Independent Research Report

Analysis of the A5/1 Cipher

William Glass

March 4, 2024

1 Abstract

This report examines the A5/1 stream cipher, a critical security mechanism core to the GSM network, focusing on its architecture, operation, and notable vulnerabilities. Developed in the 1980s, it employs a combination of Linear Feedback Shift Registers (LFSRs) to generate encryption keystreams. Initially considered secure, the cipher was later found vulnerable to various cryptanalytic attacks, such as time-memory trade-offs, underscoring the risks of predictable processes and the limitations of a 64-bit key space.

2 Introduction

Employed in 2G mobile networks, the A5/1 cipher was integral to the Global Systems for Mobile Communications (GSM) security, encrypting voice and data to safeguard communication privacy and integrity. LFSRs are at the heart of the A5/1 cipher, critical for creating pseudorandom sequences used in encryption. However, advancements in computational power and cryptanalysis have exposed vulnerabilities in systems previously deemed secure.

Notably, the A5/1 cipher, despite its sophistication, was compromised, significantly influencing current cryptographic standards and prompting the reinforcement of security measures in contemporary encryption algorithms. This report delineates the essential characteristics and operational dynamics of LFSRs within the A5/1 cipher as a foundation for a more detailed examination of GSM security mechanisms.

3 Information Theory

3.1 Galois Fields

In the context of Galois Field $GF(2)$, which operates within a binary system where only two elements exist: 0 and 1. This field forms the basis for arithmetic operations in binary logic, which directly translate to polynomial arithmetic in cryptographic applications and signal processing.

Addition in $GF(2)$ is equivalent to the XOR operation in binary logic. For any two elements a and b in $GF(2)$, their sum $a+b$ is calculated modulo 2, where $1+1=0$, $1+0=1$, and $0+0=0$. This operation is crucial for polynomial addition, where coefficients are added modulo 2.

Multiplication in $GF(2)$ is the equivalent for binary AND operation. The product of any two elements a and b is $a \times b$, with $1 \times 1 = 1$ and $0 \times a = 0$ for any a. This rule applies to polynomial multiplication, influencing the generation of terms in polynomial expressions.

3.2 Stream Ciphers

Stream ciphers are a type of symmetric cipher that use a single key (known as a secret key) for both encryption and decryption functions. These ciphers encrypt data one bit at a time by XORing each plaintext bit with a corresponding bit from a keystream. The keystream is the output produced by a function known as a keystream generator.

Keystream generators typically take a secret key of a defined bit length and, through a function, output streams of pseudorandom or random bits of arbitrary length. These bits are essential for the operation of stream ciphers. The strength of stream ciphers lies in their fast performance and minimal resource requirements during operations, making them ideal for use in systems with limited processing capabilities.

3.3 Linear Feedback Shift Registers

A LFSR is a simple (Linear) Feedback Shift Register (LFSR) design which contains a finite length of stages of n-bit length that are each capable of storing 1-bit, the individual stages are referred to as "Flip-Flops" and serve as memory cells for the register. The feedback function is the product of taps and XOR gates, taps are specific bits from the register which are then XORed to produce the output. Each stage in the LFSR has an input and an output, the output contains the current stored value. The stored values are shifted to the right at given intervals which are known as clock cycles. The LSB (least significant bit) which is the rightmost bit on the register is shifted out as the output bit, and all other preceding bits continue to shift 1-bit to the right. The feedback function produces the new MSB (most significant bit) or leftmost bit which will repeat the sequence each cycle.

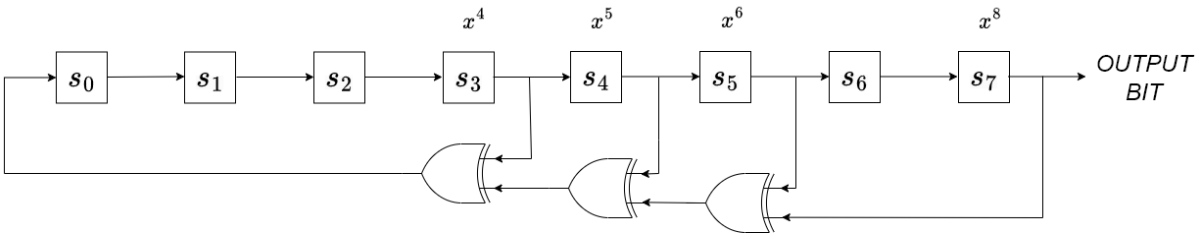


Figure 1: 8-Bit LFSR

3.3.1 Irreducible and Primitive Polynomials in LFSRs

A polynomial is irreducible if it cannot be factored into lower-degree polynomials over a finite field. In the case of Linear Feedback Shift Registers (LFSRs), a polynomial to the *n-th* degree in Galois Fields $GF(2)$ dictates that an LFSR based on this polynomial will produce a maximum length sequence (commonly known as an *m-sequence*) of $2^n - 1$ bits before repeating. For LFSRs to function effectively in stream ciphers, they must also utilize primitive polynomials. Primitive polynomials are a specialized subset of irreducible polynomials that ensure the LFSR cycles through all possible non-zero states, thus generating a sequence that is both long and appears random, which is desirable in cryptographic applications for sequence unpredictability.

The behavior of an LFSR is determined by its feedback function, which can be represented as follows:

$$B_n = A_0 B_{n-1} \oplus A_1 B_{n-2} \oplus \ldots \oplus A_{n-1} B_0 = \bigoplus_{i=0}^{n-1} A_i B_{n-1-i} \mod 2$$

Here, the feedback function B_n is computed by taking the XOR (denoted by $\oplus$) of the product of tap coefficients A_i with the corresponding bits of the register B. The choice of tap coefficients is crucial, as it directly corresponds to the terms of the primitive polynomial that characterizes the LFSR. When the polynomial is primitive, the resulting LFSR is capable of generating a pseudo-random sequence with a period of $2^n - 1$, covering all possible non-zero states of an n-bit register.

For instance, an example of an irreducible polynomial over GF(2) which is not primitive is:

$$Q(x) = x^4 + x + 1$$

While this polynomial cannot be factored into lower-degree polynomials over GF(2), it does not produce a maximum length sequence. In contrast, consider the polynomial in GF(2):

$$P(x) = x^8 + x^6 + x^5 + x^4 + 1$$

This polynomial is not only irreducible but also primitive within GF(2). The associated LFSR will produce an m-sequence of length $2^8 - 1 = 255$ bits, which exemplifies the maximal period achievable by an 8-bit LFSR. Such irreducible and primitive polynomials form the foundation for LFSRs that yield pseudo-random sequences with desirable properties for cryptographic purposes. The specific choice of polynomial depends on the requirements of the application and influences both the length and quality of the LFSR-generated sequence.

4 GSM - Global System for Mobile Communications: 2G

4.1 GSM 2G - Background

In the early days of mobile communications the 2nd Generation (2G) of mobile technology was implemented by the Global System for Mobile Communications (GSM). In 1987 GSM was formed by the consolidation of 15 European companies which signed up to the Global System for Mobile Communications Association (GSMA), and later being adopted as a mobile standard worldwide. The basic network infrastructure for GSM involves a Core Network (CN), which in essence is the backbone of the celluar network.

The CN responsibilities entail the routing, management and inter-operability with other networks. The Radio Access Network (RAN), also known as the Access Network (AN) is primarily responsible for managing the radio communication elements in a cellular network. This includes the oversight of transmission and reception of signals between User Equipment (UE), aka mobile devices, and the network infrastructure.

The Base Station Controllers (BSC) are the central management node of the RAN, which manage radio resources like channel allocation, QoS and control the radio links between multiple Base Transceiver Stations (BTS), aka cell towers, to manage the coverage area of cells and the handover process when a mobile device reallocates from one BTS to another.

GSM networks also contain two databases, a HLR (Home Location Register), the main database that stows authorized subscriber information such as the mobile number, current location and other service related provisions. Serving alongside the HLR a second database called the VLR (Visitor Location Register) which is responsible for temporary storage and handling of roaming mobile devices which have roamed from their own HLR cell coverage area.

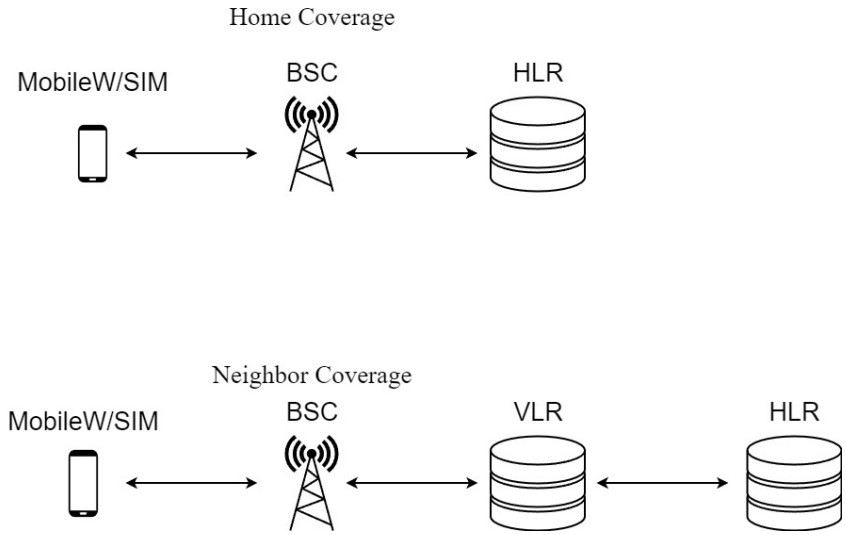


Figure 2: Basic Outline of the GSM Network

4.2 GSM 2G - Authentication Process

The SIM (Subscriber Identity Module), is a microchip which is integral to the authentication process, provided by mobile service provider. Key elements involved in the authentication process are the International Mobile Subscriber Identity (IMSI),which is a unique code that identifies the mobile subscriber, and lastly the Subscriber Authentication Key (K_i), a 128-bit value responsible for client authentication. Mobile devices are additionally identified by its International Mobile Equipment Identity (IMEI), a serial number embedded into the device.

The initiation of the authentication process can be represented as follows:

1. SIM Card and IMSI: The Subscriber Identity Module (SIM) card in the mobile device stores the IMSI, which is a unique identifier for each mobile network subscriber. When the device attempts to connect to the network, the SIM card transmits the IMSI to the mobile device.
2. Transmission of IMSI: The mobile device then forwards the IMSI to the BS.
3. Relaying IMSI to HLR: The Base Station relays the IMSI to the MSC, which then forwards it to the HLR. The HLR works in conjunction with the AuC to authenticate the subscriber. Verification and Authentication Triplets:
4. The HLR/AuC verifies the IMSI and generates five authentication triplets. Each triplet consists of:
5. $RAND$: A random number that is sent to the mobile device.
6. S_{res}: A signed response that is expected from the mobile device, calculated using the secret key K_i stored in the SIM card.
7. K_c: A ciphering key used for encrypting the communication between the mobile device and the network.

Authentication between Mobile Device MD and Mobile Operator O is the Mobile Operator

MD		O
		$\text{R}AND \leftarrow R$
SIM	$\xrightarrow{\text{Req: IMSI}}$	MSC requests HLR/AuC
MSC	$\xleftarrow{\text{Auth Data: } RAND, S_{res}, K_c}$	HLR/AuC generates $(RAND, S_{res}, K_c)$
SIM calculates S_{res} where $S'_{res} = E_{K_i}(RAND)$	$\xleftarrow{\text{Chal: } RAND}$	MSC sends $RAND$
SIM	$\xrightarrow{\text{Resp: } S'_{res}}$	MSC verifies $S'_{res} = S_{res}$
SIM derives K_c	$\xleftrightarrow{\text{secure session: } (traffic)_{K_c}}$	BSC derives K_c

These triplets are used to ensure that the subscriber's SIM card holds the correct K_i without transmitting it over the air. The mobile device uses its stored K_i to calculate the response to the $RAND$, which is then compared with the S_{res} from the HLR/AuC. If they match, the authentication is successful, allowing encrypted communication to be established using K_c.

5 Architecture of the A5/1 Stream Cipher

The A5/1 cipher is a member of the A5 family of symmetric-key algorithms. Its significance was recognized due to its widespread deployment in GSM networks for the encryption of voice communications. Designed in the 1980s, the cipher's design was initially proprietary, casting a veil of security over the operational complexities. This approach suggested a parallel between secrecy and security. Despite efforts to maintain its confidentiality, the A5/1 cipher faced scrutiny from the security community over time. It was eventually reverse-engineered, leading to public disclosure and the subsequent decommissioning of its use.

5.1 GSM Frame Structure and A5/1 Encryption

GSM conversations are transmitted in frames, sequentially ordered to manage functions such as retransmission of lost or corrupt frames, error checking, quality of service (QoS), and data transmission between the mobile device and the network operator. The A5/1 algorithm processes an input message of 228 bits

and outputs ciphertext of the same length. A single frame F comprises two 114-bit segments. Bidirectional communication involving arbitrarily chosen bits of 114-bit length, where MD is the Mobile Device and O is the Mobile Operator for a single frame F, can be depicted as:

$$F$$
$$MD \rightarrow O : \{0,1\}^{114}$$
$$MD \leftarrow O : \{0,1\}^{114}$$

This display captures the use of the GRGSM tool in monitoring a 2G cellular signal. Each line correlates with a specific frame number, providing a structured view of the communication. The hexadecimal values at the bottom represent the raw data frame as captured by the tool, meaning the hex sequence beginning at the lower part of the display corresponds to the encoded frame contents. Observing the frame bytes in Wireshark allows for a detailed analysis of the frames, enabling the identification of a specific frame in hexadecimal form as shown in the GRGSM tool by looking at offset '0030' and proceeding 11 bytes further.

```
746268 1152144: 10101010000110000010010000100110100001001011000000010000011010001000000101001110000101000011000000010100001000100
746269 1152177: 00100000010011101000010010000010000100001010100000000000101001010100001000001000000010000010010010010101000011000
746270 1152210: 00010001100110100001000000100000100001000000000101000000000000000010100000100100100100001100001001100000010000011 0
746271 1152243: 01000000010010011000010000000111010010100001000000101000001101011011100000000010001010010101100100000000000001000
746268 1152144:  05 01 03 03 49 06 1d 10 00 00 00 00 6f d8 00 00 00 00 00 00 00 00
```

Figure 3: GSM Capture of Frame Burst

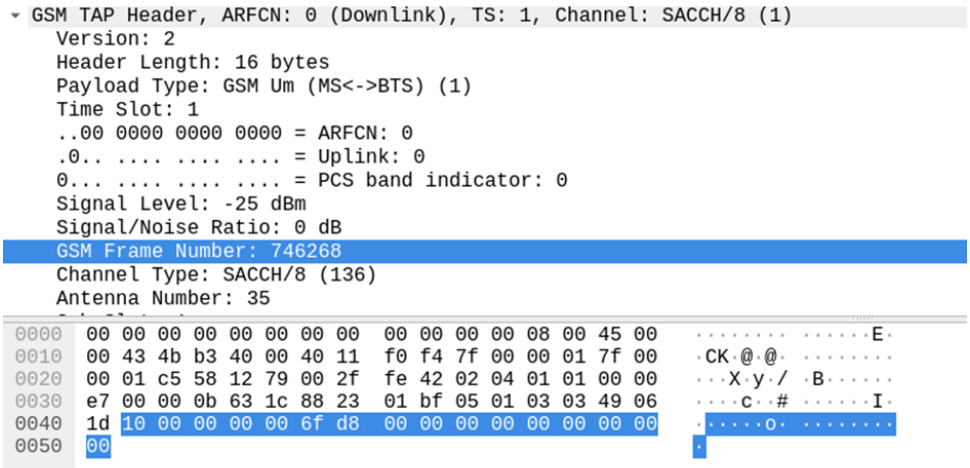


Figure 4: GSM Capture on Wireshark

5.2 A5/1: Initialization Phase

Before keystreams are generated the cipher must first undergo an initial process, known as Initialization or Key Loading. The A5/1 algorithm consists of 3 Linear Feedback Shift Registers of different lengths which provide the function for generating keystreams. The 3 variable length LFSRs (or Registers) are specifically selected to be of 19, 22, and 23 bit-lengths. These selected lengths are called taps and are associated with the irreducible polynomial and primitive factors required, to ensure unpredictibility and maximum length keystreams. Here R_j represents the 3 Registers:

1. $R_{1(X)}$: 19-bit length with feedback taps on bits $x^{18} + x^{17} + x^{16} + x^{13}$
2. $R_{2(Y)}$: 22-bit length with feedback taps on bits $y^{21} + y^{20}$
3. $R_{3(Z)}$ 23-bit length with feedback taps on bits $z^{22} + z^{21} + z^{20} + z^7$

The registers are initialized with a 64-bit ciphering key (derived from the Subscriber Authentication Key) and a 22-bit Frame Number. The process starts by ensuring the 3 LFSRs are set to an initial state of zero. This is then followed by inputting the 64-bit secret key which starts in the input at the LSB (least significant bit), from there each key bit is sequentially entered bit by bit into the LSB of the LFSRs.

Succeeding the initialization of the 64-bit K_c, the 22-Bit Frame Number F_n will now start the frame loading process in a similar operation, where each bit of the frame number is XORed with the LSB of the LFSRs. The frame number is essentially a counter that increments with each frame of communication, which in GSM is roughly every 4.615 milliseconds. It's used as an additional security measure to ensure that keystreams generated by the LFSRs will be individually unique for each communication frame.

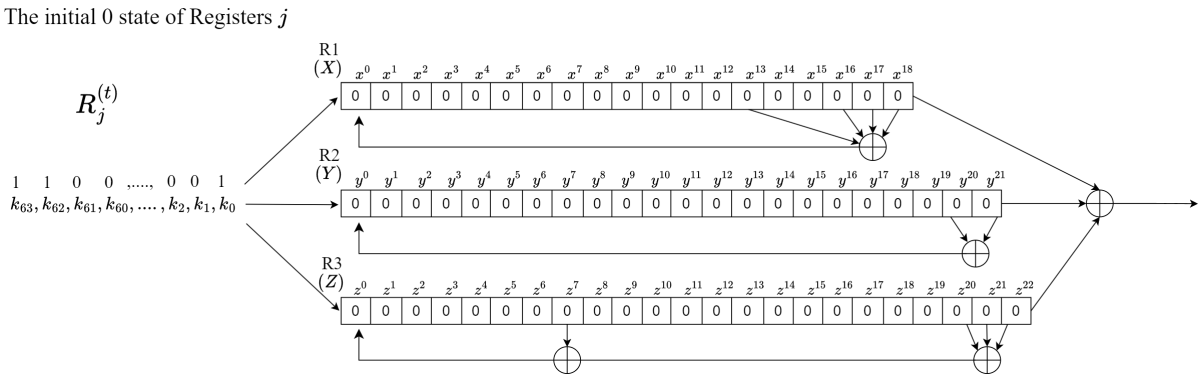
Figure 5 demonstrates the key initialization of the 3 Least Significant Bits of K_c.

K_c = 1100 1111 0010 1010 0110 0011 1000 1001 0111 0101 0011 1111 0110 1011 1101 0001
F_n = 0010 1101 1001 0111 1001 00

The initial state of Registers j are set to 0.

$$R_j^{(t)} = 0 \quad \forall j \in \{X, Y, Z\}$$

Where the increments $R_j^{(t+1)}, R_j^{(t+2)}$ denoting the discrete time steps at which each entered key bit is processed and shifts the registers.



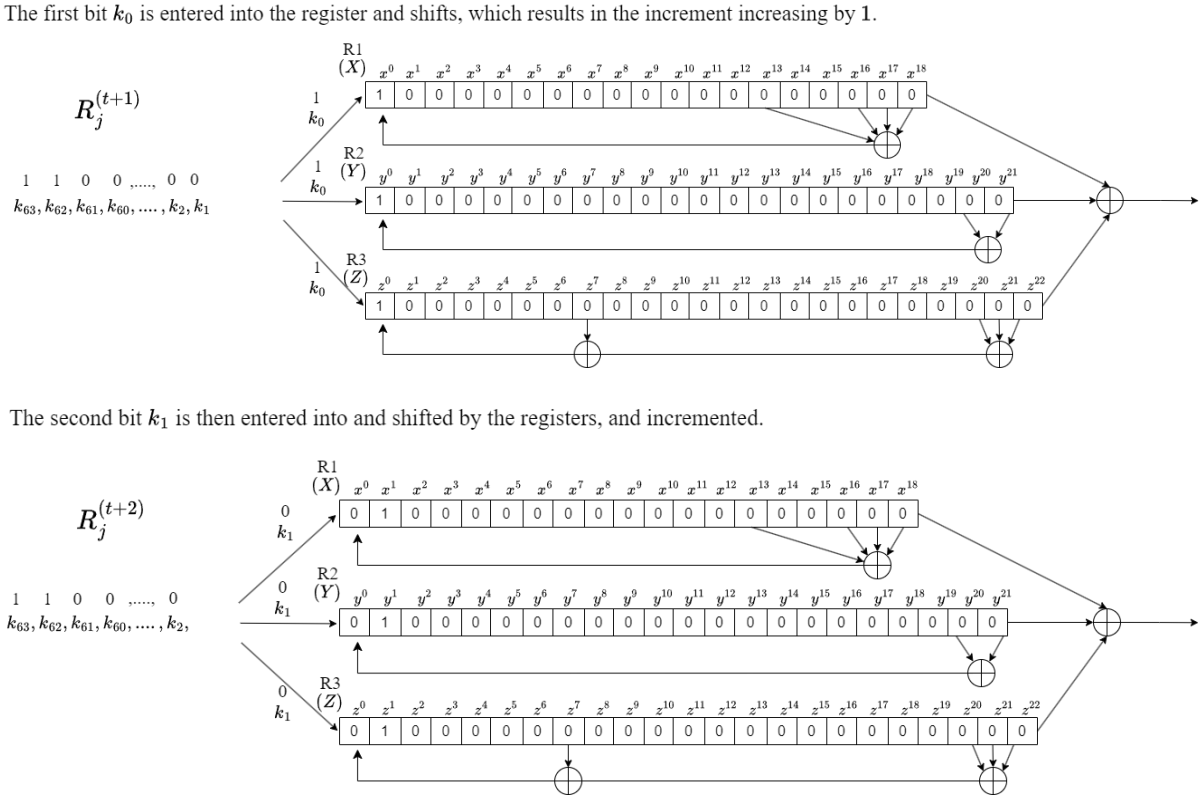


Figure 5: Diagram of A5/1 keyloading process

The bit shifting process continues through each of the cell of the register until a key bit , such as k_0 traverses the entire length of the register. For clarity only R_1 is described here as an example. Within R_1, the bits at x^{13}, x^{16}, x^{17}, x^{18} in R_1 are XORed to produce the feedback bit.

Afterwards, all bits in the register are shifted one position towards the MSB, creating empty space at the x^0 position. The calculated feedback bit occupies the position x^0 provisionally before the register updates finalize its state. Subsequently, the key loading phase continues by XORing the next key bit denoted in the below figure as k_{20}" with the feedback bit currently at x^0.

As R_1 has the shortest length at 19-bits, it serves as the primary example in these illustrations. The processes for the other two registers R_2, R_3, follow an anlogous pattern with the principal distinctions being their respected lengths and specific tap configurations. This iterative process is executed until all 64-bits of the ciphering key and all 22-bits of the frame number have been integrated during the initalization phase.

Figure 6 continues the inital keyloading process from Figure 5 to demonstrate the feedback function, with the first 20 least significant bits of the 64-bit K_c

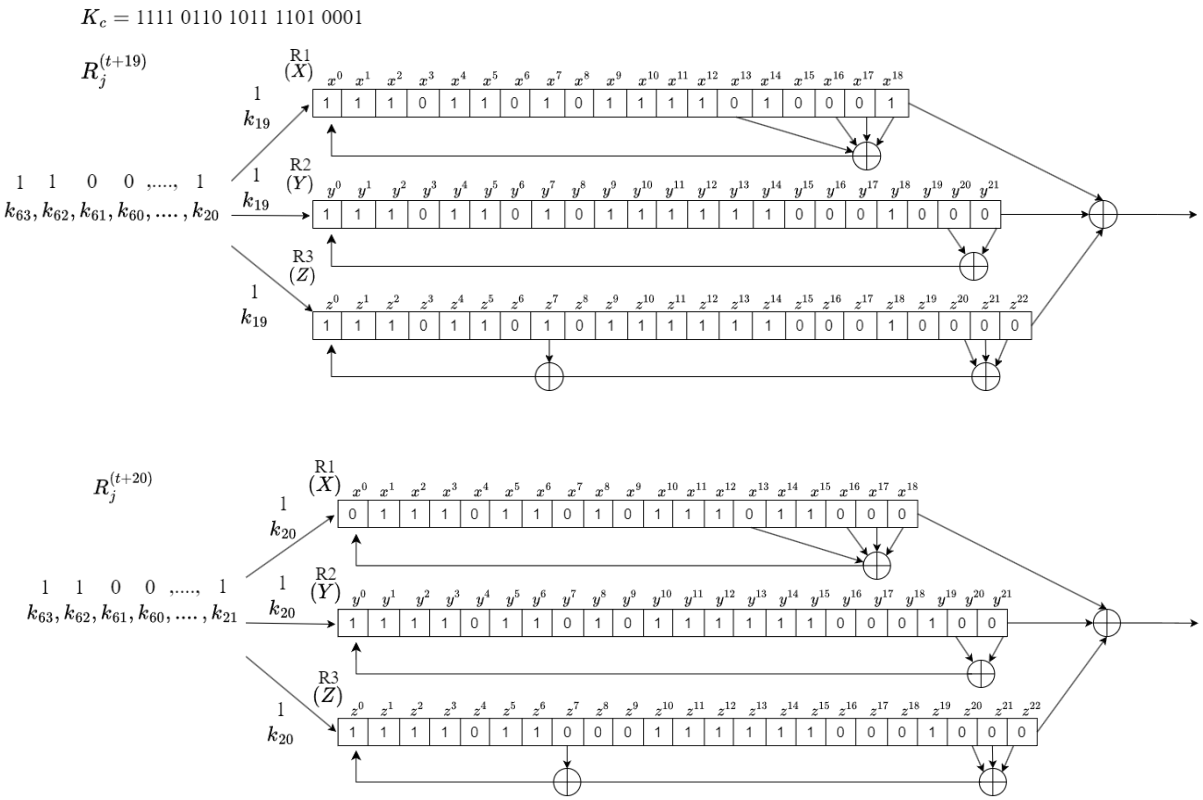


Figure 6: Feedback operation during keyloading

To further describe the feedback mechanism employed during the keyloading phase, it can be mathematically defined as followed:

1. Where the feedback bit is computed as

$$Feedback = x^{13} \oplus x^{16} \oplus x^{17} \oplus x^{18}$$

2. The computed feedback bit is XORed with k_{20}:

$$x^{'0} = k_{20} \oplus Feedback$$

3. Here $x^{'0}$ denotes the temporary state of x^0 before the state is updated.

4. The state is then updated with a shift operation, and the new state of the 0th bit is updated:

$$x^0 = x^{'0}$$

Example with R_1 the tap bits are $x^{18} = 0$, $x^{17} = 0$, $x^{16} = 0$, $x^{13} = 1$, and the key bit $k_{20} = 1$, the feedback calculation would be as follows:

$$Feedback = 0 \oplus 0 \oplus 0 \oplus 1 = 1$$
$$x^{'0} = 1 \oplus 1 = 0$$
$$x^0 = 0$$

The initialization phase iterates the feedback process, systematically integrating each of the 64-bits of the secret key into the registers. This mixing process is essential for enhancing security, as it continuously executes the feedback loop until all key bits are fed into the registers. Additionally, the 22-bit frame number is processed in a similar manner, completing the initialization. The cipher then transitions to the clocking phase.

5.3 A5/1: Clocking Phase

Majority Bit Determination
The A5/1's architecture utilizes a clocking mechanism applied to each register, known as the majority function. At the start of each cycle, the majority function evaluates specific clocking bits in each register, with x^8 in R_1, y^{10} in R_3 and z^{10} in R_2. The value (0 or 1) that appears in the majority of these clocking bits determines the majority value for that cycle.

Clocking the Majority Value
The clocked bits of the registers who match the majority value are clocked during that cycle. The implementation of the majority function applies an irregular clocking mechanism with the intention of increasing resiliency to cryptanalysis and additionally to add another layer of complexity to the generation of the keystream.

Majority Rule Example
Figure 7 below demonstrates the function of two majority values with the clocked bits. If $x^8 = 1$, $y^{10} = 1$, and $z^{10} = 0$, then the majority value for the first iteration is 1. This means R_1 and R_2 have clocked bits that match the majority value. However, R_3 does not match the majority and therefore is not clocked in the current cycle.

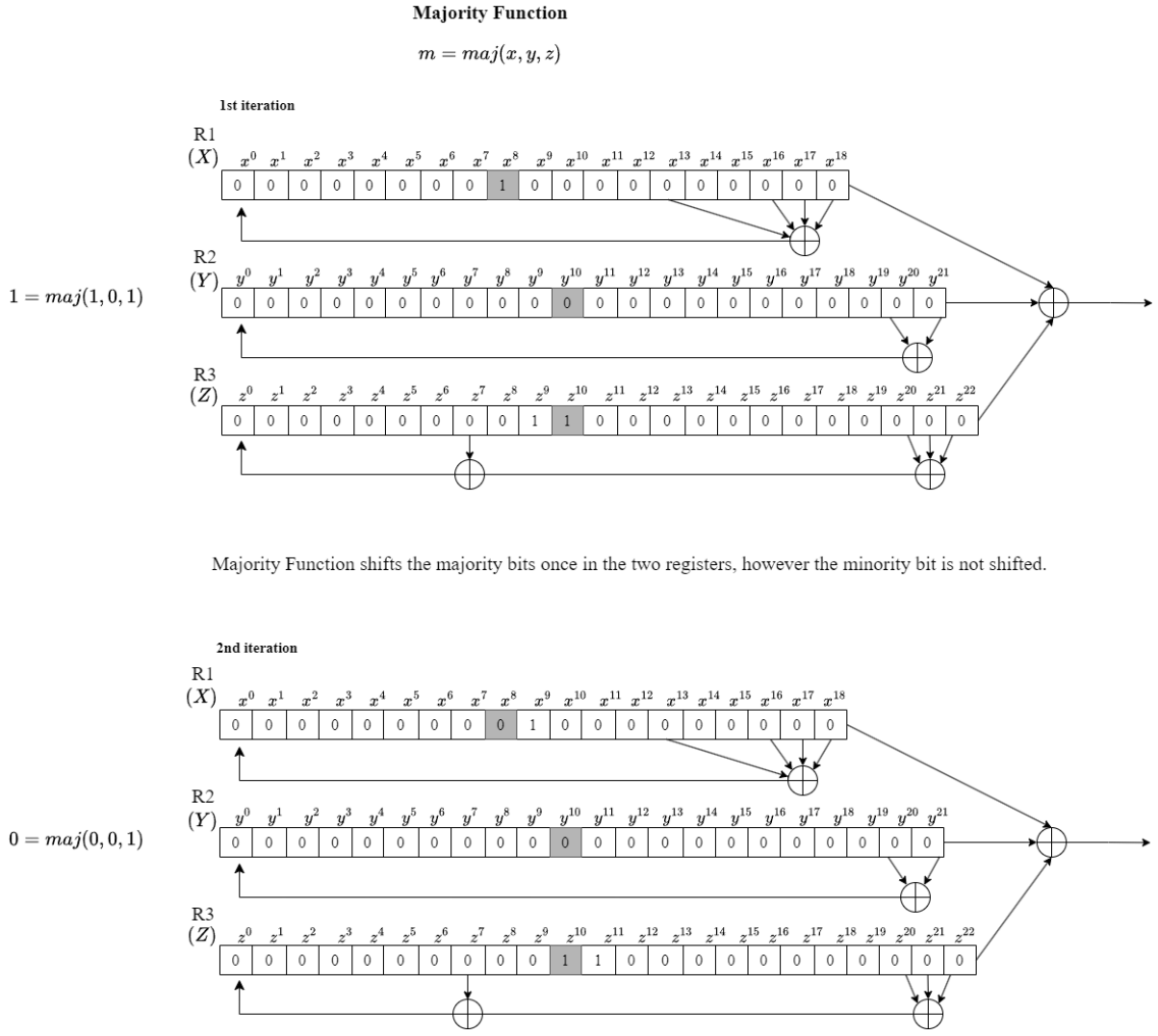


Figure 7: Majority Function

Repetition of the Majority Function
The majority rule undergoes a fixed number of cycles; specifically, the majority function is executed through 100 cycles to further diffuse any identifiable patterns after the initial key and frame number

seeding. This iterative process enhances the cipher's complexity, increasing entropy and unpredictability before the keystream generation commences.

6 Keystream Generation and Encryption

Keystream generation is critical for producing the pseudorandom sequences necessary for encryption. After initialization, the registers are clocked according to the majority rule. The keystream bit (S_i) is derived by XORing the outputs from the MSB of R_1, R_2, R_3 at positions x^{18}, y^{21}, z^{22} respectively:

$$S_i = x^{18} \oplus y^{21} \oplus z^{22}.$$

This process generates one keystream bit per cycle until the complete 228-bit keystream required for a GSM frame is obtained. Encryption involves XORing each plaintext bit (P_i) with the corresponding keystream bit (S_i) to produce the ciphertext bit (C_i). Decryption is the inverse, XORing the ciphertext bit (C_i) with the same keystream bit (S_i) to recover the plaintext bit:

$$C_i = P_i \oplus S_i$$

$$P_i = C_i \oplus S_i$$

7 Cryptanalysis and Security Weaknesses of the A5/1 Cipher

The A5/1 cipher faced a variety of cryptanalytic attacks that highlighted its vulnerabilities. Its LFSRs were susceptible to time-memory trade-off attacks. Adversaries could generate a precomputed table—or rainbow table—that cataloged the most probable internal states of the cipher. This table allowed them to determine the cipher's current state by comparing its output against the stored values.

The predictability of the initialization phase presented another vulnerability. Although the ciphering key is confidential, the frame number is publicly known. The cipher's security was not fully assessed over time; while a 64-bit key was considered secure against brute-force attacks in the 1990s, advancements in computational power have made such attacks more feasible. Additionally, implementation inefficiencies effectively reduce the key space from 64 bits to an estimated 54 bits.

The escalation in processing power transformed the hypothetical concern regarding the 64-bit key into a severe and tangible threat. Scrutiny of the A5/1 cipher's flaws and cryptanalytic efforts to exploit them have significantly contributed to the advancement of cryptographic security. These challenges underscore the importance of rigorous design, extensive testing, and the ongoing evolution of cryptographic standards to counteract new threats and technological advancements.

8 Conclusion

Exploring the A5/1 stream cipher reveals its significant impact on GSM network security, showcasing both its approach to encryption through LFSRs and its susceptibility to cryptanalytic attacks. Despite its initial success in safeguarding communications, the cipher's vulnerabilities underscore the importance of anticipating a broader range of security challenges. This examination highlights that by not subjecting the cipher to public scrutiny early on, crucial flaws were overlooked, underscoring the necessity of open evaluation in the development of secure cryptographic systems.

References

[1] Win, Myat Su Mon. "A New Approach to Feedback Shift Register." *World Academy of Science, Engineering and Technology*, vol. 48, 2008, pp. 185—189.

[2] Yahia, Muzamil Mahgoub. "Linear Complexity in LFSR-based Stream Ciphers." May 2011.

[3] Schneier, Bruce. *Applied Cryptography, Second Edition: Protocols, Algorithms, and Source Code in C*, 1996.

[4] Golic, J. "Cryptanalysis of Alleged A5 Stream Cipher." In *EUROCRYPT'97, Lecture Notes in Computer Science*, vol. 1233, Springer-Verlag, 1997, pp. 239—255.

[5] Briceno, M., Goldberg, I., Wagner, D. "A Pedagogical Implementation of A5/1." May 1999.

[6] Ekdahl, P., Johansson, T. "Another Attack on A5/1." *IEEE Transactions on Information Theory*, vol. 49, no. 1, Jan. 2003, pp. 284—289.

[7] Babbage, S. "A Space/Time Tradeoff in Exhaustive Search Attacks on Stream Ciphers." *European Convention on Security and Detection, IEEE Conference publication*, No. 408, May 1995, pp. 216—224.

[8] Goresky, Mark. "Fibonacci and Galois Representations of Feedback-with-carry Shift Registers." *IEEE Transactions on Information Theory*, vol. 48, no. 11, Nov. 2002.