An Encryption Method Based on “Zipper” Stochastic Dynamic Hashing

Guanghui Feng¹, Chunfu Zhang², Yujuan Si³, ⁴*, Liuqi Lang¹

¹School of Computing, Zhuhai College of Jilin University, Zhuhai, China
²School of Public Basic and Applied Statistics, Zhuhai College of Jilin University, Zhuhai, China
³School of Computer Science and Technology, Jilin University, Changchun, China
⁴School of Electronics, Zhuhai College of Jilin University, Zhuhai, China
*yujuan@s163.com

Abstract—This paper studies the theoretical problems and implementation methods of stream ciphers, discusses some deficiencies of conventional stream cipher algorithms, designs methods and rules for random Dynamic Hash Mapping and Bits Scrambling, applies nonlinear feedback shift register (NLFSR) technology, and a new stream cipher method with higher security strength and suitable for encryption of short information is proposed. The purpose is to apply the proposed new method of stream cipher to obtain higher encryption strength and code uniformity, and make some theories and algorithms in the research results have beneficial effects on information security in the era of big data.

1. INTRODUCTION

Stream cipher technology is a symmetric key technology. Because it’s simple to implement, the encryption speed is fast, and errors in ciphertext transmission will not cause diffusion in the plaintext, it has become an important type of cryptosystem. At present, stream cipher technology still maintains its advantages in the field of confidential institutions and mobile communications. The core idea of stream cipher is to generate a pseudo-random sequence with excellent performance, that is, a key stream for encryption and decryption.

The design of the key stream generator is the key to the security of stream ciphers. At present, the well-known stream cipher algorithms are: A5 algorithm in GSM mobile communication system [1], RC4 algorithm in IEEE 802.11 WEP standard [2], SNOW 3G algorithm in LTE system [3], E0 algorithm in Bluetooth technology [4], ZUC algorithm adopted by 3GPP as an international standard and independently developed by Chinese Academy of Sciences [5], etc.

After years of research, many scholars have mastered some methods of attacking and deciphering conventional stream ciphers. Courtois and Meier [6] proposed the application of algebraic attacks to stream cipher algorithms based on linear feedback shift registers at the European Cryptography Annual Conference. "Algebraic attack" is a new research hotspot in the field of stream ciphers. It analyzes the security of cryptosystems from a new perspective and transforms it into the problem of solving overdetermined multivariate equations. With the development of various attack methods [7-9], the design of the traditional stream cipher system has been greatly challenged and impacted.
In order to resist related attacks and algebraic attacks, the stream cipher algorithm using linear feedback shift register (LFSR) needs to use a more complex feed forward function. Nonlinear feedback shift register (NLFSR) is a pseudo-random sequence generator with high security that can effectively resist related attacks and algebraic attacks. Stream generators are widely used. For example, the algorithms Grain, Trivium and MICKEY finally recommended by the eSTREAM project all use nonlinear feedback shift registers as the key stream generator.

2. PROBLEMS OF CONVENTIONAL STREAM CIPHER

The essence of the stream cipher is to transform the plain text into binary bits according to a certain rule to obtain the cipher text. The conventional stream cipher technology is based on the exclusive OR operation of the key stream and the plaintext. Its encryption and decryption methods are single. The security strength of the stream cipher system depends entirely on the security of the key stream. There have been many attacks on key stream generators nowadays. Another problem is that for short key streams, relying solely on XOR operations cannot obtain ciphertext with better randomness and better code distribution uniformity. When the plaintext or key code distribution is biased to all "0" or all "1" in binary (that is, when there are too many bytes like 0X00 or 0XFF), there are situations where encryption fails or is easy to be attacked.

3. PRINCIPLES

The difference between this method and the traditional stream cipher encryption and decryption technology is that it does not depend on the XOR operation of the plain text and the key, but a "zipper-style" hash and position disorder of the plain text in the binary bit that depends on the key stream. This rule overcomes the defect that the traditional stream cipher encryption and decryption methods are single, and the security strength of the stream cipher system depends entirely on the security of the key stream. The existing attack techniques on stream ciphers will not be applicable to this method. In addition, the output sequence of the nonlinear feedback shift register is periodic. For a seed key of length n, the period of the output sequence can reach $2^n-1$ at the longest. We introduced a bit transformation Boolean function to scramble the key stream. As a result, a key stream with a longer period and better randomness is obtained. Therefore, this method does not require the design of an optimal nonlinear feedback shift register to achieve better average performance.

Encryption algorithm idea: In the process of scanning the plain text stream (cipher text stream) and the key stream, the position of each bit of the key stream is used as a parameter, and the value of the plain text is scrambled according to the bit operation that depends on the key stream. And according to the "zipper" hash mapping rule that depends on the key stream, it is mapped in the ciphertext space to obtain the ciphertext. "Zipper" dynamic hashing adopts the idea of n-way merge, randomly merges the binary bits that meet the conditions, thereby disturbing the arrangement order of the binary codes of the plaintext stream after bit conversion. The principle of the encryption and decryption system is shown in Fig. 1.

Figure 1. Schematic diagram of a stream cipher system based on random dynamic hashing and random position
The key stream generator is designed based on the nonlinear feedback shift register and bit transformation Boolean function. The nonlinear feedback shift register (NLFSR) is similar to the linear feedback shift register (LFSR) and is based on gate circuits. From the perspective of algebraic expressions, XOR gates are expressed as binary addition (+), and AND gates are expressed as binary multiplication (×). \( f(x_0, x_1, \ldots, x_{n-1}) \) is an \( n \)-ary Boolean function, \( n \) is a positive integer, then the \( n \)-level nonlinear feedback shift register with \( f(x_0, x_1, \ldots, x_{n-1}) \) as the feedback function is shown in Fig. 2.

![Figure 2. \( n \)-level nonlinear feedback shift register](image)

### 4. APPLICATION EXAMPLE

Dubrova listed three different types of feedback functions with algebraic order of 2 in “A list of maximum period NLFSRs”: (1) \( f(x) = x_0 \oplus x_a \oplus x_b \oplus x_c \times x_d \); (2) \( f(x) = x_0 \oplus x_a \oplus x_b \times x_c \oplus x_d \times x_e \); (3) \( f(x) = x_0 \oplus x_a \oplus x_b \oplus x_c \times x_d \oplus x_e \times x_f \). Where: \( a, b, c, d, e, h \in \{1, 2, \ldots, n-1\} \), “\( \oplus \)” is the “exclusive OR” operation (i.e. modulo 2 addition) and "\( \times \)” is "and" Operation (i.e. modulo 2 multiplication). The nonlinear feedback function used in this article is:

\[
F_1(x_0, x_1, x_2, x_3) = x_0 \oplus x_1 \oplus x_2 \times x_3
\]

When the order of the nonlinear feedback shift register is low and the period of the output sequence is short, the introduction of the bit transformation Boolean function can change the bit value of the key stream, thereby obtaining a key with a longer period and better randomness.

Example:

Assuming that the files all use the extended ASCII code (IBM extended character set) as the encoding method, the seed key \( \text{Seed key} = \{1000\}_2 \) which goes through the nonlinear feedback function \( F_1(x_0, x_1, x_2, x_3) \) shown in Fig. 3 to generate the output of sequence \( S \), which has a period length of \( 2^{4-1}=15 \). The sequence \( S \) is cyclically filled into the key stream space to obtain the sequence \( S' \), and the sequence \( S' \) is processed by the Boolean function \( F_2(j) \) to obtain the key stream \( K \). The plaintext \( M \) is subjected to bit conversion with reference to the key \( K \) to obtain \( M' \), and at the same time, the converted plaintext \( M' \) is subjected to "zipper " dynamic hashing according to the key stream \( K \), and finally the ciphertext \( C \) is obtained.

Seed key: \( \{1000\}_2 \)

\[
\begin{align*}
F_1(x_0, x_1, x_2, x_3) &= x_0 \oplus x_1 \oplus x_2 \times x_3 \\
F_2(j) &= ((C_1 + X_1x_j) \text{ MOD } Y_1 == Z_1) || (C_2 + X_2x_j) \text{ MOD } Y_2 == Z_2) \\
S &= \{1011111010110001\}_2 \\
S' &= "\text{1zb}" = \{101111101, 00110001, 01111010, 01100010\}_2 \\
K &= "\text{09}" = \{00110100, 10111011, 11110110, 11101010\}_2 \\
M &= "\text{aaaa}" = \{01100001, 01100001, 01100001, 01100001\}_2 \\
M' &= "\text{9ht}" = \{10101010, 00100101, 01101000, 01110101\}_2 \\
C &= "\text{0m}" = \{10010100, 11010000, 11011011, 00001100\}_2
\end{align*}
\]
5. ANALYSIS

In order to compare the encryption effect of this method and the traditional stream cipher method in extreme cases, a special plain text containing 256 ’a’ characters was selected to test the method and the traditional stream cipher based on XOR operation under the same key stream Method performance.

The formula for calculating the uniformity $\lambda$ of the symbol distribution in the file is as follows:

$$\lambda = \sum_{i=1}^{r} (\alpha_i - n/r)^2$$  (1)
Where:
- $\alpha_i$ represents the number of times the $i$-th character code appears in the file under a certain coding method;
- $n$ represents the length of a file, that is, the total number of characters in the file;
- $r$ represents the total amount of character encoding in a certain encoding mode, for the extended ASCII code, the value of $r$ is 256;
- $n/r$ represents the expected value of various characters appearing uniformly in all $r$ character codes.
- The smaller the $\lambda$ value, the more evenly the symbols are distributed.

We tested several $n$-bit non-linear shift registers with a seed key length of $n \in \{4, 5, 6, 7\}$. Table I shows the comparison between the ciphertext encoding amount and uniformity generated by this method and the traditional method under the effect of the best feedback function, and Table II shows the comparison of the ciphertext coding amount and uniformity generated by this method and the traditional method under the action of the general feedback function. The test results show that when the seed key length is short and the non-optimal feedback function is used, the number of codes and code distribution in the ciphertext encrypted by this method are more balanced, and the average is better than the traditional method.

### Table I. Comparison between this method and the traditional method under the action of the best feedback function

| Seed key (B) | Optimal NLFSR function $f(x)$ | Number of ciphertext codes | Evenness of ciphertext |
|--------------|-------------------------------|----------------------------|-----------------------|
|              | $\text{Our method}$ | $\text{Traditional method}$ | $\text{Our method}$ | $\text{Traditional method}$ |
| 1000         | 0, 1, 2, (1, 2) | 151 | 15 | 346 | 4114 |
| 1000         | 0, 1, 2, (1, 3) | 144 | 15 | 360 | 4114 |
| 1000         | 0, 1, 2, (2, 3) | 124 | 15 | 464 | 4114 |
| 10000        | 0, 1, 2, (2, 4) | 149 | 31 | 350 | 1864 |
| 10000        | 0, 1, 3, (1, 3) | 150 | 31 | 344 | 1864 |
| 10000        | 0, 1, 3, (1, 4) | 146 | 31 | 344 | 1864 |
| 10000        | 0, 1, 3, (2, 3) | 141 | 31 | 358 | 1864 |
| 100000       | 0, 1, 2, (1, 2) | 137 | 63 | 390 | 788 |
| 100000       | 0, 1, 2, (2, 4) | 135 | 63 | 390 | 788 |
| 100000       | 0, 1, 3, (1, 5) | 146 | 63 | 328 | 788 |
| 100000       | 0, 1, 4, (1, 4) | 142 | 63 | 362 | 788 |
| 100000       | 0, 1, 4, (1, 4) | 147 | 63 | 316 | 788 |
| 1000000      | 0, 1, 2, (2, 6) | 151 | 127 | 294 | 262 |
| 1000000      | 0, 1, 4, (1, 3) | 143 | 127 | 344 | 262 |
| 1000000      | 0, 1, 5, (1, 5) | 141 | 127 | 348 | 262 |
| 1000000      | 0, 1, 5, (3, 5) | 147 | 127 | 370 | 262 |
| 1000000      | 0, 1, 5, (4, 6) | 142 | 127 | 444 | 262 |

### Table II. Comparison of this method with traditional methods under the action of non-optimal feedback function

| Seed key (B) | Non NLFSR function $f(x)$ | Number of ciphertext codes | Evenness of ciphertext |
|--------------|----------------------------|----------------------------|-----------------------|
|              | $\text{Our method}$ | $\text{Traditional method}$ | $\text{Our method}$ | $\text{Traditional method}$ |
| 1000         | 0,1,3,(1,2) | 139 | 5 | 402 | 12852 |
| 1000         | 0,1,3,(1,3) | 28 | 7 | 2596 | 9108 |
| 1000         | 0,1,3,(2,3) | 134 | 5 | 454 | 12852 |
6. CONCLUSIONS
This paper proposes an encryption and decryption algorithm based on zipper random dynamic hashing and bits scrambling. This algorithm is consistent with the conventional stream cipher algorithm in time complexity. The important difference from the conventional method is that the security strength of this algorithm is not completely dependent on the key stream, it does not have strict requirements on the uniformity of the key stream, and its security strength is higher. In addition, the algorithm disturbs the plaintext stream and the key stream, making the algorithm applicable to short messages. The experimental results show that the code uniformity of the ciphertext generated after encryption by this algorithm is much higher than the code uniformity of the ciphertext generated by the traditional stream cipher algorithm when the seed key length is short and the non-optimal feedback function is used.

ACKNOWLEDGMENT
This work was financially supported by the Premier Key-Discipline Enhancement Scheme Supported by Guangdong Government Funds (Grant No. 2016GDYSZDXK036), And Zhuhai Laboratory of Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, And the Key Scientific and Technological Research Project of Jilin Province (Grant No. 20170414017GH), And the Foundation of "Three Levels" Backbone Teachers Team Construction Projects of Zuhai College of Jilin University.

REFERENCES
[1] Erguler I, Anarim E, “A modified stream generator for the GSM encryption algorithms A5/1 and A5/2”, European Signal Processing Conference. IEEE, 2008.
[2] Jindal P, Singh B, “RC4 Encryption-A Literature Survey”, J. Procedia Computer Science, vol.46, pp.697-705, 2015.
[3] BIKOS A N, SKLAVOS N, “LTE/ SAE security issues on 4G wireless networks”, J. IEEE Security & Privacy, vol.11(2), pp.55–62, 2013.
[4] Kitsos P, Sklavos N, Provelegios G and Skodras A, N, “FPGA-based performance analysis of stream ciphers ZUC, Snow3g, Grain V1, Mickey V2, Trivium and E0”, J. Microprocessors and Microsystems, vol.37(2), pp.235 – 245, 2013.
[5] Zhang L, Xia L, Liu Z, Jing J and Ma Y, “Evaluating the optimized implementations of SNOW 3G and ZUC on FPGA,” C. Proceedings of 2012 IEEE 11th International Conference on Trust, Security and Privacy in Computing and Communications. Piscataway: IEEE Computer Society, pp. 436 – 442, 2012.
[6] Courtois N T, Meier W, “Algebraic attacks on stream ciphers with linear feedback”, J. Proc of Eurocrypt, vol. 2636, pp. 345--359, 2003.
[7] Zhao Yuehua, Liu Wenshan, Han Mu, “A guess decision attack against Zuchongzhi algorithm,” J. Journal of Jiangsu University (NATURAL SCIENCE EDITION), vol. 36(5), pp. 578-582, 2015.

[8] Aref, M. R, Rahimi. M, Barmshory. M and Mansouri. M. H , “Dynamic cube attack on Grain-v1,” J. IET Information Security, 2015.

[9] Orumiehchiha, M. A. , Rostami, S., Shakour, E., & Pieprzyk, J, “A differential fault attack on the wg family of stream ciphers”, Journal of Cryptographic Engineering, vol.10(2), pp.189-195, 2020.