Cryptanalysis of GSM Ciphers

Khalid F. Jassim
Software Engineering Dept., AL-Rafidain University College
Baghdad, Iraq
E-mail: e_khalid961@yahoo.com

Abstract:

The stream cipher Algorithm A5 used to encrypt GSM telephone communication. The Algorithm A5 consists of three versions (A5/1, A5/2 and A5/3). In the current standard protocol the three algorithms share the same secret key. Thus in this paper we select the weakest algorithm A5/2, attack the algorithm and extract the secret key in a cipher text-only attack, using the weaknesses of the cipher and majority functions. Then using the secret key to decrypt the communication even encrypted under the algorithms A5/1 and A5/3.

Keywords: Stream Cipher, GSM Cipher, A5 Algorithm, Cryptanalysis.

1. Introduction

The GSM (Group Special Mobile) is a non-American standard for digital cellular mobile telephone. In GSM telephone communication the conversation is protected by using Stream cipher Algorithm A5 [1]. The algorithm consists of three versions:A5/1, A5/2 and A5/3. We noticed that in the current standard protocol all three ciphers share common secret key [2]. Thus to attack these ciphers you can request communication using the weakest alternative A5/2; recover the secret key in a cipher text-only attack, using cryptanalytic weak points Of the cipher and redundancy embedded into communication by error correcting codes. Recovering the secret key the attacker can then decrypt the ciphering communications even encrypted by using the algorithm versions A5/3 and A5/1 [6].Thus this paper is concentrate to attack the algorithm version A5/2.

2. Description of GSM Cipher

A5 is the stream cipher algorithm used to encrypt the link between the telephone and the base station in the GSM system. The structure of the A5/2 algorithm is shown in Fig.1. The algorithm consists of four LFSRS (SR1, SR2, SR3 and SR4), the lengths of the LFSRS are(19,22, 23 and 17) respectively.

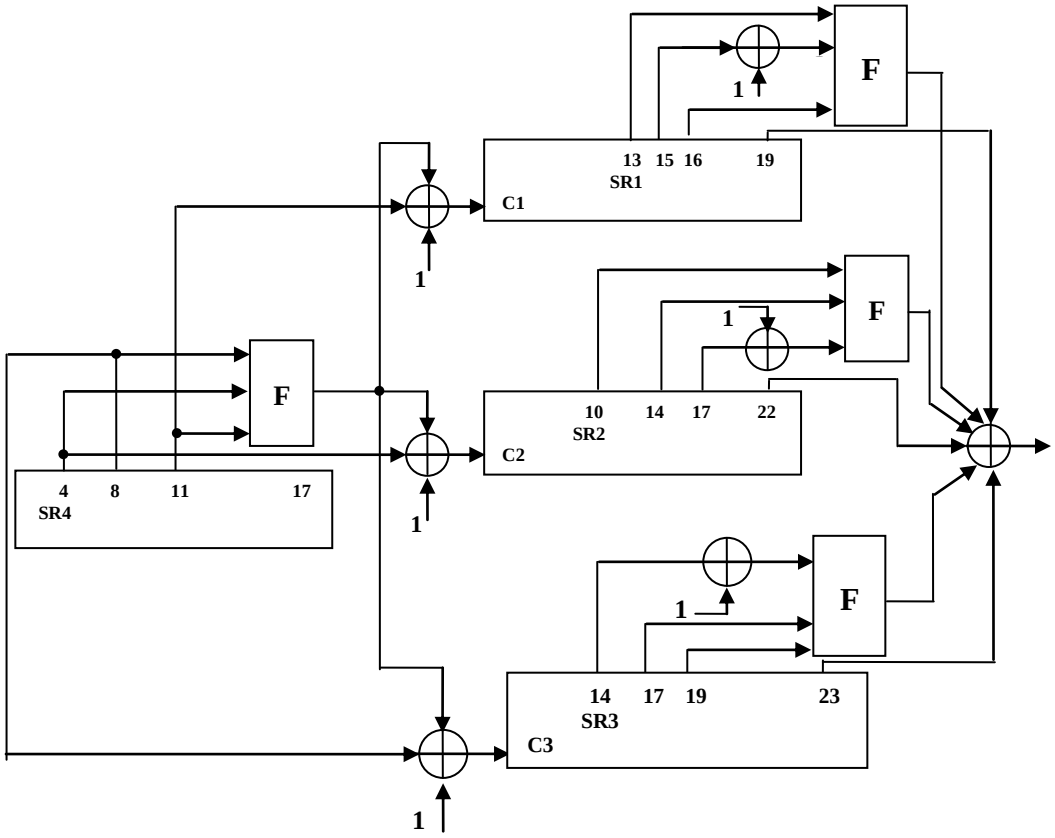

Fig.1 Structure of A5/2 Algorithm.

The feedback connections of SR1 are the positions(14,17,18,19),for SR2 the positions(21,22), for SR3 the positions(8,21,22,23) and for SR4 the positions (12, 17). The function F is the majority function defined as: F(X1, X2, X3) =X1X2+X1X3+X2X3.

X1	X2	X3	F(X1, X2, X3)
0	0	0	0
0	0	1	0
0	1	0	0
0	1	1	1
1	0	0	0
1	0	1	1
1	1	0	1
1	1	1	1

Table.1 Truth table of function F.

From Table.1 it is clear that the correlation between each of the three Inputs and the output of Function F are equal to 0.75. Such correlation is leak of information and can be used in the cryptanalytic attacks.
The output key stream generated as shown in Fig.1,where LFSR SR4 is used to control the movement of SR1, SR2 and SR3 in the stop/go technique.
In GSM system the communication consists of frames. Each frame consists of 228 bits. For every frame to be encrypted the initialization procedure takes place. The initialization depends on 64-bit secret key and 22-bit frame number (Fnum). The bits of secret key are inserted into LFSRS starting from LSB of each key byte and then the bits of frame number are inserted into the LFSRS starting from LSB. The algorithm is run for 100 clock according to stop/go technique without generating any Key stream [1].

3. Linearization

The basis of this technique is to linearize a system of nonlinear algebraic equations by assigning a new unknown variable to each monomial term that appears in the system. For a given value of the state K_t and for a given degree d, we shall let $M_d(t)$ (the monomial state) denote the GF (2) column vector with each component being a

corresponding monomial of degree d or less. The number of such monomials is $D=\sum_{i=0}^{d} C_i^{\,n} \sim C_d^{\,n}$, so $M_d(t)$ contains D components[7]. The initial monomial state M_d corresponds to the initiate state K. In general form the number of r-combinations of n objects is [9] :

$$C_r^{\,n} = n! / r!\,(n-r)!$$

Example1. If n=4 (that is, $K_t=k_3\ k_2\ k_1\ k_0$) and d=2, then there are D=11 monomials of degree <= 2: $D=C_0^{\,4} + C_1^{\,4} + C_2^{\,4} = 11$.

$M_d(t) = (m_0, m1, m2, m3,\ldots,m_{10})^T$

$=(1,k_0,k_1,k_2,k_3,k_0k_1,k_0k_2,k_0k_3,k_1k_2,k_1k_3,k_2k_3)^T$

Where the letter T denotes the transpose of the matrix to make a column vector. If K_t = 0111, then the values of the monomials are:

$M_d(t)=(1,k_0=1,k_1=1,k_2=1,k_3=0,k_0k_1=1,k_0k_2=1,k_0k_3=0,k_1k_2=1,k_1k_3=0,$
$k_2k_3=0\ )^T$

$=(1,1,1,1,0,1,1,0,1,0,0)^T$

4. Gaussian Elimination to Solve a System of Linear Equations

We will use Gaussian elimination to solve the system of linear equations .In Gaussian elimination we attempt to reduce the original system of N linear equations to triangular form. In triangular form the coefficients below the diagonal in the matrix of coefficients are all zero and the diagonal elements are all one. The coefficients above the diagonal and the components of the constant vector B no longer have their original values [8].

Example2. Consider a 3-bit LFSR as in Fig.2.

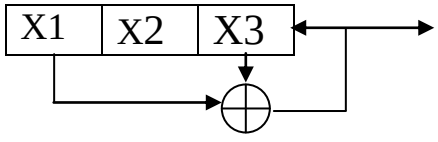

Fig.2 3-bit LFSR.

Suppose you have the following system of linear equations as an output from the above LFSR.

$$
\begin{aligned}
X1 + X3 &= 0 \\
X1 + X2 + X3 &= 1 \\
X1 + X2 &= 0
\end{aligned}
$$

Then AX = B (Where A is a matrix of coefficients ,B is constant vector)

$$
\begin{bmatrix} 101 \\ 111 \\ 110 \end{bmatrix} \begin{bmatrix} X1 \\ X2 \\ X3 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \quad \Diamond \quad \begin{bmatrix} 111 \\ 011 \\ 001 \end{bmatrix} \begin{bmatrix} X1 \\ X2 \\ X3 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}
$$

So we get the following linear equations:

$$
\begin{aligned}
X1 + X2 + X3 &= 1 \qquad \text{eq.1} \\
X2 + X3 &= 0 \qquad \text{eq.2} \\
X3 &= 1 \qquad \text{eq.3}
\end{aligned}
$$

This system of linear equations can be solved for X1,X2,X3 by solving eq.3 for X3 (X3=1),substituting this value in eq.2 (X2=1) and substituting these values in the first equation and solving for X1 (X1=1). This process is called Back Substitution. The algorithm for Gaussian elimination follows.

4.1 Algorithm for Gaussian Elimination

1. Enter the input data into Augmented matrix AUG.
2. Call subprogram GAUSS to traingularize matrix AUG
3. If a solution exists (flag is true) then
   4. Call subprogram BACK to find the solution.
   5. Print out the solution

   ELSE
   6. Print message Solution is not found.

4.2 The Augmented Matrix

To represent a system of N linear equations in N unknowns, the preferred method of representation for such system is the augmented matrix. This particular representation allows for the most concise coding of both traingularization and back substitution. Fig.3 shows the form of an augmented matrix for a system of N linear equations in N unknowns. The augmented matrix has N rows and N+ 1 column. The matrix can be

represented by 2-dimenssional array. The last column of the augmented matrix contains the constant vector B as follows [8]:

$$\begin{pmatrix} A11 & A12 & A13 & \dots & A1N & B1 \\ A21 & A22 & A23 & \dots & A2N & B2 \\ A31 & A32 & A33 & \dots & A3N & B3 \\ \dots & \dots & \dots & & \dots & \dots \\ AN1 & AN2 & AN3 & \dots & ANN & BN \end{pmatrix} \qquad \begin{pmatrix} 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \end{pmatrix}$$

(a)General Form (b) Our example

Fig.3 Original Augmented Matrix

The matrix of coefficients, A, is stored in the rest of the columns of the augmented matrix. When a system of linear equations is represented as an augmented matrix the unknowns are implicit. Our first goal is to triangularize the augmented matrix and reduce it to the following form:

$$\begin{bmatrix} 1 & a'_{12} & a'_{13} & b'_1 \\ 0 & 1 & a'_{23} & b'_2 \\ 0 & 0 & a_{13} & b'_3 \\ & & & \end{bmatrix} \qquad \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

(a) General Form (b) Our Example

Fig.4 Traingularized Augmented Matrix

So we can apply the principle of back substitution to the diagram above and find the final solution.

5. Cryptanalytic Attack

The attack starts by taking an initial state for SR4 from the set of 2^{17} possible states. Then the output sequence of SR4 is used to control the movement of SR1, SR2 and SR3. We try to construct a system of linear equations that relate the state variables of SR1, SR2 and SR3 with the output bits. The nonlinear terms caused by the nonlinear nature of majority functions can be processed by substitution of the nonlinear terms by the new variables. Thus the maximum number of variables in the system will be n=719. Due to the concentrations of the feedback

connections of SR1 and SR2 to the right of the inputs to the majority function, the input positions to this function and the last positions of these LFSRS depend on very few initial variables after every initialization process. The attack on the A5/2 algorithm consists of the following major steps:

0. We use a simulation program to represent algorithm A5/2.
1. We use 5-frames of encrypted data and their corresponding frame numbers. Store them in a file (encrypt-file).
2. Initialize SR4 by the value InistSR4:=0.
3. Set the frame number index by the value Framindx:=0.
4. Set the number of linearly independent equations by the value Eqnum:=0.
5. Select an initial state (InistSR4-th) for SR4 in simulation program.
6. Set framindx:=framindx+1; complete the initialization process, starting from the state InistSR4 of SR4, using the variables X1,X2,X3, … ,X64 to initialize SR1,SR2 and SR3, and adding the frame number (Fnum) into all the LFSRS.
7. If the end of the frame is reached, then go to step6; else Run the simulation program for one cycle, and generate the equation of the output key stream.
8. Linearize the obtained equation, by substituting the nonlinear terms by the new variables: combine this equation with its corresponding encrypted output bit from (encrypt-file); store the final equation in (equations-file). Set Eqnum:=Eqnum+1
9. If (framindx = 5 and the end of the frame is reached) then go to step 10 ; else go to step 7.
10. Check the linear equations in equation-file and exclude the equations with errors and undesired statistical properties. If we have a contradictions in our system of equations then restart the attack from step3.
11. Use Gaussian elimination to solve the system in equation-file.

6. Conclusion

In this paper, we presented steps to attack A5/2 algorithm which is used to encrypt the link between the telephone and the base station in GSM system. Linearization technique used to linearize the system of non linear equations. In Gaussian elimination we attempted to reduce the original system of linear equations to triangular form then solving the system by using Back Substitution. We conclude that the algorithm is not secure for secret and sensitive communication channels but it is suitable for commercial applications.

References:

[1]R. Anderson and M. Roe,"A5, Technical report",1994,htt://jya.com.

[2] E. Barkan, E. Biham, and N. Keller," Instant Cipher text-only Cryptanalysis of GSM encrypted Communication", in advances in Cryptology- CRYPTO 2003.

[3] E. Biham and O. Dunkelman, "Cryptanalysis of the A5/1 GSM Stream Cipher", INDOCRYPT 2000.

[4] A. Biryukov, A. Shamir, and D. Wagner, "Real Time Cryptanalysis of A5/1 on a PC", no. 1978 in Lecture Notes in Computer Science, PP. 1- 18, Springer- Verlage, 2000.

[5] M. Briceno, I. Goldberg, and D. Wagner, "A pedagogical implementation of A5/1", Technical report, 1999.

[6] J.D. Golic, "Cryptanalysis of alleged A5 Stream Cipher", EUROCRYPT 97.

[7] P. Hawkes and Gregory G. Rose, "Rewriting Variables: the Complexity of Fast Algebraic Attacks on Stream Ciphers", Phawkes@qualcomm.com.

[8] Parker D.S., Dinh L., "How to Eliminate Pivoting from Gaussian Elimination by Randomizing Instead", Technical Report, Computer Science Dept., University of California, 1995.

[9] C.L. Liu, "Introduction to Combinatorial Mathematics", McGraw –Hill, 1968.

تحليل اتصالات الهواتف الخلوية GSM المشفرة

خالد فاضل جاسم

قسم هندسة البرامجيات – كلية الرافدين الجامعة

المستخلص :-

ان خوارزمية التشفير الانسيابي A5 تستخدم لتشفير اتصالات الهواتف الخلوية GSM. تتكون خوارزمية A5 من ثلاثة انواع (A5/1 , A5/2 , A5/3). ضمن البروتوكول القياسي الحالي لاتصالات GSM تعتمد الخوارزميات الثلاثة على مفتاح سري موحد. لذلك في هذا البحث تم اختيار خوارزمية A5/2 كونها الاضعف حيث تم تحليل ومهاجمة هذه الخوارزمية باستخدام اسلوب المهاجمة بتوفر النص المشفر فقط. تم استثمار نقاط الضعف التحليلية الموجودة في النص المشفر ودوال الاغلبية والتوصل الى المفتاح السري. يمكن استخدام المفتاح السري في حل الاتصالات المشفرة التي تعتمد على الخوارزميات A5/1 , A5/3.