Algebraic analysis of SM4 cipher using SageMath

Ekaterina Maro
Southern Federal University, Chehova str., 2, Taganrog, Rostov-on-Don region, 347922, Russia

E-mail: camaro@sfedu.ru

Abstract. The article consider possibility of applying the algebraic analysis method to the Chinese national standard (SM4 cipher). Research presents approaches to algebraic analysis of the reliability of symmetric block ciphers by solving Boolean satisfiability (SAT) problem. The principle of formation of the SAT problem for algebraic analysis of SM4 cipher is proposed, as well as the parameters of the obtained SAT problems.

1. Algebraic analysis methods

In the field of information security an important task is robustness assessing of security methods for protection of confidential information. Modern information security systems use cryptographic methods and algorithm's reliability is important argument in ensuring data confidentiality. Reliability assessment of cryptographic algorithms is a complex mathematical problem that requires development of computationally efficient algorithms, complex models and methods for performing evaluation of experimental results. Algebraic analysis methods are being actively developed and improved as applied to cryptographic algorithms (especially lightweight, because of their simplified mathematical structure). The rapid increase in computational power and optimization of algorithms for solving computationally hard problems led to increasing the importance of theoretical and experimental research of the reliability of cryptographic information protection algorithms and tools.

The assessment of information security can be reduced to searching for solutions to systems of Boolean nonlinear equations and analyzing the complexity of various methods for it’s solving for many information security systems. As example, in the form of systems of Boolean nonlinear equations the following information security processes can be represented:

- audit of current state of protected object;
- validation of programs functioning;
- analysis of systems reliability for protecting confidential information (including cryptographic transformations).

One of the areas of research on cipher's reliability is the assessment of the applicability of algebraic analysis methods. Nowadays researches have presented many methods of algebraic analysis (solving systems of Boolean nonlinear equations), as example:

- reduction to linear systems (relinearization) [1],
- extended linearization (eXtended Linearization) [2],
• extended sparse linearization (eXtended Sparse Linearization) [3],
• algorithms F4, F5 [4, 5],
• Raddum-Semaev algorithm [6],
• ElimLin [7],
• SAT solving [8-12].

2. Application of algebraic analysis based on the SAT-solving to symmetric block ciphers
Robustness assessing of symmetric block encryption algorithms using methods of reduction to the SAT problem has the following main stages (figure 1):
1. The choice of method and formation of a Boolean nonlinear equations system to connect the known data (plain text / ciphertext) and the encryption key (round keys). To compose such systems for observed block ciphers, it is proposed to use bit nonlinear cipher transformation - substitution boxes and addition modulo 2^n operation.
2. Formation of a subsystem of Boolean nonlinear equations describing the round encryption keys schedule algorithm (optional; step is applied, if this procedure is used in encryption scheme).
3. Evaluation of the efficiency of the chosen method for solving the generated system of Boolean nonlinear equations. The paper proposes to use the reduction to the SAT problem and apply SAT solver CryptoMiniSat [13].
4. Searching for solving sets (finding the encryption key). The paper proposes to use free open-source mathematics software system SageMath [14].
5. Estimation of the computational complexity of the reliability analysis of encryption algorithms depending on the number of analyzed rounds and the filling of substitution boxes.
6. Estimation of the time complexity of reliability analysis depending on the parameters of the computing system (processor frequency, RAM, number of processors involved in the calculations, etc.).
7. Determination of criteria and principles for the formation of the final assessment of the strength of the cipher to the methods of reduction in the SAT-problem.

The initial data of method are:
• Structure of the investigated encryption algorithm for protecting confidential information.
• Filling of substitution transformation at investigated encryption algorithm.
• Pairs of texts: plaintext and ciphertext.

The output indicators of method are:
• Value of obtained encryption key (or several values).
• Parameters of Boolean nonlinear equations system describing investigated encryption transformation: the number of equations, unknowns and monomials.
• Computational complexity of solving a system of Boolean nonlinear equations.
• Time complexity of solving a system of Boolean nonlinear equations.
• Number of known texts pair that allow to unambiguously determine encryption key in real time.
• Required amount of RAM for carrying out computations.

3. Application of algebraic analysis based on SAT solving to the SM4 cipher
SM4 cipher [15] is a block iterative encryption algorithm used as the Chinese national standard for wireless local networks. SM4 cipher uses 128-bit blocks, a 128-bit secret key and iterative transformations in 32 rounds. The following transformations are used for SM4 cipher: four parallel used 8-bit substitution boxes (figure 2), addition modulo 2, cyclic shifts by 2, 10, 18, 24 positions. One round of encryption of the SM4 algorithm is shown in figure 3.
Form a subset of nonlinear Boolean equations for substitution primitives
Form a subset of nonlinear Boolean equations for round keys generation algorithm
Expand Boolean equations system to n-round encryption scheme

**SAT-solving**

Convert nonlinear Boolean equation system from algebraic normal form into conjunctive normal form

Estimate a number of literals and clauses

Solve CNF by SAT-algorithms

Experimental evaluation of time-complexity and required memory (RAM)

**Figure 1.** Algebraic analysis based on finding a SAT solution.

| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | a | b | c | d | e | f |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| d6 | 90 | e9 | fe | cc | e1 | 3d | b7 | 16 | b6 | 14 | c2 | 28 | fb | 2c | 05 |
| 2b | 67 | 9a | 76 | 2a | be | 04 | C3 | aa | 44 | 13 | 26 | 49 | 86 | 06 | 99 |
| 9c | 42 | 50 | f4 | 91 | ef | 98 | 7a | 33 | 54 | 0b | 43 | ed | cf | ac | 62 |
| e4 | b3 | 1c | a9 | e9 | 08 | e8 | 95 | 80 | df | 94 | fa | 75 | 8f | 3f | a6 |
| 47 | 07 | a7 | fc | f3 | 73 | 17 | ba | 83 | 59 | 3c | 19 | e6 | 85 | 4f | a8 |
| 68 | 6b | 81 | b2 | 71 | 64 | da | 8b | f8 | eb | 0f | 4b | 70 | 56 | 9d | 35 |
| 1e | 24 | 0e | 5e | 63 | 58 | d1 | a2 | 25 | 22 | 7c | 3b | 01 | 21 | 78 | 87 |
| d4 | 00 | 46 | 57 | 9f | d3 | 27 | 52 | 4c | 36 | 02 | e7 | a0 | c4 | c8 | 9e |
| ae | bf | 8a | d2 | 40 | c7 | 38 | b5 | a3 | f7 | f2 | ce | f9 | 61 | 15 | a1 |
| 9e | 0e | ae | 5d | a4 | 9b | 34 | 1a | 55 | ad | 93 | 32 | 30 | f5 | 8c | b1 | e3 |
| 1d | f6 | e2 | 2e | 82 | 66 | ca | 60 | c0 | 29 | 23 | ab | 0d | 53 | 4e | 6f |
| b4 | 6b | 37 | 45 | de | fd | 8e | 2f | 03 | ff | 6a | 72 | 6d | 6c | 5b | 51 |
| c8 | d1 | b | af | 92 | bb | dd | bc | 7f | 11 | d9 | 5e | 41 | 1f | 10 | 5a | d8 |
| d0 | a | c1 | 31 | 88 | a5 | cd | 7b | bd | 2d | 74 | d0 | 12 | b8 | e5 | b4 | b0 |
| e8 | 99 | 69 | 97 | 4a | 0c | 96 | 77 | 7e | 65 | b9 | f1 | 09 | e5 | 6e | c6 | 84 |
| f | 18 | f0 | 7d | ec | 3a | dc | 4d | 20 | 79 | ee | 5f | 3e | d7 | eb | 39 | 48 |

**Figure 2.** SM4 substitution box.
Figure 3. SM4 encryption algorithm scheme.

For modeling the algebraic analysis of the SM4 cipher it should be noted that the main transformation used to form the system of equations is a nonlinear substitution in S-boxes. Assume an $s \times s$-bit S-box. Denote the number of all compositions of input and output block bits as $t$. Hence, at least $t-2^s$ linearly independent equations exist that hold with the probability of 1.

For instance, a $s \times s$-bit S-box Boolean nonlinear equations are defined as follows (1):

$$
\sum_{i,j,k=0}^{s-1} \alpha x_i y_j y_k + \sum_{i,j,k=0}^{s-1} \beta x_i x_j y_k + \sum_{i,j=0}^{s-1} \gamma x_i y_j + \sum_{i,j=0}^{s-1} \delta y_i y_j + \sum_{i=0}^{3} \lambda x_i + \sum_{i=0}^{3} \omega y_i + \eta = 0
$$

where $\alpha, \beta, \gamma, \delta, \lambda, \omega, \eta$ are binary coefficients;

$x_i$ and $y_i$ are correspondingly $i$-th bit of input and output vector of the S-box;

$s$ is length of input and output vector.

The number of possible monomials with the degree of 3 or less is calculated according to formula (2):

$$
t = \binom{2s}{3} + \binom{2s}{2} + 2s + 1
$$

i.e. for the SM4 cipher $8 \times 8$-bit S-box we obtain $t$ number of monomials:

$$
\eta = \binom{16}{3} + \binom{16}{2} + 16 + 1 = 697.
$$

Therefore, it is possible to find $r \geq t-2^s = 441$ linearly independent cubic equations. System of Boolean equations for SM4 substitution box includes following equations (figure 4).

SM4 cipher S-box transformation is presented as 441 linearly independent equations with the number of monomials not exceeding 697. One round of SM4 encryption algorithm (taking into account that the only nonlinear operation is substitution in S-boxes and other transformations lead to permutation of bits) will be specified as a system of 1764 cubic linearly independent equations with 64 unknowns (round key $r_k$ and S-boxes output vector). For the full-round version of SM4 (32 rounds) were formed 56448 cubic equations with 2048 unknowns (32 $r_k$ and 32 S-boxes output vector).

After the generation of the system of Boolean nonlinear equations, we substitute input and output vectors of the S-box through known texts pairs (plaintexts and ciphertexts) using knowledge of encryption algorithm structure (known data used for replacement are presented in figure 5):
\[ \text{Input } S = P_{t+i_1} \oplus P_{t+i_2} \oplus P_{t+i_3} \oplus rk, \]
\[ \text{Output } S = P_t \oplus Y_{\ll 2} \oplus Y_{\ll 10} \oplus Y_{\ll 18} \oplus Y_{\ll 24} \oplus Ct, \]

where Input S - concatenation of input vectors of 4 S-boxes;
Output S- concatenation of output vectors of 4 S-boxes;
P_t- plaintext;
Ct – ciphertext;
rk – secret round key;
i – the index of the current block, ranging from 0 to 4.

Figure 4. System of Boolean equations for SM4 substitution box.

Figure 5. Structure of dependencies between SM4 cipher plaintext/ciphertext and input/output bits of S-boxes.
At this stage, a system of Boolean equations presented in algebraic normal form (ANF) is obtained. To use existed SAT-solvers we should convert formed system of Boolean equations (in the ANF) to conjunctive normal form (CNF). To represent the resulting system in CNF, we use the anf2cnf conversion library [16-18]. Consider an example of converting an arbitrary system of Boolean equations into CNF. Let a system of equations of the form [18]:

\[ a \cdot b \oplus b \cdot c \oplus b \oplus d = 0, \]
\[ b \cdot c \oplus c \oplus a = 0, \]

where \( a, b, c, d \) - binary unknowns, \( \oplus \) - addition operation modulo two (XOR), \( \cdot \) - bitwise multiplication operation (AND).

Following the algorithm described above, the system of equations is reduced to a linear form by introducing additional unknowns \( z_1 \) and \( z_2 \):

\[ z_1 = a \cdot b, \]
\[ z_2 = b \cdot c, \]
\[ z_1 \oplus z_2 \oplus b \oplus d = 0, \]
\[ z_2 \oplus c \oplus a = 0. \]

We represent each equation of the transformed system in CNF:

1. \( z_1 = a \cdot b \) represents in form:

\[ z_1 \lor \bar{a} \lor b = 1, \]
\[ \bar{z}_1 \lor a = 1, \]
\[ \bar{z}_1 \lor b = 1. \]

2. \( z_2 = b \cdot c \) represents in form:

\[ z_2 \lor \bar{b} \lor \bar{c} = 1, \]
\[ \bar{z}_2 \lor b = 1, \]
\[ \bar{z}_2 \lor c = 1. \]

3. \( z_1 \oplus z_2 \oplus b \oplus d = 0 \) represents in form:

\[ \bar{z}_1 \lor z_2 \lor b \lor d = 1, \]
\[ z_1 \lor \bar{z}_2 \lor b \lor d = 1, \]
\[ z_1 \lor z_2 \lor b \lor \bar{d} = 1, \]
\[ \bar{z}_1 \lor \bar{z}_2 \lor b \lor \bar{d} = 1, \]
\[ \bar{z}_1 \lor z_2 \lor \bar{b} \lor \bar{d} = 1, \]
\[ z_1 \lor z_2 \lor \bar{b} \lor \bar{d} = 1. \]

4. \( z_2 \oplus c \oplus a = 0 \) represents in form:

\[ z_2 \lor c \lor \bar{a} = 1, \]
\[ \bar{z}_2 \lor \bar{c} \lor a = 1, \]
\[ \bar{z}_2 \lor c \lor a = 1, \]
\[ \bar{z}_2 \lor \bar{c} \lor \bar{a} = 1. \]

Then the system of equations in CNF should be solved by SAT-solver algorithm. We chose to apply CryptoMiniSat SAT-solver for solving cryptotask. We made experiment at environment SageMath [14] by built-in CryptoMiniSat SAT-solver.
The total time to solve one-round SM4 cipher Boolean nonlinear system using the SAT-solver CryptoMiniSat in SageMath environment required 39.47 seconds on PC (IntelCore i5 2.8 GHz, 8 GB).

4. Conclusion

The paper presents a description of an assessment method of symmetric block ciphers robustness based on reduction to the SAT problem. An algorithm for representing the SM4 encryption scheme as a system of nonlinear Boolean equations in algebraic normal form is described. The software implementation of presented algorithm of generating nonlinear Boolean equations systems for substitution transformation and addition modulo 2 operation of SM4 symmetric encryption standard is carried out. We describe reduction of SM4 cipher to SAT problem and present experimental results on search timings for one-round SM4 encryption using CryptoMiniSat SAT solver (built-in at SageMath environment). The SAT-solving process are required 39.47 seconds on PC (IntelCore i5 2.8 GHz, 8 GB) for one-round SM4 cipher, which is presented as system of 1764 cubic linearly independent equations with 64 unknowns.

The proposed method for performing algebraic analysis based on the SAT-solving approach can be adapted for the further analysis of various developed and applied cryptographic systems built on the basis of substitution boxes and addition modulo 2 operation.

References

[1] Kipnis A and Shamir A 1999 Cryptanalysis of the HFE public key cryptosystem by relinearization Advances in Cryptology–Crypto'99 1666 19-30
[2] Courtois N, Klimov A, Patarin J and Shamir A 2000 Efficient algorithms for solving over defined systems of multivariate polynomial equations EUROCRYPT 392-407
[3] Courtois N and Pieprzyk J 2002 Cryptanalysis of block ciphers with over defined systems of equations ASIACRYPT 267-87
[4] Faugere J-C 1999 A New Efficient Algorithm for Computing Grobner bases (F4) Journal of Pure and Applied Algebra 139(1-3) 61-88
[5] Faugere J-C 2002 A New Efficient Algorithm for Computing Grobner Bases without Reduction to Zero (F5) ISSAC 75-83
[6] Raddum H and Semaev I 2008 Solving Multiple Right Hand Sides linear equations Journal of Designs, Codes and Cryptography 49(1-3) 147-60
[7] Courtois N and Bard G V 2007 Algebraic Cryptanalysis of the Data Encryption Standard In IMA Int. Conf., Springer 4887 152-69
[8] Bard G, Courtois N and Jefferson C 2007 Efficient Methods for Conversion and Solution of Sparse Systems of Low-Degree Multivariate Polynomials over GF(2) via SAT-Solvers In presented at ECRYPT workshop Tools for Cryptanalysis http://eprint.iacr.org/2007/024.pdf
[9] Bard G V 2009 Algebraic Cryptanalysis (Switzerland: Springer)
[10] Een N and Sorensson N MiniSat 2.0. An open-source SAT solver package http://www.cs.chalmers.se/Cs/Research/FormalMethods/MiniSat/
[11] Semenov A A, Zaikin O S, Bespalov D V and Ushakov A A 2008 SAT-approach in cryptanalysis of some stream encryption systems Computational technologies 13(6) (in Russian)
[12] Semenov A A 2015 Algorithms for solving the problem of Boolean satisfiability (SAT) and their application in cryptanalysis PHDays
[13] Soos M, Nohl K and Castelluccia C 2009 Extending SAT Solvers to Cryptographic Problems Theory and Applications of Satisfiability Testing - SAT 2009 Lecture Notes in Computer Science 5584 ed O Kullmann (Heidelberg: Springer)
[14] SageMath the Sage Mathematics Software System (Version 6.10) The Sage Developers https://www.sagemath.org
[15] Diffie W and G Ledin 2008 SMS4 Encryption Algorithm for Wireless Networks IACR Cryptol ePrint Arch. 329
[16] Albrecht M 2010 Tools for Algebraic Cryptanalysis Proceedings of the ECRYPT Workshop on Tools for Cryptanalysis 13-4
[17] Soos M 2010 anf2cnf script released http://www.msoos.org/2010/09/anf2cnf-script-released/
[18] Soos M 2010 ANF to CNF conversion http://www.msoos.org/2010/07/anf-to-cnf-conversion/