A DATA CONFIDENTIALITY APPROACH TO SHORT MESSAGE SERVICE (SMS) ON ANDROID

KAUNG HTET MYINT

M.C.Sc. **JANUARY 2019**

A DATA CONFIDENTIALITY APPROACH TO SMS
ON ANDROID

By

KAUNG HTET MYINT

B.C.Sc. (Hons:)

A Dissertation Submitted in Partial Fulfillment of the Requirements
For the Degree
of
Master of Computer Science
(M.C.Sc.)

University of Computer Studies, Yangon.

JANUARY 2019

ACKNOWLEDGEMENTS

Firstly, I would like to express my gratitude and sincere honor to all persons who gave me a lot of help and support for the successful completion of this thesis.

I would like to express my gratitude to **Prof. Dr. Mie Mie Thet Thwin,** the Rector, the University of Computer Studies, Yangon, for her kind permission to do this thesis.

I am heartily thankful to my supervisor, **Dr. Tun Myat Aung**, Professor and Head of Network Security Lab, the University of Computer Studies, Yangon, for his encouragement, invaluable guidance and strong technical support from the initial to the final level enabled me to develop an understanding of the subject.

My heartfelt thanks and respect go to **Dr. Thi Thi Soe Nyunt**, Professor and Head of Faculty of Computer Science, for her kind administrative support as a course coordinator throughout the development of the thesis.

I am indebted to **Daw Ni Ni San**, Lecturer, Department of Language, University of Computer Studies, Yangon, for her guidance and editing the thesis from the language point of view.

It is a pleasure to thank my family because I would have not finished this thesis without the support of my family members who has always been there for me whenever I need them, the encouragement they gave to keep me going and their love to empower me.

I would also thank our teachers and friends from the University of Computer Studies, Yangon, for their thoughtful and help for my seminar presentation. I wish to acknowledge with deep gratitude for the valuable guidance received from our teachers who gave advice for the study.

Lastly, I offer my regards and blessings to all of those who supported me in any respect during the completion of the thesis.

Statement of Originality

I hereby certify that the work embodied in this thesis is the result of original research and has not been submitted for a higher degree to any other University or Institution.

8/1/2019
Date

Kaung
Kaung Htet Myint

ABSTRACT

Short Message Service (SMS) is a text messaging service component of mobile communication systems. It uses standardized communications protocols to exchange short text between mobile devices. SMS does not have any built-in procedure to offer security for the text transmitted as data. Most of the applications for mobile devices are designed and developed without taking security into consideration. As confidentiality is the original purpose of cryptology, the proposed system is introduced a data confidentiality approach to SMS on Android. It includes SMS network architecture as well as cryptographic protocols as theory background and it also deals with design, implementation and confidentiality assessment of RC4 stream cipher for SMS data confidentiality on mobile networks. In practical use, SMS messages are not encrypted by default during transmission. To secure the message, the receiver needs to put password. In the thesis, RC4 stream cipher is implemented by using Key-Scheduling Algorithm (KSA) and Pseudo-Random Generation Algorithm (PRGA) in Java programming language.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF EQUATIONS

CHAPTER 1
INTRODUCTION

1.1 Introduction to Data Confidentiality and Short Message Service

Data confidentiality means allowing only authorized users to access protected data. It is the integral part of information security. It is concerned with both storage and transmission of information. The user has to protect the sensitive information from spiteful actions while sending Short Message Service (SMS). An encipherment security system can be used for the security of data confidentiality.

SMS is a means of delivering short messages across the mobile networks. It initially stores and then sends messages among mobiles. The text-only message from the sending device is kept in a central short message center. It subsequently forwards it to the receiving mobile. SMS is currently supported by GSM (Global System for Mobiles), CDMA (Code Division Multiple Access) and TDMA (Time Division Multiple Access).

The principal aim of SMS is to transmit text messages from one mobile phone to another. It is of multiple benefits to everyday activities. However, it is not presumed to be a safe and secure means when confidential information is sent by the normal SMS services. These days, there exist a number of security issues and vulnerabilities related to SMS [5][7]. Thus it is crucial to prevent the SMS message from being unlawfully intercepted by undesirable parties as well as to make sure that the message originates from the lawful sender.

The art of encoding and decoding ciphers is called cryptology. These days, private-key encryption schemes make two parties who possess confidential information able to interact over a public network, while preventing a hostile eavesdropper from accessing the contents of their communication. State-of-the-art cryptology, on the other hand, commands a much larger scope. Data privacy cryptology is now associated with other purposes such as data integrity, user authentication or more complicated security concerns.

In addition, cryptology is now also concerned with the public key setting, in which communicating users do not have any prior confidential data. Finally, the discipline of cryptography currently covers more than just the fundamentals, now ranging from the study of the rudiments of mathematics to the design and analysis of complicated systems.

Cryptology nowadays is used almost anywhere. If you buy some goods online by credit card, you have invariably used cryptography to make sure that your credit card number cannot be observed by an eavesdropper in the meantime. Apart from encryption, if you have updated Windows or Mac software, then unwittingly, you have depended on digital signatures to check the authenticity of the update. Finally, Bitcoin, the latest type of digital currency, uses encryption techniques, for its security.

Cryptography is so successful that a number of its underlying concepts have spread to other areas of computer security. To illustrate, formal threat modeling is emphasized and proofs of security within well-defined models have become possible. The introduction will deal with two basic settings of cryptography and their related primitives. The first primary setting is the private key setting in which the communicating parties possess prior secret information. Private-key encryption systems ensure confidentiality of their interaction and message authentication codes. This ensures data integrity and data cannot be inappropriately modified as a result.

The second setting dealt with is the public-key setting in which interacting parties do not have common secret information and events. Public-key encryption systems and digital signatures handle the analysis concerns. Various techniques are utilized to construct these from the relevant parameters in the private-key setting.

Cryptography refers to the science of converting ordinary information into unintelligible form. A cryptographic system changes a plain text into an encoded text using a key produced by a cryptographic algorithm. A stream cipher that is employed to protect internet traffic as part of the Secure Socket Layer (SSL) is RC4 [8]. Data confidentiality for SMS sent over mobile networks can be protected by RC4 stream cipher.

1.2 Motivation

The main purpose of the thesis is to create a data confidentiality technique to SMS (Short Message Service) on Android. SMS features in mobile communication systems. It uses standard communication protocols to send and receive short texts between mobile phones and other devices. SMS does not incorporate a procedure to accord security for the text sent as data. The majority of the applications for mobile devices are designed and implemented without taking security into account. SMS messages are not normally encrypted by default.

Confidentiality is the notion of making sure that data is not made accessible or exposed to unauthorized people. Encryption is the main approach to confidentiality. Both symmetric and asymmetric encryption can be employed. As confidentiality was the original purpose of cryptology, this thesis is introduced a data confidentiality approach to SMS on Android. It encompasses SMS network architecture as well as cryptographic protocols as theory background and it also deals with design, implementation and confidentiality assessment of RC4 stream cipher for SMS data confidentiality on mobile networks.

1.3 Objectives of the Thesis

The aim of the thesis is to offer data secrecy during the transmission of SMS messages so as to prevent them from being unlawfully intercepted by illegal parties and to ensure the authenticity of the message from the genuine sender.

1.4 Organization of the Thesis

The purpose of the thesis is to ensure data confidentiality during the period of SMS message transmission so as to protect the SMS message from illegal interception by hostile parties and to ensure that the message comes from the legitimate sender. The thesis is organized as follows:

Chapter 1 is Introduction, motivations, objectives and organization of the thesis. Chapter 2 discusses fundamental concepts of SMS technology, SMS mobile networks, introduction to cryptography and RC4 cipher. Chapter 3 presents design and development of mobile application techniques utilized to ensure data confidentiality of SMS messages sent on mobile networks. Chapter 4 deals with the usage of statistical test suites in measuring data confidentiality. Chapter 5 concludes the proposed system by data confidentiality level of pseudorandom number sequence produced by RC4 cipher and by recommending RC4 cipher suitable for use in data confidentiality of SMS messages sent on mobile networks.

CHAPTER 2
BACKGROUND THEORY

Nowadays, there exist a number of security issues and vulnerabilities related to SMS [5][8]. Security Goals, Security attacks, Basic Concepts of SMS Technology, Message Flow of SMS Network, SMS Mobile Network Communication System, SMS Packet Format and Cryptography are explained in this chapter.

2.1 Security Goals

There are three types of security goals such as confidentiality, integrity and availability. Authorized users have right to use data are meant confidentiality. Integrity stands as guarantee that data is truthful as well as precise. Availability means an assurance to consistent right to use data by intended users.

2.1.1 **Confidentiality**

Confidentiality is the notion of making sure that data is not made accessible or exposed to unauthorized people and is approximately comparable with secrecy. Measurements approximated to confirm confidential information is intended to avoid useful data from getting unauthorized users, as creating certain which authorized users can catch it. Access will be limited to intended person to interpret the facts in query. The facts to be classified in accordance with the quantity and damaged type are public. This damaged type drop into unauthorized users. Strict procedures can be fulfilled in accordance with those classifications.

A good example of system for confidentiality exists as an account number when banking online. A public process of safeguarding confidentiality is data encryption. Personal Identifications and passwords establish a usual process. Other process exits as biometric verification and security tokens.

2.1.2 **Integrity**

Sustaining the reliability, correctness, and credibility of information over its whole lifetime cycle are known as integrity. Information should not be altered in transfer. This information cannot be changed by unintended hands must be confirmed.

Measures undertaken to ensure integrity contain authorizations of file and operator access controls. To avoid incorrect alterations, unintentional removal by official person becoming a trouble may be used by version control. Checksums, for verification of integrity may be done by some data. To return the unnatural facts to its right form, redundancies must be accessible.

2.1.3 **Availability**

Availability maintains a properly working structure situation that is allowed of software conflicts. Availability is finest guaranteed by thoroughly sustaining all hardware, acting hardware maintenances directly when required. To keep current with all necessary system upgrades, it is significant. Data generated and kept by an association wants to be accessible to official entities.

2.2 Security Attacks

Three security goals (confidentiality, integrity, and availability) can be endangered by security attacks which were divided into three groups as shown in Figure 2.1.

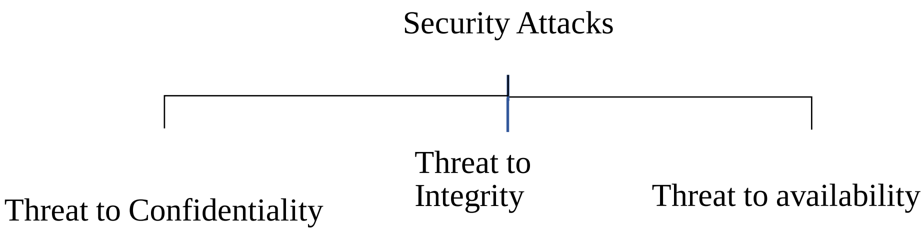


Figure 2.1 Taxonomy of attacks with relation to security goals

Threat to confidentiality is snooping and traffic analysis. Unauthorized access to or interception of data is called as snooping. A file transferred through the internet may have confidential information exists as an example. The transmission may be intercepted by an unauthorized entity and use the contents of own benefit. The data can be made non-intelligible to the intercepter by using encipherment techniques to avoid snooping.

Traffic analysis is Encipherment of data may create it non-intelligible for the intercepter, however, the intercepter can attain another type of data by checking online circulation. The intercepter can discover the electronic address (Eg: e-mail) of the users can be found as example. Couples of requests and responses can be collected to guess the nature of transaction.

Threat to integrity is modification, masquerading, replaying and repudiation. Threat to availability is denial of service.

2.3 Basic Concepts of SMS Technology

A text messaging service component of most telephones, internet, and mobile-device systems is known as SMS (short message service). Standardized communication protocols are used to permit smart phones to transfer short text messages. Short Message Service is also commonly referred to as a "text message". The user can conduct a message of up to 160 characters to another device with a SMS. In SMS, longer messages will automatically be fragmented up into several parts. This type of text messaging is supported by most cell phones.

To conduct SMS, the following steps are needed.

1. Launch the Messages application on your phone.
2. Tap on the Compose Message button.
3. Enter the phone number or name of the contact you want to text.
4. Type your message.
5. Hit Send.

The formal name for text messaging is SMS. Short Message Service is a way to conduct short, text-only messages from one phone to another. These messages are usually conducted over a cellular data network.

By cooperating with the cellular network, Short Message Service transmits text messages from one phone to other phone. These devices require Short Messaging Entities (SMEs). These are starting points (sender) and end points (receiver) for SMS messages. They never connect directly with each other [3]. They always connect with a Short Message Service Center (SMSC).

A mobile telephone can be an SME. Computer contained messaging software, which can connect directly with the SMSC of the service source, can be an SME. Two types of SMS messages conditional on the character of the device in the network are Mobile-originated (MO) messages and Mobile-terminated (MT) messages. The mobile phone sends MO messages to the SMSC and receives MT messages. These MO and MT messages are encrypted in a different way during conduction. Short Message Service Center’s functions are displayed in Figure 2.2 [9].

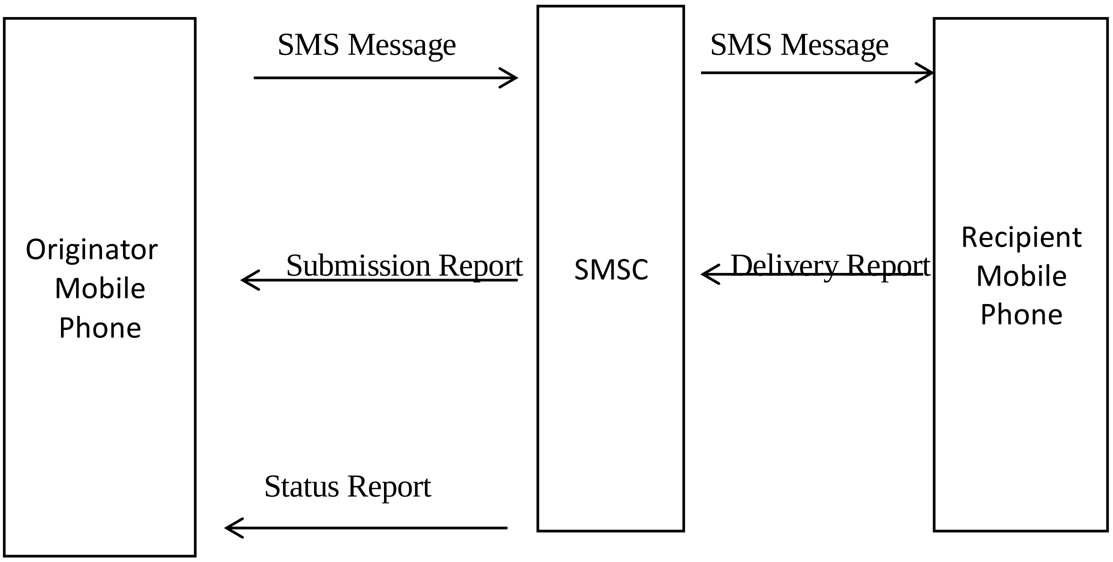


Figure 2.2 Short Message Service Center’s functions

2.4 Message Flow of SMS Network

First of all, the mobile Sender handovers the short message to Network Operator. Then, Network Operator acknowledges “Short Message” to the Sender. After that, The Network Operator also conducts “Short Message” to the SMS Center.

SMS Center acknowledges “Short Message” to the Network Operator. SMS Center sends the “Short Message” to the Network Operator too. Network Operator acknowledges the “forward Short Message” to the SMS Center.

Then, Network Operator conducts the “Short Message” to the Receiver. Receiver acknowledges the “Short Message” to the Network Operator.

The Message Flow of SMS network is as shown in Figure 2.3.

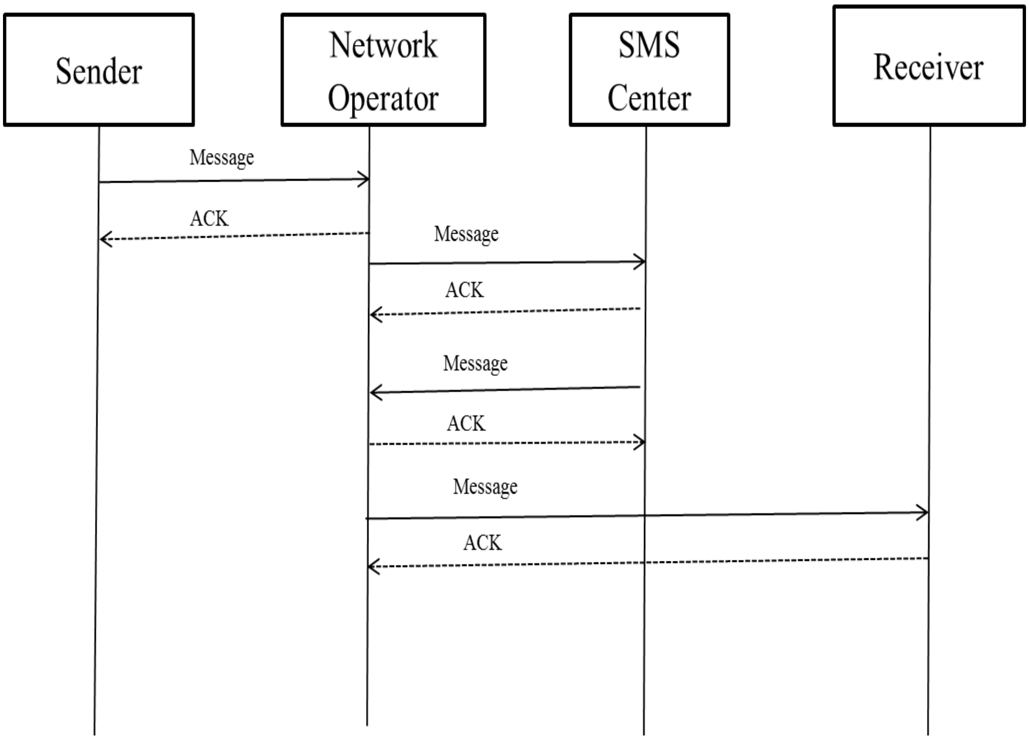


Figure 2.3 Sequence Diagram of the Message Flow of SMS Network

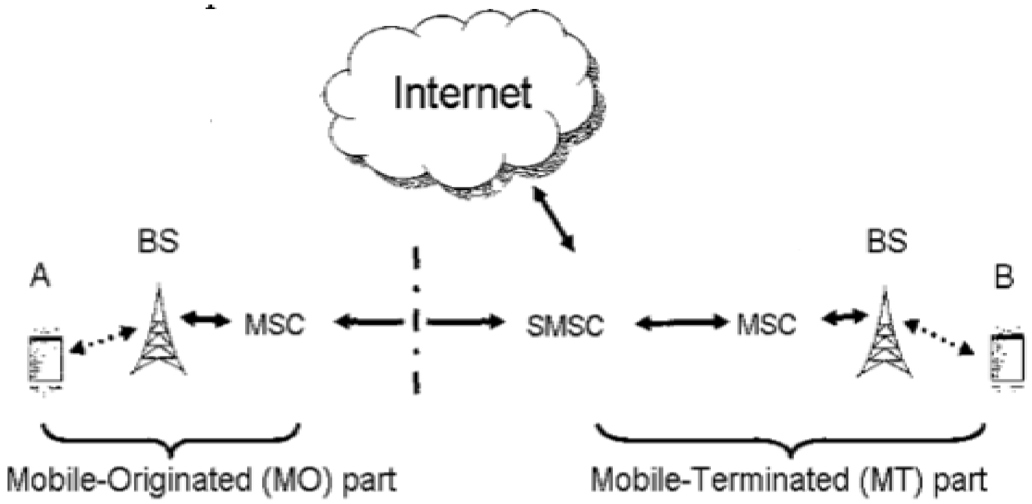
2.5 SMS Mobile Network Communication System

The Common Channel Signaling System 7 (SS7) conveys SMS messages. A worldwide standard that describes the processes and procedures for exchanging data amongst network components of wire line and wireless phone carriers known as SS7 [3]. These components use the SS7 procedure to give-and-take control data for call system, movement control, etc.

Theoretically, the common SMS mobile network architecture contains of two parts known as Mobile Originating (MO) part and Mobile Terminating (MT) part. The phone of the sender that gives the wireless structure for network part and Mobile Switching Center changes all circulation into and out of the structure in spite of the source are known as MO.

The other part contains an improper location and the termination of MSC for the phone, as well as a central stock and onward server is called SMS Centre. It is accountable for receiving information and keeping information [9].

Mobile Network Architecture for SMS communication is shown in Figure 2.4.



A - Sender

B – Receiver

BS – Base Station

MSC – Mobile Switching Center

SMSC – Short Message Service Center

Figure 2.4 Mobile Network Architecture for SMS Communication

2.6 Cryptography

Cryptography is the art and science of storing information safe. It uses calculation to encrypt and decrypt information. Cryptography can keep sensitive data or convey it through uncertain networks (Internet). Therefore, only intentional user can understand it. It is related with the procedure of changing ordinary plain text into unintelligible text and vice-versa. It is a technique of keeping and transferring information to individual form for whom it is intended can read and process it

Cryptography is the art and science of keeping message secure. It makes encoding and decoding ciphers. It is the science of information and communication security [1]. Nowadays, private-key encryption schemes make two parties who possess confidential information able to interact over a public network, while preventing a hostile eavesdropper from accessing the contents of their communication. State-of-the-art cryptology, on the other hand, commands a much larger scope. Data privacy cryptology is now associated with other purposes such as data integrity, user authentication or more complicated security concerns.

In addition, cryptology is now also concerned with the public key setting, in which communicating users do not have any prior confidential data. Finally, the discipline of cryptography currently covers more than just the fundamentals, now ranging from the study of the rudiments of mathematics to the design and analysis of complicated systems.

Cryptology nowadays is used almost anywhere. If you buy some goods online by credit card, you have invariably used cryptography to make sure that your credit card number cannot be observed by an eavesdropper in the meantime. Apart from encryption, if you have updated Windows or Mac software, then unwittingly, you have depended on digital signatures to check the authenticity of the update. Finally, Bitcoin, the latest type of digital currency, uses encryption techniques, for its security.

Cryptography is so successful that a number of its underlying concepts have spread to other areas of computer security. To illustrate, formal threat modeling is emphasized and proofs of security within well-defined models have become possible. The introduction will deal with two basic settings of cryptography and their related primitives.

The first primary setting is the private key setting in which the communicating parties possess prior secret information. Private-key encryption systems ensure confidentiality of their interaction and message authentication codes. This ensures data integrity and data cannot be inappropriately modified as a result.

The second setting dealt with is the public-key setting in which interacting parties do not have common secret information and events. Public-key encryption systems and digital signatures handle the analysis concerns. Various techniques are utilized to construct these from the relevant parameters in the private-key setting.

Cryptography refers to the science of converting ordinary information into unintelligible form. A cryptographic system changes a plain text into an encoded text using a key produced by a cryptographic algorithm. A stream cipher that is employed to protect internet traffic as part of the Secure Socket Layer (SSL) is RC4 [8]. Data confidentiality for SMS sent over mobile networks can be protected by RC4 stream cipher.

2.6.1 Cryptographic goals

Current cryptography concerns itself with the following four objectives:

1) **Confidentiality** (means allowing only authorized users to access protected information)
2) **Integrity** (without the change being detected, the information can be unchanged in storing or transfer between creator and intentional user)
3) **Non-repudiation** (sender of the information cannot refute at a later stage his or her intentions in the making or transmission of the information)
4) **Authentication** (the creator and intentional user can approve respectively their character and the endpoint of the information).

2.6.2 Basic Cryptographic Techniques

It is essential to know the main terms used in cryptology. The common terminology is described in table 2.1.

Table 2.1 Basic Cryptographic Techniques

Term	Meaning
Plaintext	The message to be sent securely from the source to the intended endpoint of the message.
Encryption	The procedure of hiding a data in disguising its material.
Ciphertext	Encrypted information that is sent over an uncertain communication medium.
Decryption	It is a procedure to revert ciphertext into plain text.
Cryptography	The art and science of storing information safe.
Cryptanalysis	The art and science of damaging ciphertext.
Cryptanalysts	Practitioners of cryptanalysis.

In plaintext or cleartext, the users can read and understand without any special actions. The procedure of hiding a data in disguising its material is called encryption**.** Encrypted information that is sent over an uncertain communication medium is called ciphertext. Decryption is a procedure to revert ciphertext into plain text.

Plaintext is usually denoted by M. Ciphertext is usually denoted by C. Like plaintext, ciphertext is also a stream of binary data. Encryption function E( ), functions on M to create C is as follow:

$$E(M)=C \tag{2.1}$$

In the reverse procedure, the decryption function D( ) functions on C to create M is as follow:

$$D(C)=M \tag{2.2}$$

Since the entire point of encrypting and then decrypting a data is to recover the original plaintext as in equation 2.3.

$$D(E(M))=M \tag{2.3}$$

Cryptography exits as the art and science of storing information safe. It uses calculation to encrypt and decrypt information. Cryptography can keep sensitive data or convey it through uncertain networks (Internet). Therefore, only intentional user can understand it.

Cryptanalysis is the art and science of damaging ciphertext while cryptography is the art and science of storing information safe.

2.6.3 Cryptographic Algorithms

Cryptography stands a technique of keeping and transferring information to individual form for whom it is intended can read and procedure it.

Cryptographic algorithms can be separated into:

- Symmetric key algorithms.
- Asymmetric key algorithms.

Symmetric key algorithms have the things that similar secret keys are used for encryption and decryption. They are termed as private key algorithms [1]. Asymmetric key algorithms use two unlike keys: public key for encryption and private key for decryption.

Asymmetric encryption is also recognized as public-key cryptography. Asymmetric encryption varies from symmetric encryption mainly in that two keys are used for encryption and decryption [1]. Most common asymmetric encryption algorithm stands as Rivest-Shamir-Adleman (RSA). Two forms of symmetric-key algorithm are known as block cipher and stream ciphers. A block cipher functions on blocks of data which are 8 to 16 bytes long. The plaintext is broken into blocks by block ciphers and function on each block independently. A block cipher is used if key reusability is wanted. The encryption role is identical for every single block. The "pad" is known as a key stream in cryptography circles. Padding is adding extra bytes to an incomplete block of data to make the data complete. A block cipher primitive is used in such a way that it turns successfully as a symmetric key cipher in some types of operation [1].

A stream cipher stands a symmetric key cipher where plaintext digits are merged with a key-stream. An algorithm that makes a pad founded on the key is used by a stream cipher. The XOR operation for encryption and decryption is used by stream ciphers. A stream cipher is used for speed. A true key stream would be random; a stream cipher produces pseudo-random values and technically can’t be called a pad [1]. Diagram of Stream Cipher is displayed in Figure 2.5.

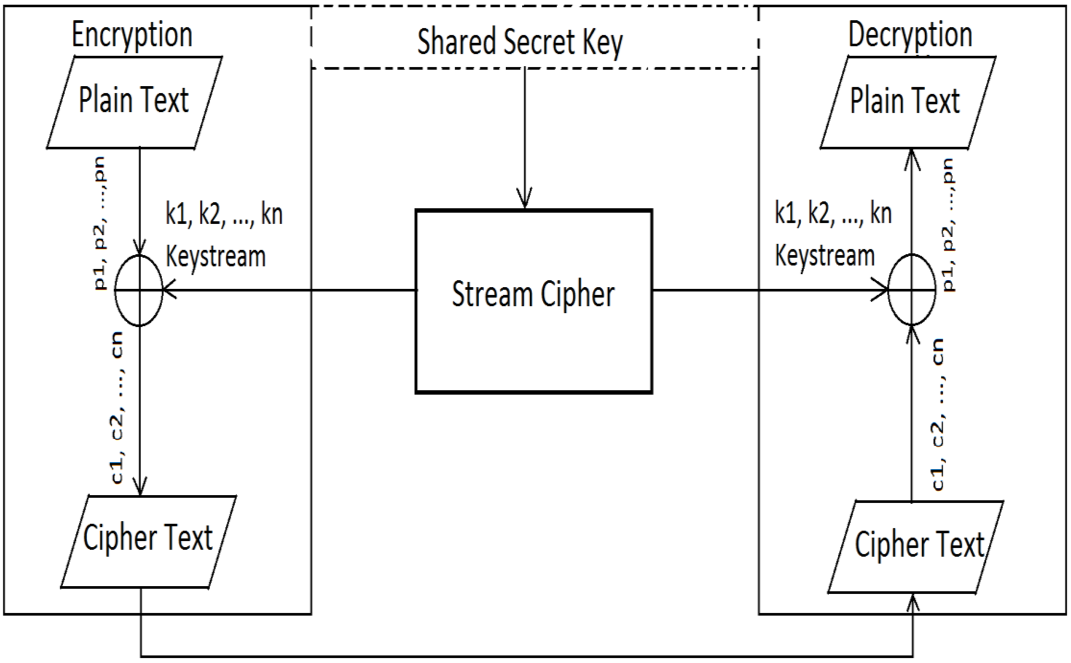


Figure 2.5 Stream Cipher

2.7 Key Generation

The creation of a secure key is known as symmetric-key encipherment. Different symmetric-key ciphers require keys of different sizes. The selection of the key must be based on a symmetric approach to evade a security leak. This suggests that there is a requirement for random (or pseudorandom) number generator.

2.8 Random Numbers

In the practice of encryption for various network security uses, random numbers show a vital part. Random numbers are used in network security algorithms founded on cryptography. These applications provide increase to two different and not essentially well-matched necessities for a classification of random numbers: randomness and unpredictability.

Randomness

- Uniform distribution: Distribution of numbers in order should be constant; that refers as, the rate of existence of each numbers should be roughly the same.
- Independence: No value in the sequence can be conditional from the others.

Unpredictability

- The necessity is not so much that the arrangement of statistics be statistically random but that the consecutive numbers of the arrangement are unpredictable in application such as mutual confirmation and session key generation.

Pseudo-random Number Generator

A pseudo-random number generator stands a set of rules for making an arrangement of numbers that estimates the possessions of random number. The arrangement is not truly unpredictable in that is totally confirmed by a comparatively minor set of original standards, termed as pseudo-random number generator's state. Most of them create arrangements which are consistently disseminated by any of numerous tests.

2.9 Rivest *Cipher* 4 (RC4 Cipher)

Rivest *Cipher* 4 (RC4) is very popular because it is simple and it can be very fast. It is an adjustable stream size key cipher included bytes focused on processes. It is founded on the practice of unplanned arrangement. Ron Rivest designed RC4 in 1987[8]. It is a standard of IEEE 802.11 within WEP (Wireless Encryption Protocol) and makes a key stream.

RC4 (recognized as ARC4 or ARCFOUR denotation Alleged *RC4*) stands a stream *cipher in cryptography*. While significant for it is simple and very fast speed, many susceptibilities have been exposed in *RC4,* interpreting it unsafe.

RC4 makes bits of a pseudo-random stream (a key-stream). As with any stream cipher, these can be used for the procedure of hiding a data in disguising its material (encryption) by merging it with the message to be sent securely from the source to the intended endpoint of the message (plaintext) using bit-wise exclusive-or. A procedure to revert ciphertext into plain text (decryption) is executed the similar way.

A stream cipher stands a symmetric key cipher where plaintext digits are merged with a key-stream. A block cipher primitive is used in such a way that it turns successfully as a symmetric key cipher in some types of operation.

RC4 is byte-oriented stream cipher in which a byte (8bits) of a plaintext is exclusive-ored with a byte of key to make a byte of ciphertext. The secret key, from which the one-byte keys in the stream are created, can contain anywhere from 1 to 256 bytes.

One of the popular stream ciphers in the world is RC4 cipher. RC4 cipher includes two parts. They are (1) Key-scheduling algorithm (KSA) and (2) Pseudo-random generation algorithm (PRGA).

(1) Key-scheduling Algorithm (KSA)

The key-scheduling algorithm is used to start up the *arrangement* in the range "S". The number of bytes in the key is called "***keylength***". It exits in the range 1 to 256. At first, the array box "S" is started up to the identity arrangement. The array box "S" is then handled for 256 repetitions in a similar way to the main Pseudo-random Number Generator, but also combines in bytes of the key at the same time.

The key-scheduling algorithm (KSA) [1] is as follow:

```
Begin
for i from 0 to 255
     S[i] :=i
endfor
j:=0
for i from 0 to 255
     j:=(j+S[i]+key[I mod keylength]) mod 256
     swap values of S[i] and S[j]
endfor
end
```

Figure 2.6 **Key-Scheduling Algorithm (KSA)**

(2) Pseudo-Random Generation Algorithm (PRGA)

The arrangement is started with a variable length key, characteristically between 40 and 2048 bits, via the key-scheduling *algorithm* (KSA). The stream of bits is created using the ***pseudo-random generation algorithm*** (***PRGA***).

RC4 creates a key-stream. After that, the stream of bits is created by a PRGA. It amends the condition and outputs a byte of the key-stream. PRGA is displayed in Figure 2.7.

```
Begin
     i :=0
     j:=0
     while GeneratingOutput:
     i=(i+1) mod 256
     j=(j+S[i]) mod 256
     swap values of S[i] and S[j]
     K:= S[(S[i]+S[j]) mod 256]
     output K
      endwhile
end
```

Figure 2.7 P*seudo*-R*andom Generation Algorithm* (*PRGA*).

The RC4 algorithm works in two phases, key setup and ciphering. Key setup is the first and most difficult phase of this algorithm. During an N-bit key setup (N being your key length), the encryption key is used to make an encrypting variable using two arrangements, state and key, and N-number of combining operations. These combining operations include swapping bytes, modulo operations, and other formulas. The key setup creates the encryption variable one time. The encryption variable goes into the ciphering phase, where it is XORed with the plain text message to make an encrypted message. After the receiver gets the encrypted message, he decrypts it by XORing the encrypted message with the identical encrypting variable. General logic design structure of RC4 cipher is displayed in Figure 2.8.

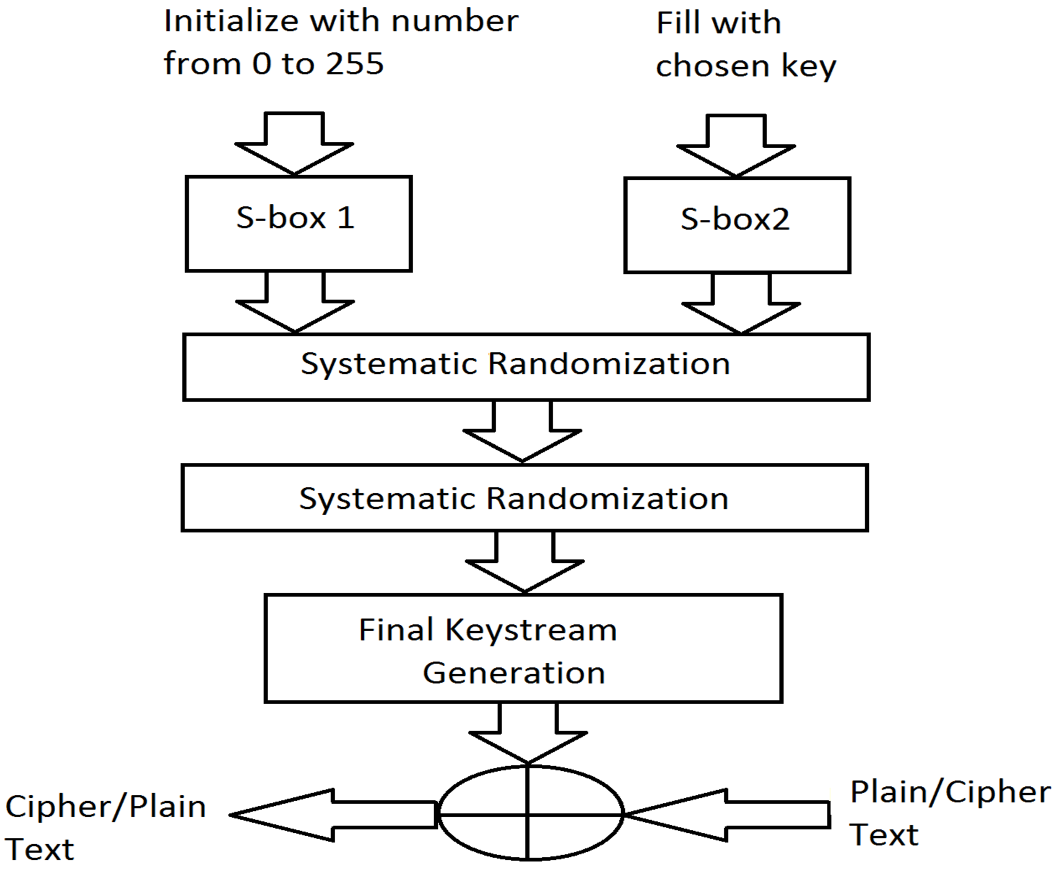


Figure 2.8 General Logic Design Structure of RC4 cipher

RC4 Strengths

It is difficult to know which place in the table is used to choose each worth or value in the arrangement (sequence). It is also difficult to know where any worth is in the table. The user can use a specific RC4 key only one time. RC4 can encrypt about 10 times faster than DES (Data Encryption Standard).

RC4 only needs 256 bytes of computer memory for the algorithm's state array, k bytes of memory for the RC4 key, and a small amount of memory for the integer values for the I and j array indices since RC4 only uses byte manipulations. Modulo 256 operation required by the algorithm can be accomplished with a bitwise AND and 255 to have a smaller memory as well.

CHAPTER 3

DESIGN AND USER INTERFACES

This system is implemented as a crypto system. It is the combination of the symmetric stream cipher RC4 algorithm for confidentiality and SMS Mobile applications for data security.

3.1 Overview of the Proposed System

To work with the system,

The sender needs to perform the following steps:

- Open to enter the *SendSMS* mobile application.
- Insert the phone number and password.
- Compose the plain text.
- Press the Send Message button.

The receiver needs to perform the following steps:

- Open to enter the *ReseiveSMS* mobile application.
- Write telephone number of the writer and the identical password used by the writer.
- Press the Receive Message button.
- Establish two applications: *SendSMS* and *ReceieveSMS*.

The proposed system uses SMS and the symmetric key algorithm RC4 stream cipher. If the sender sends the cipher to the receiver, the user will be defined the password and it is used as the input. Then, RC4 key scheduling-algorithm starts up RC4's internal pseudo-random number generator.

The generator produces key stream to be used for XOR'ing with the plaintext. The output is the cipher text. SMS uses to embed the encrypted data and the data is travelling over the network. Before entering the system, the users need to share the RC4 keys.

On the receiver side, the cipher is received, and RC4 decryption algorithm is decrypted to get the plain text. The same key is used for both sender as well as receiver.

3.2 Design of the Proposed System

RC4 stream cipher is implemented by using *Key-Scheduling Algorithm* (KSA) and *Pseudo-Random Generation Algorithm* (PRGA) in Java programming language. To offer data secrecy during the transmission of SMS message, to prevent them from being unlawfully intercepted by illegal parties and to ensure the authenticity of the message from the genuine sender, the intended user needs to put phone number of sender and password.

The design for implementation of two android mobile applications: *SendSMS* and *ReceieveSMS*. For *SendSMS* mobile application, at first password is used in RC4 cipher to create key stream and it is XORed with SMS plain text to come out cipher text. The *Sending* process sends the cipher text to the phone number acknowledged by this application. Similarly, in *ReceieveSMS* mobile application the *Receiving* process receives the cipher text from the phone number acknowledged by this application. Then the cipher text is XORed with the key stream created by RC4 cipher that applies the same password to come out original SMS plain text.

The design of two smart phone applications: *SendSMS* and *ReceieveSMS*. *SendSMS* and *ReceieveSMS* diagrams are displayed in Figure 3.3.

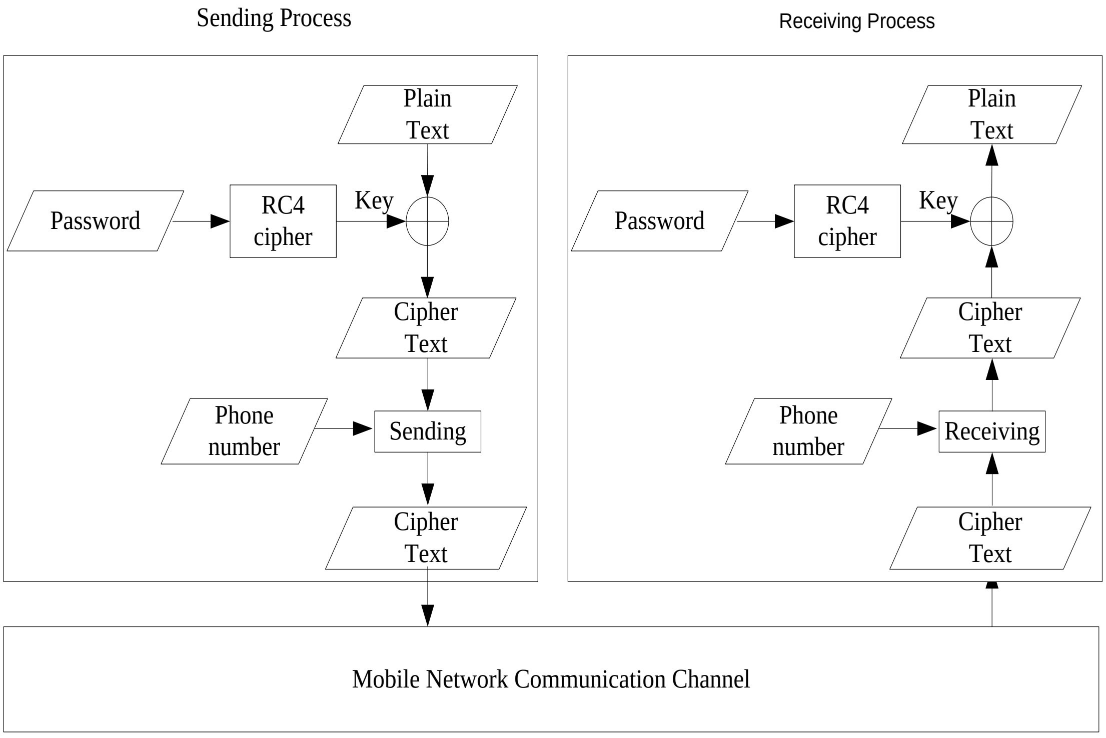


Figure 3.1 *SendSMS* and *ReceiveSMS* diagrams

3.3 The Data Flow Diagram of SendSMS and ReceiveSMS

The data flow diagram of these two mobile applications is displayed in Figure 3.4. SendSMS mobile application receives SMS plain text, password and phone number of the receiver as inputs and comes out cipher text. The cipher text is passed through mobile network communication channel. ReceieveSMS mobile application receives cipher text that passed through mobile network communication channel, password and phone number of the sender as inputs and comes out SMS plain text.

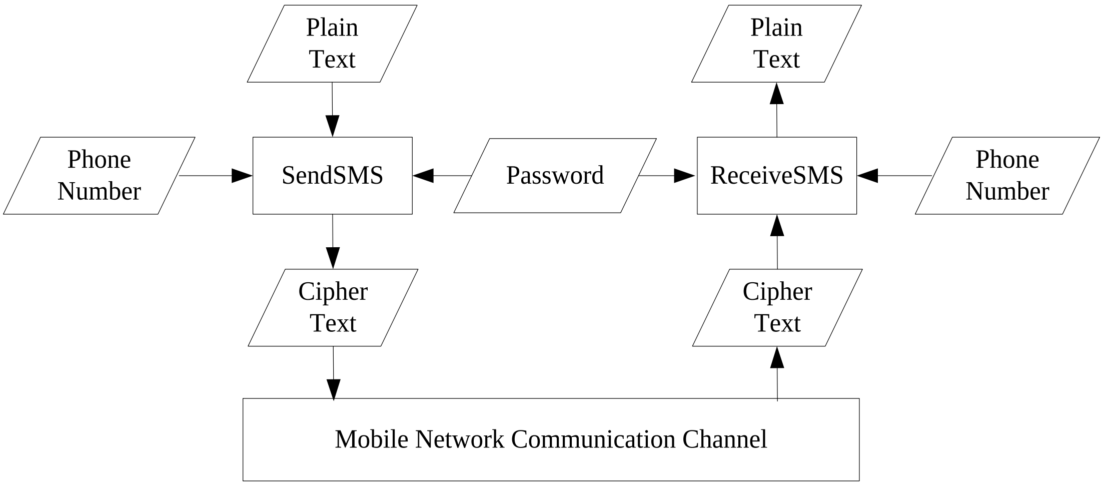


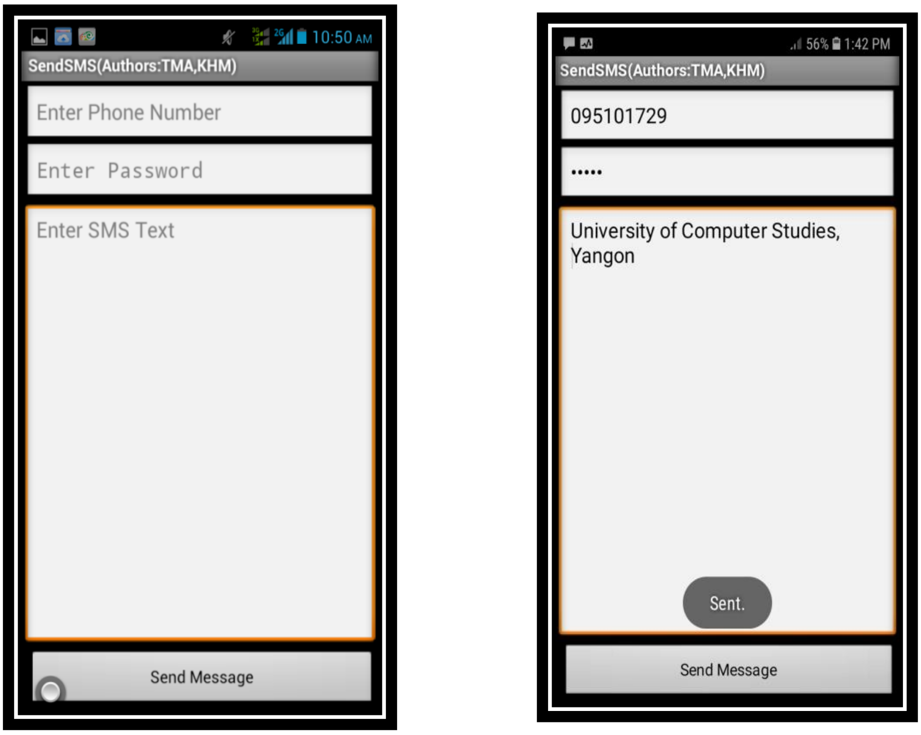
Figure 3.2 The Data Flow Diagram

3.4 User Interfaces

User interfaces for implemented system are *SendSMS* mobile application and *ReceiveSMS* mobile application on smart phone.

3.4.1 SendSMS Mobile Application

The RC4 is presented to provide SMS security for mobile users. The system is developed by using Eclipse's Android Developer and the system can run on Android version 2.0 and above. The implemented system consists of two applications, *SendSMS* mobile application and *ReceiveSMS* mobile application. To start the system, the user must have Android OS and already installed the applications. After starting the *SendSMS* mobile application, the display screen of the system is appeared, as in Figure 3.3 (a) and the user can use the system. It includes two textboxes, one text area and a button. The first text box is Enter Phone Number to insert the phone number the user wants to send the text message. The second text box is Enter Password to insert the password. The letters in any language and numbers can be included in password. The text area is Enter SMS Text to input the message the user wants to send. The message can be included letters, numbers and other signs. There is also a button for sending SMS message. The system shows a message alert Sent when the user fills the phone number, password, SMS text message and clicks Send SMS button. When all texts are completed, the display screen can be seen. The message is cancelled when the user forget to fill any of the text boxes or a text area. When all text boxes are completed and correctly filled, the user can see as in Figure 3.3 (b). The user clicks the Send Message button and the message is sent to the phone number the user entered. When the user reads the message arriving in original SMS, the encrypted SMS appears in original SMS.


SendSMS(Authors:TMA,KHM)

Enter Phone Number

Enter Password

Enter SMS Text

SendSMS(Authors:TMA,KHM)

095101729

•••••

University of Computer Studies, Yangon

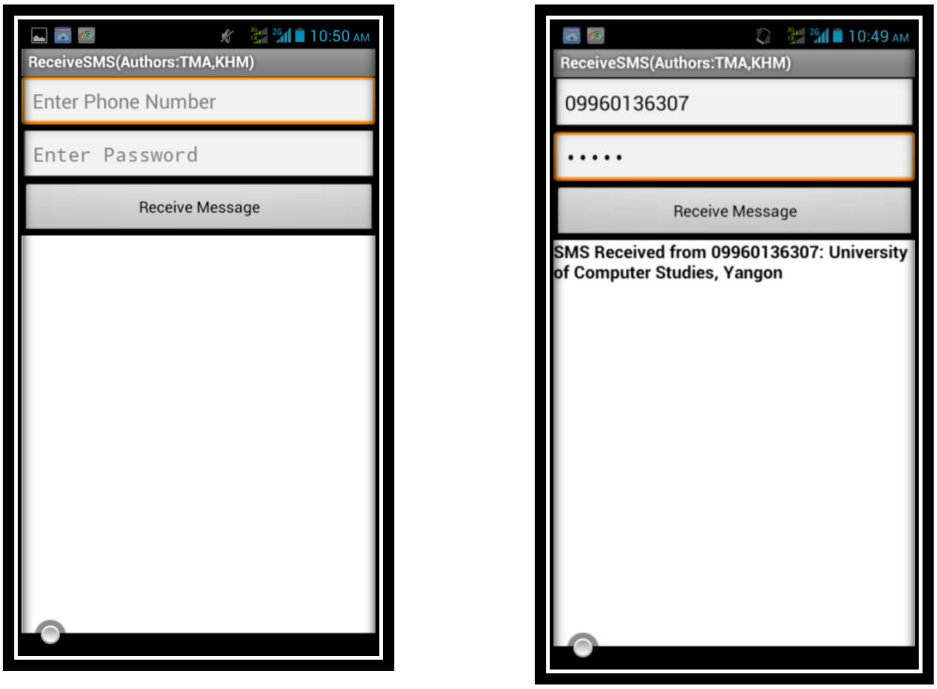
Sent.

(a)Plain User Interface **(b) User Interface with Sample Data**

Figure 3.3 System User Interfaces for SendSMS Mobile Application

3.4.2 ReceiveSMS Mobile Application

To access the message, the receiver must also install the ReceiveSMS mobile application in the receiver's android OS. When the message is arrived to the ReceiveSMS mobile application, the user must enter the mobile phone number that exactly the same number and password to offer data secrecy during the transmission of SMS message, to prevent them from being unlawfully intercepted by illegal parties and to ensure the authenticity of the message from the genuine sender. Then, press Receive Message, the decrypted SMS text message is shown as in Figure 3.4 (b).



09960136307

•••••

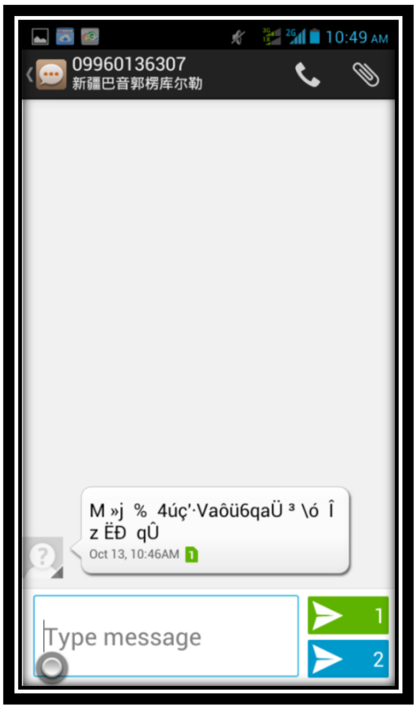
SMS Received from 09960136307: University of Computer Studies, Yangon

(a) Plain User Interface **(b) User Interface with Sample Data**

Figure 3.4 System User Interfaces for ReceiveSMS Mobile application

3.4.3 Encrypted Cipher Text in Original SMS

After sending from the SendSMS mobile application, the encrypted cipher text of the SMS message that pass through the SMS network is appeared as in Figure 3.5 and the user can see in the original SMS as an unintelligent gibberish cipher text which encrypt the plain text. Therefore, the user cannot read the text message from the SendSMS mobile application without using ReceiveSMS mobile application. If the user sends the wrong phone number that leads to unintended person, the cipher text only received in the original SMS.



09960136307
新疆巴音郭楞库尔勒

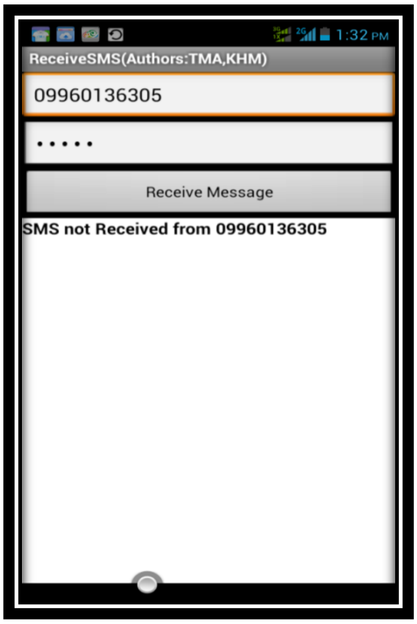
M »j % 4úç'·Vaôü6qaÜ ³ \ó Î
z ËĐ qÛ

Oct 13, 10:46AM

Figure 3.5 Cipher Text Message in Original SMS

3.4.4 Filling Wrong Phone Number

In ReceiveSMS mobile application, if the user fills the wrong phone number, the warning message “SMS not Received from that phone number” shown in Figure 3.6 is received.



09960136305

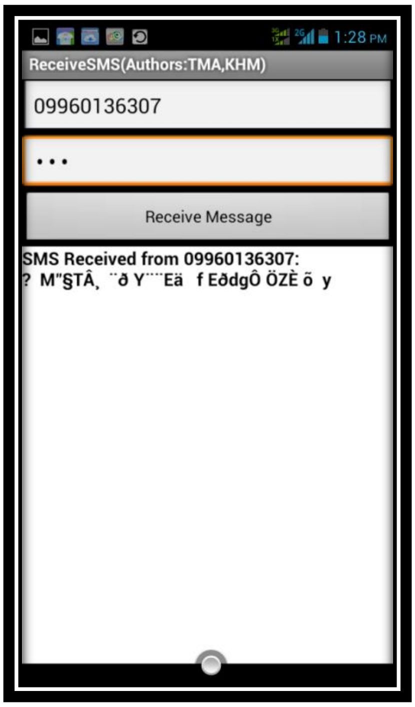
•••••

SMS not Received from 09960136305

Figure 3.6 System User Interfaces when filling wrong phone number

3.4.5 Filling Incorrect Password

After sending from the SendSMS mobile application, the encrypted cipher text of the SMS message is appeared as in the original SMS as an unintelligent cipher text. If the user enters the incorrect password, the wrong encrypted cipher text of the SMS message is appeared as shown in Figure 3.7.



09960136307

•••

SMS Received from 09960136307:
? M"§TÂ¸ ¨ð Y¨¨¨Eä f EðdgÔ ÖZÈ õ y

Figure 3.7 System User Interfaces when filling incorrect password

3.4.6 Mobile Applications User Interfaces

Mobile application user interfaces are SendSMS and ReceieveSMS mobile applications. SendSMS mobile application is used by creator and ReceieveSMS mobile application is used by intended person. The creator must input phone number of the intended person, password and SMS message to SendSMS smart phone application and press Send Message button. The intended person must input phone number of the creator and the same password used by the creator to ReceiveSMS mobile application and press Receive Message button. Then SMS message of the creator is shown in the window screen of ReceiveSMS smart phone application.

CHAPTER 4
MEASUREMENT OF CONFIDENTIALITY

RC4 stands a stream cipher in cryptography and it makes pseudorandom number arrangement for data confidentiality. While significant for it is simple and very fast speed, many susceptibilities have been exposed in *RC4*, interpreting it unsafe. The quality of data confidentiality relies on randomness of pseudorandom number arrangement made by RC4 and it can be measured by 15 statistical tests recognized by NIST, National Institute of standards and Technology. [4]

4.1 The Statistical Tests

4.1.1 The Frequency (Monobit) Test

The objective of this test is to decide whether the number of ones and zeros in a sequence are roughly the same as would be estimated for a actually random sequence. Meanings of each variant are described in Table 4.1.

The function Frequency (n), where:

Table 4.1 The function Frequency (n) and Test Statistic

Symbol	Meaning
n	The length of the bit string.
ε	The sequence of bits made by the RNG or PRNG being tried; this is a global structure at the time of the function is named; ε= $\varepsilon_1\varepsilon_2, \ldots ,\varepsilon_n$.
The test Statistic S_{obs}	The absolute value of the sum of the Xi (Xi =2ε-1= ±1) in the sequence is separated by square root of sequence's length.

The reference dissemination for the test statistic is, z = mod ($S_{obs)}$/√2. Test statistic is half normal for large n. If the sequence is unplanned, then the plus and minus ones will be likely to terminate one another out. Therefore, the test statistic will be about 0. The test statistic will be likely to be larger than zero if there are too many ones or zeroes.

The test statistic for the watched sum sobs is provided by

$$s_{obs} = \frac{|Sn|}{\sqrt{n}}. \qquad 4.1$$

The P-value is provided by

$$\text{P-value} = \text{erfc}(\frac{\text{sobs}}{\sqrt{2}}). \quad 4.2$$

The tried sequence will be recognized as random if the P-value ≥ 0.01, if not it is believed to be non-random.

4.1.2 Frequency Test within a Block

The objective of this test is to decide whether the frequency of ones in a M-bit block is roughly M/2, as would be predictable under an assumption of randomness. For block size M=1, this test degenerates to trial 1, the Frequency (Monobit) test. Meanings of each variant are described in Table 4.2.

The function Block Frequency (M,n), where:

Table 4.2 The function Block Frequency (M,n) and Test Statistic

Symbol	Meaning
M	The length of each block.
n	It is already described in The Frequency Test.
ε	It is already described in The Frequency Test.
The test Statistic x^2(obs)	A measure of how well the watched proportion of ones within a provided M-bit block compares the predictable proportion (1/2).

The reference dissemination for the test statistic exits as a χ^2 distribution. Decide the proportion π_i of ones in each *M*-bit block by the equation.

$$\pi_i = \frac{\sum_{j=1}^{M} \varepsilon(i-1)M+j}{M}, \text{for } 1 \le i \le N. \quad 4.3$$

Calculate the x^2statistic:

$$x^2(\text{obs}) = 4M \sum_{i=1}^{N} (\pi_i - \frac{1}{2})^2. \quad 4.4$$

The P-value is then provided by

$$\text{P-value} = \text{igamc}(\frac{N}{2}, \frac{x^2(\text{obs})}{2}). \quad 4.5$$

If the calculated *P-value* is < 0.01, then determine that the sequence exits as non-random. If not, determine that the sequence exits as random.

4.1.3 The Runs Test

The objective of the runs test is to decide whether the number of runs of ones and zeros of various lengths is as predictable for a random sequence. In specific, this test decides whether the oscillation between such zeros and ones exists as too fast or too slow. Meanings of each variant are described in Table 4.3.

The function, Runs(n), where:

Table 4.3 The function Runs(n) and Test Statistic

Symbol	Meaning
n	It is already described in The Frequency Test.
ε	It is already described in The Frequency Test.
The test Statistic V_n (obs)	The total number of runs (i.e., the total number of zero runs + the total number of one runs) through all n bits.

The reference dissemination for the test statistic exits as a χ2 distribution. The Runs test conveys out a Frequency test as a requirement.

$$\pi = \frac{\sum_j \varepsilon_j}{n} \qquad 4.6$$

Calculate the test statistic

$$V_n(\text{obs}) = \sum_{k=1}^{n-1} r(k) + 1 \qquad 4.7$$

where $r(k)=0$ if $\varepsilon_k = \varepsilon_{k+1}$and $r(k)= 1$, otherwise.Calculate

$$\text{P-value} = \text{erfc}\left(\frac{|v_n(obs) - 2n\pi(1-\pi)|}{2\sqrt{2n}\ \pi(1-\pi)}\right). \qquad 4.8$$

If the calculated P-value is < 0.01, then determine that the sequence exits as non-random. If not, determine that the sequence exits as random.

4.1.4 Test for the Longest-Run-of-Ones in a Block

The objective of this test is to decide whether the length of the longest run of one's within the tried sequence is constant with the length of the longest run of ones that would be predictable in a random sequence. An irregularity in the predictable length of the longest run of one's indicates that there is also an irregularity in the predictable length of the longest run of zeroes. Meanings of each variant are described in Table 4.4.

The function Longest-Run-of-Ones(n), where:

Table 4.4 The function Longest-Run-of-Ones(n) and Test Statistic

Symbol	Meaning
n	It is already described in The Frequency Test.
ε	It is already described in The Frequency Test.
M	It is already described in The Frequency Test within a Block.
N	The number of blocks; chosen according to the value of M.
The test Statistic x^2(obs)	A measure of how well the watched longest run length within M-bit blocks compares the predictable longest length within M-bit blocks.

The reference dissemination for the test statistic stands a χ^2 distribution.

$$x^2 (obs) = \sum_{i=0}^{K} \frac{(vi_i - N\pi_i)^2}{N\pi_i}. \quad 4.9$$

$$P\text{-}value = \text{igamc}(\frac{K}{2},\frac{x^2(obs)}{2}). \quad 4.10$$

If the calculated *P-value* is < 0.01, then determine that the sequence exits as non-random. If not, determine that the sequence exits as random.

4.1.5 The Binary Matrix Rank Test

The objective of this test is to check for linear dependence among fixed length substrings of the original sequence. Meanings of each variant are described in Table 4.5. Rank (n), where:

Table 4.5 The function Rank(n) and Test Statistic

Symbol	Meaning
n	It is already described in The Frequency Test.
ε	It is already described in The Frequency Test.
M	The number of rows in each matrix. For the test complement, M is 32. If other values of M are used, new estimates need to be calculated.
Q	The number of columns in each matrix. For the test complement, Q is 32. If other values of Q are used, new estimates need to be calculated.
The test statistic	A measure of how well the watched number of ranks of several different orders compares the predictable number of ranks under an

x^2(obs)	assumption of randomness.

The reference dissemination for the test statistic stands a χ^2 distribution.

$$x^2(\text{obs})=\frac{(F_M-0.2888N)^2}{0.2888*2}+\frac{(F_{M-1}-0.5776N)^2}{0.5776N}+\frac{(N-F_M-F_{M-1}-0.1336N)^2}{0.1336N} \qquad 4.11$$

$$\text{P-Value}=e^{\frac{-x^2(obs)}{2}}. \qquad 4.12$$

If the calculated *P-value* is < 0.01, then determine that the sequence exits as non-random. If not, determine that the sequence exits as random.

4.1.6 The Discrete Fourier Transform Test

The Objective of this test is to identify periodic features (i.e., repetitive patterns that are near each other) in the tried arrangement that would point out a deviation from the hypothesis of randomness. The purpose is to identify whether the number of peaks exceeding the 95% threshold is significantly different than 5%. Meanings of each variant are described in Table 4.6.

DiscreteFourierTransform(n), where:

Table 4.6 The function DiscreteFourierTransform(n) and Test Statistic

Symbol	Meaning
n	It is already described in The Frequency Test.
ε	It is already described in The Frequency Test.
The test Statistic D	The normalized difference between the watched and the predictable number of frequency components that are beyond the 95 % threshold.

The reference dissemination for the test statistic exits as the normal distribution.

$$d=\frac{(N_1-N_0)}{\sqrt{\frac{n(.95).05)}{4}}}. \qquad 4.13$$

$$\text{P-value}=\text{erfc}(\frac{|d|}{\sqrt{2}}) \qquad 4.14$$

If the calculated *P-value* is < 0.01, then determine that the sequence exits as non-random. If not, determine that the sequence exits as random.

4.1.7 The Non-overlapping Template Matching Test

The objective of this test is to identify generators that create too many existences of a given non-periodic (aperiodic) pattern. An m-bit window is applied to search for a particular m-bit pattern for this test. If the arrangement is not set up, the window slips one bit position. If the pattern is found, the window is reset to the bit after the established pattern, and the search resumes. Meanings of each variant are described in Table 4.7. The function is NonOverlappingTemplateMatching (m,n)

Table 4.7 The function NonOverlappingTemplateMatching (m,n) and Test Statistic

Symbol	Meaning
m	The length in bits of each template. The template stands as the target string.
n	It is already described in The Frequency Test.
ε	It is already described in The Frequency Test.
B	The m-bit template to be compared; B exits as a string of ones and zeros (of length m)
M	The length in bits of the substring of ε to be tried.
N	The number of independent blocks.
The test Statistic x^2(obs)	A measure of how well the watched number of template "hits" compares to the expected number of template "hits" (under an assumption of randomness).

$$\mu=\frac{(M-m+1)}{2^m}. \qquad 4.15$$

$$\sigma^2 = M\left(\frac{1}{2^m}-\frac{2m-1}{2^{2m}}\right). \qquad 4.16$$

$$x^2(\text{obs})=j = \sum_{j=1}^{N}\frac{(W_j-\mu)^2}{\sigma^2}. \qquad 4.17$$

$$\text{P-value}=\text{igamc}\left(\frac{N}{2},\frac{x^2(obs)}{2}\right). \qquad 4.18$$

If the calculated *P-value* is < 0.01, then determine that the sequence exits as non-random.

If not, determine that the sequence exits as random.

4.1.8 The Overlapping Template Matching Test

The objective of the test is to use an m-bit window to search for a specific m-bit pattern. As with the non-overlapping test, if the pattern is not found, the window slides one bit position. The difference between this test and the non-overlapping test is that when the pattern is found, the window slips only one bit before resuming the search. Meanings of each variant are described in Table 4.8.

OverlappingTemplateMatching (m,n)

Table 4.8 The function OverlappingTemplateMatching (m,n) and Test Statistic

Symbol	Meaning
m	The length in bits of the template.
n	It is already described in The Frequency Test.
ε	It is already described in The Frequency Test.
B	The m-bit template to be compared.
K	The number of degrees of freedom.
M	The length in bits of a substring of ε to be tried.
N	The number of independent blocks.
The test Statistic x^2(obs)	It is already described in The Non-overlapping Template Matching Test.

The reference dissemination for the test statistic exits as the χ^2 distribution.

$$\lambda=\frac{(M-m+1)}{2^m} \quad 4.19$$

$$\eta=\frac{\lambda}{2} \quad 4.20$$

$$x^2(\text{obs})=\sum_{i=0}^{5}\frac{(v_i-N\pi_i)}{N\pi_i} \quad 4.21$$

$$\text{P-value}=\text{igamc}(\frac{5}{2},\frac{x^2(obs)}{2}). \quad 4.22$$

If the calculated *P-value* is < 0.01, then determine that the sequence exits as non-random. If not, determine that the sequence exits as random.

4.1.9 Maurer's "Universal Statistical" Test

The objective of the test is to identify whether or not the sequence can be significantly compressed without loss of data.

A significantly compressible sequence exits as non-random. Meanings of each variant are described in Table 4.9.

Universal(*L*, *Q*, *n*), where

Table 4.9 The function Universal (*L, Q, n*) and Test Statistic

Symbol	Meaning
L	The length of each block.
Q	The number of blocks in the initialization sequence.
n	It is already described in The Frequency Test.
ε	It is already described in The Frequency Test.
The test Statistic f_n	The sum of the $\log_2$ distances between matching *L*-bit templates,

The reference dissemination for the test statistic exits as the half-normal distribution (a one-sided variant of the normal distribution) as is also the case for the Frequency test.

Calculate test statistic:

$$\sigma = \mathrm{c}\sqrt{\frac{variance(L)}{K}} \qquad 4.23$$

$$\mathrm{c}=0.7-\frac{0.8}{\mathrm{L}} + \left(4 + \frac{32}{L}\right)\frac{K^{-\frac{3}{L}}}{15}. \qquad 4.24$$

$$\mathrm{f_n}=\frac{1}{\mathrm{K}}\sum_{\mathrm{i=Q+1}}^{\mathrm{Q+K}} \log_2\left(\mathrm{i}-\mathrm{T_j}\right) \qquad 4.25$$

$$\text{P-value}=\mathrm{erfc}\left(\left|\frac{\mathrm{f}_n-\mathrm{expectedValue(L)}}{\sqrt{2}\sigma}\right|\right). \qquad 4.26$$

If the calculated *P-value* is < 0.01, then determine that the sequence exits as non-random. If not, determine that the sequence exits as random.

4.1.10 The Linear Complexity Test

The objective of this test is to use linear complexity to test for randomness. If the initial state of Linear Feedback Shift Registers (LFSR) is (ε*L-1*, ..., ε*1*, ε*0*), then the output sequence, (ε*L*, ε*L+1*, …), fulfills the following recurrent formula for $j \geq L$

$$\varepsilon_j=(c_1\varepsilon_{j-1}+ c_2\varepsilon_{j-2} +\ldots+ c_L\varepsilon_{j-L}) \bmod 2 \qquad 4.27$$

c1, …, *cL* exist as coefficients of the connection polynomial corresponding to a given LFSR. An LFSR is said to make a provided binary sequence if this sequence exits as the output of the LFSR for some initial state.

Random sequences are characterized by longer LFSRs. An LFSR that is too short implies non-randomness. Meanings of each variant are described in Table 4.10.

LinearComplexity(*M*, *n*), where:

Table 4.10 The function Linear Complexity (*M, n*) and Test Statistic

Symbol	Meaning
M	The length in bits of a block.
n	It is already described in The Frequency Test.
ε	It is already described in The Frequency Test.
K	The number of degrees of freedom.
The test Statistic x^2(obs)	A measure of how well the watched number of occurrences of fixed length LFSRs compares the predictable number of occurrences under an assumption of randomness.

The reference dissemination for the test statistic exits as the χ^2 distribution.

$$x^2(\text{obs}) = \sum_{i=0}^{K} \frac{(v_i - N\pi_i)^2}{N\pi_i} \qquad 4.28$$

$$\text{P-value} = \text{igamc}\left(\frac{K}{2}, \frac{x^2(\text{obs})}{2}\right) \qquad 4.29$$

If the calculated *P-value* is < 0.01, then determine that the sequence exits as non-random. If not, determine that the sequence exits as random.

4.1.11 The Serial Test

The objective of this test is to decide whether the number of existences of the *m*-bit overlapping patterns 2^m is roughly the same as would be predictable for a random sequence. Random sequences have uniformity; it means, every *m*-bit pattern has the same chance of appearing as every other *m*-bit pattern. For *m* = 1, the Serial test is equivalent to the Frequency test. Meanings of each variant are described in Table 4.11.

Serial *(m,n)*, where:

Table 4.11 The function Serial*(m,n)* and Test Statistic

Symbol	Meaning
m	The length in bits of each block.
n	It is already described in The Frequency Test.

ε	It is already described in The Frequency Test.
$\nabla\varphi_m^2$ (obs) and $\nabla^2\varphi_m^2$(obs)	A measure of how well the watched frequencies of m-bit patterns compare the predictable frequencies of the m-bit patterns.

The reference dissemination for the test statistic exits as the χ^2 distribution.

$\nabla\varphi_m^2 = \varphi_m^2 - \varphi_{m-1}^2$ 4.30

$\nabla^2\varphi_m^2 = \varphi_m^2 - 2\varphi_{m-1}^2 + \varphi_{m-2}^2$ 4.31

P-value1 = igamc(2^{m-2},$\nabla\varphi_m^2$) 4.32

P-value2= igamc(2^{m-3},$\nabla^2\varphi_m^2$) . 4.33

If the calculated *P-value* is < 0.01, then determine that the sequence exits as non-random. If not, determine that the sequence exits as random.

4.1.12 The Approximate Entropy Test

The objective of the test is to match the frequency of overlapping blocks of two consecutive/adjacent lengths (*m* and *m+1*) against the predictable result for a random sequence. Meanings of each variant are described in Table 4.12.

Approximate Entropy (*m,n*), where:

Table 4.12 The function Approximate Entropy (*m,n*) and Test Statistic

Symbol	Meaning
m	The length of each block – the first block length used in the test. *m+1* exits as the second block length used.
n	The length of the entire bit sequence.
ε	It is already described in The Frequency Test.
The test Statistic x^2(obs)	A measure of how well the watched value of *ApEn*(*m*) matches the expected value.

The reference dissemination for the test statistic exits as the χ^2 distribution.

$\varphi^{(m)}=\sum_{i=0}^{2^m-1}\pi_i \log \pi_i$, where $\pi_i = C_j^3$, j=$\log_2$ I 4.34

x^2=2n[log2-ApEn(m)] where ApEn(m)=$\varphi^{(m)}$- $\varphi^{(m+1)}$ 4.35

P-value=igamc(2^{m-1},$\frac{x^2}{2}$). 4.36

If the calculated *P-value* is < 0.01, then determine that the sequence exits as non-random. If not, determine that the sequence exits as random.

4.1.13 The Cumulative Sums Test

The objective of the test is to decide whether the cumulative sum of the partial sequences occurring in the tested sequence is too large or too small relative to the expected behavior of that cumulative sum for random sequences. This cumulative sum may be known as a random walk. For a random sequence, the trips of the random walk should be near zero. For certain types of non-random sequences, the trips of this random walk from zero will be large. Meanings of each variant are described in Table 4.13. Cumulative Sums (*mode,n*), where:

Table 4.13 The function Cumulative Sums (*mode,n*) and Test Statistic

Symbol	Meaning
n	It is already described in The Frequency Test.
ε	It is already described in The Frequency Test.
Mode	A switch for using the test either forward through the input sequence or backward through the sequence.
The test Statistic z	The largest trip from the origin of the cumulative sums in the corresponding (-1, +1) sequence.

The reference dissemination for the test statistic exits as the normal distribution.

Calculate the test statistic $z = max 1 \leq k \leq n|Sk|$, that $max 1 \leq k \leq n|Sk|$ exits as the largest of the absolute values of the partial sums Sk.

$$\text{P-value}=1-\sum_{k=\frac{(\frac{-n}{z}+1)}{4}}^{\frac{(\frac{n}{z}-1)}{4}}[\Phi\left(\frac{(4k+1)z}{\sqrt{n}}\right)-\Phi\left(\frac{(4k-1)z}{\sqrt{n}}\right)]+\sum_{k=\frac{(\frac{-n}{z}-3)}{4}}^{\frac{(\frac{n}{z}-1)}{4}}[\Phi\left(\frac{(4k+3)z}{\sqrt{n}}\right)-\Phi\left(\frac{(4k+1)z}{\sqrt{n}}\right)] \qquad 4.37$$

If the calculated *P-value* is < 0.01, then determine that the sequence exits as non-random. If not, conclude that the sequence exits as random.

4.1.14 The Random Excursions Test

The objective of this test is to decide if the number of visits to a particular state within a cycle deviates from what one would expect for a random sequence. This test is focus on the number of cycles having exactly *K* visits in a cumulative sum random walk. This test is really a series of eight tests (and conclusions), one test and decision for each of the states: -4, -3, -2, -1 and +1, +2, +3, +4. Meanings of each variant are described in Table 4.14. Random Excursions (*n*), where:

Table 4.14 The function RandomExcursions(*n*) and Test Statistic

Symbol	Meaning
n	It is already described in The Frequency Test.
ε	It is already described in The Frequency Test.
Test statistic x^2(obs)	For a provided state *x*, a measure of how well the watched number of state visits within a cycle compare the predictable number of state visits within a cycle, under an assumption of randomness.

The reference dissemination for the test statistic exits as the χ^2 distribution.

$$x^2(\mathrm{obs})=\sum_{k=0}^{5}\frac{(v_k(x)-J\pi_k(x))^2}{J\pi_k(x)}. \qquad 4.38$$

$$\text{P-value}=\mathrm{igamc}\left(\frac{5}{2},\frac{x^2(obs)}{2}\right). \qquad 4.39$$

4.1.15 The Random Excursions Variant Test

The objective of this test exits as to identify deviations from the predictable number of visits to various states in the random walk. This test is focused on the total number of times that a particular state is visited (i.e., occurs) in a cumulative sum random walk. This test exits as really a series of eighteen tests (and conclusions), one test and decision for each of the states: -9, -8, …, -1 and +1, +2, …, +9. Meanings of each variant are described in Table 4.15. Random Excursions Variant (*n*), where:

Table 4.15 The function Random Excursions Variant (*n*) and Test Statistic

Symbol	Meaning
n	Bit string’s length; available as a parameter during the function call.
ε	It is already described in The Frequency Test.
The test	For a provided state *x*, the total number of times that the provided

Statistic ξ	state is visited during the entire random walk.

The reference dissemination for the test statistic exits as the half normal (for large *n*). (Note: If ξ is disseminated as normal, then |ξ| is disseminated as half normal.) If the sequence exits as random, then the test statistic will be about 0. If there are too many ones or too many zeroes, then the test statistic will be large.

ξ (x) exits as the total number of times that state x occurred across all J cycles.

$$\text{P-value}=\text{erfc}\left(\frac{|\xi(x)-J|}{\sqrt{2J(4|x|-2)}}\right). \qquad 4.40$$

If the calculated *P-value* is < 0.01, then determine that the sequence exits as non-random.

If not, determine that the sequence exits as random.

4.2 Implementation of Frequency Test and Runs Test

4.2.1 Frequency Test

Objective of the Test is to decide whether the number of ones and zeros in a arrangement are just about the same as would be predictable for a actually random sequence.

Description of the test is the tests change the sequence ε into a new sequence X, such that $X_i=2\varepsilon_i -1 = \pm 1$. The calculation of this sequence is assumed by

$$S_n = X_1 + X_2 + \dots + X_n. \qquad 4.41$$

If ε= 1011010101,then n=10

and S_n =1+ (-1)+1+1+(-1)+1+(-1)+1+(-1)+1=2.

The test statistic for the observed sum s_{obs} is assumed by

$$s_{obs} = \frac{|s_n|}{\sqrt{n}}. \qquad 4.42$$

$s_{obs}=(\frac{|2|}{\sqrt{10}})$=.632455532

The P-value is assumed by

$$\text{P-value} = \text{erfc}\left(\frac{s_{obs}}{\sqrt{2}}\right). \qquad 4.43$$

erfc(z) is the complementary error function

$\text{erfc}(\frac{.632455532}{\sqrt{2}})$=0.527089

Decision Rule

The verified sequence will be recognized as random if the P-value ≥ 0.01, if not it is non-random.P-value=0.527089 ≥0.01. Therefore, The sequence is random.

4.2.2 Runs Test

Objective of the test is to define whether the number of runs of ones and zeros of various lengths is as predictable for a random sequence. In particular, this test defines whether the oscillation between such zeros and ones is too fast or too slow.

Calculate the pre-test proportion π of ones in the input sequence:

$$\pi = \frac{\sum_j \varepsilon_j}{n} \qquad 4.44$$

If ε=1001101011, then n=10 and $\pi = \frac{6}{10} = \frac{3}{5}$.

Define if the prerequisite Frequency test is approved: If it can be displayed that $|\pi - \frac{1}{2}| \geq$ r, then the Runs test need not be executed. If the test is not appropriate, then the P-value is set to 0.0000. For this test, $r = \frac{2}{\sqrt{n}}$ has been predefined in the testcode. $|\pi - \frac{1}{2}| = 0.1 <$ $r = \frac{2}{\sqrt{10}} = 0.63246$, and the test is not run. Since the observed value π is within the particular bound, the runs test is appropriate.

Calculate the test statistic

$$V_n(obs) = \sum_{k=1}^{n-1} r(k) + 1 \qquad 4.45$$

r(k)=0 if $\varepsilon_k = \varepsilon_k$+1and r(k)= 1 otherwise. Since ε= 1 00 11 0 1 0 11, then V10(obs)=(1+0+1+0+1+1+1+1+0)+1=7.

$$\text{P-value} = \text{erfc}\left(\frac{|v_n(obs) - 2n\pi(1-\pi)|}{2\sqrt{2n}\,\pi(1-\pi)}\right). \qquad 4.46$$

$$\text{P-value} = \text{erfc}\left(\frac{7-2*10*\frac{3}{5}(1-\frac{3}{5})|}{2\sqrt{20}\,\frac{3}{5}(1-\frac{3}{5})}\right) = 0.147232$$

Decision Rule

If the calculated P-value is < 0.01, then define that the sequence is non-random. If not, define that the sequence is random. P-value=0.147232≥0.01.The sequence is random.

4.2.3 Results of Testing

Frequency test examines the numbers of occurences of the bits in the keystream. Runs test examines the independence of the keystream bits.

Let the keystream 1011010101 be tested by Frequency test and Runs test. The result of Frequency test is SUCCESS because the numbers of occurences of bits- zero and one in the keystream are equal. The result of Runs test is FAILURE because the adjacent bits of the keystream are dependent.

Keystream : 1011010101

Frequency Test SUCCESS   p_value = 0.527089

Runs Test   FAILURE   p_value = 0.005658

Let the keystream 1001101011 be tested by Frequency test and Runs test. The result of Frequency test is SUCCESS because the numbers of occurences of bits- zero and one in the keystream are equal. The result of Runs test is SUCCESS because the adjacent bits of the keystream are independent.

Keystream : 1001101011

Frequency Test   SUCCESS   p_value = 0.527089

Runs Test   SUCCESS   p_value = 0.147232

In practical use, the following plain text is encrypted by using the following RC4 keystream. The confidentiality of the RC4 keystream is measured by using the Frequency Test and Runs Test. We found that their results are SUCCESS. Therefore, the confidentiality of RC4 stream cipher may be strong for SMS Security.

Plain Text:University of Computer Studies, Yangon (UCSY)

Password:APPLE

Keystream:hlvUùi+ESÑ;?é½þ0???Äã¥?)ß©Ñn??¸´O¹ÿa0_o?é6

CipherText:=#?X,'¨ü??½_ðåæ°?×´z«Üg¸í'áÕ!Þ?wI,É°

Password:APPLE

Keystream:hlvUùi+ESÑ;?é½þ0???Äã¥?)ß©Ñn??¸´O¹ÿa0_o?é6

Plain Text:University of Computer Studies, Yangon (UCSY)

RC4 Keystream(byte) : hlvUùi+ESÑ;?é½þ0???Äã¥?)ß©Ñn??¸´O¹ÿa0_o?é6

RC4Keystream(bit):0110100001101100011101100101010101101001001010 110100010101010011001110110011000000101001000000110110111000001011001 11111010011110110000100110000010111110001110001101111100110110

Frequency Test SUCCESS p_value = 0.763025

Runs Test SUCCESS p_value = 0.446643

CHAPTER 5
CONCLUSION AND FUTURE EXTENSION

The purpose of this thesis is to offer data secrecy during the transmission of SMS messages so as to prevent them from making unlawfully interrupted by unlawful parties and to certify the authenticity of the data from the genuine creator.

5.1 Conclusion

Nowadays, SMS is of multiple benefits to everyday activities. However, it is not presumed to be a secure and protected means while confidential data is sent by the normal SMS services. These days, a number of security issues and vulnerabilities related to SMS were existed. Thus it is crucial to protect the SMS message from making unlawfully interrupted by undesirable parties as well as to make sure that the message originates from the lawful sender. There are answers such as encrypted SMS should be considered if there is a necessity to send sensitive information via SMS.

The art of encoding and decoding ciphers is termed cryptology. Data privacy cryptology is now associated with other purposes such as data integrity, user authentication or more complicated security concerns. Cryptology is used almost anywhere currently. If you buy some goods online by credit card, you have invariably used cryptography to make sure that your credit card number cannot be detected by an eavesdropper in the meantime. Apart from encryption, if the user has updated Windows or Mac software, then unwittingly, the user has depended on digital signatures to check the authenticity of the update. Finally, Bitcoin, the latest type of digital currency, uses encryption techniques, for its security.

Cryptography is so successful that a number of its underlying concepts have spread to other areas of computer security. Cryptograph refers to the science of changing normal data into unintelligible form. A cryptographic system changes a plain text into an encoded text using a key produced by a cryptographic algorithm. A stream cipher that is employed to guard mobile network traffic as part of the Secure Socket Layer (SSL) is RC4. Data confidentiality for SMS sent via internet can be protected by RC4 stream cipher.

In this thesis, pseudo-random number arrangement made by RC4 stream cipher is calculated by Frequency Test and Run Test. The pseudorandom number arrangement may be measured to be random with a confidence of 99% giving to P-value of every single test. For that reason, it is suggested that the user should use the pseudorandom number arrangement made by RC4 steam cipher for data confidentiality of SMS message on Android.

5.2 Limitations

The system can only support the cryptography portion of security of RC4 algorithm. The user must have a phone sim card and mobile device that runs on Android OS. This system can only support the encryption of text messages. The system does not consider about other types of messages.

5.3 Further Extension

In the digital world currently, the encryption of SMS message grows into more and more essential. Future work will support the encryption of other types of messages, image or song or video file extensions. There are many features that can be appended to the system, to increase the system security, integrity, efficiency and functionally with other method and protocol. To be more secure, RC4 keys can be able to encrypt by using hybrid cryptosystem that is combination of the symmetric key algorithm and the asymmetric key algorithm.

REFERENCES

[1] Behrouz Forouzan, “Cryptography and Network Security”, International Edition, McGrawHill, ISBN:978-007-126361-0, 2008.

[2] Mary Agoyi and Devrim Seral, “SMS Security: An Asymmetric Encryption Approach”, IEEE 6th International Conference on Wireless and Mobile Communications, 2010. Valencia, Spain.

[3] Medani1, Gani1, Zakaria, Zaidan and Zaidan, “Review of mobile short message service security issues and techniques towards the solution”, Academic Journals of Scientific Research and Essays, ISSN 1992-2248, March, 2011, Volume. 6(6).

[4] National Institute of Standards and Technology’s publication “A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications”, Special Publication 800-22, 2010.

[5] Neetesh Saxena and Ashish Payal, “Enhancing Security System of Short Message Service for M-Commerce in GSM”, International Journal of Computer Science & Engineering Technology (IJCSET) ISSN : 2229-3345, 2011, Volume 2, No 4.

[6] Ozeki NG “ The world’s most reliable SMS Gateway Software” Hungary.

[7] Sharad Kumar Verma and D.B. Ojha, “An Approach to Enchance the Mobile SMS Security”, Journal of Global Research in Computer Science ISSN 2229-371X, May 2014, Volume 5, No. 5.

[8] Vaishali Singh and Shriddha Shrivastawa, “RC4 Stream Cipher Design for Data Security”, International Journal of Advance Research in Science and Engineering ISSN 2319-8346, Volume 6,Issue 5, 2017.

[10] Veena K.Katankar and V.M.Thakare, “Short Message Service using SMS Gateway”, International Journal on Computer Science and Engineering, 2010, Volume 2, No. 4.