Submission 4: Final Report

Course Code: EEE3097S

Authors: Bradley Carthew, Nathanael Thomas

Student Numbers: CRTBRA002, THMNAT011

Date: 31 October 2021

1. Admin Documents

1.1. Contributions

Contributor	Tasks	Sections	Page Numbers
Nathanael Thomas, THMNAT011	 Link GitHub repository Make Timeline Compile the specifications Contributed to the paper design Need for simulation based validation Steps taken to execute the simulations Explanation of the encryption and decryption of the old/simulated data Compilation of the results obtained from the old data from IMU and execution of the algorithms on the entire system Encryption and decryption algorithm execution for the data obtained from the old IMU data Explaining the process taken to execute the compression and encryption as well as decryption and decompression of the data in relation to system as a whole Encryption and decryption algorithm execution for the data obtained from the sense hat B Explanation of the encryption of the data obtained from the sense hat B Change of the specifications Future plans around the project Contribute to the conclusion Referencing compilation 	1.3 1.4 2.3 3. 4.1 4.2 4.3.3 4.4.1 4.4.3 5.3.1 5.3.3 6.2 6.3 7. 8.	3 3 4-7 7-11 11 11-12 15 16-17 20-21 21-23 23 27 28 29 29-30 30-31
Bradley Carthew, CRTBRA002	 Compile contributions table Get screenshot of Trello page Discuss requirements Generate list of requirements Contribute to paper design Compile experimental setup for system and compression in validation using simulated or old data section (4.). Compile results for compression in validation using simulated or old data section (4.) 	1.1 1.2 2.1 2.2 3. 4.3.1 4.3.2 4.4.2	1- 2 2 3 3-4 7-11 12-13 14 18-19

 Need for hardware based validation Steps taken to execute hardware based validation 	5.1 5.2	21 21
 Compile experimental setup for compression in validation using different IMU section (5.). 	5.3.2	23
 Compile results for system and compression in validation using different IMU section (5.) Analyse the ATPs Contribute to the conclusion Compile list of references 	5.4.1 5.4.2 6.1 7. 8.	23-27 27 28 29-30 30-31

Table 1: A table showing the contribution of each of the team members.

1.2. Project Management Tool

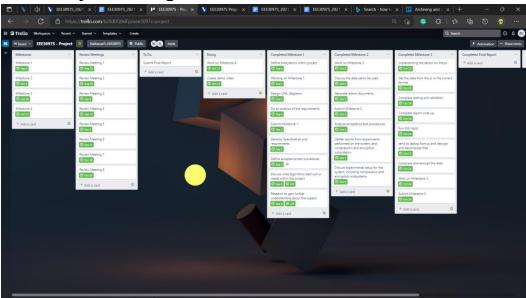


Figure 1: Snapshot of our Trello page.

Link to Trello: https://trello.com/b/Xd0Q9dFp/eee3097s-project

1.3. Link to our GitHub Repository

https://github.com/bradcarthew1/THMNAT011_CRTBRA002_EEE3097S

1.4. Timeline

Below is a basic Timeline demonstrating our progress throughout the development and implementation stages of our project.

	PROJECT NAME IMU simulation		PROJECT START DATE 16/8/2021	PROJECT END DATE 31/10/2021																						
Task ID	Task Description	Task Duration	Start Date	End Date	Aug,16,2021	22,	Aug, 23, 2021	29,	Aug, 30, 2021	Sep,5,2021	Sep,12,2021	Sep,13,2021	19	20,2	,26	V	ന	4,	7,2	Oct, 11, 2021	Oct, 17, 2021	Oct,18,2021	Oct,24,2021	Oct,25,2021	31	Nov,1,2021
1	Milestone 1(Paper Design)	20	Aug,16,2021	Sep,05,2021																				\neg	Т	
2	Milestone 2 (progress report 1)	30	Sep,05,2021	Oct,05,2021																						
3	Milestone 3 (progress report 2)	19	Oct,05,2021	Oct,24,2021																						
4	Milestone 4 (final report)	7	Oct,24,2021	Oct,31,2021																						
5	Demo video	7	Oct,24,2021	Oct,31,2021																						

Figure 2: Timeline of our progress taken from our Trello account.

The only delay was during milestone 3 when waiting to receive the sense hat B. This pushed back progress a little. But had not exponential effects on completing the milestone before the due date.

See the link to our Trello page for a more in-depth look at our milestones' progress.

2. Requirement Analysis

2.1. Discussion of Requirements

The initial requirements given to us at the start of the project, in the problem statement, are listed below:

- Req1: The IMU to be used is ICM-20649. However, we shall not provide you these IMUs. So you shall need to find out ways to design your IP without the real hardware!
- Req2: Oceanographers have indicated that they would like to be able to extract at least 25% of the Fourier coefficients of the data. Make sure that your compression satisfies this.
- Req3: In addition to reducing the amount of data we also want to reduce the amount
 of processing done in the processor (as it takes up power which is limited). Try to
 minimize the computation required for your IP.

2.2. Requirements

Below is the list of the derived requirements, they have been split up into hardware and software subsystems for easy reading.

Hardware Subsystem Requirements:

- An alternative processing unit must be used to implement our design, because the IMU on the SHARC Buoy will not be made available
- The processing unit must be able to generate datasets using on-board sensors
- The processing unit must have a power source
- The processing unit must be able to connect to a network
- The processing unit must be able to implement and execute compression and encryption algorithms

Software Subsystem Requirements

Compression

- Data after decompression must be the same as the original dataset
- Data must be compressed to a size smaller than that of the original file's size
- Compression algorithm must include the compression and decompression of datasets
- Compression must not be strenuous on the processing unit
- Compression must execute quickly
- Compression must not be too complex
- At least 25% of the Fourier coefficients must be accessible

Encryption

- Data outputted should be the same as the data that was inputted
- Encryption algorithm must include the encoding and decoding of dataset
- Encryption should be fast
- Encryption should not be strenuous on the processing unit
- Must be able to encrypt large sets of data
- The encryption algorithm must not be complex
- The dataset must be encrypted so as to eradicate data leaks

2.3. Specifications

Below is a list of desired specifications derived from the requirements (above, in Section 2.2). The specifications are split up into hardware subsystem and software subsystem specifications for easy reading. For each specification, an acceptance test procedure was developed so as to check that the specifications are met.

Hardware Subsystem Specifications

Requirement	Specification	Acceptance Test Procedure
	Make use of a Raspberry Pi Zero W for processing and attach the Sense-HAT B to provide sensor functionality.	

The processing unit must be able to generate datasets using on-board sensors	On-board sensors must measure yaw, pitch, roll, magnetic fields (in x, y and z directions), gyroscopic movement (in x, y and z directions) and acceleration (in x, y and z directions).	Move the Sense-HAT B around in different directions and check to see that the yaw, pitch, roll, gyroscope and acceleration values are changing. Bring a magnet close to the Sense-HAT B to check that the magnetometer is functioning correctly.
The processing unit must have a power source	Make use of a USB to USB mini for connection from a laptop to Raspberry Pi Zero W to provide power.	Check to see that the green light on the Raspberry Pi Zero W and the red light on the Sense-HAT B turn on.
The processing unit must be able to connect to a network	Connect the Raspberry Pi Zero W to a laptop hotspot (WiFi adapter 802.11a/b/g wireless mode must be set to 2.4 GHz 802.11g).	Check to see that the device has been added to a list of devices connected to the laptops mobile hotspot.
The processing unit must be able to implement and execute compression and encryption algorithms.	The Raspberry Pi Zero W must be able to implement and execute gzip compression and AES encryption algorithms.	Compile and run the AES encryption and gzip compression algorithms on different datasets. Check that the files contents have been altered.

Table 2: List of requirements, specifications and acceptance test procedures for the hardware subsystem.

Software Subsystem Specifications Compression

Requirement	Specification	Acceptance Test Procedure			
Data after decompression must be the same as the original dataset	The contents of the file, after executing the gunzip algorithm, must be the same as the contents of the original file.	Use the windows powershell to run a bash script to check that the contents of the two files are the same. A message must be displayed indicating whether the file contents are the same or not the same.			
Data must be compressed to a size smaller than that of the original file's size	The file size after executing the gzip algorithm must be to 50% smaller than the original file size.	Use the "Is -Ih" command in the terminal to show the file size after compression and compare this to the original file size.			

Compression algorithm must include the compression and decompression of datasets	Make use of the gunzip and gzip commands to decompress and compress data, respectively.	Compile and run the gzip compression algorithms on different datasets. Check that the files contents have been altered.
Compression must not be strenuous on the processing unit	The average CPU usage after running the gzip (or gunzip) algorithm must not exceed 50%.	Use the "sudo iostat -c 2 2" command in the terminal to show the average CPU usage before and after compression and decompression. An average CPU usage spike over 50% indicates that the algorithm is too power hungry, and anything below that indicates that the CPU usage is within valid bounds.
Compression must execute quickly	The gzip (or gunzip) algorithm execution time must not exceed 3 seconds.	Use a timer on a smartphone to check that after running the algorithm the execution time remains under the 3 second threshold.
Compression must not be too complex.	Make use of the gzip compression algorithm.	Implement different compression algorithms and run tests to see which algorithm executes faster.
At least 25% of the Fourier coefficients must be accessible.	At least 25% of the Fourier coefficients of the dataset must be made available after executing the gzip and gunzip compression commands.	Use the windows powershell to run a bash script to check that the contents of the two files are the same. If they are identical this indicates that 100% of the Fourier coefficients are available for processing.

Table 3: List of requirements, specifications and acceptance test procedures for the compression software subsystem.

Encryption

Requirement	Specification	Acceptance Test Procedure					
	executing the AES decryption algorithm, must	Use the windows powershell to run a bash script to check that the contents of the two files are the same. A message must be displayed indicating whether the file contents are the same or not the same.					

Encryption algorithm must include the encoding and decoding of dataset	Make use of the AES encryption and AES decryption algorithms (Code Cracker algorithm) to encode and decode data respectively.	Compile and run the AES encryption and decryption algorithms on different datasets. Check that the files contents have been altered.				
Encryption should be fast	The AES decryption and encryption algorithm execution time must not exceed 3 seconds.	Use a timer on a smartphone to check that after running the algorithm the execution time remains under the 3 second threshold.				
Encryption should not be strenuous on the processing unit	The average CPU usage after running the AES encryption and decryption algorithm must not exceed 50%.	Use the "sudo iostat -c 2 2" command in the terminal to show the average CPU usage before and after encryption and decryption. An average CPU usage spike over 50% indicates that the algorithm is too power hungry, and anything below that indicates that the CPU usage is within valid bounds.				
Must be able to encrypt large sets of data	The AES algorithm must be able to encrypt and decrypt a file containing around 15000 samples.	Run the encryption and decryption algorithm on a file with 15000 samples, check to see that the contents have been altered and that a message is displayed saying the file has been encrypted or decrypted successfully.				
The encryption algorithm must not be complex	Make use of an AES encryption algorithm implemented using C++.	Implement the algorithm using different coding languages and run tests to see which language executes faster.				

Table 4: List of requirements, specifications and acceptance test procedures for the encryption software subsystem.

3. Paper Design

3.1. Comparison of Compression and Encryption Algorithms Compression

Compression is the process of reducing the size of data for the purposes of saving bandwidth, distributing data and saving space. [6]

There are two main methods used to compress data:

Lossy Compression [8]

Lossy compression involves reducing the size of data by getting rid of unnecessary or less relevant information. The amount of data is greatly reduced resulting in an improved data compression ratio. [6]

Advantages [7]

File size is greatly reduced

Disadvantages [7]

- Is irreversible original data cannot be completely restored
- Quality of the image is greatly reduced with a higher ratio of compression

Examples of Algorithms

- Discrete Wavelet Transform (DWT) involves a Fourier analysis and decomposes a signal (set of data) into sinusoidal basis functions of different frequencies.
- Discrete Cosine Transform (DST) decorrelates data, encodes each transform coefficient independently without losing compression efficiency. [9]

Lossless Compression [8]

Lossless compression involves reducing the size of data without any loss of data and is done by removing unnecessary metadata. [7]

Advantages [7]

- No loss in the quality of the data
- File size is reduced slightly

Disadvantages [7]

Results in larger files than lossy compression

Examples of Algorithms

- Arithmetic Coding a type of entropy encoding technique, that takes a file composed of symbols (eight-bit characters) and converts these symbols to a floating-point number that is greater than or equal to zero and less than one. [10]
- Huffman Coding involves assigning a variable-length code to different input characters, this code length is related to how frequently characters are used. [11]
- Lempel Ziv Welch (LZW) reads a sequence of symbols, groups the symbols into strings and converts the strings into codes. [11]

Encryption

Encryption can be defined as the process of converting readable data into encoded format. Encryption is used to provide privacy to the data as the encrypted data needs a personalised key either public or private or the use of a cypher depending of what was set/chosen by the person who had encrypted the data. [5]

There are two main types of encryption algorithms:

Symmetric Encryption

Makes use of a single key to encrypt and decrypt data.

Advantages [5]

- Simpler than asymmetric encryption
- Faster performance
- Less power used in computation

• Key length is smaller of size 128-256-bit length

Disadvantages [5]

- As there is only one key it needs to be kept secret
- Less private compared to asymmetric encryption

Three examples of algorithms [12]

- Two-fish can encrypt a key of bits of length 256 bits. Only one key is required. Fastest symmetric technique and is openly available.
- AES Advanced Encryption Standard, includes three block cyphers AES –
 128,192,256. Depending on which cypher block is used, the cypher will encrypt and
 decrypt using cryptographic keys of the size of 128,192 or 256 bits, in blocks of 128
 bits.
- DES- Data Encryption Standard it uses blocks of plaintext of a specific size 64-bits and will return an encrypted block of the same size.

Asymmetric Encryption

Defined as the encryption and decryption of data with the use of key pairs, known as public and private keys. Even though asymmetric encryption offers greater security through authentication. The downfall is that is easy to go in one direction (encrypt) but harder to retrieve the original [5]

Advantages [5]

- Provides better security then symmetric encryption
- Authentication, the data is only seen by the person who is supposed to see it

Disadvantages [5]

- Can be a very slow process
- High risks involved if one were to lose their private key. No longer able to decrypt the messages
- If anyone else were to know your private key they have access to all your messages, security breach
- Difficult to retrieve original data

Three examples of algorithms [5]

- RSA makes use of prime factorisation. By multiplying two large prime number and to decrypt will find the original prime number.
- Diffie-Hellman Allows for two parties to with no prior knowledge to have a shared key established between them over an insecure channel
 - ECC Elliptic curve cryptography is an encryption technique using a public key. This technique is based elliptic curve theory.

3.2. Feasibility Analysis

The major deciding factor around the feasibility is that the SHARC Bouy runs on a small battery therefor to reduce battery consumption processing must be kept minimal.

As the data from the SHARC Buoy is relatively large and doesn't require complex privacy protection, we have decided on using symmetric encryption. The basis behind this decision was that the symmetric encryption was more effective on large chunks of data. Required less computational processing when compared to asymmetric encryption. On deciding which encryption algorithm would be most effective in completing the encryption, we looked at three possible encryption algorithms. Those being Two-fish, Advanced Encryption Standard and Data Encryption Standard. Upon further discussion. The decision was made that the Two-fish encryption algorithm would be used. The reason behind this decision being that the two-fish is the fastest encryption algorithm, meaning that the processing time is reduced. With the reduction in processing time. Less computational processing required, battery consumption minimal. Besides, it being the fastest encryption algorithm it also only requires one key and the key can be of lengths of up to 256 bits. This algorithm is also available to the public making it easily accessible. This is why it was concluded that the two fish algorithm would be best fit for what was required. After decompression, it is required that a minimum of 25% of the Fourier coefficients of the data be available. This makes it possible for lossless compression algorithms to be used. An advantage of using lossless compression is that the data in the file is not altered when compared to lossy compression techniques. As the main goal is getting reliable and accurate data the choice of lossless compression is preferred. The three main compression algorithms considered, are arithmetic coding, Huffman coding and Lempel-Ziv Welch. It was decided that the arithmetic coding algorithm will be used. This is mainly because it has a quicker execution time but also because it has a better complexity when compared to Huffman coding. Huffman coding has a complexity of O(n^2) and arithmetic coding has a complexity of O(nlogn).

3.3. Possible Bottlenecks

The phenomenon of a Bottlenecks can be defined by single component within a system limiting its performance and capacity. Within this system, the Bottlenecks can stem from components within the system. The main being the IMU, in our case the Raspberry Pi zero. The possible Bottlenecks that come from the use of the raspberry pi and the network over which the data is exchanged/ sent over are:

- The speed at which the Pi is able to process the data / algorithms.
- The bandwidth over which the data is transferred. Determines how much processing is required. Which in turn dictates the power consumption
- Inefficiencies within the algorithms used, due to the fact as they are open-source algorithm, they are not case specific and are designed for the general case.

3.4. Inter-Subsystem and Inter-Sub-Subsystem Interactions

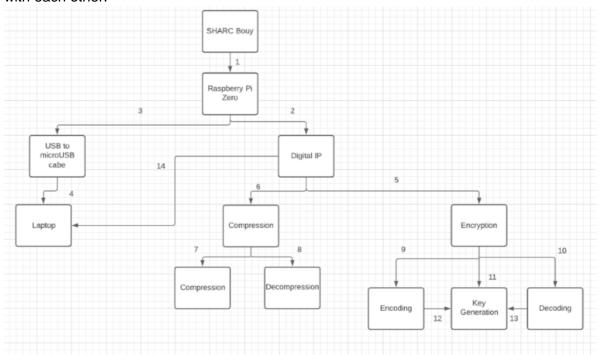
The following interactions will be explained making use of the UML diagram in question 3.4. The inter-subsystem interactions are indicated by the key "IS" and the inter-sub-subsystems by "ISS".

- 1. The Raspberry Pi Zero W emulates the IMU found on the SHARC Buoy. (IS)
- 2. The digital IP (intellectual property) runs on the raspberry pi zero w interface. (IS)
- 3. The Raspberry Pi Zero W is connected to the laptop via the USB cable. (IS)
- 4. The Laptop provides power to the raspberry pi through the USB cable. (IS)
- 5. The digital IP initiates encryption (IS)
- 6. The digital IP initiates the compression. (IS)
- 7. Via the arithmetic coding algorithm the data is compressed. (IS)

- 8. After compression occurs, the arithmetic coding algorithm decompresses the data. (IS)
- 9. After the Digital IP initiates encryption, data is then encrypted using encryption algorithms. (IS)
- 10. After the data has been sent, the encryption algorithm will perform the decryption of the data (IS)
- 11. The encryption algorithm will generate a key to allow for the encryption and decryption processes to occur. The encryption key is needed for any user who would like access to the data (IS)
- 12. The encryption algorithm will fetch the encryption key to check if the user has access to the data. If the user does the encryption algorithm is executed (ISS)
- 13. The encryption algorithm will fetch the encryption key to check if the user has access to the data. If the user does the encryption algorithm will then de-encrypt the data. (ISS)
- 14. The laptop is used to display the output after compression and encryption algorithms have executed. (ISS)

3.5. UML Diagram of System

The below diagram describes how the inter-subsystems and inter-sub-subsystems interact with each other.



4. Validation using Simulated or Old Data

4.1. Need for Simulation-Based Validation

The reason behind simulation based validation is that this provides us with a benchmark that we are then able to compare the hardware based validation to. The simulation-based validation will provide us with the most accurate results as it is directly related to the SHARC BUOY as the data is the one from the BUOY. When comparing with hardware we will be able to determine how accurately the data produced matched that of the simulations. It is a cost

effective approach when dealing with the data. As it only requires a computer and raspberry pi and no other external components and hardware

4.2. Steps

Below are the steps taken during the software-based validation:

- All the csv file provided were converted into .txt file for easier access across all platforms and easier formatting.
- The data was moved to our local git repository.
- Using our virtual machine which ran ubuntu linux. Using the "git pull" we were able to get all the data required.
- On the virtual machine the data was then compressed and encrypted
- The data was then pushed back to the local git repository.
- Using the raspberry pi through a laptop interface the data from the git repository was pulled.
- Now the compressed and encrypted data files were decrypted and decompressed.
- From there the changed files were then compared to the original files to see if there
 were any changes.
- Even though the report only shows the one data sets results. We repeated the method on different data sets of different sizes and they all displayed the same result where the files were all identical.

•

The data being used in this section is the real-world data obtained from an IMU on board a SHARC Buoy. The data is in the form of a .csv file, however it has been converted into a text file format (.txt) for ease of use when executing compression and encryption algorithms. The datasets have an average sample size of 15000, where data entries were made every 10 milliseconds.

4.3. Experimental Setup

4.3.1. System

The overall system requires that encrypted and compressed data can be sent to the processing block (in our case, the Raspberry Pi Zero) where it will be decrypted and decompressed.

The sub system is required to:

- Compress and encrypt data from a remote location
- Send the data to the processing block in real-time
- Decrypt and decompress the data on the processing block
- Data integrity after decryption and decompression must be upheld (file contents must not have changed).
- The file size of the data must have reduced in size to around 50% of the original file size
- At least 25% of the Fourier coefficients must be extracted
- The CPU and RAM usage must be kept to a minimum in order to reduce battery consumption

- Use a virtual machine to run Ubuntu Linux this would act as a remote server.
- The raspberry pi would simulate the IMU.

- The size of the data files will be checked using the following command in terminal "Is
 -Ih".
- Then the file data 1.txt was compressed using the command "gzip data_1.txt"
- Using the command "Is-Ih" the size of the file will be checked to see if the size of the file has actually decreased. This will indicate that the algorithm has worked. As well as to check that the file size has been reduced to less than 50%.
- After the encryption.cpp file has been compiled using the command "g++ encryption.cpp -o encrypt" the following command was run "./encrypt" which will prompt the user to enter the file that needs to be encrypted the file entered will be data_1.txt.gz this will then encrypt the file.
- Then the contents of the file will be checked using "sudo joe data_1.txt.gz" this will show if the data within the file has been encrypted.
- From there the data will be added, committed and pushed to the git repository.
- The raspberry pi will be setup and it will pull from the git repository. Using the git pull command.
- Next the avg CPU usage will be checked "sudo iostat -c 2 2" This is used to check that the CPU usage is below 50%
- This will take not of what the average CPU usage is when the pi is idle
- Now that the data file is on the raspberry pi. The decryption.cpp will be executed similarly to the encryption.cpp using the command "g++ decryption.cpp -o decrypt" after this has been executed and compiled properly
- The command "./decrypt" will be run this will then prompt the user to enter the file they
 want to decrypt. To which the file "data_1.txt.gz" will be entered after the decryption is
 complete.
- Next the avg CPU usage will be checked "sudo iostat -c 2 2" this was done to check how the average CPU usage changed after this command was run. This is used to check that the CPU usage is below 50%
- Next the command "Is -Ih" was run to check that the size of the file had not changes since the decryption algorithm was executed.
- Next the command "gunzip data_1.txt.gz" was run this will then decompress the file thus returning the original file.
- Next the avg CPU usage will be checked "sudo iostat -c 2 2" was run to check how the CPU usage was affected after the decompression was done. This is used to check that the CPU usage is below 50%
- The command "Is-Ih" was executed to check the size of the file and compare it to the size of the original data file. This will be an indicator to show that all process have worked and the original file can be retrieved from all these processes.
- Lastly this file will be taken onto a windows operation based laptop to compare the integrity of the files. Using the command "if((Get-FileHash "C:\Users\Adarsh\Desktop\EEE3097S\CHeck\Data1.txt").hash -eq (Get-FileHash "C:\Users\Adarsh\Desktop\EEE3097S\CHeck\Data2.txt").hash) {"Both the files are identical"} else {"Both file are NOT identical"}" in windows powershell. This command will return if the files are identical or not.
- If the files are identical this will mean that the fourier plot will be the same. So the
 fourier coefficients will be the same value. If they are not identical the data will be sent
 to matlab and the coefficient checked and a to what degree the coefficient is decreased
 to.

4.3.2. Compression [CRTBRA002]

The compression subsystem requires that data be compressed and decompressed.

The sub system is required to:

- Compress the data
- Decompress the data
- Data integrity after decompression must be upheld (file contents must not have changed).
- The file size of the data must have reduced in size to around 50% of the original file size
- The CPU and RAM usage must be kept to a minimum in order to reduce battery consumption
- At least 25% of the Fourier coefficients must be extracted

- The size of the data files will be checked using the following command in terminal "Is
 -Ih". Check the size of the file.
- Then the file data_1.txt was compressed using the command "gzip data_1.txt"
- Using the command "Is-Ih" the size of the file will be checked to see if the size of the file has actually decreased. This will indicate that the algorithm has worked. As well as to check that the file size has been reduced to less than 50%.
- Next the avg CPU usage will be checked "sudo iostat -c 2 2" This is used to check that the CPU usage is below 50%
- Next the command "gunzip data_1.txt.gz" was run this will then decompress the file thus returning the original file.
- Next the avg CPU usage will be checked "sudo iostat -c 2 2" was run to check how the CPU usage was affected after the decompression was done. This is used to check that the CPU usage is below 50%
- The command "*ls-lh*" was executed to check the size of the file and compare it to the size of the original data file. To check if the data has been restored to the original size.
- Lastly this file will be taken onto a windows operation based laptop to compare the integrity of the files. Using the command " if((Get-FileHash "C:\Users\Adarsh\Desktop\EEE3097S\CHeck\Data1.txt").hash -eq (Get-FileHash "C:\Users\Adarsh\Desktop\EEE3097S\CHeck\Data2.txt").hash) {"Both the files are identical"} else {"Both file are NOT identical"}" in windows powershell. This command will return if the files are identical or not.
- If the files are identical this will mean that the fourier plot will be the same. So the
 fourier coefficients will be the same value. If they are not identical the data will be sent
 to matlab and the coefficient checked and seen to what degree the coefficient is
 decreased to.

4.3.3. Encryption [THMNAT011]

The encryption subsystem requires that data be encrypted and decrypted.

The sub system is required to:

- Encrypt the data
- Decrypt the data
- Data integrity after decryption must be upheld (file contents must not have changed).
- The CPU and RAM usage must be kept to a minimum in order to reduce battery consumption

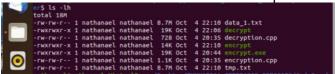
- The size of the data files will be checked using the following command in terminal "Is -Ih". Check the size of the file.
- Next the avg CPU usage will be checked "sudo iostat -c 2 2" this is to be used to check the CPU usage at the idle state.
- After the encryption.cpp file has been compiled using the command "g++ encryption.cpp -o encrypt" the following command was run "./encrypt" which will prompt the user to enter the file that needs to be encrypted the file entered will be data_1.txt.gz this will then encrypt the file.
- Then the contents of the file will be checked using "sudo nano data_1.txt" this will show if the data within the file has been encrypted.
- Next the avg CPU usage will be checked "sudo iostat -c 2 2" This is used to check that the CPU usage is below 50%
- Now that the data file is on the raspberry pi. The decryption.cpp will be executed similarly to the encryption.cpp using the command "g++ decryption.cpp -o decrypt" after this has been executed and compiled properly
- The command "./decrypt" will be run this will then prompt the user to enter the file they
 want to decrypt. To which the file "data_1.txt" will be entered after the decryption is
 complete.
- Next the avg CPU usage will be checked "sudo iostat -c 2 2" This is used to check that the CPU usage is below 50%
- Then the contents of the file will be checked using "sudo nano data_1.txt" this will show if the data within the file has been decrypted.
- Lastly this file will be taken onto a windows operation based laptop to compare the integrity of the files. Using the command " if((Get-FileHash "C:\Users\Adarsh\Desktop\EEE3097S\CHeck\Data1.txt").hash -eq (Get-FileHash "C:\Users\Adarsh\Desktop\EEE3097S\CHeck\Data2.txt").hash) {"Both the files are identical"} else {"Both file are NOT identical"}" in windows powershell. This command will return if the files are identical or not.

4.4. Results

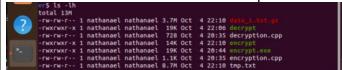
4.4.1. System

These are the results from the remote computer (running Ubuntu Linux) during file compression and encryption:

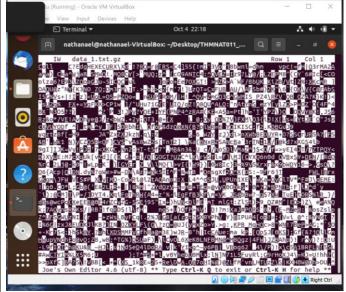
- Is -Ih #Check file size before compression



- gzip data_1.txt #Compression
- Is -Ih #Check file size after compression



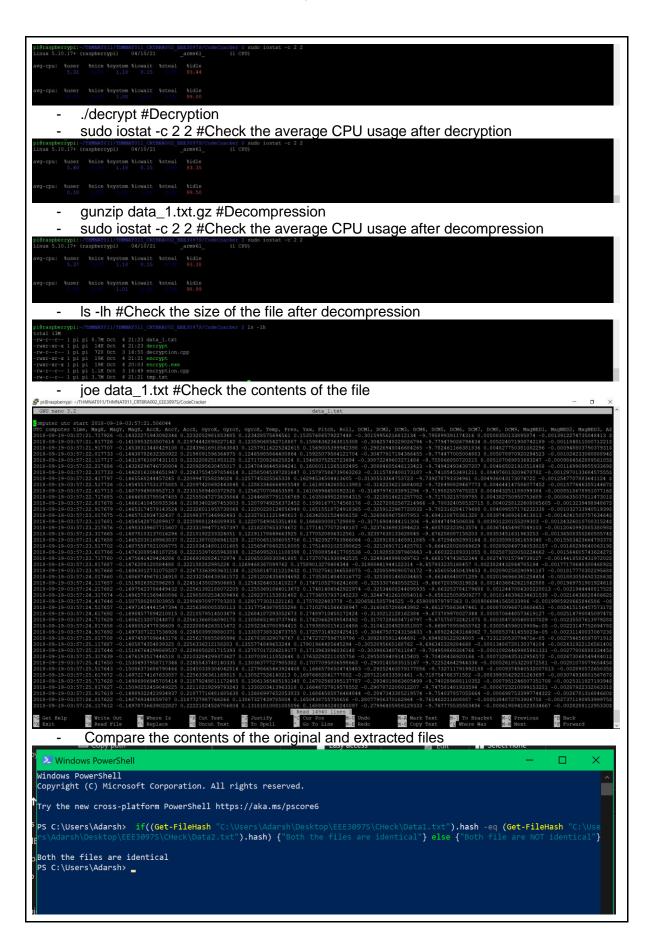
- ./ encrypt #Encryption
- joe data_1.txt.gz #Check the contents of the file after encryption



Then from there the data was added and committed to our public GitHub repository where it will be pulled by the Raspberry Pi Zero W.

These are the results that were collected from the Raspberry Pi Zero W after decryption and decompression:

- sudo iostat -c 2 2 #Check the average CPU before decryption and decompression



4.4.2. Compression [CRTBRA002]

These are the results of the independent testing of the compression algorithm on the Raspberry Pi Zero W:

Is -lh #Check the file size before compression

- sudo iostat -c 2 2 #Check the average CPU usage before compression

```
pi@raspberrypi:~/THMNAT011/THMNAT011_CRTBRA002_EEE3097S/CodeCracker $ sudo iostat -c 2 2
Linux 5.10.17+ (raspberrypi) 04/10/21 __armv61_ (1 CPU)

avg-cpu: %user %nice %system %iowait %steal %idle
3.23 0.00 1.24 0.21 0.00 95.32

avg-cpu: %user %nice %system %iowait %steal %idle
0.50 0.00 1.00 0.00 98.51
```

- gzip data_1.txt #Compression
- sudo iostat -c 2 2 #Check the average CPU usage after compression

- Is -Ih # Check the file size after compression

nano data_1.txt.gz #Check the file contents after compression



4.4.3. Encryption [THMNAT011]

These are the results of the independent testing of the encryption algorithm on the Raspberry Pi Zero W:

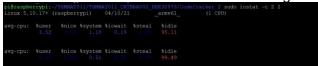
- Is -Ih #Check the file size before encryption



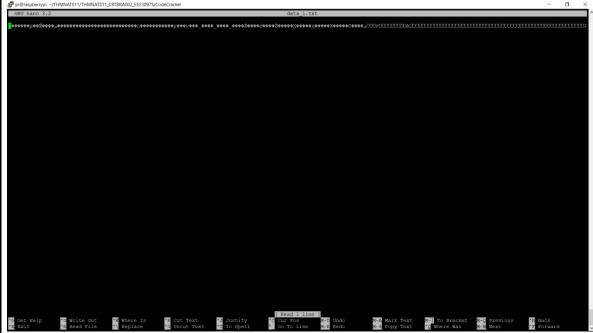
- sudo iostat -c 2 2 #Check the average CPU usage before encryption



- ./encrypt #Encryption
- sudo iostat -c 2 2 #Check the average CPU usage after encryption



- nano data_1.txt #Check the file contents after encryption



- gunzip data_1.txt.gz #Decryption
- sudo iostat -c 2 2 #Check the average CPU usage after decryption



- Compare the contents of the original and extracted files

5. Validation using a different-IMU

5.1. Need for Hardware-Based Validation

The reason behind hardware based validation is to check that the software based system can be implemented in the real-world and is not just a valid solution from a software perspective. It is also vital that we are able to generate our own data from the surrounding environment in real-time and not just use data provided to us, as the SHARC Buoy is required to generate data from the wave movement and pancake ice in real-time too. The software design also needs to be feasible and cost-effective in reality, this is another reason why a hardware solution must be implemented.

5.2. Steps

Below are the steps taken during the hardware-based validation:

- Generated data using the attached Sense-HAT B
- Compress and encrypt the data on the Raspberry Pi Zero W
- Send the compressed and encrypted data across to a remote computer using GitHub
- Decrypt and decompress the data on the remote laptop
- The original file and extracted file are compared to check for data loss
- Average CPU usage is monitored throughout the execution of the algorithms
- Files of varying sizes were used to test the algorithms (only one case was shown in the results section to save space and avoid repetition).

The data being used in this section is the data generated from the Sense-HAT B attached to the Raspberry Pi Zero. The data is extracted in the form of a text file format (.txt) for ease of use when executing compression and encryption algorithms. The datasets can have a sample size of any size, this is determined by the user. However the larger the dataset, the longer it will take to generate the file. File generation has an average rate of 2.5 samples per second.

5.3. Experimental Setup

5.3.1. System

The overall system requires that data from a an IMU sensor be encrypted and compressed on a processing block (in our case, the Raspberry Pi Zero) and sent to a remote location where it can be decrypted and decompressed.

The sub system is required to:

- Compress and encrypt data from on the processing block
- Send the data to a remote location in real-time

- Decrypt and decompress the data on the remote computer
- Data integrity after decryption and decompression must be upheld (file contents must not have changed).
- The file size of the data must have reduced in size to around 50% of the original file size
- At least 25% of the Fourier coefficients must be extracted
- The CPU and RAM usage must be kept to a minimum in order to reduce battery consumption

- Use "sudo python ICM20948.py" to run the program that generates data from the sensors on the IMU. After this command is run the user must enter the sample size in the next line for the program to begin (for example, "15000"). To append this data to a file, use the command "sudo python ICM20948.py > 15k_data.txt".
- Use a virtual machine to run Ubuntu Linux this would act as a remote server.
- The raspberry pi would simulate the IMU.
- The size of the data files will be checked using the following command in terminal "Is
 -Ih".
- Then the file data_1.txt was compressed using the command "gzip 15k_data.txt"
- Using the command "Is-Ih" the size of the file will be checked to see if the size of the file has actually decreased. This will indicate that the algorithm has worked. As well as to check that the file size has been reduced to less than 50%.
- After the encryption.cpp file has been compiled using the command "g++ encryption.cpp -o encrypt" the following command was run "./encrypt" which will prompt the user to enter the file that needs to be encrypted the file entered will be data_1.txt.gz this will then encrypt the file.
- Then the contents of the file will be checked using "sudo joe 15k_data.txt.gz" this will show if the data within the file has been encrypted.
- From there the data will be added, committed and pushed to the git repository.
- The raspberry pi will be setup and it will pull from the git repository. Using the git pull command.
- Next the avg CPU usage will be checked "sudo iostat -c 2 2" This is used to check that the CPU usage is below 50%
- This will take not of what the average CPU usage is when the pi is idle
- Now that the data file is on the raspberry pi. The decryption.cpp will be executed similarly to the encryption.cpp using the command "g++ decryption.cpp -o decrypt" after this has been executed and compiled properly
- The command "./decrypt" will be run this will then prompt the user to enter the file they
 want to decrypt. To which the file "15k_data.txt.gz" will be entered after the decryption
 is complete.
- Next the avg CPU usage will be checked "sudo iostat -c 2 2" this was done to check how the average CPU usage changed after this command was run. This is used to check that the CPU usage is below 50%
- Next the command "Is -Ih" was run to check that the size of the file had not changes since the decryption algorithm was executed.
- Next the command "gunzip 15k_data.txt.gz" was run this will then decompress the file thus returning the original file.

- Next the avg CPU usage will be checked "sudo iostat -c 2 2" was run to check how the CPU usage was affected after the decompression was done. This is used to check that the CPU usage is below 50%
- The command "Is-Ih" was executed to check the size of the file and compare it to the size of the original data file. This will be an indicator to show that all process have worked and the original file can be retrieved from all these processes.
- Lastly this file will be taken onto a windows operation based laptop to compare the integrity of the files. Using the command " if((Get-FileHash "C:\Users\Adarsh\Desktop\Git\THMNAT011_CRTBRA002_EEE3097\check\15k_data.txt").ha sh -eq (Get-FileHash "C:\Users\Adarsh\Desktop\Git\THMNAT011_CRTBRA002_EEE3097\check\data1.txt").hash) {"Both the files are identical"} else {"Both file are NOT identical"}" in windows powershell. This command will return if the files are identical or not.
- If the files are identical this will mean that the fourier plot will be the same. So the
 fourier coefficients will be the same value. If they are not identical the data will be sent
 to matlab and the coefficient checked and a to what degree the coefficient is decreased
 to.

Comparison of the sub-systems individually the tests were run purely on the raspberry pi.

5.3.2. Compression [CRTBRA002]

To avoid repetition in the experimental setup, this section will be left out, therefore refer to Section 4.3.2 for the independent setup of the compression algorithm. This conclusion was made because there is little to no variation in the independent setup of the compression algorithm.

5.3.3. Encryption [THMNAT011]

To avoid repetition in the experimental setup, this section will be left out, therefore refer to Section 4.3.3 for the independent setup of the encryption algorithm. This conclusion was made because there is little to no variation in the independent setup of the encryption algorithm.

5.4. Results

5.4.1. System



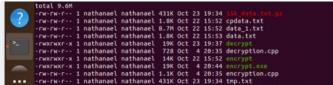


Then from there the data was added and committed to our public GitHub repository where

it will be pulled by a remote laptop running Ubuntu Linux.

These are the results that were collected from the laptop after decryption and decompression:

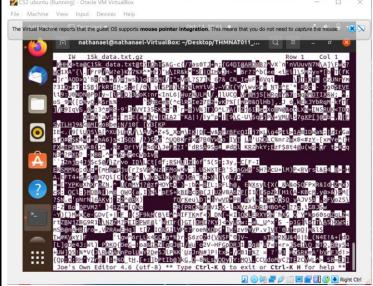
- Is -Ih #Check the file is there and data is compressed.



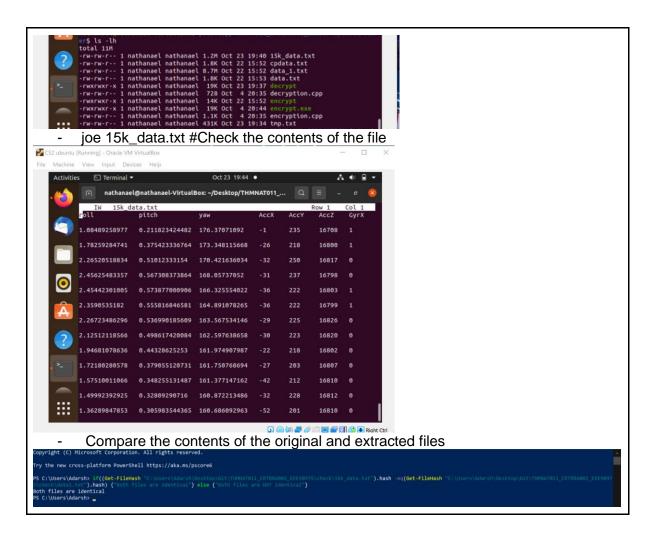
- joe 15k_data.txt.gz #Check the contents of the compressed and encrypted file



- ./decrypt #Decryption
- joe 15k_data.txt.gz #Check the contents after decryption



- gunzip 15k_data.txt.gz #Decompression
- Is -Ih #Check the size of the file after decompression



5.4.2. Compression [CRTBRA002]

To avoid the repetition of results this section will be left out, therefore refer to Section 4.4.2 for the independent testing of the compression algorithm. This conclusion was made because the compression algorithm yields the same results for data sets of varying sizes and this can be seen from the system results derived in Sections 4.4.1 and 5.4.1.

5.4.3. Encryption [THMNAT011]

To avoid the repetition of results this section will be left out, therefore refer to Section 4.4.3 for the independent testing of the encryption algorithm. This conclusion was made because the encryption algorithm yields the same results for data sets of varying sizes and this can be seen from the system results derived in Sections 4.4.1 and 5.4.1.

6. Consolidation of ATPs and Future Plan

6.1. Analysis of ATPs

Description	Acceptable	Not Acceptable	ATP Met [Y/N]	Comment
Amount of Fourier coefficients extracted	More than 25%	Less than 25%	Υ	The original file is identical to the extracted file (according to the windows powershell). This validates that 100% of the Fourier coefficients are available because there is no change to the data in the extracted file.
Size of data set after compression	Data set is reduced to 50% or more of original size	Data set is not reduced to 50% or more of the original size	Y	The original size of the file containing the data was 1.2Mb after gzipping the file thus compressing it the size of the file was 431 kb. This shows the the file has been reduced to 35.91% of its original size
Integrity of the data set after encryption	Outputted data set values are the same as the original values	Outputted data set values differ from the original values	Y	The integrity of the data is still there as after checking using the windows powershell it is seen that both files are identical.
Integrity of the data set after compression	Data set values remain the same as original values	Data set values differ from the original values	Υ	From the figures showing the windows powershell. With the result being that the files are identical.
CPU consumption	CPU usage is less than 50%	CPU usage is more than 50%	Υ	As can be seen in the figures above the average CPU usage in % barely went above 1.5 and the max when viewing one of the text files resulted in the average CPU usage to reach 2.05.

Table 5: Demonstrates if acceptance test procedures have been met or not.

6.2. Change to Specifications

The system functioned as expected and the results demonstrated that the specifications and requirements were met, therefore it was determined that there was no need to change the specifications set out in the previous report.

6.3. Future Plan

Some factors we would have to take into consideration when implementing our system in the real-world system (a SHARC Buoy):

- We would need to figure out a method that will automate the entire system as it will be
 placed in the SHARC buoy in a remote location in the Antarctic, and won't be able to
 be accessed directly using a wired connection.
- A replacement of the GitHub repository to something that makes use of the Iridium satellite networks.
- As our Raspberry Pi Zero W and Sense-HAT B are not waterproof and cannot operate
 in extreme cold, we would need to work on creating a casing that can protect it from
 water damage and the ice cold Antarctic temperatures.
- Looking for an external power source that will be able to last for extended periods of time as our system will be left for months on end, before the power source can be replaced.
- Looking for a better sensor HAT and processing unit that would be able to function at temperatures below zero.

7. Conclusion

The driving factor behind the compression of our data was because sending large data files would result in higher power consumption, and transferring data using Iridium satellite networks is very expensive. Because the data is also supposed to be confidential, the encryption proved to be an important component of the design. Only people who had access to a key would be able to decrypt and read the data.

The algorithm that was chosen to execute encryption and decryption was the AES algorithm and this allowed for for easy implementation and design due to it being a relatively simple algorithm to design. LZ77 and Huffman coding was chosen for compression and decompression, using the gzip command to execute it. It proved to have the best computation time when compared to other algorithms. Because this implementation was already available as a command on both the Raspberry Pi Zero W and Linux based system it proved to work efficiently and effectively. No data was lost during compression or decompression. The algorithms also follow lossless compression and this made it easy to determine how many Fourier coefficients were made available after compression and decompression. From the results, it can be seen that the file before the compression, encryption, decryption and decompression and the file after this process where identical. This shows that the algorithms chosen didn't result in any of the data being changed.

The reasons behind why we chose C++ as the language of choice, was because it was the fastest and most efficient when compared to other languages (algorithms were able to execute in less than 3 seconds). Whereas, languages such as python took more than 30 mins to execute the encryption and compression. This would not allow us to meet our requirement concerning minimal computation so as to not result in heavy power consumption.

All the ATPs were met, and system requirement and specifications were implemented as stated in the allocated sections above.

Recommendations:

 Run software-based and hardware-based simulations in varying temperature conditions to identify the operating range of the system.

- Attach a the device to a buoy in a pool so as to simulate some of the conditions that the real-world SHARC buoys would experience, this would make data generation more accurate.
- Automate file generation, compression and encryption so as to better emulate the realworld situation (the SHARC Buoy will be in a remote location and can't be accessed directly).

8. References

- [1] Codescracker.com. 2021. *C++ Program to Encrypt and Decrypt a File*. [online] Available at: https://codescracker.com/cpp/program/cpp-program-encrypt-file.htm [Accessed 5 October 2021].
- [2] B. Carthew and N. Thomas, "Submission 2: First Progress Report," Oct. 2021.
- [3] Waveshare.com. 2021. Sense HAT (B) Waveshare Wiki. [online] Available at: https://www.waveshare.com/wiki/Sense_HAT_(B) [Accessed 24 October 2021].
- [4] Product.tdk.com. 2021. World's First Wide-Range 6-Axis MEMS MotionTracking™ Device for Sports and High Impact Applications. [online] Available at: https://product.tdk.com/system/files/dam/doc/product/sensor/mortion-inertial/imu/data_sheet/ds-000192-icm-20649-v1.0.pdf [Accessed 24 October 2021].
- [5] J. Thakkar, "Types of Encryption: 5 Encryption Algorithms & How to Choose the Right One", *Hashed Out by The SSL Store™*, 2021. [Online]. Available: https://www.thesslstore.com/blog/types-of-encryption-encryption-algorithms-how-to-choose-the-right-one/. [Accessed: 03- Sep- 2021].
- [6] "Lossy compression and Lossless compression algorithms", *PeaZip file archiver utility, free RAR ZIP software*, 2021. [Online]. Available: https://peazip.github.io/lossless-compressionlossy-compression.html. [Accessed: 03- Sep- 2021].
- [7] "Lossy vs Lossless | Differences, Types, Benefits & Examples", *Teach Computer Science*, 2021. [Online]. Available: https://teachcomputerscience.com/lossy-vs-lossless/. [Accessed: 03- Sep- 2021].
- [8]"Difference Between Lossy Compression and Lossless Compression", *BYJUS*, 2021. [Online]. Available: https://byjus.com/gate/difference-between-lossy-compression-and-lossless-compression/. [Accessed: 05- Sep- 2021].
- [9]S. Khayam, "The Discrete Cosine Transform(DCT): Theory and Application", *Egr.msu.edu*, 2003. [Online]. Available: https://www.egr.msu.edu/waves/people/Ali_files/DCT_TR802.pdf. [Accessed: 03- Sep- 2021].
- [10]M. Nelson, "Data Compression With Arithmetic Coding", *Mark Nelson*, 2014. [Online]. Available: https://marknelson.us/posts/2014/10/19/data-compression-with-arithmetic-coding.html. [Accessed: 03- Sep- 2021].
- [11]"Huffman Coding", *Tutorialspoint.com*. [Online]. Available: https://www.tutorialspoint.com/huffman-coding. [Accessed: 05- Sep- 2021].
- [12]I. News and I. News, "5 Common Encryption Algorithms and the Unbreakables of the Future", *StorageCraft Technology*, *LLC*, 2021. [Online]. Available: https://blog.storagecraft.com/5-common-encryption-algorithms/. [Accessed: 05- Sep- 2021].