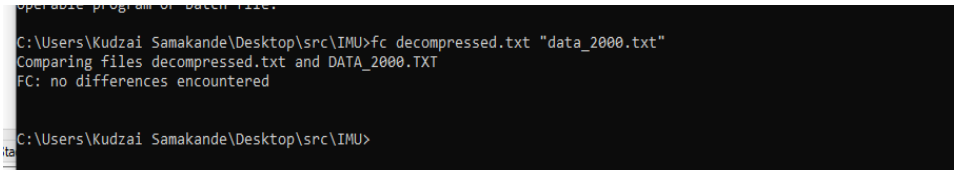
Input and output of the decompression block

The input is a txt file
The output is also txt file

The sample rate is 0.3 s/ sample…..each sample is 180/2000kb
Sample rate in kn/s = 0.09/0.3= 0.3kb/s

Comparing file to see if it matches after decompression

Windows command line is used to check if the original and the decompressed files are the same. The command is fc decompressed.txt "data_2000.txt" and showed that the file was the same.



No differences were encountered

Comments

As expected the compression ratio increases with increasing compression level.
Increased compression level takes more time trying to find matches which in turn decreases speed.
High compression ratios are achieved at higher levels because the algorithms are using larger windows to find matches.
With increasing file size the compression speed and ratios increase for each level this is attributed to the fact that larger files have greater chances of matches, there are many matches therefore the compression ratios are larger.

Compression Blocks ATPs

ATP	Met
Compression Speed >= Sampling Speed(0.3Kb/s)	✓
Compression ratio > 1.5	✓
Decompressed file identical to the original file	✓

Encryption Block

We used AES encryption from the pycrypto library and an algorithm adapted from M.A. Zia [7]. The encryptor class contains all the methods necessary to encrypt and decrypt files. This program recursively runs through each file in the same directory (except for the files containing the python code). When the encrypt_all_files method is called, a timer is run until every file has been encrypted. A second timer is used similarly for the decrypt_all_files method. Our ATPs for the encryption block are encryption/decryption speeds < 10% of data recording speed, and all files must remain exactly the same after encrypting and decrypting them.

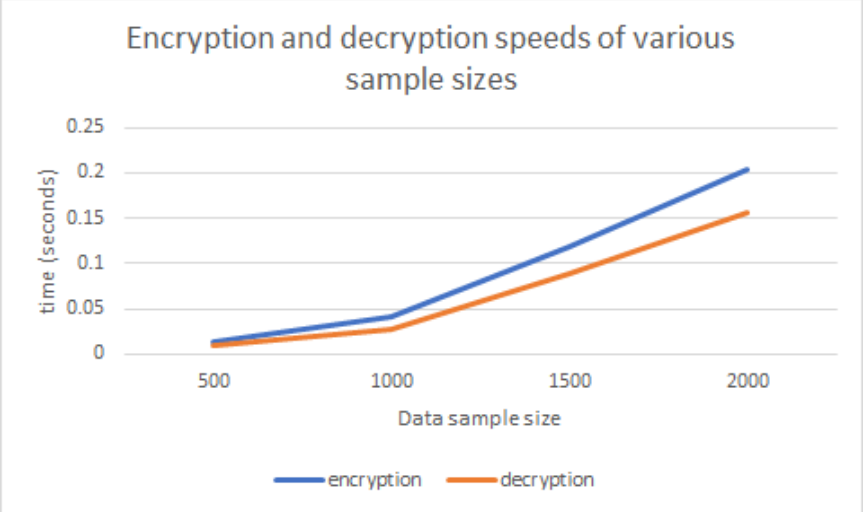
Input and Output of the encryption block

The inputs are all the files of any type in the same directory as the program
The outputs are similarly named files except they will have a .enc extension appended to the end and the original files are deleted.

When decrypting, .enc files are deleted and the files are all saved again with their original names.

Results

Encryption and decryption was done using the sense hat data. This includes the comparison between 500, 1000, 1500 and 2000 samples of data we recorded from the HAT.



Comment on results

Testing the encryption and decryption speeds on the different sample sizes produced expected results. Each increase in sample size produced an increase in both encryption and decryption time. At a sample size of 500, the speeds were very similar. As the sample size increases these times diverge too with encrypting taking longer than decrypting.

These results are only limited to sample sizes tested. The sense HAT can only record in roughly 0.2-0.3 second intervals so we could not run tests on larger file sizes as it would take too long to produce the larger files. However, we can extrapolate that our system will still function the same. Even at 2000 samples, taking roughly 10 minutes to record, encryption and decryption speeds are still very quick, encryption taking 0.2046 seconds and decryption taking 0.1564 seconds. For larger files, encryption and decryption speeds should remain quick compared to the time taken to record, meaning no bottleneck will occur between batches of data when encrypting.

Comparing file to see if it matches after decryption:

A comparison test was done on each sample size of data to test that after decrypting the file remains the same as before decryption (comparison function explained in Experiment, subheading System)

Test	File remains exactly the same
500 samples	✓
1000 samples	✓
1500 samples	✓

2000 samples	✓

Encryption Block ATPs

ATP	Met
Processing Speed < 10% of data recording speed (per sample size)	✓
Decrypted file identical to the original file	✓

The encryption block has met all the ATPs we designed for.

System

The system is a combination of 2 two subsystems which are compression and encryption blocks. Our system is designed to run linearly through recording, compressing, encrypting, decrypting, decompressing and finally a comparison of the data before and after. The objective of the experiment is to test how fast our system performs as a whole and to check if the end file is identical to the original file.

Method

To do this, three classes are used; one to handle compression and all its methods, one to handle encryption and all its methods, and one to compare a file's data after the first four stages. Below shows the terminal message when running the program:

```
1. Press '1' to record samples.
2. Press '2' to compress all files in the directory.
3. Press '3' to encrypt all files in the directory.
4. Press '4' to decrypt all files in the directory.
5. Press '5' to decompress all files in the directory.
6. Press '6' to compare pre-compression/encryption to
post-decryption/decompression.
7. Press '7' to exit.
```

When '1' is pressed, it will ask the user how many samples they want in the batch. For testing our program we limited the data files to sample sizes of 500, 1000, 1500 and 2000. This is because it takes between 0.2-0.3 seconds per recording. Using really large sample sizes would take a long time to capture, thus we decreased this to speed up testing and validating of the system. If this program were implemented for the buoy, the sample size can be selected to any desired value. The sampling rate can also be decreased by adding "time.sleep(x)" to the code, however the maximum sampling rate is still limited to roughly 0.2 seconds between readings.

It is designed such that each number is pressed sequentially, although this does allow for one to try only compressing and decompressing (by entering 2 then 5) or only encrypting and decrypting (by entering 3 then 4). Comparisons can still be done when only compressing/decompressing or

only encrypting/decrypting. After each process, the time it took is printed to the terminal. When the program functions for compression, encryption, decryption and decompression are executed, it will act on every other file in the same directory except for the python files.

File Comparison Check

The comparison class is used to test that the files remain the exact same as the raw data data files after running through the system. To do this, the comparison class opens every file and saves them as text in a variable (pre_text). This is done before the user is able to input to the program. Once the user has compressed, encrypted, decrypted and then decompressed the other files, the comparison test can be initiated by pressing 6. This will open up the files again and save them as text in a new variable (post_text). The pre_text and post_text variables are then compared. If there is any discrepancy the program will tell the user that there was loss and print to the terminal “There was loss after decryption and decompression.” If there is no difference found, the program will print to the screen “Files are the same after decryption and decompression.”

This comparison method can be adapted in the code to test individually per file or if you only want to test pre/post encryption or pre/post compression. With our methods of compression and encryption used, it is expected that no loss will occur.

System Results

The system was designed to test compression then encryption. Our code cannot compress files that have already been encrypted. Below shows the system results for compression then encryption.

Table#: Results from running encryption followed by compression

File Size(Kb)	Encryption Time(s)	Compression Time(s)	Output File Size(Kb)	Total Time(s)	Processing Speed(KB/s)	Differences Registered
44 (500 samples)	0.114	0.2021	21	0.316	139	0
90 (1000 samples)	0.24	0.4679	40	0.7079	127	0
134 (1500 samples	0.37	0.6350	61	1.005	133	0
179 (2000 samples)	0.57	0.8971	79	1.4671	122	0

Table#: Results from running compression followed by encryption

File	Encryption	Compression	Output	Total	Processing	Differences

Size(Kb)	Time(s)	Time(s)	File Size(Kb)	Time(s)	Speed(KB/s)	Registered
44 (500 samples)	0.0152	0.2021	23	0.2173	202.49	0
90 (1000 samples)	0.0374	0.4679	43	0.5053	178.11	0
134 (1500 samples	0.1097	0.6350	64	0.7447	179.94	0
179 (2000 samples)	0.2103	0.8971	83	1.1074	161.64	0

Comments

Encryption followed by compression has the worst performance. However the end files are smaller compared to the files if we did compression and then encryption after. The smaller file is attributed to the fact that encryption might have increased chances of many matches which makes compressed files smaller.

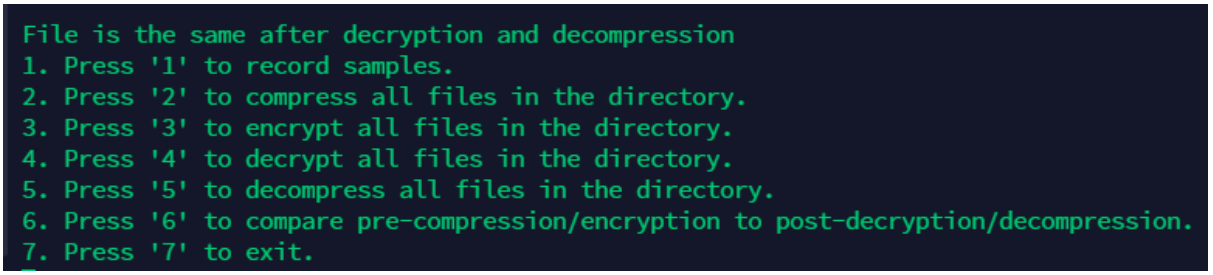
The compression speed for the st sequence is the worst because both algorithms are operating on the file of roughly the same size. If compression starts 1st then the encryption algorithm is working on a much smaller file which results in a speed up.

As expected, times increase for larger file sizes however the processing speed remains somewhat constant. At small file size, i.e. 500 samples, this processing speed is the fastest, however it is highly unlikely that real implementation would require batches of such small sizes. Processing speed of 1500 samples is slightly greater than that of 1000 samples. This means there is an optimum number of sample sizes for great procession speeds.

However, it is expected that using much larger files will slow down this processing speed.

Lastly the decrypted and decompressed file is compared against the original file, no differences were registered..

Output of tests after decrypting and decompressing



This shows that our system maintains all data after processing it. The system we are going to implement will start by compressing data, encrypt it and on the other hand after transmission there will be decryption 1st followed by decompression.

System ATPs

ATP	Met
Processing Speed > sampling rate(0.6kB/s)	✓
System throughput = < modem broadband(amount of data that the communication lines can handle at a time)	✓
Encryption Time =<Compression time (encryption must be faster or equal to the rate at which files are being compressed to avoid backlog)	✓
CPU time //check processing power	Not checked
Compressed and Encrypted file at most ⅔ (67%) the size of the original file	✓
Decrypted and decompressed file identical to the original file.	✓

ATP Consolidation

Below is an extract from our Sub Systems ATPs from paper design

Data compression

Requirements	Specifications	Acceptance Test Procedure	Decision
Fast compression and decompression	Compression speed greater than 50MB/s	The decompressed data must be identical to the original raw data.	Test Passed
Great compression ratio	compression ratio > 1.5		

Data encryption

Requirements	Specifications	Acceptance Test Procedure	Decision
Use encryption library	AES encryption library for Raspberry Pi	Decrypt the transmitted data and compare it to raw input	Test Passed
Must be able to receive output from compression as input and must be able to send output after encryption to pc	Use python scripts to handle data being sent between the sub-subsystems	Data files need to be transmitted properly without fault.	Test Passed

These ATPs needed to be redesigned because they were vague and very limited. They were also not organised properly in a coherent ATP section

Below is an extract from our first progress report ATPS

ATP	Met
Compression speed < 10 seconds for entire dataset	ATP met
Compression ratio >= 1.5	ATP met
Decompression speed < 10 seconds for entire dataset	ATP met
Encryption speed < 10 seconds for entire dataset	ATP met
One block should not operate more than 5 times slower than the other so that bottleneck does not occur	ATP met

Our ATPs were more defined but still not exact for our system. They still lacked detail and further explanation.

Below is an extract from our second progress report ATPs

Requirements	Specifications	ATP	Decision
Power consumption should be minimized	Ip is power by 11mA rechargeable batteries		Not tested
Fast compressions and decompression	Compression speed >> sampling rate	Compression speed < 10 seconds for entire dataset Decompression speed < 10 seconds for entire dataset	ATP met
Reduced file sizes for transmission	Compressed file<iridium bandwidth	Compression ratio >= 1.5	ATP met
Fast encryption and decryption	Encryption speed =< Compression Speed (making use of AES encryption)	Encryption speed < 10 seconds for entire dataset	ATP met
Secure files for transmission	The system should be able to compress then encrypt the files with no bottleneck	One block should not operate more than 5 times slower than the other so that bottleneck does not occur	ATP met
Minimal data loss	Data should be preserved after decryption and decompression	No data loss	ATP met

These ATPs are much more defined, with included requirements and specifications for each ATP. As we progressed through our design, we needed to refine our acceptance tests. This helped us focus on defining our system more, testing for specific requirements and presenting the results in an easy to read format.

Final System ATPs

Requirements	Specifications	ATP	Old data	ICM20948
Perform minimum computations to save power	Use at most 75% of the pi cores	Pi connected to a 1500mAh battery must run for 20hrs	Not Tested	
The system shall digitally sample the movement and orientation of the SHARC BUOY. A proper sensor is needed to collect acceleration and orientation status of the BUOY periodically.	Use ICM-20649 IMU to capture information and an ARM based microcontroller (raspberry Pi)	Subject the system to movement, shaking, orientation and magnetic field and check if the measured data is changing accordingly.		Test Passed
Retain the lowest 25% Fourier coefficients of the sampled data	Algorithm to perform perform frequency analysis and discard the the upper 95% Fourier coefficients of the sampled data	Compare the sampled data and the output of the system in the frequency domain and check if the output data has the 25% Fourier coefficients	Not Tested, this requirement was changed during the progression of the project. It was no longer required.	
The system must pass compressed and encrypted data to the transmission module	Use compression libraries eg Zlib or GZip Use AES encryption libraries	Check if the file size to be transmitted is considerably lower than the file size of the sampled data Inspect if the encrypted file if difficult to read/decipher	Test Passed	Test Passed
The transmitted data should be recovered easily	Implement decompression and encryption algorithms	Compare the decrypted and decompressed file with the original file. There must be 0 differences	Test Passed	Test Passed
System`s data processing speed must keep up with the sampling rate to avoid backlogs and data losses	Compression rate plus encryption rate must be greater than the sampling rate	Processing speed in(Kb/s) = <Sampling rate(Kb/s)	Test Passed	Test Passed

Future Plan

We wanted to design a modular system that is easy to read, adapt and implement. This was our purpose in designing our program to run sequentially from user input. The user can select the sample size of the data to record, the user can select to compress and/or encrypt, and the user can easily modify the code to run their own test. Therefore, in future use of this program the user will be able to incorporate our work into the system it's being used for with little difficulty. The program can be modified to run all the operations one after the other more automatically and the number of samples in one batch can be set to the desired size for use on the SHARC buoy.

Conclusion

The project went on according to the timeline. All Deadlines were met and each individual delivered their subsystems as promised. All the anticipated challenges/bottlenecks were overcomed. The systems integrated well and performed up to speed.

The software system as a whole met all its objectives which are to compress, encrypt, decrypt and decompress files.

On compression great compression ratios were achieved, and the encrypted files are very secure for transmission.

Tests were performed and the end file was identical to the original file. During the two tests, simulated tests were faster because we were using a laptop with many cores compared to the raspberry pi which we used with the HAT.

In conclusion the designed sub system is ready for the next stage, integration with other subsystems of the SHARC BUOY and then deployment.

References

[1] J. N. Jacobson, "SHARC Buoy - Robust firmware design for a novel, low-cost autonomous platform for the Antarctic Marginal Ice Zone in the Southern Ocean," Nation Research Foundation, Cape Town, 2021.

[2] (2019). Understanding Zlib [Online]. Available: https://www.euccas.me/zlib/

[3] Raul Fraile: How GZIP compression works | JSConf EU 2014. Available: https://www.youtube.com/watch?v=wLx5OGxOYUc&ab_channel=JSConf

[4] O Shadura and B Bockelman 2020 J. Phys.: Conf. Ser. 1525 012049. Available: https://iopscience.iop.org/article/10.1088/1742-6596/1525/1/012049/pdf.

[5] B. Daniel, “Symmetric vs. Asymmetric Encryption: What's the Difference?,” Trenton Systems Blog, 4 May 2021. [Online]. Available: https://www.trentonsystems.com/blog/symmetric-vs-asymmetric-encryption. [Accessed 4 September 2021].

[6] J. Thakkar, “Types of Encryption: 5 Encryption Algorithms & How to Choose the Right One,” The SSL Store, 22 May 2020. [Online]. Available: https://www.thesslstore.com/blog/types-of-encryption-encryption-algorithms-how-to-choose-the-right-one/. [Accessed 4 September 2021].

[7] M. A. Zia, "Python File Encryptor," Javapocalypse, 20 January 2018. [Online]. Available: https://github.com/the-javapocalypse/Python-File-Encryptor/blob/master/script.py. [Accessed 30 September 2021].

[7] Waveshare, "Sense HAT (B)," Waveshare, 21 July 2021. [Online]. Available: https://www.waveshare.com/wiki/Sense_HAT_(B). [Accessed 20 October 2021].