



The University of Nottingham

Faculty of Computer Science and Engineering

School of Electrical and Electronic Engineering

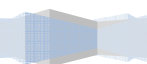
Efficient Compression and Encryption of SMS Messages

Author : Amar Faiz Zainal Abidin (001723)
Supervisor : Dr. Lim Wee Gin
Moderator : Dr. Mutasim Nour

Abstract

Recently, the usage of SMS (Short Message Service) has not only been limited to public consumers. Its potential as Business-to-Business (B2B) communication, as well as Machine-to-Machine (M2M) communication has been a hot topic among business and Information Technology (IT) sectors. The major drawbacks of SMS technology that it has maximum of 160 characters for one SMS and lack of security feature for B2B communication.

This thesis is written to address both problems above by exploring 3 new ideas in SMS technology. First, to take advantage of current SMS Protocol Data Unit (PDU) which can be utilized to let the programmer to devise a scheme to compress the SMS and at the same time add encryption feature to the make the sent SMS less vulnerable to attack from hackers. Second, is to introduce the reader to a range of compression methods and implement them in SMS technology. Third, to explore the encryption technology and the feasibility study of implementing them as a part of SMS technology.



Acknowledgement

I would like to take this opportunity to express my gratitude to my supervisor Dr. Lim Wee Gin for his guidance and assistance in the completion of this project. I would also like to extend my appreciation to last year's University of Nottingham Malaysia Campus Final Year Project Student, Lau Yee Kuan, who has laid down the thesis for this project. My special thanks to my colleagues who are Sudheeran Sivathanan (second year Electrical and Electronic), Adam Aris (second year Electrical and Electronic) and Ganeish Velmurugan (first year Electrical and Electronic). These people have helped me to the completion of this project and thesis. Last but not least, I would also like to thank my family members for their moral support throughout the duration of this project.

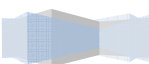


Table of Contents

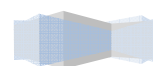
Chapter 1: Introduction

1.0	Overview	07
1.1	Short Message Service (SMS)	07
1.2	Data Compression	08
1.3	Data Encryption	09
1.4	Hardware	09
1.5	Project Outline	10
1.6	Project Objectives	11
1.7	Project Deliverable	12
1.8	Thesis Structure	12

Chapter 2: Literature Review

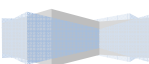
2.0	Overview	13
2.1	Compression	13
2.1.1	Uncompressed and compressed data	13
2.1.2	Huffman coding	17
2.1.2.1	Generating Huffman tree	17
2.1.2.2	Encoding and decoding the Huffman Coding	19
2.1.2.3	Discussion on Huffman Coding	19
2.1.3	Shannon-Fano Coding	19
2.1.3.1	Generating Shannon-Fano Tree	19
2.1.3.2	Encoding and decoding the Shannon-Fano Coding	21
2.1.3.3	Discussion on Shannon-Fano Coding	21
2.1.4	Arithmetic Coding	22
2.1.4.1	Encoding and decoding the Arithmetic Coding	22
2.1.4.2	Discussion on Arithmetic Coding	25
2.1.5	Modified Arithmetic Coding	26
2.1.5.1	Modified Arithmetic coding and decoding	26
2.1.5.2	Discussion on Modified Arithmetic Coding	30
2.2	Encryption	31
2.2.1	Unencrypted and Encrypted Data	31
2.2.2	Symmetric Key Encryption	32
2.2.2.1	Traditional Encryption	32
2.2.2.1.1	Polybius Checkerboard Encryption	32
2.2.2.1.2	Blaise de Vigenere Encryption	32
2.2.2.1.3	Wheatstone-Playfair Encryption	34
2.2.2.1.4	Discussion on Traditional Encryption	35
2.2.2.2	Advances Encryption Standard (AES)	35
2.2.3	Asymmetric Key Encryption	38
2.2.3.1	RSA Encryption	39
2.2.3.1.1	Generating public key and private key	40
2.3	Communication	40
2.3.1	AT Command	42
2.3.2	Protocol Data Unit (PDU)	43
2.3.2.1	Parameter Descriptions	43

2.3.2.1.1 Service Centre Address (SCA)	44
2.3.2.1.2 Protocol Data Unit Type (First Octet)	44
2.3.2.1.3 Message Reference Number (TP-MR)	44
2.3.2.3.4 Destination Address (DA) Originator Address (OA)	44
2.3.2.3.5 Protocol Identifier (TP-PID)	44
2.3.2.3.6 Service Center Time Stamp (STCS)	45
2.3.2.3.7 Validity Period (TP-VP)	45
2.3.2.3.8 User Data Length (TP-UDL) and User Data (TP-UD)	46
2.3.3 Example on using AT Command (GSM Modem) and PDU to send single and concatenated SMS	46
2.3.3.1 Sending single SMS in Text format	48
2.3.3.2 Sending single SMS in PDU format	49
2.3.3.3 Sending concatenated SMS in PDU format	50
2.3.3.4 Receiving single SMS in Text format	53
2.3.3.5 Receiving single SMS in PDU format	54
2.3.3.6 Receiving concatenated SMS in PDU format	55
Chapter 3: Methodology	
3.0 Overview	58
3.1 Compression	60
3.1.1 Modeling	60
3.1.2 Compression Dictionary	63
3.1.3 Compression Method	66
3.2 Encryption	66
3.3 Communication	67
3.4 The software	68
Chapter 4: Testing, Result and Discussion	
4.1 Testing	70
4.2 Results and Discussion	71
4.2.1 Compression	71
4.2.2 Encryption	77
4.2.3 Communication	78
Chapter 5: Conclusion	
5.1 Compression	81
5.2 Encryption	82
5.3 Communication	82
5.4 Future Development	83
Reference	84



Abbreviation

SMS	Short Message Service
B2B	Business-to-Business
M2M	Machine-to-Machine
PDU	Protocol Data Unit
IT	Information Technology
MCMC	Malaysia Communications and Multimedia Commission
B2C	Business-to-Consumer
ETSI	European Telecommunications Standard Institutes
GSM	Global System for Mobile Communication
ASCII	American Standard Code for Information Interchange
SMSC	Short Message Service Centre
C2M	Consumer-to-Machine
MP3	Moving Picture Expert Groups Layer 3 Audio
AT	Advanced Technology
COM	Component Object Model
OS	Operating System
UCS2	2 bytes Universal Character Set
UCS4	4 bytes Universal Character Set
IC	Integrated Circuit
ATM	Automated Teller Machine
PIN	Personal Identification Number
MAC	Message Authentic Code



Chapter 1

Introduction

1.0 Overview

Due to the aggressive marketing of telecom providers in Malaysia in recent years, the cost of sending of SMS has been so affordable. This has led to an increase of SMS users. According to a study conducted by Malaysia Communications and Multimedia Commission (MCMC), SMS still plays a dominant role among Malaysians [1]. The usage of SMS among telecom subscribers is 84.9% from total telecom subscribers.

The same trend has been seen in the developing market such as Asia and Africa. This attracts the business sectors to use SMS as an alternative medium for electronic mail (email), to exchange information in B2B and M2M. A report from Cellular News in September 2006 clearly stated the heavy usage of SMS in B2B [2]: "95% of respondents reported receiving business communications via SMS while 75% received email communication for business purposes. The study was conducted by Webchek using their SMS channel to conduct the research."

The drawback of SMS technology as a major medium of communication in B2B is the length of SMS, which relatively shorter than email and it is not secure. This project tries to solve both problems and allow SMS to become more appealing for communication tool between B2B and Business-to-Consumer (B2C). In collaboration with an industrial partner, Mobitek Sdn. Bhd., this project tries to achieve a compression ratio of 50% under optimized conditions and encrypt the SMS to make it less vulnerable to hackers.

This project consists of three main parts:

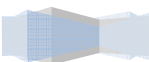
1. Compression – Apply the current compression method to reduce the size of the SMS.
2. Encryption – Perform encryption using conventional encryption method to make the SMS more secure.
3. Communication - Construct and send the appropriate format of SMS to cater for both single SMS and concatenated SMS.

1.1 Short Message Service (SMS)

Short Message Service (SMS) history started way back in 1992, where it was first used to exchange short text between telecom subscribers in Europe [3]. Due to its simplicity and cost effectiveness compared to a phone call, the usage of SMS has increased to surpass conventional phone calls. Then a standard format of SMS has been set by European Telecommunications Standard Institutes (ETSI), which is documented in GSM 03.40 [4] and GSM 03.38 [5].

According to this format, the SMS has the following characteristics:

1. A single short message can be up to 160 characters (for a 7-bit ASCII character), 140 characters (for a 8-bit ASCII character) or 70 characters (for 16-bit Unicode).



2. SMS is a store and forward service, i.e. short messages are not sent directly from sender to recipient, but via SMS Centre (SMSC) instead. With this, SMS is able to have a special feature of having delivery status report.
3. SMS can be sent through three modes: Block mode, Text mode and PDU mode.

The main advantages of SMS compared to Internet technology are:

1. **Reliable:** The mobile network is rarely down, compared to Internet that has a higher probability of having downtime.
2. **Coverage:** mobile telecommunication coverage is much larger than the static Internet coverage.
3. **Fast:** SMS is able to reach the device/user immediately in case of any emergency.

The practical applications of SMS in B2B, consumer-to-machine (C2M) and M2M are:

1. **Home Automation:** The user can control the electrical appliances from anywhere. For example, user can switch on the air-conditioner before he reaches home.
2. **ATM machine:** In the case of any problems with the ATM machines, central monitoring system will be notified immediately via SMS.
3. **Vending Machine management:** The Vending Machine can send SMS when there is a shortage of cans.

1.2 Data Compression

According to the author of A Guide to Data Compression Methods, David Salomon, data compression is the process of converting an input data stream into another data stream that has a smaller size [6]. It reduces redundancy in a message to decrease the size of the message [7].

Data compression consists of two parts which are modeling and coding (compression method):

1. **Modeling:** This part is where the effectiveness of the algorithm to detect the text and create rules that allows the coding to compress the text effectively.
2. **Coding:** This part where the application of scholar compression method, such as Huffman and Arithmetic coding being implemented.

Figure below gives a better idea of the relationship modeling and coding [8]:

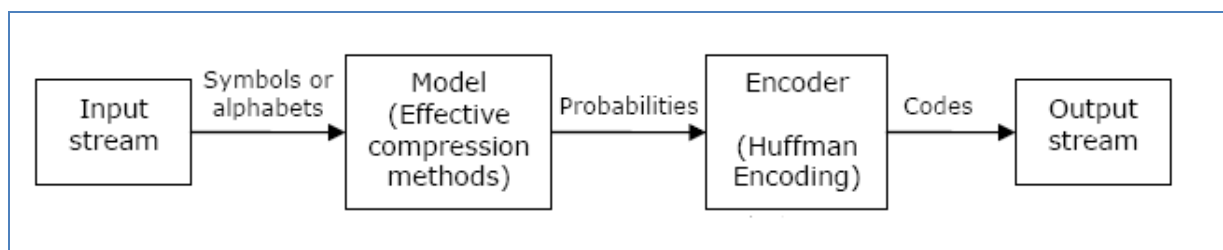
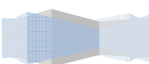


Figure 1.1: Relationship between modeling and coding in data compression

There are two types of data compression which are:



1. Lossless data compression: The exact original data will be obtained after decompression. Use in text and spreadsheet. Example, Huffman coding and Arithmetic coding.
2. Lossy data compression: There will be a certain loss of accuracy of the original data after decompression. Used in sound and image. For example MP3,

Compressing SMS text is far more challenging than compressing a text file in computer due to several constraints. This will be touched in detail in Chapter 3.

1.3 Data Encryption

Encryption can be defined as the process of converting information into a form unintelligible to anyone except holders of a specific cryptographic key [9]. In layman's term, encryption scrambles the characters inside the text.

A number of encryption methods have been introduced recently with different approaches on how to 'scramble' the text. However, they have some characteristics in common. These are listed below:

1. Rules: The rules that need to follow by the sender and recipient for a successful encryption and decryption.
2. Key: The secret text that the sender and recipient has set, so if other person who know the rules to encrypt and decrypt the text will not be able to read the cipher text because the text is encrypted using the key.

There has been a great movement in business sectors, to employ encryption technology into their information area. This to protect their privacy and restricts unauthorized personal to manipulate the information.

1.4 Hardware

Figure below shows the only hardware used in this project which is Wavecom GSM Modem that will be used to send the compressed and encrypted SMS.



Figure 1.2: Wavecom GSM Modem

This GSM Modem can be communicates using AT Command in Hyper Terminal, a software included in Windows XP Operating System (OS). Another approach is to write a program and that can communicate with GSM modem directly using the COM Port.

1.5 Project Outline

Before going into the objectives of the project, figure below illustrates the overall of this project all about:

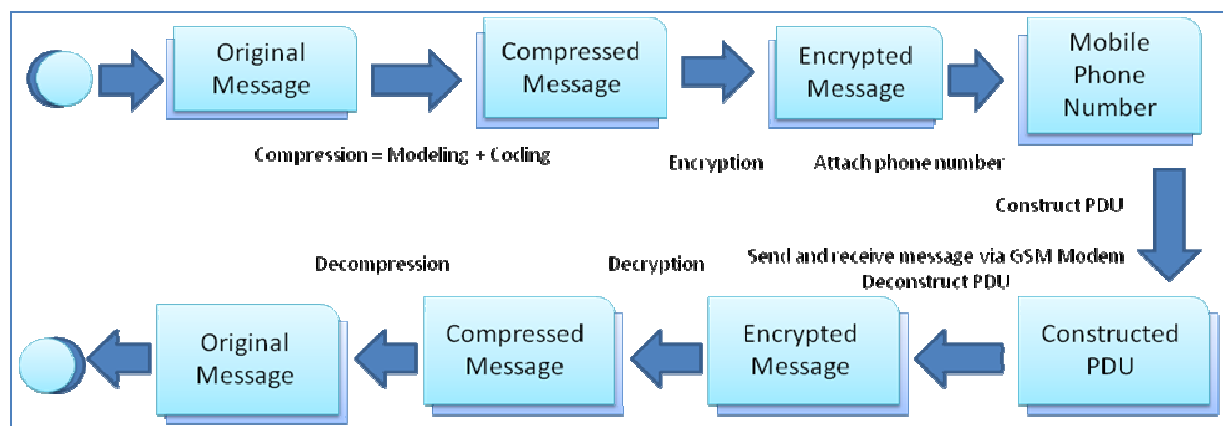


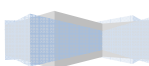
Figure 1.3: Illustration of the project outline

The following describes the block diagram above:

1. Perform effective compression of the message using the combination of modeling and coding.
2. The compressed message then being encrypts using a reliable encryption method.
3. The encrypted message will be converts to a defined SMS format in PDU format so it can be send to the SMSC.
4. The encrypted SMS then is send to the recipient using the GSM modem.
5. After that the SMS will be retrieved from the SMSC to the GSM modem.
6. From the defined SMS format, the PDU will be converted back to the encrypted message.
7. Then the encrypted message will be decrypt to obtain the compressed message.
8. Last step is to decompress the compressed message back to the original message.

1.6 Project Objectives

Figure 1.4 shows the objectives of the project set by the supervisor of this project:



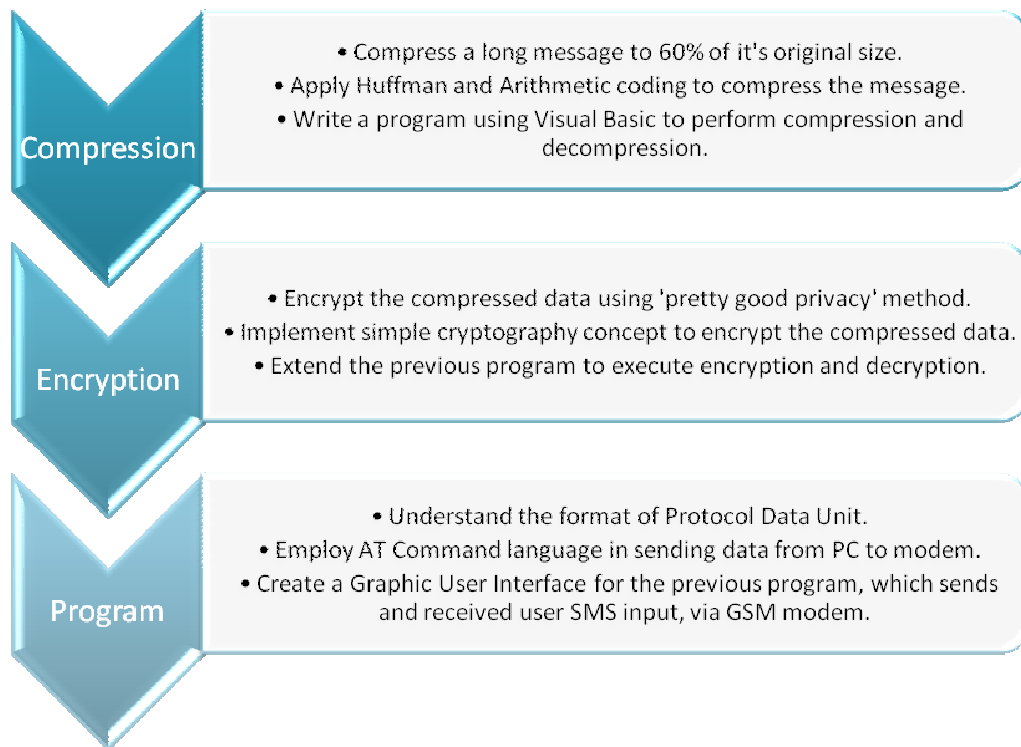
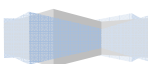


Figure 1.4: Complete objectives of the project

1.7 Project Deliverable

Upon completion of the project, the built program should cover all the objectives mentioned in Figure 1.4. At the same time, the program should also satisfy the 3F factors stated in Figure 1.5.



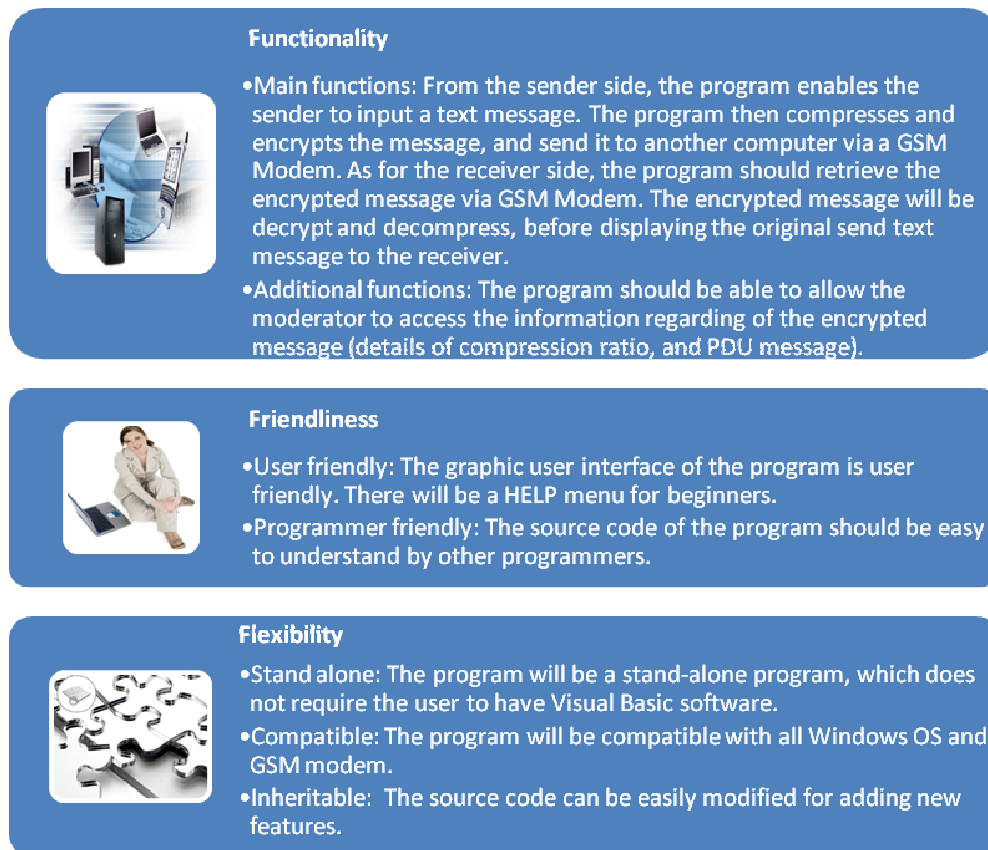


Figure 1.5: Proposed outcome of the project

1.8 Thesis Structure

This chapter has gives a glimpse through of the work present, objectives and proposed outcomes. The rest of this thesis will be devoted on the project.

Chapter 2 provides the necessary knowledge and information required to complete the project. This chapter is presented in a textbook alike (including the example and discussion).

Chapter 3 explains the details on the implementation of the literature review in Chapter 2, into the real project. This chapter also explained on the testing that had been done on the program written.

Chapter 4 displays the result of the project and also discussion of the project outcome.

Finally, Chapter 5 provides the verdict of the thesis and the possible future development of the project.

