ISSN 1000-9825, CODEN RUXUEW
Journal of Software, Vol.19, No.5, May 2008, pp.1256−1264
DOI: 10.3724/SP.J.1001.2008.01256

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel/Fax: +86-10-62562563

对一种新的序列密码结构的密码分析*

黄小莉[1,2+], 武传坤[1]

[1](中国科学院 软件研究所 信息安全国家重点实验室,北京 100190)

[2](中国科学院 研究生院,北京 100049)

Cryptanalysis of a New Stream Cipher Structure

HUANG Xiao-Li[1,2+], WU Chuan-Kun[1]

[1](The State Key Laboratory of Information Security, Institute of Software, The Chinese Academy of Sciences, Beijing 100190, China)

[2](Graduate University, The Chinese Academy of Sciences, Beijing 100049, China)

+ Corresponding author: E-mail: huangxiaoli@is.iscas.ac.cn

Huang XL, Wu CK. Cryptanalysis of a new stream cipher structure. *Journal of Software*, 2008,19(5):1256−1264. http://www.jos.org.cn/1000-9825/19/1256.htm

Abstract: This paper studies the security of a newly proposed stream cipher structure based on linear feedback shift register, nonlinear feedback shift register and filter Boolean functions. A distinguishing attack on this structure is presented. An example is illustrated to show the effectiveness of this attack. This new attack suggests that this new stream cipher structure has potential security weakness.

Key words: cryptanalysis; stream cipher; distinguishing attack; hypothesis testing; eSTREAM

摘 要: 对新提议的一种基于线性反馈移位寄存器、非线性反馈移位寄存器和过滤布尔函数的序列密码结构的安全性进行了研究,对这种结构给出了一种区分攻击.举例子说明了此攻击的有效性.这种新的攻击表明,此种新的序列密码结构存在潜在的安全弱点.

关键词: 密码分析;序列密码;区分攻击;假设检验;eSTREAM

中图法分类号: TP309 **文献标识码**: A

1 Introduction

Stream ciphers are important encryption algorithms in symmetric encryption system. There are two kinds of stream ciphers: synchronous stream ciphers and self-synchronous stream ciphers. Many stream ciphers are used widely in real life. The security of stream ciphers has been paid attention to for many years because of their importance. Therefore the design and analysis of stream ciphers have been hot research topics. Many old stream ciphers are considered insecure because they are broken or nearly broken. Therefore efforts have been made in recent years to find secure and fast stream ciphers.

* Supported by the National Natural Science Foundation of China under Grant No.60673068 (国家自然科学基金); the National Basic Research Program of China under Grant No.2004CB318004 (国家重点基础研究发展计划(973))

Received 2006-10-26; Accepted 2007-03-28

In 1999, the European Commission developed a project NESSIE[1]. The main purpose of this project is to call for good stream cipher primitives that are obtained after an open call and evaluation. But after several rounds of evaluation no secure enough stream ciphers were selected.

Many designs of old stream ciphers are based on linear feedback shift registers (LFSR) and Boolean functions. LFSRs have many advantages, so they are used widely in different circuits. For many years LFSRs have been one of the most important components for constructing keystream generators. But LFSRs have a big disadvantage, that is, the internal states are linearly dependent. This property can be exploited successfully by algebraic attack. In 2002, Courtois N. successfully broke the stream cipher Toyocrypt using algebraic attack for the first time[2]. Later, he successfully broke stream ciphers based on LFSRs and Boolean functions also using algebraic attack[3,4]. All these facts made the security of old stream ciphers suspected.

In 2004, ECRYPT—the European Network of Excellence for Cryptology, started a new stream cipher project called eSTREAM[5]. The objective of this project is to call for secure and fast stream cipher algorithms. There were 35 candidate stream ciphers submitted in 2005. Many candidate stream ciphers of the project avoid using old designs but adopt new design. Some designs use LFSRs, nonlinear feedback shift registers (NLFSR) and Boolean functions such as Grain[6]. It is necessary to analyze the security of these candidate stream cipher structures.

Distinguishing attack is a method often used in analyzing the security of stream ciphers such as Refs.[7–9]. The purpose of distinguishing attack is to try to distinguish the observed keystream from a truly random sequence. Distinguishing attack is often weaker than key recovery attack. In some cases distinguishing attack can be turned into key recovery attack. Generally, there are two manners to achieve distinguishing attack: the first one is distinguishing probability distribution of keystream output from a uniform distribution directly; the second one is using linear approximation equations with noticeable biases and parity checks.

To the authors' knowledge, there are a few analytical results available about the security of stream cipher structure based on LFSRs, NLFSRs and filter Boolean functions. Maximov A. analyzed the security of this structure[10]. Berbain C. *et al*. cryptanalyzed the stream cipher Grain[11]. In this paper, the security of this new stream cipher structure will be analyzed. A distinguishing attack on this structure based on the idea similar to that of Ref.[8] is given, however, we use a new method different from Ref.[8]. The method in Ref.[8] can not be used to the structure considered in this paper. In Ref.[8], the entries of the parity check equation are considered as a vector. In this paper, the input variables of the filter Boolean function are considered as a vector. This vector is written as the bit-wise XOR of two vectors. Then all these vectors which correspond to the parity check equations are written as the bit-wise XOR of two vectors. Using the parity check equations we can deduce that the distribution of all these vectors is nonuniform. This is the key to achieve our distinguishing attack. Our attack implies that the newly proposed stream cipher structure is vulnerable to distinguishing attack.

This paper is organized as follows. In Section 2, some necessary preparations for our attack are given. In Section 3, our distinguishing attack is described in detail. In Section 4, an example is demonstrated to show the validity of our method, and Section 5 gives our conclusion.

2 Preliminaries

2.1 Description of structure

In this paper we consider binary additive synchronous stream ciphers based on LFSR, NLFSR and filter Boolean function or some other structures that can be reduced to this structure. The structure is shown in Fig.1.

Let the length of LFSR be n, the feedback polynomial of LFSR be $f(x)$ and the internal state of LFSR be $s_t,s_{t+1},\ldots,s_{t+n-1}$ at time t, where $t\geq 0$. $s_0,s_1,\ldots,s_{n-1}$ denote the initial state. Let the length of NLFSR be m, the feedback

polynomial of NLFSR be $g(x)$ and the internal state of NLFSR be $b_t,b_{t+1},\ldots,b_{t+m-1}$ at time t, where $b_0,b_1,\ldots,b_{m-1}$ are the initial state. $h(x)$ denotes the filter Boolean function. Its input is some internal state bits of LFSR and NLFSR. z_t is the keystream output at time t. It is the output of the filter Boolean function of some internal state bits at time t. In Fig.1, the broken line denotes optional data flow, that is, the internal state of NLFSR may or may not be affected by the internal state of LFSR.

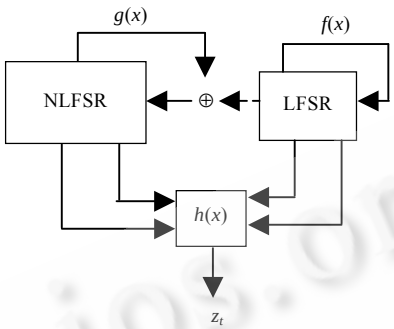

Fig.1 Structure based on LFSR, NLFSR and filter Boolean function

2.2 Hypothesis testing

There are two hypotheses: H_0 denotes the hypothesis that the obtained data comes from the considered cipher and H_1 denotes that the data comes from a random source. Let $X_1,X_2,\ldots,X_k$ be k independent and identically distributed (i.i.d.) random variables from the alphabet X. The distribution of random variable X_i is denoted by $D(x_i)=Pr(X_i=x_i)$, where Pr denotes the probability, x_i denotes the observed value, $1\le i\le k$. The distribution of $(X_1,\ldots,X_k)$ is denoted by $D(x_1,\ldots,x_k)=Pr(X_1=x_1,\ldots,X_k=x_k)$. We use D_0 to denote the distribution under the hypothesis H_0 and D_1 under H_1. Then we can denote the two hypotheses by H_0:$D=D_0$ and H_1:$D=D_1$.

In a hypothesis test, there are two important things that have to be considered. One is how to perform the test optimally. The other is how many samples are needed to be observed. The well-known Neyman-Pearson lemma gives the answer of how to accomplish the optimal test.

Lemma 1. Let $X_1,X_2,\ldots,X_k$ be drawn i.i.d. according to mass function D. Consider the decision problem corresponding to the hypotheses H_0:$D=D_0$ versus H_1:$D=D_1$. Let $C=\left\{(x_1,\ldots,x_k)\middle|\frac{D_0(x_1,\ldots,x_k)}{D_1(x_1,\ldots,x_k)}>c\right\}$, where c is a fixed nonnegative number. Let $\alpha=D_0(\overline{C})$ and $\beta=D_1(C)$ be the error probabilities corresponding to C, where $\overline{C}$ is the complement of C. Let α' and β' be any other set with associated error probabilities. If $\alpha'\le\alpha$, then $\beta'\ge\beta$. This says the critical region C is optimal.

For more detailed content about hypotheses testing, refer to Ref.[12]. Set $\alpha=\beta$ in our attack, so $c=1$[8]. Then the inequality in the expression of C as described above can be rewritten as

$$A=\sum_{i=1}^{k}\left(\log_2\frac{D_0(x_i)}{D_1(x_i)}\right)>0\ .$$

In order to compute the needed sample number, the notion of statistical distance is introduced in Ref.[13].

Definition 1. The statistical distance between two distributions D_0 and D_1 defined over the finite alphabet X is defined as

$$\varepsilon=\sum_{x\in X}|D_0(x)-D_1(x)|\ .$$

If the distributions are smooth, then the sample number N that we need to observe satisfies $N\approx\varepsilon^{-2}$[13].

3 Our Attack

The first step of our attack is to find low-weight multiples of the linear feedback polynomial. The *weight* of a binary polynomial is defined as the number of its nonzero coefficients. These multiples generate the same linear output as that of the original linear feedback polynomial. The goal is to obtain a number of polynomial multiples of low-weight and of as small degree as possible. This method has been used frequently in cryptanalysis. Let ⊕ denote the operation of binary bit-wise XOR and ⊕ denote the sum of bit-wise XOR in the following section.

3.1 Finding low-weight polynomial multiples

There are some literatures on finding low-weight multiples of a polynomial. In Ref.[14] an algorithm of finding w-weight multiples of the polynomial $f(x)$ of degree d and its time complexity are given. We adopt the method in Ref.[14]. It is stated that the critical degree when the multiples of weight w begin to appear is $(w-1)!^{1/(w-1)}2^{d/(w-1)}$. The polynomial residue algorithm of finding polynomial multiples in Ref.[14] is as follows.

Input: A binary polynomial $f(x)$, maximum degree n of polynomial multiples and a positive integer k;

(1) Compute and store the residues $x^i \bmod f(x)$, $1\le i\le n$.

(2) Compute and store all the residues $x^{i_1}\oplus\ldots\oplus x^{i_k} \bmod f(x)$ for all $\binom{n}{k}$ combinations, $1\le i_1\le\ldots\le i_k\le n$.

(3) Find all the 0 and 1 matches of all the residues from step (2) by fast sorting.

Output: All the polynomial multiples of $f(x)$ of degree at most n and of weight at most $2k+1$.

The time complexity of the algorithm is about $O(T\log_2 T)$ with $T=n^k/k!$. In order to obtain $N_{w,n}$ polynomial multiples of weight w, the expected n satisfies $n=((w-1)!N_{w,n})^{1/(w-1)}2^{d/(w-1)}$. So the corresponding T is

$$T=\begin{cases}\dfrac{((2k)!N_{w,n})^{1/2}}{k!}2^{d/2}, & w=2k+1\\ \dfrac{((2k-1)!N_{w,n})^{k/(2k-1)}}{k!}2^{dk/(2k-1)}, & w=2k\end{cases}.$$

The second step of our attack is to make a distinguishing attack. In the second step two cases are considered.

3.2 The case that uses one low-weight polynomial multiple

Suppose that a polynomial multiple of weight w of the feedback polynomial $f(x)$ has been found. This multiple corresponds to a parity check equation of weight w

$$s_t\oplus s_{t+c_1}\oplus\ldots\oplus s_{t+c_{w-1}}=0 \tag{1}$$

Assume that the input variables of the filter Boolean function $h(x)$ at time t are the internal state bits s_{t+a_1}, $s_{t+a_2},\ldots,s_{t+a_u}$, b_{t+b_1}, $b_{t+b_2},\ldots,b_{t+b_v}$, where a_i,b_j are the positions of the LFSR and NLFSR, $1\le u\le n$, $1\le v\le m$, $t\ge 0$. Then the keystream output at time t can be denoted as $z_t=h(s_{t+a_1},s_{t+a_2},\ldots,s_{t+a_u},b_{t+b_1},b_{t+b_2},\ldots,b_{t+b_v})$.

These input variables are written as vectors $(s_{t+a_1},s_{t+a_2},\ldots,s_{t+a_u},b_{t+b_1},b_{t+b_2},\ldots,b_{t+b_v})$. Let

$$\boldsymbol{S}_t=\left(s_{t+a_1},s_{t+a_2},\ldots,s_{t+a_u},\underbrace{0,0,\ldots,0}_{v}\right) \tag{2}$$

$$\boldsymbol{B}_t=\left(\underbrace{0,0,\ldots,0}_{u},b_{t+b_1},b_{t+b_2},\ldots,b_{t+b_v}\right)$$

$$\boldsymbol{Z}_t=(z_t,z_{t+c_1},\ldots,z_{t+c_{w-1}}) \tag{3}$$

Then $(s_{t+a_1},s_{t+a_2},\ldots,s_{t+a_u},b_{t+b_1},b_{t+b_2},\ldots,b_{t+b_v})=\boldsymbol{S}_t\oplus\boldsymbol{B}_t$. Using the above notations, Eq.(3) can be denoted as

$$\boldsymbol{Z}_t=(h(\boldsymbol{S}_t\oplus\boldsymbol{B}_t),h(\boldsymbol{S}_{t+c_1}\oplus\boldsymbol{B}_{t+c_1}),\ldots,h(\boldsymbol{S}_{t+c_{w-1}}\oplus\boldsymbol{B}_{t+c_{w-1}})) \tag{4}$$

From the parity check Eq.(1), we can obtain $s_{t+c_{w-1}}=\bigoplus_{i=0}^{w-2} s_{t+c_i}$, where c_0=0. Then through Eq.(2) the following equations are satisfied

$$\boldsymbol{S}_{t+c_{w-1}}=\left(s_{t+c_{w-1}+a_1},s_{t+c_{w-1}+a_2},\ldots,s_{t+c_{w-1}+a_u},\underbrace{0,0,\ldots,0}_{v}\right)=\left(\bigoplus_{i=0}^{w-2}s_{t+c_i+a_1},\ldots,s_{t+c_i+a_u},\underbrace{0,0,\ldots,0}_{v}\right)$$
$$=\bigoplus_{i=0}^{w-2}\left(s_{t+c_i+a_1},\ldots,s_{t+c_i+a_u},\underbrace{0,0,\ldots,0}_{v}\right)=\bigoplus_{i=0}^{w-2}\boldsymbol{S}_{t+c_i}.$$

So Eq.(4) can be written as

$$\boldsymbol{Z}_t=(h(\boldsymbol{S}_t\oplus\boldsymbol{B}_t),h(\boldsymbol{S}_{t+c_1}\oplus\boldsymbol{B}_{t+c_1}),\ldots,h(\bigoplus_{i=0}^{w-2}\boldsymbol{S}_{t+c_1}\oplus\boldsymbol{B}_{t+c_{w-1}})) \quad (5)$$

Now we show that the distribution of $\boldsymbol{Z}_t$ is nonuniform in most cases. From Eq.(5) it is obvious that the vector $(\boldsymbol{S}_t\oplus\boldsymbol{B}_t, \boldsymbol{S}_{t+c_1}\oplus\boldsymbol{B}_{t+c_1},\ldots,\boldsymbol{S}_{t+c_{w-1}}\oplus\boldsymbol{B}_{t+c_{w-1}})$ is determined completely by the vectors $\boldsymbol{S}_t, \boldsymbol{S}_{t+c_1},\ldots,\boldsymbol{S}_{t+c_{w-2}},\boldsymbol{B}_t,\boldsymbol{B}_{t+c_1},\ldots,\boldsymbol{B}_{t+c_{w-1}}$. There are $2^{(u+v)w-u}$ possible values of $\boldsymbol{S}_t, \boldsymbol{S}_{t+c_1},\ldots,\boldsymbol{S}_{t+c_{w-2}},\boldsymbol{B}_t,\boldsymbol{B}_{t+c_1},\ldots,\boldsymbol{B}_{t+c_{w-1}}$ and $2^{(u+v)w}$ possible values of $(\boldsymbol{S}_t\oplus\boldsymbol{B}_t, \boldsymbol{S}_{t+c_1}\oplus\boldsymbol{B}_{t+c_1},\ldots,\boldsymbol{S}_{t+c_{w-1}}\oplus\boldsymbol{B}_{t+c_{w-1}})$. This means that when we go through all the possible values of $\boldsymbol{S}_t, \boldsymbol{S}_{t+c_1},\ldots,\boldsymbol{S}_{t+c_{w-2}},\boldsymbol{B}_t,\boldsymbol{B}_{t+c_1},\ldots,\boldsymbol{B}_{t+c_{w-1}}$, we cannot get all the possibilities of $(\boldsymbol{S}_t\oplus\boldsymbol{B}_t, \boldsymbol{S}_{t+c_1}\oplus\boldsymbol{B}_{t+c_1},\ldots,\boldsymbol{S}_{t+c_{w-1}}\oplus\boldsymbol{B}_{t+c_{w-1}})$. So the distribution of $(\boldsymbol{S}_t\oplus\boldsymbol{B}_t, \boldsymbol{S}_{t+c_1}\oplus\boldsymbol{B}_{t+c_1},\ldots,\boldsymbol{S}_{t+c_{w-1}}\oplus\boldsymbol{B}_{t+c_{w-1}})$ is nonuniform. Generally speaking, the distribution of $\boldsymbol{Z}_t$ which is generated by these nonuniformly-distributed vectors is also nonuniform. Therefore the nonuniform distribution of $\boldsymbol{Z}_t$ can be used to make a distinguishing attack. The detailed steps of our distinguishing attack are as follows:

1. Find a low-weight multiple of the linear feedback polynomial $f(x)$. Let the weight be w.
2. Compute the distribution $D_0(\boldsymbol{Z}_t)$.
3. Compute the sample number N needed to be observed.
4. For t=0,1,…,N, compute $A=\sum_{t=0}^{N}\left(\log_2\frac{D_0(\boldsymbol{Z}_t)}{2^{-w}}\right)$.
5. If A>0, then output "cipher"; otherwise output "random".

We can compute $D_0(\boldsymbol{Z}_t)$ by building a trellis to lower computational complexity like Ref.[15].

3.3 Creating a trellis

Now we give a simple example to explain how to create a trellis to compute $D_0(\boldsymbol{Z}_t)$. Assume that the filter Boolean function is

$$h(x_1,x_2,x_3)=x_1x_2\oplus x_1x_3\oplus x_1\oplus x_2.$$

If a 3-weight multiple of the linear feedback polynomial is considered, we can write

$$\boldsymbol{Z}_t=(z_t,z_{t+c_1},z_{t+c_2})=(h(\boldsymbol{S}_t\oplus\boldsymbol{B}_t),h(\boldsymbol{S}_{t+c_1}\oplus\boldsymbol{B}_{t+c_1}),h(\boldsymbol{S}_t\oplus\boldsymbol{S}_{t+c_1}\oplus\boldsymbol{B}_{t+c_2})),$$

where $\boldsymbol{S}_t$=(s_{t+a_1}, s_{t+a_2},0), $\boldsymbol{B}_t$=(0,0, b_{t+b_1}). The vectors $(\boldsymbol{S}_t\oplus\boldsymbol{B}_t, \boldsymbol{S}_{t+c_1}\oplus\boldsymbol{B}_{t+c_1},\boldsymbol{S}_t\oplus\boldsymbol{S}_{t+c_1}\oplus\boldsymbol{B}_{t+c_2})$ are determined completely by the vectors $\boldsymbol{S}_t, \boldsymbol{S}_{t+c_1},\boldsymbol{B}_t,\boldsymbol{B}_{t+c_1},\boldsymbol{B}_{t+c_2}$. There are $2^{3\times3-2}=2^7$ possibilities in all. Denote

$$0_h=\{(x_1,x_2,x_3)|h(x_1,x_2,x_3)=0\},$$
$$1_h=\{(x_1,x_2,x_3)|h(x_1,x_2,x_3)=1\}.$$

Then

$$0_h=\{(0\ 0\ 0),(0\ 0\ 1),(1\ 0\ 1),(1\ 1\ 1)\},$$
$$1_h=\{(0\ 1\ 0),(0\ 1\ 1),(1\ 0\ 0),(1\ 1\ 0)\}.$$

Figure 2 shows how to compute $D_0(\boldsymbol{Z}_t)$ based on a trellis. For example, when $\boldsymbol{Z}_t$=(1 0 1), according to 0_h and 1_h, all the possibilities of $\boldsymbol{S}_t\oplus\boldsymbol{B}_t, \boldsymbol{S}_{t+c_1}\oplus\boldsymbol{B}_{t+c_1}$ and $\boldsymbol{S}_t\oplus\boldsymbol{S}_{t+c_1}\oplus\boldsymbol{B}_{t+c_2}$ are as follows

$$\boldsymbol{S}_t \oplus \boldsymbol{B}_t = 0\ 1\ 0,\ 0\ 1\ 1,\ 1\ 0\ 0,\ 1\ 1\ 0;$$
$$\boldsymbol{S}_{t+c_1} \oplus \boldsymbol{B}_{t+c_1} = 0\ 0\ 0,\ 0\ 0\ 1,\ 1\ 0\ 1,\ 1\ 1\ 1;$$
$$\boldsymbol{S}_t \oplus \boldsymbol{S}_{t+c_1} \oplus \boldsymbol{B}_{t+c_2} = 0\ 1\ 0,\ 0\ 1\ 1,\ 1\ 0\ 0,\ 1\ 1\ 0.$$

Then the possible $\boldsymbol{S}_t$, $\boldsymbol{S}_{t+c_1}$, $\boldsymbol{B}_t$, $\boldsymbol{B}_{t+c_1}$ are shown in Fig.2. The line segment between two vectors means that they are possible. The number 2 between two vectors means that there are 2 possibilities for $\boldsymbol{B}_{t+c_2}$. The line segment between two vectors without numbers means that there is only one possibility. A curve joined end to end from possible $\boldsymbol{B}_{t+c_1}$ to $\boldsymbol{S}_{t+c_1}$, then to $\boldsymbol{S}_t$ and then to $\boldsymbol{B}_t$ is a possibility of $(\boldsymbol{S}_t \oplus \boldsymbol{B}_t, \boldsymbol{S}_{t+c_1} \oplus \boldsymbol{B}_{t+c_1}, \boldsymbol{S}_t \oplus \boldsymbol{S}_{t+c_1} \oplus \boldsymbol{B}_{t+c_2})$. The number of all these curves is the number of possible $(\boldsymbol{S}_t \oplus \boldsymbol{B}_t, \boldsymbol{S}_{t+c_1} \oplus \boldsymbol{B}_{t+c_1}, \boldsymbol{S}_t \oplus \boldsymbol{S}_{t+c_1} \oplus \boldsymbol{B}_{t+c_2})$. When $\boldsymbol{Z}_t$=(1 0 1), the number is 20. Then $D_0(1\ 0\ 1)=20/2^7$.

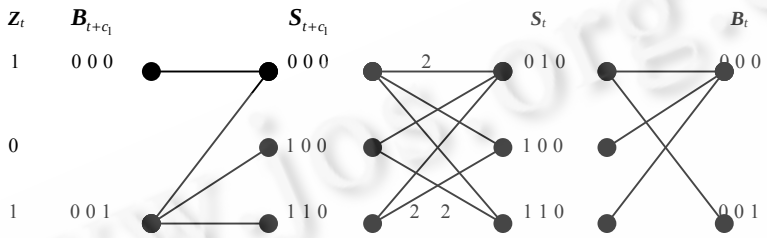

Fig.2 The trellis used to compute $D_0(\boldsymbol{Z}_t)$

Using this method, we can easily compute the distribution $D_0(\boldsymbol{Z}_t)$ as shown in Table 1 .

Table 1 The distribution $D_0(\boldsymbol{Z}_t)$

z_t	z_{t+c_1}	z_{t+c_2}	$D_0(z_t, z_{t+c_1}, z_{t+c_2})$
0	0	0	$20/2^7$
0	0	1	$12/2^7$
0	1	0	$12/2^7$
0	1	1	$20/2^7$
1	0	0	$12/2^7$
1	0	1	$20/2^7$
1	1	0	$20/2^7$
1	1	1	$12/2^7$

It is also necessary to compute the number N that we need to observe in our attack. According to $D_0(\boldsymbol{Z}_t)$, the statistical distance ε between $D_0(\boldsymbol{Z}_t)$ and $D_1(\boldsymbol{Z}_t)$ is

$$\varepsilon = \sum_{z_t} |D_0(\boldsymbol{Z}_t) - 1/8| = 1/4 .$$

Then $N \approx \varepsilon^{-2} = 2^4$. The number N can further be reduced by the method described below.

3.4 The case that uses more than one low-weight polynomial multiple

In this subsection, a method that can decrease N largely is described. More than one low-weight polynomial multiple can be used simultaneously to reduce N. Suppose we have found two w-weight parity check equations

$$s_t \oplus s_{t+c_1} \oplus \ldots \oplus s_{t+c_{w-1}} = 0,$$
$$s_t \oplus s_{t+c_w} \oplus \ldots \oplus s_{t+c_{2w-2}} = 0.$$

We use the notations z_t, $\boldsymbol{S}_t$ and $\boldsymbol{B}_t$ to denote the same meaning as Section 3.2. Now let

$$\boldsymbol{Z}_t = (z_t, z_{t+c_1}, \ldots, z_{t+c_{2w-2}}) \quad (6)$$

Then Eq.(6) can be denoted as

$$\boldsymbol{Z}_t = (h(\boldsymbol{S}_t \oplus \boldsymbol{B}_t), h(\boldsymbol{S}_{t+c_1} \oplus \boldsymbol{B}_{t+c_1}), \ldots, h(\boldsymbol{S}_{t+c_{w-2}} \oplus \boldsymbol{B}_{t+c_{2w-2}})) \quad (7)$$

Like Section 3.2, we can get $\boldsymbol{S}_{t+c_{w-1}} = \bigoplus_{i=0}^{w-2} \boldsymbol{S}_{t+c_1}$, $\boldsymbol{S}_{t+c_{w-2}} = \boldsymbol{S}_t \oplus \bigoplus_{i=0}^{2w-3} \boldsymbol{S}_{t+c_1}$, where $c_0=0$. Therefore Eq.(7) can be written as

$$\boldsymbol{Z}_t=(h(\boldsymbol{S}_t\oplus\boldsymbol{B}_t),h(\boldsymbol{S}_{t+c_1}\oplus\boldsymbol{B}_{t+c_1}),\ldots,h(\bigoplus_{i=0}^{w-2}\boldsymbol{S}_{t+c_1}\oplus\boldsymbol{B}_{t+c_{w-1}}),h(\boldsymbol{S}_{t+c_w}\oplus\boldsymbol{B}_{t+c_w}),\ldots,h(\boldsymbol{S}_t\oplus\bigoplus_{i=0}^{2w-3}\boldsymbol{S}_{t+c_1}\oplus\boldsymbol{B}_{t+c_{2w-2}})) \quad (8)$$

Using the same principle as that of in Section 3.2 we know that the distribution of $\boldsymbol{Z}_t$ is nonuniform in general. The detailed process of our distinguishing attack in this case are as follows:

1. Find two low-weight multiples of the linear feedback polynomial $f(x)$. Let the weight be w.
2. Compute the distribution $D_0(\boldsymbol{Z}_t)$.
3. Compute the sample number N needed to be observed.
4. For t=0,1,…,N, compute $A=\sum_{t=0}^{N}\left(\log_2\frac{D_0(\boldsymbol{Z}_t)}{2^{1-2w}}\right)$.
5. If A>0, then output "cipher"; otherwise output "random".

When more than two low-weight polynomial multiples are used, our attack is based on the same idea.

4 Example

An example will be given to show the validity of our attack. Suppose the considered cipher structure is the one that we discuss in this paper. Assume that the filter Boolean function $h(x)$ is 6-input, 4-degree and 1-resilient as

$$h(x)=x_6(x_1x_2x_5\oplus x_1x_2x_3\oplus x_1x_2x_4\oplus x_2x_5\oplus x_1x_5\oplus x_1x_4\oplus x_3x_4\oplus x_2x_4\oplus x_5\oplus x_3\oplus 1)\oplus$$
$$x_5(x_1x_2x_4\oplus x_2x_4\oplus x_1x_4\oplus x_4\oplus 1)\oplus x_4(x_1x_3\oplus x_1\oplus x_3)\oplus x_1x_2x_3\oplus x_2,$$

where x_1, x_2, x_3, x_4 correspond to four internal state bits of LFSR at time t and x_5, x_6 correspond to two internal state bits of NLFSR at time t.

Here we first consider a 3-weight multiple of the linear feedback polynomial. For a 3-weight parity check equation, from the foregoing notations, according to Eq.(5) there is the following equation

$$\boldsymbol{Z}_t=(z_t, z_{t+c_1}, z_{t+c_2})=(h(\boldsymbol{S}_t\oplus\boldsymbol{B}_t),h(\boldsymbol{S}_{t+c_1}\oplus\boldsymbol{B}_{t+c_1}),h(\boldsymbol{S}_t\oplus\boldsymbol{S}_{t+c_1}\oplus\boldsymbol{B}_{t+c_2})),$$

where $\boldsymbol{S}_t=(s_{t+a_1}, s_{t+a_2}, s_{t+a_3}, s_{t+a_4}, 0, 0)$, $\boldsymbol{B}_t=(0,0,0,0,b_{t+b_1}, b_{t+b_2})$. We compute $D_0(\boldsymbol{Z}_t)$ using the method explained in Section 3.3. The distribution $D_0(\boldsymbol{Z}_t)$ is shown in Table 2. The statistical distance ε_1 between $D_0(\boldsymbol{Z}_t)$ and $D_1(\boldsymbol{Z}_t)$ is $\varepsilon_1=2^{-7}$. So the needed sample number is $N_1\approx\varepsilon_1^{-2}=16384$ to make distinguishing attack.

Table 2 $D_0(\boldsymbol{Z}_t)$ under the attack in Section 3.2

z_t	z_{t+c_1}	z_{t+c_2}	$D_0(z_t, z_{t+c_1}, z_{t+c_2})$
0	0	0	$2032/2^{14}$
0	0	1	$2064/2^{14}$
0	1	0	$2064/2^{14}$
0	1	1	$2032/2^{14}$
1	0	0	$2064/2^{14}$
1	0	1	$2032/2^{14}$
1	1	0	$2032/2^{14}$
1	1	1	$2064/2^{14}$

If two 3-weight multiples of the linear feedback polynomial are considered, we can write

$$\boldsymbol{Z}_t=(z_t, z_{t+c_1}, z_{t+c_2}, z_{t+c_3}, z_{t+c_4})$$
$$=(h(\boldsymbol{S}_t\oplus\boldsymbol{B}_t),h(\boldsymbol{S}_{t+c_1}\oplus\boldsymbol{B}_{t+c_1}),h(\boldsymbol{S}_t\oplus\boldsymbol{S}_{t+c_1}\oplus\boldsymbol{B}_{t+c_2}),h(\boldsymbol{S}_{t+c_3}\oplus\boldsymbol{B}_{t+c_3}),h(\boldsymbol{S}_t\oplus\boldsymbol{S}_{t+c_3}\oplus\boldsymbol{B}_{t+c_4})).$$

The distribution $D_0(\boldsymbol{Z}_t)$ is computed in Table 3. According to this distribution the statistical distance $\varepsilon_2\approx0.0083$ between $D_0(\boldsymbol{Z}_t)$ and $D_1(\boldsymbol{Z}_t)$. So the needed sample number $N_2\approx\varepsilon_2^{-2}\approx14263$ to make distinguishing attack. It is easy to see that this result is better than N_1.

Table 3 $D_0(\boldsymbol{Z}_t)$ under the attack in Section 3.4

z_t	z_{t+c_1}	z_{t+c_2}	z_{t+c_3}	z_{t+c_4}	$D_0(\boldsymbol{Z}_t)$
0	0	0	0	0	$129152/2^{22}$
0	0	0	0	1	$130944/2^{22}$
0	0	0	1	0	$130944/2^{22}$
0	0	0	1	1	$129152/2^{22}$
0	0	1	0	0	$130944/2^{22}$
0	0	1	0	1	$133248/2^{22}$
0	0	1	1	0	$133248/2^{22}$
0	0	1	1	1	$130944/2^{22}$
0	1	0	0	0	$130944/2^{22}$
0	1	0	0	1	$133248/2^{22}$
0	1	0	1	0	$133248/2^{22}$
0	1	0	1	1	$130944/2^{22}$
0	1	1	0	0	$129152/2^{22}$
0	1	1	0	1	$130944/2^{22}$
0	1	1	1	0	$130944/2^{22}$
0	1	1	1	1	$129152/2^{22}$
1	0	0	0	0	$133248/2^{22}$
1	0	0	0	1	$130944/2^{22}$
1	0	0	1	0	$130944/2^{22}$
1	0	0	1	1	$133248/2^{22}$
1	0	1	0	0	$130944/2^{22}$
1	0	1	0	1	$129152/2^{22}$
1	0	1	1	0	$129152/2^{22}$
1	0	1	1	1	$130944/2^{22}$
1	1	0	0	0	$130944/2^{22}$
1	1	0	0	1	$129152/2^{22}$
1	1	0	1	0	$129152/2^{22}$
1	1	0	1	1	$130944/2^{22}$
1	1	1	0	0	$133248/2^{22}$
1	1	1	0	1	$130944/2^{22}$
1	1	1	1	0	$130944/2^{22}$
1	1	1	1	1	$133248/2^{22}$

5 Conclusion

The security of a recently proposed new structure of stream ciphers has been studied in this paper. This structure is based on LFSR, NLFSR and a filter Boolean function. A distinguishing attack against this structure is given. An example is presented to show the validity of our attack, but our method is invalid for Grain because $D_0(\boldsymbol{Z}_t)$ is uniform. But if $D_0(\boldsymbol{Z}_t)$ is nonuniform when another filter Boolean function is used, our method is valid. Our results demonstrate that the new stream cipher structure may suffer heavier security threats than it has been revealed, and our method should be taken into account in designing new structures of stream ciphers. According to this attack, it is necessary to ensure that $D_0(\boldsymbol{Z}_t)$ is uniform in designing stream ciphers based on LFSR, NLFSR and filter Boolean functions. Otherwise, it suffers the distinguishing attack as described in this paper.

Acknowledgement We thank the anonymous referees for their useful comments.

References:

[1] NESSIE. New european schemes for signatures, integrity, and encryption. http://www.nessie.eu.org/nessie/

[2] Courtois N. Higher order correlation attacks, XL algorithm and cryptanalysis of toyocrypt. In: Lee PJ, Lim CH, eds. Proc. of the Information Security and Cryptology (ICISC 2002). LNCS 2587, Berlin: Springer-Verlag, 2003. 182−199.

[3] Courtois N, Meier W. Algebraic attacks on stream ciphers with linear feedback. In: Biham E, ed. Advances in Cryptology—Eurocrypt 2003. LNCS 2656, Berlin: Springer-Verlag, 2003. 345−359.

[4] Courtois, N. Fast algebraic attack on stream ciphers with linear feedback. In: Boneh D, ed. Advances in Cryptology—Crypto 2003. LNCS 2729, Berlin: Springer-Verlag, 2003. 176−194.

[5] eSTREAM. The ECRYPT stream cipher project. http://www.ecrypt.eu.org/stream/

[6] Hell M, Johansson T, Meier W. Grain—A stream cipher for constrained environments. The ECRYPT Stream Cipher Project Report 2005/010, 2005. http://www.ecrypt.eu.org/stream/

[7] Englund H, Hell M, Johansson T. Correlation attacks using a new class of weak feedback polynomials. In: Roy B, Meier W, eds. Proc. of the Fast Software Encryption (FSE 2004). LNCS 3017, Berlin: Springer-Verlag, 2004. 127−142.

[8] Englund H, Johansson T. A new simple technique to attack filter generators and related ciphers. In: Handschuh H, Hasan A, eds. Proc. of the Selected Areas in Cryptography (SAC 2004). LNCS 3357, Berlin: Springer-Verlag, 2005. 39−53.

[9] Englund H, Johansson T. A new distinguisher for clock controlled stream ciphers. In: Gilbert H, Handschuh H, eds. Proc. of the Fast Software Encryption (FSE 2005). LNCS 3557, Berlin: Springer-Verlag, 2005. 181−195.

[10] Maximov A. Cryptanalysis of the "grain" family of stream ciphers. In: Proc. of the 2006 ACM Symp. on Information, Computer and Communications Security (ASIACCS 2006). New York: ACM Press, 2006. 283−288.

[11] Berbain C, Gilbert H, Maximov A. Cryptanalysis of grain. In: Proc. of the Workshop Record of the State of the Art of Stream Ciphers (SASC 2006, WR18). Leuven, 2006. http://www.ecrypt.eu.org/stvl/sasc2006/

[12] Cover T, Thomas JA. Elements of Information Theory, Wiley Series in Telecommunication. New York: Wiley, 1991.

[13] Coppersmith D, Halevi S, Jutla CS. Cryptanalysis of stream ciphers with linear masking. In: Yung M, ed. Advances in Cryptology —Crypto 2002. LNCS 2442, Berlin: Springer-Verlag, 2002. 515−532.

[14] Golić JD. Computation of low-weight parity-check polynomials. Electronic Letters, 1996,32(21):1981−1982.

[15] Leveiller S, Boutros J, Guillot P, Zémor G. Cryptanalysis of nonliear filter generators with {0,1}-metric Viterbi decoding. In: Honary B, ed. Proc. of the Cryptography and Coding 2001. LNCS 2260, Berlin: Springer-Verlag, 2001. 402−414.

HUANG Xiao-Li was born in 1978. She is a Ph.D. candidate at the Institute of Software, the Chinese Academy of Sciences and Graduate School of the Chinese Academy of Sciences. Her current research areas are design and analysis of stream ciphers.

WU Chuan-Kun was born in 1964. He is a professor at the Institute of Software, the Chinese Academy of Sciences and a CCF senior member. His research areas are design and analysis of cryptographic algorithms, design and analysis of network security protocols.