CAPITAL UNIVERSITY OF SCIENCE AND
TECHNOLOGY, ISLAMABAD

Improvement in the Interception Vulnerability Level of Encryption Mechanism in GSM

by

Reshail

A thesis submitted in partial fulfillment for the
degree of Master of Science

in the

Faculty of Engineering
Department of Electrical Engineering

2018

I dedicate this to my

Maternal Grandmother: Sabir Jan (Late)

Parents: Shahnaz Sardar & Raja Javed Iqbal Abbasi

Brothers: Muhammad Ali Javed Abbasi & Dr. Khalil Ur Rehman (Brother-in-Law)

Sisters: Dr. Amra Khalil Ur Reham & RM. Rabia Javed Abbasi

Nephew: Muhammad Mustafa Khalil Ur Rehman

CAPITAL UNIVERSITY OF SCIENCE & TECHNOLOGY
ISLAMABAD

CERTIFICATE OF APPROVAL

Improvement in the Interception Vulnerability Level of Encryption Mechanism in GSM

by

Reshail

MEE161011

THESIS EXAMINING COMMITTEE

S. No.	Examiner	Name	Organization
(a)	External Examiner	Dr. Tauseef Jamal	PIEAS, Islamabad
(b)	Internal Examiner	Dr. Imtiaz Ahmed Taj	CUST, Islamabad
(c)	Supervisor	Dr. Noor Muhammad Khan	CUST, Islamabad

Supervisor Name
Dr. Noor Muhammad Khan
October, 2018

Dr. Noor Muhammad Khan
Head
Dept. of Electrical Engineering
October, 2018

Dr. Imtiaz Ahmed Taj
Dean
Faculty of Engineering
October, 2018

Author's Declaration

I, **Reshail** hereby state that my MS thesis titled "**Improvement in the Interception Vulnerability Level of Encryption Mechanism in GSM**" is my own work and has not been submitted previously by me for taking any degree from Capital University of Science and Technology, Islamabad or anywhere else in the country/abroad.

At any time if my statement is found to be incorrect even after my graduation, the University has the right to withdraw my MS Degree.

(Reshail)

Registration No: MEE161011

Plagiarism Undertaking

I solemnly declare that research work presented in this thesis titled "*Improvement in the Interception Vulnerability Level of Encryption Mechanism in GSM*" is solely my research work with no significant contribution from any other person. Small contribution/help wherever taken has been dully acknowledged and that complete thesis has been written by me.

I understand the zero tolerance policy of the HEC and Capital University of Science and Technology towards plagiarism. Therefore, I as an author of the above titled thesis declare that no portion of my thesis has been plagiarized and any material used as reference is properly referred/cited.

I undertake that if I am found guilty of any formal plagiarism in the above titled thesis even after award of MS Degree, the University reserves the right to withdraw/revoke my MS degree and that HEC and the University have the right to publish my name on the HEC/University website on which names of students are placed who submitted plagiarized work.

(Reshail)

Registration No: MEE161011

Acknowledgements

In the name of Allah, the Most Gracious and the Most Merciful Alhamd-o-lillah, all praises to Allah for the strengths and His blessing in completing this thesis. Special appreciation goes to my supervisor Dr. Noor Muhammad Khan, of his supervision and constant support. I am lucky to have such a kindhearted supervisor. I must say he is the seasoned and renowned professional with his vast knowledge and significant experience in the field of Telecommunication systems and a huge thanks to him for his support.

I also want to acknowledge my ARWIC colleagues Mr. Yasir Mirza and Mr. Saud Yousafzai for their support and assistance. My acknowledgement also goes to all the technicians and office staffs of Telecommunication Systems lab for their co-operations.

A huge gratitude for my friends and fellows for their company, support and encouragement through all circumstances, no matter how difficult. Especially I show my courtesy towards Tanzela Irshad, Haris Farooq, Waleed Farooq, Sajjad Arshad, umer daraz , anum munir and others.. I have been greatly supported by my beloved parents, teachers and friends. I owe a huge debt for their endless love, prayers and Encouragement.

Abstract

Due to day-to-day advancements in the technology, cellular mobile operators around the world are facing higher data-rate demands and novel security issues. The emerging issues are basically the result of the evolving features of technology advancement. Therefore, there is an urgent need for specific modifications in the existing security mechanisms used in cellular communication systems, in order to enable them to face evolving security challenges.

This thesis focuses on analysis the existing security mechanisms used in GSM (Global system for mobile) and proposes necessary modifications in them in order to enhance security. In GSM generally three linear feedback shift registers are used to fragment key in data encryption mechanism. In the dissertation, it is proposed that in order to improve the security, the number of registers is increased, then the number of combinations also increases. As combinations increase, the vulnerability of intercepting the encryption key decreases that enhances the security of the system.

Moreover, a closed-form formula is developed to calculate the number of combinations of the session key if it is fragmented into a number of registers. The results are plotted to show the trends of the number of combinations and the vulnerability level of interception with respect to the number of registers. It is observed from the simulations that when the number of registers is increased, the number of combinations also increases; on the other hand the vulnerability level of interception decreases. In this way, personal data or information is protected and hacking of data becomes difficult.

The derived formula can be used for the analysis of security mechanism in GSM, wireless sensor networks and internet of things (IOT).

Contents

List of Figures

List of Tables

Abbreviations

AMPS	Advance mobile phone systems
GSM	Global system for mobile
3GPP	3rd Generation partnership project
UMTS	Universal mobile telecommunications system
SAGE	Security algorithms group of experts
MS	Mobile Station
BTS	Base Transceiver Station
VLR	Visitor location Register
SIM	Subscriber Identity module
SRES	Signed response
LFSR	Linear feedback shift registers
D2D	Device to Device Communication

Symbols

K_c	Session key
F_n	Frame number
K_i	Authentication key
P_b	Vulnerability level of interception
C_b	Total number of combinations of bits

Chapter 1

Introduction

1.1 Evolution of Mobile Communications

Communication is very important in our daily lives. To convey one's message anytime and anywhere, one's connection with another party is of major concern. In ancient times, people use fire, smoke and carrier pigeons for sending their messages. With the passage of time, it is noted that there must be a fast carrier so that our messages can be received with less time consuming and with less loss of actual information. Marconi invented a radio in 1880s. Radio is the invention which is much faster than other methods to convey message to anyone. To transfer data from one end to another, the medium and its properties are very important. The mobile telephone service (MTS) was introduced in 1946. This radio connection was simplex- only one party could speak at one time. The drawback of this technique was its high power requirement for which the instruments carried large batteries. To overcome this drawback, Advance Mobile Phone system (AMPS) was introduced. AMPS contained 666 paired voice channels and analog-frequency based modulation. In 1991s, the AMPS were modified as Narrowband AMPS.NAMPS employed a digital modulation, voice compression and time-division multiple access. However, the need for more capacity paved way to the introduction of Global System for Mobile Communications (GSM). GSM is the first system that permits

a cellular user in one European country to operate in another one [1].

GSM (Global system for Mobile communication) proved to be a fast growing and the most popular telecommunication systems. The First network of GSM was 2G and it was opened in Finland. Due to the advancement in the technologies, the evolving in 2G was expanded continuously. Currently, 3G cellular systems are deployed worldwide. The 3G standards were developed by ITU (International Telecommunication Union) under the name of UMTS (Universal Mobile Telecommunication Systems

Later on, the most important companies in telecommunication system connected to 3GPP program. The main role of this program was to produce the specification for a 3G system. 3GPP organization was the only program which bears the responsibility of the 3G system. The radio spectrum allocation for UMTS is same Europe and Japan, but it was different in the United States i.e. IMT-2000 is allocated to 2G. The 3G proposal was compatible with IS-95B systems, means they can both exist in the same spectrum at the same time. Later on, these allocations are extended. There are many ways to enhance the infrastructure. Further specifications for GSM have been transferred by 3GPP groups [2].

1.2 Security in GSM

In mobile communications, data is transferred from one mobile end to another mobile or fixed end. This data is personal and confidential. It must be sent in such a way that hacking of data is almost impossible. In this regard, GSM being one of the most popular mobile communication systems must be a secure network to retain confidentiality of data from transmitter to receiver. Since in GSM, transmitter and receiver communicate with each other wirelessly; therefore, an intruder may use air interface to hack data of legitimate user in the network. Hence, GSM ought to be secured from such attacks through the validations of legitimacy of its users.

Many cryptographic algorithms are used to secure the data. GSM has a special way to secure data:

1. Authentication
2. Encryption

The three major algorithms used in GSM are A3, A5, and A8. These are introduced by Security Algorithms Group of Experts (SAGE). The three algorithms as mentioned earlier are explained as follows:

- **A3 Algorithm**
  This algorithm is used to authenticate the user by authenticating its SIM card in GSM network.
- **A5 Algorithm**
  After authentication of the user, this algorithm is used to encrypt voice and data. This algorithm has 4 versions. A5/0 is dummy cipher and it is weak algorithm, A5/1 is used in some parts of the world. The newest version is A5/3 which is based on KGCORE function for 3GPP mobile communications.
- **A8 Algorithm**
  This algorithm is used for generating symmetric encryption keys that are used in A5/1 or A5/2 [3].

There are three main components related to GSM Security:

- Base Transceiver Station (BTS)
- Mobile Station (MS)
- Visitor location Register (VLR)

1.2.1 Base Transceiver Station (BTS)

The main function is to generate a Random number (RAND) of 128bits and transfer it to the Mobile station. Its working also includes receiving the SRES and transmitting it to the VLR.

1.2.2 Mobile Station (MS)

The MS uses A3 and A8 algorithms. Both algorithms used 128bits Random number (RAND) and the authentication key (K_i) stored in a SIM.A8 algorithm produces 64 bit session key which is used in encryption and A3 algorithm is used to get SRES of 32bits.

1.2.3 Visitor Local Register (VLR)

The working of VLR is same as MS it contains A3 and A8 algorithms. It produces 32 bit SRES and received SRES form BTS produced by MS are matches its own signed response with received SRES, if they are same so Authentication is possible. As shown in Figure 1.1 given below.

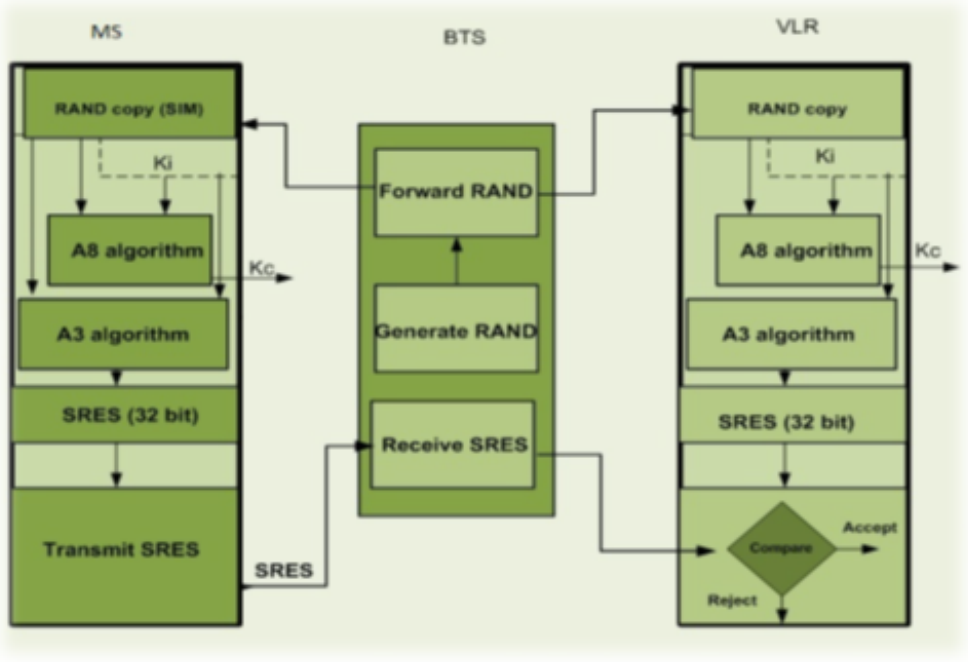

FIGURE 1.1: Security of GSM.

1.3 Authentication Center

Authentication is the first most important step in GSM security. The GSM authentication center is used to check authenticity of every SIM card available to users. If authentication fails, then the user is blocked from using the network. On the other hand, when authentication is achieved, the user is allowed to communicate with the network. In the early 1990s, the initial security key of 64 bits seemed to be reasonable, but due to rapid improvements in computing power and successful attack of brute forces (Programs that decode key or passwords). Length or size of authentication key was made to be 128 bits. As we know that the authentication key (K_i) is the most essential component in the authentication process. Therefore, it cannot be transmitted over the air without any protection. It is recommended that it must be placed in SIM card in MS.

1.3.1 Authentication Process

The A3 algorithm is used to authenticate the user either it is legal for communicating with the network or the user is blocked.

A3 Algorithm

The signed response (SRES) of 32-bit from MS is formed by using the steps so that the subscriber is allowed to communicate with the GSM network:

- Firstly, the BTS generates 128bits Random number and then transmits it to the mobile station.
- Using A3 algorithm, the cell phone encrypts a Random number (RAND) and as a result 32 bit signed response is generated. As shown in Figure 1.2 given below.
- In parallel, the VLR can calculate SRES at the same time because it has a copy of A3 algorithm.

- The VLR matches the value of receiving SRES with its own calculated SRES value.
- If both the values are similar, the authentication is accepted and subscriber can use the GSM.
- If both the values are dissimilar, the connection is not accepted and error is sent to the cell phone.

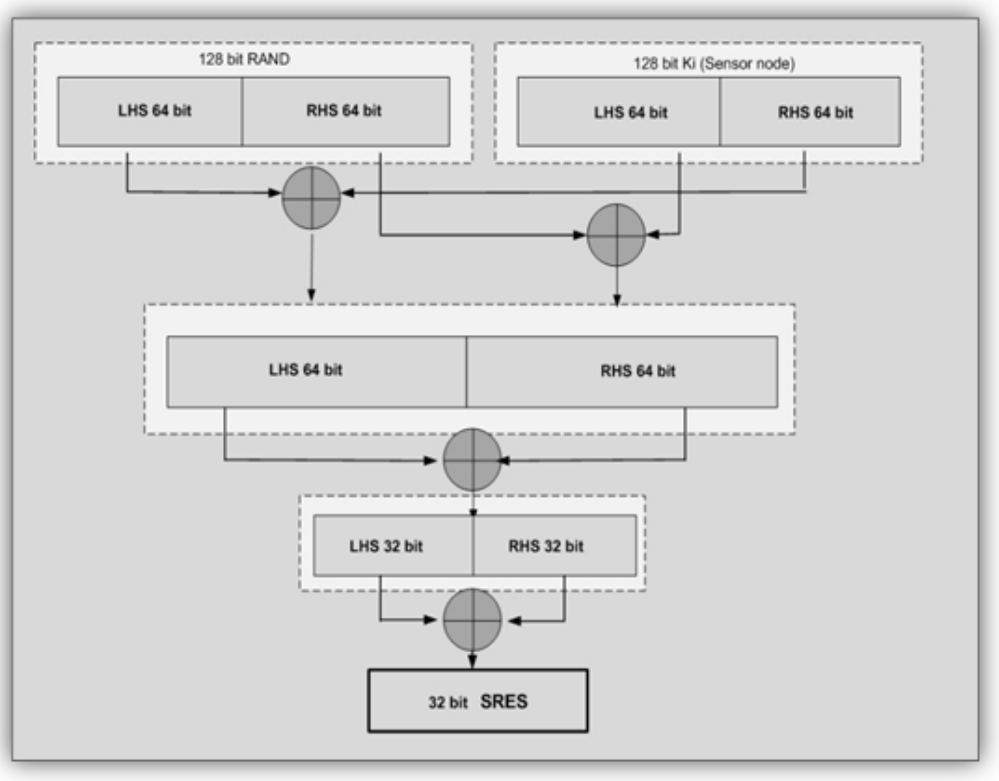

FIGURE 1.2: Block diagram of A3 Algorithm generating 32-bits SRES.

1.4 Encryption in GSM

The second important step in GSM security is encryption. For this purpose, session key, denoted by K_c plays a vital role in securing the data. Both parties, i.e the BTS and MS must have similar K_c in order to encrypt or decrypt the data. The session key (K_c) generated by the MS with the help of A8 algorithm, by taking its authentication key K_i and the random number RAND sent by the BTS. Then

the process of encrypting the data through session key is carried out using A5 algorithm.

1.4.1 GSM A5/1 Algorithm

As discussed earlier, the A5/1 algorithm is utilized for the encryption of data in GSM. For this purpose, 64 bits K_c generated by A8 algorithm is used with a 22-bit frame number denoted as F_n. During the call, the frame number changes, but session key (K_c) remains the same. Using K_c and 22-bit frame number, the A5/1 algorithm outputs a 228-bit keystream for the duplex link. 114 bits of this 228-bit keystream are used for MS to BTS link while the rest of 114 bits are used for BTS to MS link. In order to perturb the position of bits, the A5/1 algorithm uses three registers called linear feedback shift registers (LFSR) of length 19, 22, 23 bits. These registers are denoted by R1, R2, and R3 in our case as presented in figure 1.3. After when the bits are moved to three shift registers, a majority function is applied to select two registers as successors. The step by step process of shift register utilization, majority function application and successor selection is shown in Figure 4. After the bits are shifted to the respective registers, a majority rule is applied to elect the successor. For this purpose, bits from position C1, C2 and C3 are picked from R1, R2 and R3 registers respectively. The majority rule picks two successors based on bit one or bit zero. The position, having maximum numbers of 1s or 0s will become the successor like in our example the register R1 and register R3 are the successors as there are two 1s and one 0. In the next step, the two registers R1 and R3 are stepped forward. In register R1, bits are at the position 13th, 16th ,17th and 18th are taken and XORed with each other. The result is saved at LSB by shifting the register values to right side. Similarly, from the register R3, the bits from position 7th , 20th , 21st and 22nd are picked and XORed with each other saving the result at LSB.

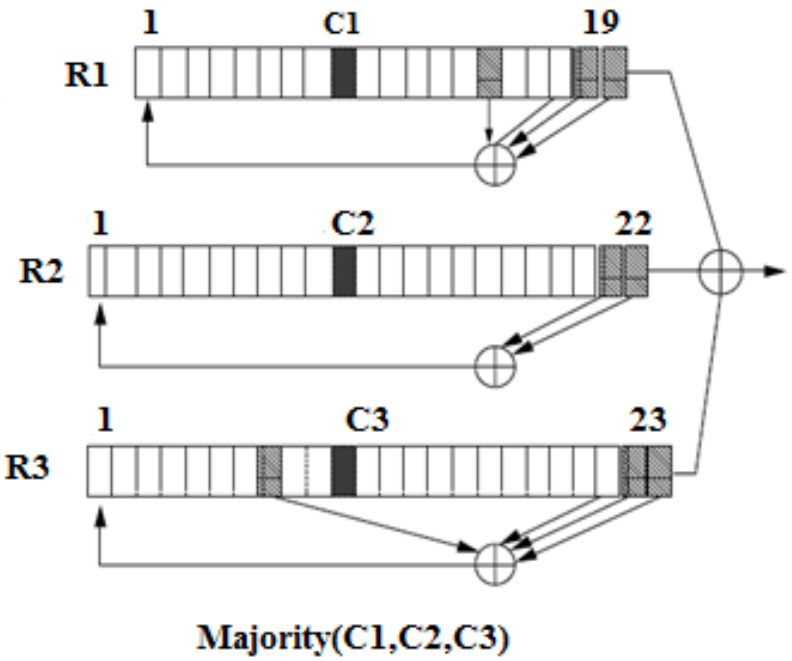

FIGURE 1.3: A5/1 stream cipher.

1.4.2 Generation of Key Stream Bits

The following steps are used to generate keystream bits from frame counter F_n and session key Kc. K_c.

Step 1: The three registers are first initialized as zeroed; the clock will continue till 64 cycles are completed. In this step, each bit of session key (K_c) is XORed in parallel with the MSB's (Most significant bits) of each register.

Step 2: Now the three registers are clocked for additional cycles with a stop or go clock control. In the second step, 22-bits of frame number F_n are again XORed in parallel with the MSB's of the three registers respectively.

Step 3: In this step, the three registers are clocked for 100 additional cycles with a stop or go control. Clocking follows the majority rule. Majority bit is determined based on clocking bit of registers (LFSR1 clocking bit: C1=8th bit, LFSR2 clocking bit: C2=10th and LFSR3 clocking bit: 10th bit). If clocking bit of any register is the same as the majority bit, the register is clocked. This process is done just to mix the Frame number (Fn) and session key (K_c) together.

Step 4: In step 4, the three registers are clocked 288 times with stop or go control so that two 114-bit sequences of the output keystream are produced. As shown in Figure 1.4. The output keystream bits of each register at every cycle are XORed with 114bits of the plaintext [4].

FIGURE 1.4: Block Diagram of A5/1.

1.5 Research Objective

In GSM, the data of each user is protected through A5/1 algorithm that usually works by dividing 64-bit session key into three registers, i.e. R1, R2 and R3. The registers are then incorporated with a predefined rule to generate encrypted data. Although, the encryption mechanism of GSM is secure; however, no probabilistic model has yet been formulated to analyze the probability of breaking or interception of K_c. Therefore, an expression must be developed to analyze the security of the encrypted data. Furthermore, the security of A5/1 algorithm can be enhanced by employing more than three registers with a predefined rule.

1.6 Thesis Organization

The rest of the thesis is organized as follows. Chapter 2 represents a detailed overview of related work in the field of GSM security along with the issues and problem statement in the same field. Chapter 3 represents various encryption/decryption mechanisms used in GSM. Chapter 4 proposes a probabilistic model for interception in GSM followed by conclusion in chapter 5.

Chapter 2

Literature Survey

2.1 Literature Survey

As technologies increase day by day, the encryption or decryption mechanism must be improved in such a way that our data. Many encryption or decryption mechanisms are introduced that are useful in wireless mobile communication systems.
In [5], a comprehensive overview on security of mobile phone networks, evaluation of attacks and defense is presented. The paper addresses downgrade attacks and mitigation of denial of service by smart protocol. Numerous attacks and error across the three mobile network generations are discussed. Cryptography will help the data to be secure and confidential. Cryptographic systems are designed in such a way that everything is known publicly and just focused on the secrecy of the key. The following table 2.1 shows the weak cryptography. The authors also discussed the attacks by breaking either the key on SIM card that is the authentication key (K_i) or the session key (K_c).
In GSM there is no network authentication. Due to this, GSM has security issues, therefore; some authentication protocols must be introduced in GSM. A high quality mode of operation like downgrade protection scheme needs to be suggested in future to secure pre-authentication traffic. In future, a secure algorithm like Device to Device communication (D2D) is used in such a way that the data is securely

	Cipher	Type	Effective(nom.) Key length	Attachable
2G	A5/0	Null Cipher	-	●
	A5/1+ Comp128v ½	LSFR Keystream	54(64) bits	●
	A5/1+ Comp128v 3	LSFR Keystream	64 bits	●
	A5/2	LSFR Keystream	40(64) bits	●
	A5/3	KASUMI	64(128) bits	◐
	A5/4	KASUMI	128 bits	◐
	GEA1	LSFR keystream	64(96) bits	?
	GEA2	LSFR keystream	64(125) bits	?
	GEA3	KASUMI	64(128) bits	◐
	GEA4	KASUMI	128 bits	◐
3G	UEA0	Null Cipher	-	●
	UEA1	KASUMI	128 bits	◐
	UEA2	SNOW 3G	128 bits	○
4G	EEA0	Null Cipher	-	●
	EEA1	SNOW 3G	128 bits	○
	EEA2	AES	128 bits	○
	EEA3	ZUC	128 bits	○

? No information ◐ At acks known, but not pract cal

○ Not at achable ● At acks with commodity hardware

TABLE 2.1: Overview of cipher.

communicated from one point to another.

In [6], the authors presented the GSM components, its security model; attack on A5/2 algorithm and with this efficient attack how other algorithms are hacked. They also discussed how to protect data from these types of attacks. The most efficient attack was the use of the fake base station as presented in figure 2.1. It is not difficult for a fake base station to get information from an original base station. Man-in-the-middle is an active attacker; with the help of this attacker, a person can easily decrypt the information or encrypted data.

To prevent data from attack majority functions that were used to produce output by taking bits from different registers must take more bits so the degree of equations will increase. By doing this, solving those equations are very difficult, so data will be secure. After initialization, the values stored in all registers allow the attack, not setting the values in those registers would increase complexity of an attack. Another way to save from attack is that before encryption error corrections codes are used.

The major flaw is the clocking function for the A5/2 is non linear when introduced

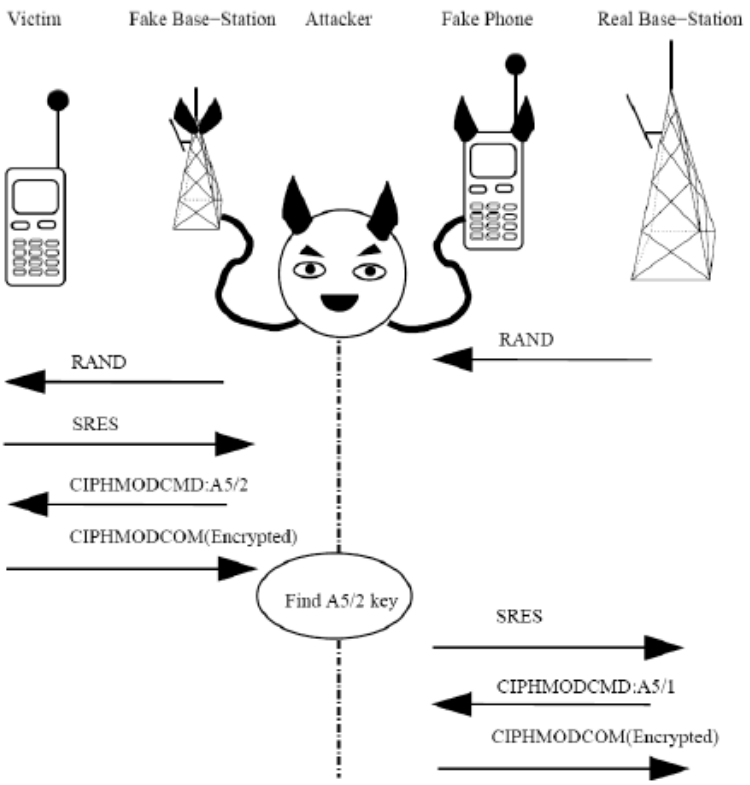

FIGURE 2.1: Man-in-the-middle attack.

register R4, the system become non-quadratic and construction breaks down. Another flaw is that only the network is authenticated not the cell phone so attack becomes easy. By using the same session key (K_c) for authentication and encryption, the cipher becomes weaker and attack become easier.

In [7], the authors allowed better flexibility in term of communication, increasing security and privacy. For authentication process in GSM algorithm A3 was used and for encryption mechanism A8 was used. The SRES form was used for authentication and in encryption of voice data. After simulation, it was observed that this proposed scheme was able to detect higher probability of errors or attacks and this scheme is based on less power consumption and low cost than existing schemes.

In [8], the authors described an encryption mechanism for GSM communication and various attacks on GSM protocol. As soon as possible cryptographic algorithms are replaced by GSM operators or more protected 3rd generation system. The paper discussed the attacks on algorithm A5/1 and A/2 and designed a more

secure A5/3 algorithm. The paper also demonstrated Man-In-The-Middle attack, the authentication mechanism and encryption mechanism.

In [9], the researchers have discussed A5 algorithm, which consists of three registers of total length 64 bit session key (K_c) and 22-bit public key. First, they have constructed three registers R1, R2, R3. These registers have a binary clocking sequence which is obtained from clock-sequence. The main objective of this paper was to reconstruct 64-bit session key from keystream, produced by 22-bit public key and session key (K_c). The objective of the attack was to know key streams for finding the internal state of LFSR. If we know the secret key from initial states of LFSR attacking become much easier. By increasing the internal memory of A5 algorithm security level will be increased. Another way of increasing the security of A5 algorithm is to place a clock in two bits this will make the mechanism more secure.

In [10], the authors described a new attack on A5/1 algorithm. During the first two minutes of conversation the output of A5/1 algorithm requires for the first attack, and key can be computed within one second. In second attack the output occurs after two seconds and obtains the key just in few minutes. These two attacks are of different type of time-memory tradeoff. In this paper, three LFSR are used to form a keystream of 228 bits from 64-bit session key and 22-bit frame number.

In [11], the author presented the real world attack on the internal state of A5/1 algorithm within 6hours. It uses a low hardware cost device. To make the attack possible there must be 64 consecutive bits of keystream. Another contribution of this paper is that the attack optimization is improved up to 13%in computation time.

In [12], the researchers described the method for evaluation of stream cipher strength. It gave the result of the linear weakness of the arbitrary keystream generator. The method used in the paper is called linear cryptanalysis of Stream cipher and based on circuit approximation of finite-state machine. It discussed linear correlation, used for the attack on the initial state of keystream. The linear cryptanalysis of arbitrary keystream generators and irregularly clocked control

shift registers based on linear or non-linear feedback, gave linear statistical weakness. Finally, they concluded that linear cryptanalysis attack is easy much easier one this simple number of linear feedback shift registers (LFSR).

In [13], the authors discussed about the security of the data that how keystream is converted into plaintext and the way to construct self synchronization on the stream cipher so that data can be more secure. The cipher text symbol depends on both the previous and current plaintext. This paper proposed a way to construct the key from any secure stream cipher in stream cipher mode (SCM). In this mode the message length can be large.

In [14], presented linear equations, appreciable to keystream generators producing short keystream sequences are presented. As a result of this attack, the linear correlations in Bluetooth combiner on output and input are characterized to reconstruct a 128-bits secret key. The modified Bluetooth stream cipher is more resistant to linear cryptanalysis.

In [15], the authors demonstrated the attacks of binary keystream generators based on irregularly and regularly clocked shift registers with and without memory. By using regularly clock control linear feedback shift registers could hardly deal with large correlation noise. Linear feedback shift registers (LFSR) combined with memory less function to resist the correlation. The linear feedback shift registers combined with memory based on clock-controlled seems to be more secure.

In [16], the weakness of linear cryptanalysis of MUGI with two objects, to find linear statistical distinguisher and to reconstruct the secret key is described. MUGI also known as buffer is a specific keystream generator for stream cipher. The buffer does not provide a good statistic of output sequences.

In [17], the authors presented a new, more secure linear feedback shift registers (LFSR). In this method, the structure used for linear feedback shift registers is combined with LFSR keystream to obtain a stream generator. Security can be obtained by using Boolean function. This introduces a non linear structure similar to a non-linear filter generator.

In [18], the mechanism of LFSR, its classification and properties such as statistic behavior, period and linear complexity which play a vital role in stream cipher

design is discussed. Different techniques are also discussed to show non linearity in generating sequence to make it cryptographically more secure.

In [19], the researchers described the attacks on stream cipher and clock-controlled stream ciphers. It also demonstrated the new design word-oriented stream cipher using dynamic feedback control and showed analysis result of its performance and security. This stream cipher offers a high performance encryption or decryption mechanism for software implementations.

In [20], the researchers discussed about the several families of keystream generators which resist attacks. In these types, the number of known bits is used to synthesize generator and is called synthesizing algorithm. They also described other attacks such as correlation attacks, differential cryptanalysis attacks, and linear Cryptanalysis attacks.

However, no probabilistic analysis was provided that can quantify the vulnerability level of interception of security keys.

2.2 Problem Formulation and Proposed Solution

In order to make the GSM based cellular communication more secure and protected, there must be a non-conquerable or at least hard-to-conquer encryption mechanism to avoid interception by the third party or malicious machines. The K_c which is obtained from the A8 algorithm is used to encrypt the personal data before it is sent. If the K_c is more secure, then there would be less chance of hacking the private or personal data.

If we increase the length of session key K_c which is currently 64 bits, it would certainly be more secure and there will be less chance of losing the data interception. Secondly, as discussed earlier, by placing the 64 bits in three different registers is one of the mechanisms used to perturb K_c before it is exploited in the process of encryption.

Keeping in view the diversity in data interception techniques, various options for more data security in GSM need to be explored. One of such options may include by increasing the length of K_c from 64 to 128, or any possible length. But this

will result in requiring changes in various levels, which may not be possible at this stage. The second option is to increase the number of shift registers from 3 to any possible odd numbers (5, 7, 9 etc.) by keeping the size of K_c 64 bits. Increasing the number of the shift register will not require any substantial architectural change in the overall security mechanism of GSM but will certainly decrease the vulnerability level of interception of security keys.

In order to confirm the strength of the modified version of the session key for perturbation, we aim to develop a probabilistic model that finds the vulnerability level of the interception of Kc in GSM and evaluates the vulnerabilities of the encrypted data.

2.3 Research Methodology

The solution to the problem proposed in this thesis is the output of extensive literature reviews of books, conference and journal papers. We derived a probabilistic model and implemented it in Matlab® for simulations. The model was refined by tuning the variables and narrowing down the parameters in order to get optimum results. Moreover the results were obtained after extensive simulations by setting up different security environments in order to validate the strength of proposed scheme.

2.4 Thesis Contribution

We have developed formulation based on analytical expression and tried to reduce the vulnerability level of code breaking with the help of this new formula.

Chapter 3

Encryption Mechanism in GSM

3.1 Existing GSM Encryption Technique

A5/1 encryption algorithm is a well known encryption algorithm used in GSM. In A5/1 algorithm a session key (Kc) and frame number of 22 bits are used for the process of encryption. As discussed in chapter 1, the A5/1 uses three linear feedback shift registers usually named as R1, R2 and R3 of length 19, 22 and 23 bits respectively. The taped bits of each register are XORed together and the output of each register is again XORed that represents one keystream bit. The tap positions of each register are pre-determined. For example the bit positions of R1 are 13, 16, 17 and 18, the taps positions of R2 are at 20 and 21 and the bit positions of R3 are at 20, 21, 22 as shown in figure 3.1.

The registers are clocked in a stop/go manner using majority rule. Each register has clocking bit i.e. bit 8 for R1, bit 10 for R2 and R3. At each cycle, the majority function is applied on clocking bits. After the majority function, the successive registers are selected for onward process. By using session key (K_c) and frame counter F_n we get keystream bits. The three linear feedback shift registers are initialized as zeroes, the registers are clocked for 64 cycles and then each bit of K_c is XORed in parallel with the previous XORed bits of each register. After this, 22 bit frame number is XORed with feedback of each register. This process

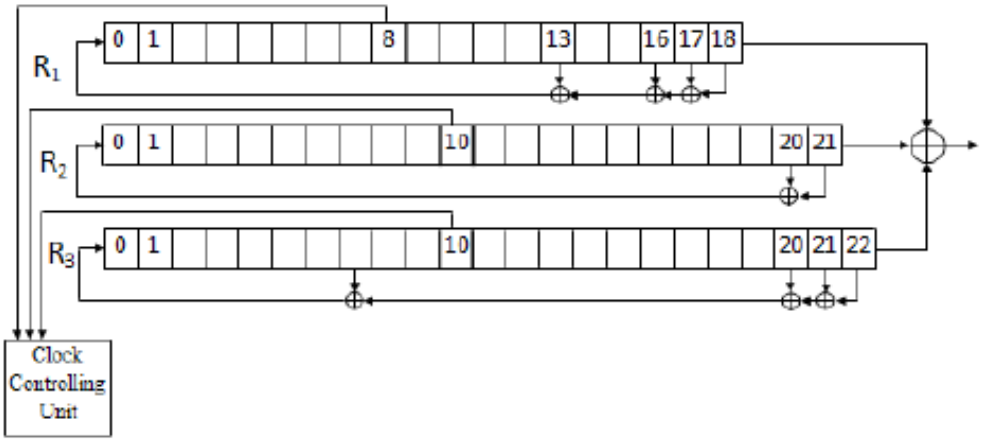

FIGURE 3.1: A5/1 Stream Cipher.

continues up to 22 cycles. The registers are then clocked for 100 additional cycles followed by the majority rule. Finally, the registers are clocked for 228 cycles and output of each register is XORed and two 114-bit sequences are produced named as keystream bits. This process takes place using three registers; however, we can increase the number of registers, in order to enhance security.

The number of registers can be increased to protect the data; the number of registers must be odd, i.e. 5, 7 or 9 etc., because if the number of registers is more even than the majority function will not work and there comes a tie.

3.2 Proposed Modifications in GSM Encryption Mechanism

As discussed earlier, the security of data always remains a primary concern by the research community. The evolving strategies and system manipulation are also a great challenge for GSM technology. The standard GSM technology uses three linear feedback shift registers as a first step in the process of data encryption. Thus, there is a need for improving encryption techniques in such a way that the security mechanism can be implemented in GSM without any adverse effects. For this purpose, we propose a new data security design that can utilize more than

three shift registers. Moreover, we develop a model for the calculation of the possible number of registers that can be used to accommodate the 64 bit key.
Three scenarios are presented in this section labeled as encryption with five registers, encryption with seven linear feedback shift registers and encryption with nine registers in section 3.2.1, 3.2.2, 3.2.3 respectively. The majority function is used to select the registers for further perturbation is carried out to make it harder for the attacker to find the arrangement of bits in a register. If we have 3 Linear feedback shift registers (LFSR) each register having length 8 bits, the working of majority function is as shown in Figure 3.2

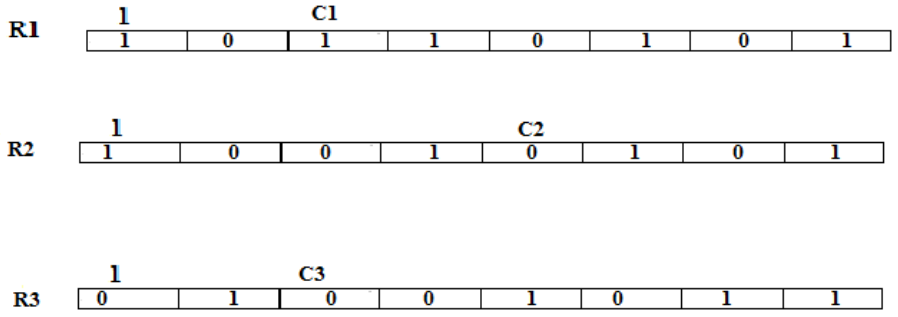

FIGURE 3.2: Majority rule.

$$\mathrm{M} = \mathrm{Majority}(R_1(3), R_2(5), R_3(3)) \tag{3.1}$$

$$\mathrm{M} = \mathrm{Maj}(1, 0, 0) \tag{3.2}$$

From the figure, it is clear that R2 and R3 linear feedback shift registers are successors. Others steps are almost the same to make keystream bit for each direction for BTS to MS and for MS to BTS.

3.3 Example Scenario

In our example, we increase the number of registers from 3 to 5, 7 and 9, that our data is more secure and protected. The possibility of loss of data will be reduced

by increasing the number of registers keeping the length of K_c constant.

3.3.1 Encryption with 5 Shift Registers

While increasing the number of registers, one thing must be kept in mind that the minimum number of bits will not be less than 2 bits. Each register contains more than 2 bits, then XORed operation can be applied. For example, if we divide the bits of K_c into 5 registers, the first 4 registers contain 13 bits each and 5th register contains 12 bits.

The Registers are named as R1, R2, R3, R4 with a length of 13 bits each and R5 having length 12 bits. The output of all registers is XORed with each other and represents one keystream. The predefined bits of each register are XORed and entered to the left side of each register. The taps of R1, R2, R3, and R4, which are XORed are at 4th, 6th, 12th and 13th bit positions. The taps of R5 which are XORed, are at bit positions 11 and 12. The registers are clocked using majority rule. The registers R1, R2, R3 and R4 all have clocking bit at position 5. The 5th register has a clocking bit, at position 6, as presented in figure 3.3. The clocking bit of all 5 registers is noted and majority bit is determined at each cycle. The only registers are clocked which agrees with the majority bit. When a register is clocked, the taps are XORed and result is stored in the leftmost side of the register. The output of each register is XORed together and as a result keystream is produced.

3.3.2 Encryption with 7 Shift Registers

In this scenario, the seven linear feedback shifts registers R1, R2, R3, R4, R5, R6 and R7 are used. The registers from R1 to R6 have uniform length 9 bits. The 7th register is of length 10 bits. The XORed operation is applied on 5th, 7th, 8th and 9th bit positions of each linear feedback shift registers R1, R2, R3, R4, R5 and R6. The bits of linear feedback shift register R7 which are XORed with each

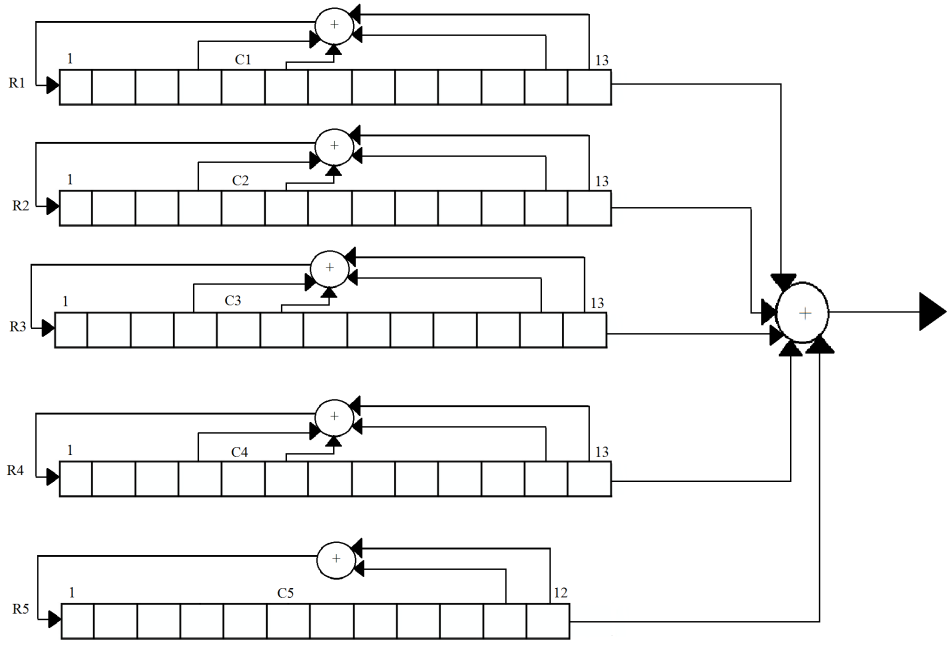

FIGURE 3.3: 5 Linear feedback shift registers (LFSR).

other is at 9th and 10th position. The registers R1, R2, R3, R4, R5 and R6 all have clocking bit at position 3. The clocking bit of 7th register is at bit position 4. When a register is clocked, then its taps are XORed and result is stored in leftmost bit of the linear feedback shift register. And then output of each register is XORed to generate keystream bit. As shown in the figure 3.4.

3.3.3 Encryption with 9 Shift Registers

In this example, Kc is kept constant, i.e, 64 bits and these 64 bits are divided into 9 registers. These registers R1, R2, R3, R4, R5, R6, R7 and R8 are of length 7 bits. The last 9th register is of length 8 bits. The bits which are XORed, are at positions 4 , 5 , 6 and 7, of each of the linear feedback shift register (LFSR) as shown in figure 3.5. The 9th register has taps at 7th and 8th bit positions. The clocking bit of each of first 8 registers is at same position i.e. 2nd position and 9th register clocking bit is at bit position 3 as shown in the figure 3.5.

The same procedure is applied on all registers as follows in A5/1 algorithm for GSM to get the keystream at the end.

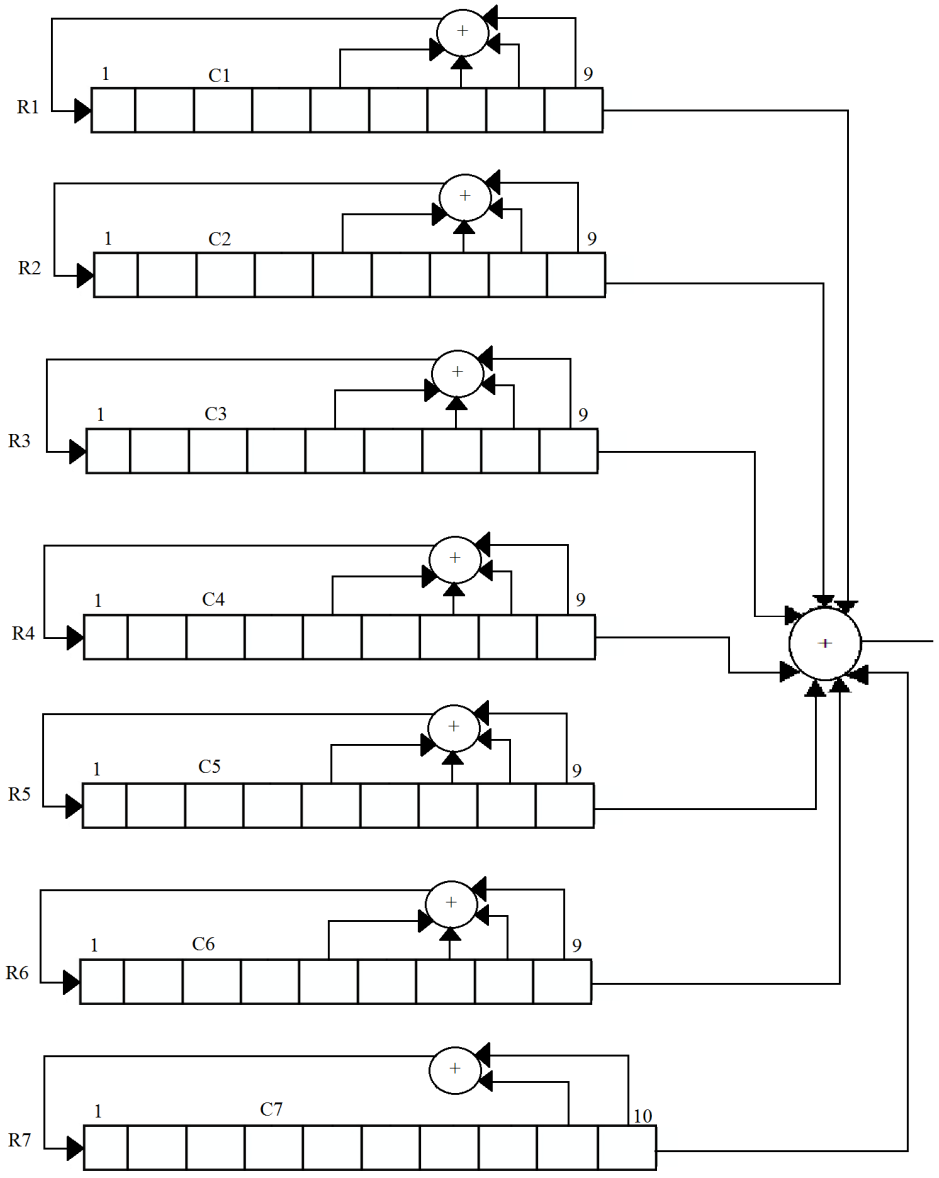

FIGURE 3.4: 7 Linear feedback shift registers (LFSR).

If we increase the length of registers from 3 to the maximum possibility than personal or confidential data is more secure and protected through encryption mechanism. The registers are divided in such a way that the length of any one of the registers will not be less than or equal to 2 bits because in A5/1 algorithm XORed operation cannot be applied properly. Along with this, it must be kept in mind that the number of registers must be odd like 3, 7, 9etc., because if the number of registers is even majority rule will not apply, as in some cases tie may happen. When we increase the number of registers in an odd pattern,

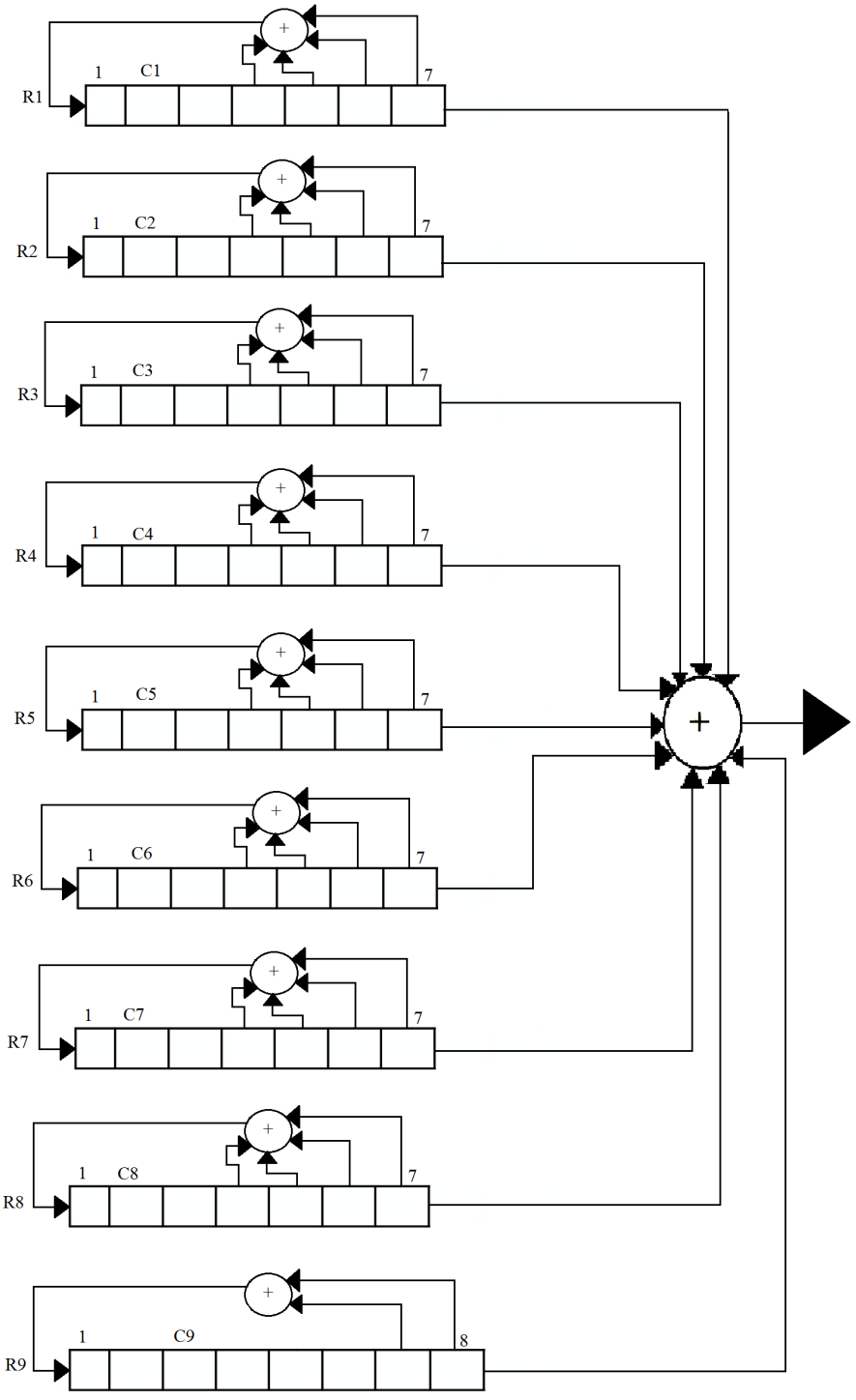

FIGURE 3.5: 9 Linear feedback shift registers (LFSR).

security of data is increased, this can be proved by method of vulnerability level of interception. In the next chapter 4, we discuss how we can find the vulnerability level of interception of different registers and explain that the vulnerability level of interception decreases with increasing the number of registers to a specific number.

Chapter 4

The Proposed Probabilistic Model for Interception in GSM

In this chapter 4, we develop a probabilistic model for interception in GSM. In section 4.1, we discuss the total number of possible combinations of the register contents of the session key K_c and its vulnerability level of interception while keeping the length of the session key constant. However, the number of registers used in this method is not limited to three as in GSM. Section 4.2 presents the formulation of probabilistic model and section 4.3 presents results and discussions. The results have been obtained for varied the number of registers and combinations of K_c.

4.1 Vulnerability Level of Interception of Session Key (K_c) when Multiple Registers are Used

As discussed earlier, the GSM technology uses, A5/1 algorithm for encryption. In this process, the session key (K_c) is divided into three linear feedback shift registers. This process is used for the encrypting of voice data during voice communication. As discussed in chapter 3, we proceed the same data encryption by increasing the number of shift registers.

The vulnerability level of interception of interception for the proposed scheme is given below:

$$P_b = \frac{1}{C_b} \tag{4.1}$$

Where

P_b is vulnerability level of interception of interception.

C_b is total number of combination of bits.

For example, if we have two bit combinations of the encryption key, the vulnerability level of interception of the session key will be 50%. This implies that if the number of combinations (C_b) is increased, the vulnerability level of interception (P_b) of the data will be decreased and vice versa. Likewise in GSM, for the same length of session key (K_c), if the number of registers is increased, the vulnerability level of interception will be decreased.

In our research, we derive a formula for the total number of combinations after increasing the number of registers from three registers up to a possible odd number. In this case, the length of a session key (K_c) will remains constant. As we increase the number of registers in odd fashion, possibility of combinations of bits will increase. Through this technique the vulnerability of the information bits is decreased, resulting in more secure data.

A generated formula for the total number of combinations of n-bits encryption key is segregated in D distinct registers can be written as;

$$C_b = \left[\prod_{d=1}^{D-1} \frac{(n-2D+d)}{d}\right] \times \left[\sum_{i=0}^{n} \binom{n}{i}\right] \tag{4.2}$$

where

C_b is total number of possible combinations of bits

n is the number of bits of session key (K_c)

D is the number of distinct registers where it can take values 3,5,7,9

d is the index that starts from 1 and goes up to $D-1$ values.

Each register can take length of minimum two bits or more. The derivation of this formula is given in Appendix.A.

4.1.1 Example

If we have three registers i.e. $D = 3$
The total numbers of bits in a session key (K_c) is n= 8
Then, the total number of combinations is:

$$\begin{aligned} C_b &= \left[\frac{8-6+1}{1} \times \frac{8-6+2}{2}\right] \times \left[\binom{8}{0} + \binom{8}{1} + \binom{8}{2} + \cdots + \binom{8}{8}\right] \\ &= 6 \times 256 \\ &= 1536 \end{aligned}$$

Hence, the vulnerability level of interception of breaking 8-bit session key segregated in the three registers becomes:

$$P_b = \frac{1}{1536} \tag{4.3}$$

$$= 6.5 \times 10^{-4} \tag{4.4}$$

This example shows that by increasing the number of registers, the total number of combinations will also increase. This means that when bits' combinations increase, then the vulnerability level of intercepting the data decreases; or in other words, there will be less chances for the interceptor to get the session key.
As far as the relationship of the vulnerability level of interception with the number of registers is concerned, by increasing the number of registers the number of combinations of bits increases and the vulnerability level of interception decreases. In other words, with an increase in the number of combinations, the vulnerability level of interception decreases and our data is more secure and protected.

4.2 Results and Discussion

In the coming subsections, we plot the graphs for the total number of combinations and the vulnerability level of interception, when registers are increased. The graphs show that when the number of registers are increased, the number of combinations of bits is also increased, but the vulnerability level of interception of the session key (K_c) decreases.

4.2.1 Number of Combinations

The relationship between the number of registers and the number of combinations is shown in Fig. 4.1. The graph shows that if we increase the number of registers, the total number of bit combinations also increases. After a specific point while increasing the number of registers, the combinations of bits start decreasing. Keeping the length of the session key constant, all possible registers verses combinations are shown in Fig. 4.1. There are a maximum number of combinations at a specific number of registers, i.e. 19. The total number of combinations decreases, even if the number of registers is increased, beyond the number 19.

4.2.2 Vulnerability Level of Interception

In Fig 4.2, the graph shows that with an increase in the number of registers, the vulnerability level of interception decreases and the data or personal information gets more secure and protected. This means when registers are increased, personal or confidential data is secured in such a way that hacking of data is not as much easier. When the vulnerability level of interception decreases, the data becomes less vulnerable. Hence it is proved that when registers in GSM are increased, the number of bit combinations also increased but the vulnerability level of interception decreases. The personal data is secured in such a way that hacking or theft of information becomes less vulnerable.

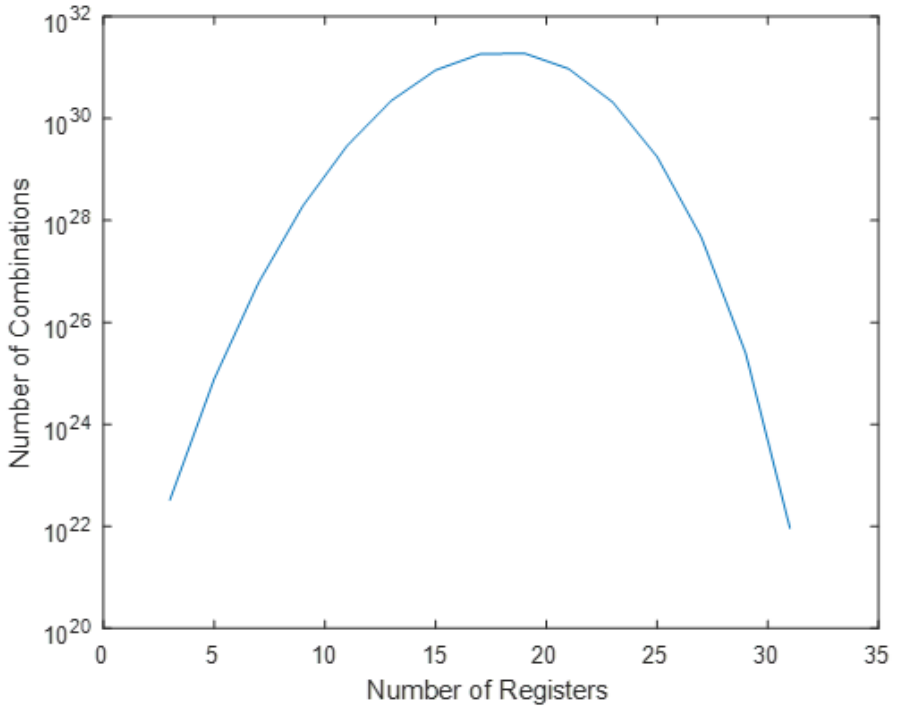

FIGURE 4.1: Possible number of Combinations.

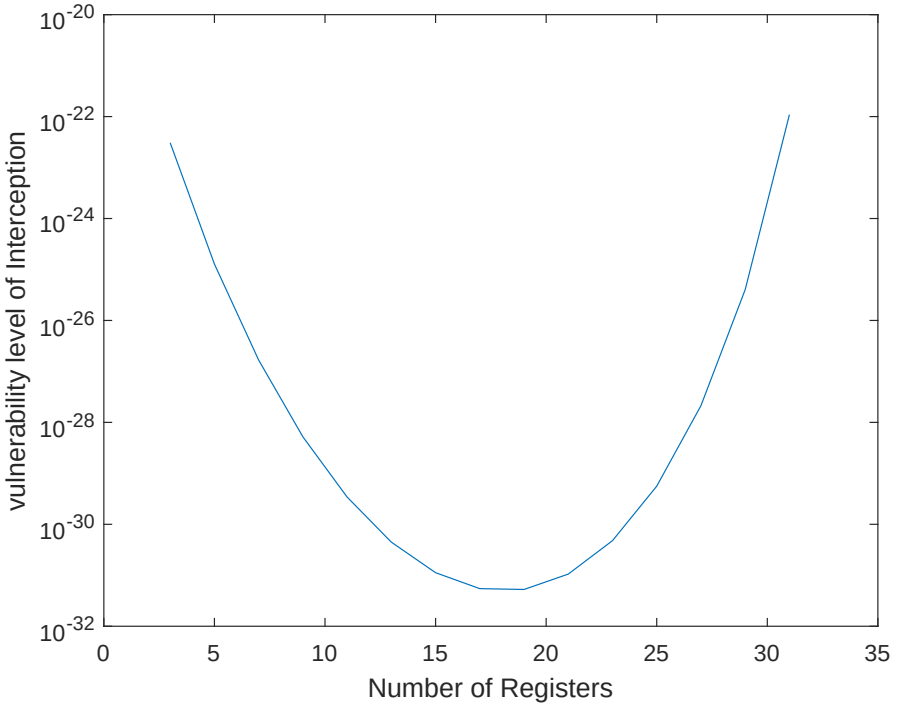

FIGURE 4.2: Vulnerability level of interception.

4.2.3 The Vulnerability Level of Interception, If a Session Key Length is Increased

From Fig. 4.3, it is shown that when session key is increased, number of combinations increases and the vulnerability level of interception decreases. The graph itself given below is nonlinear but it is linear on logarithmic scale.

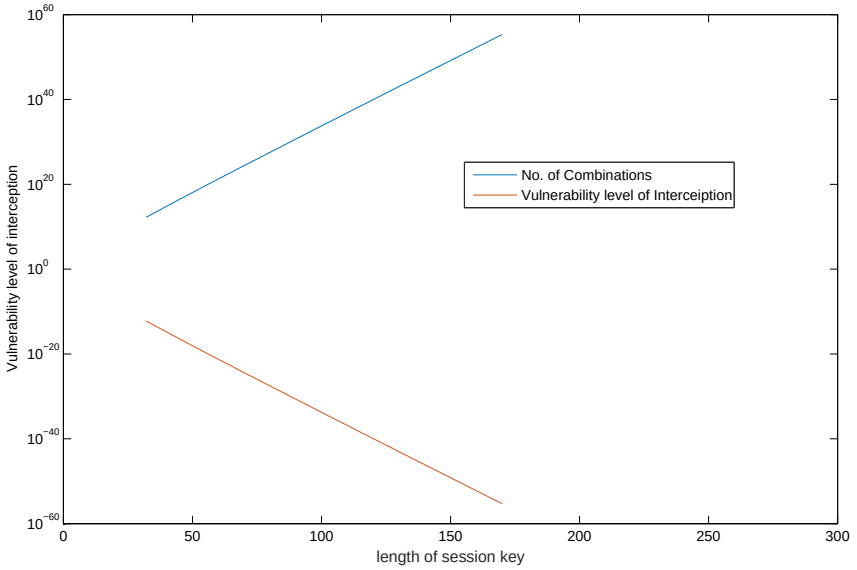

FIGURE 4.3: Vulnerability level of interception and number of combinations verses session key.

Chapter 5

Conclusion and Future Work

In this chapter, the conclusion of the thesis is presented and future prospectives of the proposed research are discussed.

5.1 Conclusion

It has been observed that the security of GSM can be improved if we increase the number of linear feedback shift registers. In this thesis, analysis has been presented that the security of GSM is increased by increasing the number of registers. A modification in the encryption mechanism of GSM has been proposed. It has been observed that when the number of registers is increased, the number of combinations also increase. As combinations increase, the vulnerability level of interception decreases that enhances the security of the system.

Moreover, a closed-form formula has been developed to calculate the number of combinations of the session key if it is fragmented into a number of registers.The results have been plotted to show the trends of the number of combinations and probability of interception with respect to the number of registers. It has been observed from the simulations that when the number of registers is increased, the number of combinations also increase; on the other hand the vulnerability level of

interception decreases. In this way, personal data or information is protected and hacking of data becomes difficult.

5.2 Future Work

The proposed modifications in the encryption mechanism of GSM can be used to enhance the security of 3G+ cellular standards. The derived formula can be used to analyze then encryption mechanisms in GSM, wireless sensor networks and internet of things (IOT).

Bibliography

[1] T. S. Rappaport *et al.*, *Wireless communications: principles and practice*. Prentice Hall PTR New Jersey, 1996.

[2] J. Korhonen, *Introduction to 3G Mobile Communications*. Artech House, Norwood, MA, USA, 2001.

[3] A. Raheem, *An investigation into authentication security of GSM algorithm for mobile banking*. Anchor Academic Publishing (aap_verlag), 2014.

[4] M. Y. Rhee, *Mobile communication systems and security*. John Wiley & Sons, 2009.

[5] D. Rupprecht, A. Dabrowski, T. Holz, E. Weippl, and C. Ppper, "On security research towards future mobile network generations," *IEEE Communications Surveys Tutorials*, pp. 1–1, 2018.

[6] M. Stausholm and M. Dahl, "Insecurity of GSM communication," pp. 1–11, 2006.

[7] M. Saud Khan and N. M. Khan, "Low complexity signed response based sybil attack detection mechanism in wireless sensor networks," *Journal of Sensors*, vol. 2016, 2016.

[8] E. Barkan, E. Biham, and N. Keller, "Instant ciphertext-only cryptanalysis of GSM encrypted communication," *Journal of Cryptology*, vol. 21, no. 3, pp. 392–429, 2008.

[9] J. D. Golić, "Cryptanalysis of alleged A5 stream cipher," in *International Conference on the Theory and Applications of Cryptographic Techniques*, vol. 1233. Berlin, Heidelberg: Springer, 1997, pp. 239–255.

[10] A. Biryukov, A. Shamir, and D. Wagner, "Real time cryptanalysis of A5/1 on a pc," in *International Workshop on Fast Software Encryption*, vol. 1978. Berlin, Heidelberg: Springer, 2000, pp. 1–18.

[11] T. Gendrullis, M. Novotnỳ, and A. Rupp, "A real-world attack breaking A5/1 within hours," in *International Workshop on Cryptographic Hardware and Embedded Systems*, vol. 5154. Washington, D.C., USA: Springer, 2008, pp. 266–282.

[12] J. D. Golić, "Linear cyptanalysis of stream ciphers," in *International Workshop on Fast Software Encryption*, vol. 632. Berlin, Heidelberg: Springer, 1994, pp. 154–169.

[13] ——, "Modes of operation of stream ciphers," in *International Workshop on Selected Areas in Cryptography*, vol. 2012. Berlin, Heidelberg: Springer, 2000, pp. 233–247.

[14] J. D. Golić, V. Bagini, and G. Morgari, "Linear cryptanalysis of bluetooth stream cipher," in *International Conference on the Theory and Applications of Cryptographic Techniques*, vol. 2332. Berlin, Heidelberg: Springer, 2002, pp. 238–255.

[15] J. D. Golić, "On the security of shift register based keystream generators," in *International Workshop on Fast Software Encryption*, vol. 809. Berlin, Heidelberg: Springer, 1993, pp. 90–100.

[16] ——, "A weakness of the linear part of stream cipher mugi," in *International Workshop on Fast Software Encryption*, vol. 3017. Berlin, Heidelberg: Springer, 2004, pp. 178–192.

[17] P. Deepthi and P. Sathidevi, "Hardware stream cipher based on lfsr and modular division circuit," *International Journal of Electronics, Circuits and Systems*, vol. 2, no. 4, pp. 224–232, 2008.

[18] F. Masoodi, S. Alam, and M. Bokhari, "An analysis of linear feedback shift registers in stream ciphers," *International Journal of Computer Applications*, vol. 46, no. 17, pp. 46–49, 2012.

[19] S. Kiyomoto, T. Tanaka, and K. Sakurai, "A word-oriented stream cipher using clock control," in *SASC 2007 workshop record*, 2007, pp. 260–274.

[20] A. Klapper, "On the existence of secure keystream generators," *Journal of cryptology*, vol. 14, no. 1, pp. 1–15, 2001.

Appendix A

Appendix

In order to develop the mathematical formulation for the number of combinations formed with an arbitrary key size n and segmentation group registers,D, the approach of mathematical induction was followed.
For this purpose, a unique sequence of combinations was formed by taking various odd numbers of segmentation group registers,D, with a fixed key size of n =8 and n =16 respectively.

A.1 Taking n =8

Taking n =8 with D=3, following possible combinations of the contents of the group registers appear: Resulting sequence of combinations for n=8 is 6. It was

(2, 2, 4) (2, 4, 2) (4, 2, 2) (2, 3, 3) (3, 2, 3) (3, 3, 2)

observed that D larger than 3 was not possible.

A.2 Taking n =16

With n =16, the only possible group registers are D =3, 5, 7. The following possible combinations of the contents of the group registers appear:

(2, 2, 12)	(2, 12, 2)	(12, 2, 2)	(3, 2, 11)	(3, 11, 2)	(11, 3, 2)	(11, 2, 3)
(2, 3, 11)	(2, 11, 3)	(4, 2, 10)	(4, 10, 2)	(10, 4, 2)	(2, 4, 10)	(2, 10, 4)
(5, 2, 9)	(5, 9, 2)	(9, 5, 2)	(9, 2, 5)	(2, 5, 9)	(2, 9, 5)	(6, 2, 8)
(6, 8, 2)	(8, 6, 2)	(8, 2, 6)	(2, 6, 8)	(2, 8, 6)	(7, 2, 7)	(7, 7, 2)
(2, 7, 7)	(6, 6, 4)	(6, 4, 6)	(4, 6, 6)	(6, 7, 3)	(6, 3, 7)	(3, 6, 7)
(3, 7, 6)	(7, 6, 3)	(7, 3, 6)	(5, 4, 7)	(5, 7, 4)	(4, 5, 7)	(4, 7, 5)
(7, 5, 4)	(7, 4, 5)	(5, 5, 6)	(5, 6, 5)	(5, 6, 5)	(6, 5, 5)	(5, 8, 3)
(5, 3, 8)	(8, 5, 3)	(8, 3, 5)	(3, 5, 8)	(3, 5, 8)	(5, 9, 2)	(5, 2, 9)
(9, 2, 5)	(9, 5, 2)	(2, 5, 9)	(2, 9, 5)	(4, 4, 8)	(4, 8, 4)	(8, 4, 4)
(3, 3, 10)	(3, 10, 3)	(10, 3, 3)				

A.3 With n=16 and D=5

The following possible combinations of the contents of the group registers appear:

A.4 When n =16 and D =7

The following possible combinations of the contents of the group registers appear:

It was observed any number larger than 7 was not possible. Resulting sequence of combinations: 66, 210, 27.

(3, 3, 3, 3, 4)	(3, 3, 3, 4, 3)	(3, 3, 4, 3, 3)	(3, 4, 3, 3, 3)	(4, 3, 3, 3, 3)
(3, 3, 4, 4, 2)	(3, 3, 4, 2, 4)	(3, 3, 2, 4, 4)	(4, 4, 3, 3, 2)	(4, 4, 3, 2, 3)
(4, 4, 2, 3, 3)	(2, 3, 3, 4, 4)	(2, 3, 4, 3, 4)	(2, 3, 4, 4, 3)	(2, 4, 3, 3, 4)
(2, 4, 3, 4, 3)	(2, 4, 4, 3, 3)	(3, 4, 3, 4, 2)	(3, 4, 3, 2, 4)	(3, 4, 2, 3, 4)
(3, 4, 2, 4, 3)	(3, 4, 4, 2, 3)	(3, 4, 4, 3, 2)	(4, 3, 3, 4, 2)	(4, 3, 3, 2, 4)
(4, 3, 2, 3, 4)	(4, 3, 2, 4, 3)	(4, 3, 4, 2, 3)	(4, 3, 4, 3, 2)	(4, 2, 3, 3, 4)
(4, 2, 3, 4, 3)	(4, 2, 4, 3, 3)	(3, 2, 3, 4, 4)	(3, 2, 4, 3, 4)	(3, 2, 4, 4, 3)
(4, 4, 4, 2, 2)	(4, 4, 2, 4, 2)	(4, 4, 2, 2, 4)	(4, 2, 4, 4, 2)	(4, 2, 4, 2, 4)
(4, 2, 2, 4, 4)	(2, 4, 4, 4, 2)	(2, 4, 4, 2, 4)	(2, 4, 2, 4, 4)	(2, 2, 4, 4, 4)
(3, 3, 3, 2, 5)	(3, 3, 3, 5, 2)	(3, 3, 5, 3, 2)	(3, 3, 2, 3, 5)	(3, 3, 2, 5, 3)
(3, 3, 5, 2, 3)	(2, 3, 3, 3, 5)	(2, 3, 3, 5, 3)	(2, 3, 5, 3, 3)	(3, 2, 3, 3, 5)
(3, 2, 3, 5, 3)	(3, 2, 5, 3, 3)	(3, 5, 3, 3, 2)	(3, 5, 3, 2, 3)	(3, 5, 2, 3, 3)
(5, 3, 3, 3, 2)	(5, 3, 3, 2, 3)	(5, 3, 2, 3, 3)	(2, 5, 3, 3, 3)	(5, 2, 3, 3, 3)
(3, 3, 2, 2, 6)	(3, 3, 2, 6, 2)	(3, 3, 6, 2, 2)	(2, 2, 3, 3, 6)	(2, 2, 3, 6, 3)
(2, 2, 6, 3, 3)	(2, 3, 3, 2, 6)	(2, 3, 3, 6, 2)	(2, 3, 6, 3, 2)	(2, 3, 6, 2, 3)
(2, 3, 2, 3, 6)	(2, 3, 2, 6, 3)	(3, 2, 3, 2, 6)	(3, 2, 3, 6, 2)	(3, 2, 6, 3, 2)
(3, 2, 6, 2, 3)	(3, 2, 2, 6, 3)	(3, 2, 2, 3, 6)	(2, 6, 3, 3, 2)	(2, 6, 3, 2, 3)
(2, 6, 2, 3, 3)	(6, 2, 3, 3, 2)	(6, 2, 3, 2, 3)	(6, 2, 2, 3, 3)	(3, 6, 3, 2, 2)
(3, 6, 2, 3, 2)	(3, 6, 2, 2, 3)	(6, 3, 3, 2, 2)	(6, 3, 2, 3, 2)	(6, 3, 2, 2, 3)
(2, 2, 2, 3, 7)	(2, 2, 2, 7, 3)	(2, 2, 7, 2, 3)	(2, 2, 7, 3, 2)	(2, 2, 3, 7, 2)
(2, 2, 3, 2, 7)	(3, 7, 2, 2, 2)	(7, 3, 2, 2, 2)	(2, 3, 2, 2, 7)	(2, 3, 2, 7, 2)
(2, 3, 7, 2, 2)	(3, 2, 2, 2, 7)	(3, 2, 2, 7, 2)	(3, 2, 7, 2, 2)	(2, 7, 2, 2, 3)
(2, 7, 2, 3, 2)	(2, 7, 3, 2, 2)	(7, 2, 2, 2, 3)	(7, 2, 2, 3, 2)	(7, 2, 2, 3, 2)
(2, 2, 2, 2, 8)	(2, 2, 2, 8, 2)	(2, 2, 8, 2, 2)	(2, 8, 2, 2, 2)	(8, 2, 2, 2, 2)

A.5 Mathematical Induction on Resulting Sequences

The total bit combinations consist of two parts

$$C_b = P_1 P_2 \tag{A.1}$$

From inspection, it can be easily observed that, P_2 depends on n and does not depend on D. P_2 reveals that it is a binomial mutation with formulation

$$P_2 = \sum_{i=0}^{n} \binom{n}{i} \tag{A.2}$$

Moreover, P_1 seems to be dependent on the value of D and n. From the above combinations developed manually for $n = 8$ and $n = 16$, for various numbers of group registers, the following sequences of combinations appeared; for $n = 8$,

(2, 2, 2, 4, 6)	(2, 2, 2, 6, 4)	(2, 2, 6, 2, 4)	(2, 2, 6, 4, 2)	(2, 2, 4, 2, 6)
(2, 4, 2, 2, 6)	(2, 4, 2, 6, 2)	(2, 4, 6, 2, 2)	(4, 2, 2, 2, 6)	(4, 2, 2, 6, 2)
(4, 2, 6, 2, 2)	(2, 6, 2, 2, 4)	(2, 6, 2, 4, 2)	(2, 6, 4, 2, 2)	(6, 2, 2, 2, 4)
(6, 2, 2, 4, 2)	(6, 2, 4, 2, 2)	(4, 6, 2, 2, 2)	(6, 4, 2, 2, 2)	(2, 2, 2, 5, 5)
(2, 2, 5, 2, 5)	(2, 2, 5, 5, 2)	(5, 5, 2, 2, 2)	(2, 5, 2, 2, 5)	(2, 5, 2, 5, 2)
(2, 5, 5, 2, 2)	(5, 2, 2, 2, 5)	(5, 2, 5, 2, 2)	(2, 2, 5, 4, 3)	(2, 2, 5, 3, 4)
(2, 2, 3, 5, 4)	(2, 2, 3, 4, 5)	(2, 2, 4, 3, 5)	(2, 2, 4, 5, 3)	(2, 3, 2, 5, 4)
(2, 3, 2, 4, 5)	(2, 3, 4, 2, 5)	(2, 3, 4, 5, 2)	(2, 3, 5, 3, 4)	(2, 3, 5, 4, 2)
(3, 2, 2, 5, 4)	(3, 2, 2, 4, 5)	(3, 2, 4, 2, 5)	(3, 2, 4, 5, 2)	(3, 2, 5, 2, 4)
(3, 2, 5, 4, 2)	(2, 4, 2, 5, 3)	(2, 4, 2, 3, 5)	(2, 4, 3, 5, 2)	(2, 4, 3, 2, 5)
(2, 4, 5, 3, 2)	(2, 4, 5, 2, 3)	(4, 2, 2, 5, 3)	(4, 2, 2, 3, 5)	(4, 2, 5, 2, 3)
(4, 2, 5, 3, 2)	(4, 2, 3, 2, 5)	(4, 2, 3, 5, 2)	(2, 5, 2, 4, 3)	(2, 5, 2, 3, 4)
(2, 5, 3, 2, 4)	(2, 5, 3, 4, 2)	(2, 5, 4, 3, 2)	(2, 5, 4, 2, 3)	(5, 2, 2, 4, 3)
(5, 2, 2, 3, 4)	(5, 2, 3, 2, 4)	(5, 2, 3, 4, 2)	(5, 2, 4, 2, 3)	(5, 2, 4, 3, 2)
(5, 4, 2, 2, 3)	(5, 4, 2, 3, 2)	(5, 4, 2, 2, 3)	(4, 5, 2, 2, 3)	(4, 5, 2, 3, 2)
(4, 5, 3, 2, 2)	(5, 3, 2, 2, 4)	(5, 3, 2, 4, 2)	(5, 3, 4, 2, 2)	(3, 5, 2, 2, 4)
(3, 5, 2, 4, 2)	(3, 5, 4, 2, 2)	(4, 3, 2, 2, 5)	(4, 3, 2, 5, 2)	(4, 3, 5, 2, 2)
(3, 4, 2, 2, 5)	(3, 4, 2, 5, 2)	(3, 4, 5, 2, 2)		

(2, 2, 2, 2, 2, 2, 4)	(2, 2, 2, 2, 2, 4, 2)	(2, 2, 2, 2, 4, 2, 2)	(2, 2, 2, 4, 2, 2, 2)
(2, 2, 4, 2, 2, 2, 2)	(2, 4, 2, 2, 2, 2, 2)	(4, 2, 2, 2, 2, 2, 2)	(2, 2, 2, 2, 2, 3, 3)
(2, 2, 2, 2, 3, 2, 3)	(2, 2, 2, 3, 2, 2, 3)	(2, 2, 3, 2, 2, 2, 3)	(2, 3, 2, 2, 2, 2, 3)
(3, 2, 2, 2, 2, 2, 3)	(2, 3, 3, 2, 2, 2, 2)	(2, 3, 2, 3, 2, 2, 2)	(2, 3, 2, 2, 3, 2, 2)
(2, 3, 2, 2, 2, 3, 2)	(2, 3, 2, 2, 2, 2, 3)	(2, 2, 3, 3, 2, 2, 2)	(2, 2, 3, 2, 3, 2, 2)
(2, 2, 2, 2, 3, 3, 2)	(2, 2, 2, 3, 2, 3, 2)	(3, 2, 2, 2, 2, 3, 2)	(3, 2, 2, 2, 3, 2, 2)
(3, 2, 2, 3, 2, 2, 2)	(3, 2, 3, 2, 2, 2, 2)	(2, 2, 3, 2, 2, 3, 2)	

$P_1 = 66$ and for $n = 16$, $P_2 = 66, 210, 27$.

Using Mathematical induction approach, P_1 seems to be a product of multiple terms that depend on a series with a local variable d ranging from 1 to $D - 1$. As a rule, it was clear that the value of D cannot be taken equal to 2 and bits in each register can also not be taken lesser than or equal to 2. Therefore, it seems $(n - 2D + d)/d$ must be a basic unit of a product with d ranging from 1 to $D - 1$. Using a tedious process of hit and trials the following formula is developed in terms of n and D. Thus P_1 becomes,

$$P_1 = \left\lceil \prod_{d=1}^{D-1} \frac{(n - 2D + d)}{d} \right\rceil \tag{A.3}$$

In case of the occurrence of floating point, a floor value must be taken to ensure an integer value. Therefore, the total number of combinations formula becomes;

$$C_b = P_1 P_2 \tag{A.4}$$

$$C_b = \left[\prod_{d=1}^{D-1} \frac{(n - 2D + d)}{d} \right] \times \left[\sum_{i=0}^{n} \binom{n}{i} \right] \tag{A.5}$$