Non-linear Shift Register-based Bent Combiner Cryptanalysis Results

Hope4Sec Team

Hope4Sec, Tallinn, Estonia
contact@hope4sec.eu

June 5, 2023

Abstract

This white paper presents a *Concept Cipher* called *Cipherbent6* which captures the properties of Non Linear Feedback Shift Registers (NLFSRs) combined by a highly non linear Boolean function. This concept cipher enables to test a proprietary cryptanalysis whose first results are presented. This cryptanalysis, yet at its very beginning, already outperforms known cryptanalyses.

Keywords: Combinatorial Cryptanalysis, Nonlinear Shift Register, Stream Cipher, Algebraic Attack, Bent function, Achterbahn.

1 Introduction

Stream ciphers [14] are one of the main cryptographic primitives used in symmetric cryptography. At the origin, the first stream ciphers were built with Linear Feedback Shift Registers (LFSRs), where linearity is meant in the register update function while the combining function is meant to be non-linear to break the intrinsic linear properties of the sequences produced by the register. The most common design is the combiner generator, depicted in Figure 1 to which most of other the stream cipher designs can be reduced.

Most stream ciphers use combined or non-linearly filtered LFSRs [14]. However, their security has been questioned over the years. The natural evolution of these systems is towards the use of Non Linear Feedback Shift Registers (NLFSRs) [9] and Boolean functions with good cryptographic properties according to the general structure in Figure 1. If for the latter a substantial body of knowledge exists [2] – with however a large number of problems still open as soon as the number of variables exceeds ten – the study of NLFSRs is still in its infancy. To date, for example, no NLFSRs are known in maximum period for lengths greater than 31 [5] when looking for rather simple, sparse feedback polynomials (a desirable property when dealing with implementation aspects). But the

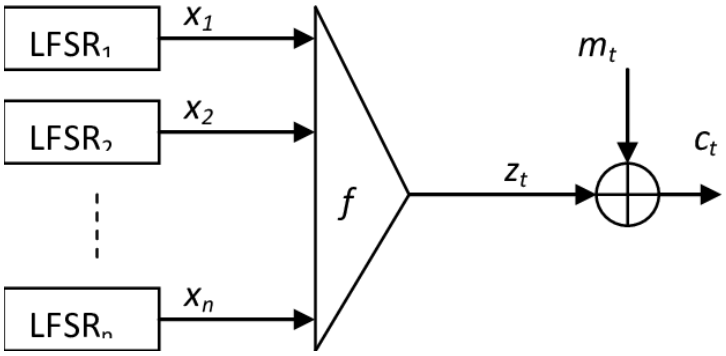

Figure 1: General Structure of LFSR-based Combiner

clever combination of these registers can lead to systems with effective security. The two best examples are Achterbahn [8] and Trivium [7].

Attacking this new class of systems remains an open problem [17]. For real systems, no effective cryptanalysis is known that could seriously question their security. Algebraic or statistical attacks by correlation are no longer applicable. It is therefore necessary to consider a radically different approach that is neither algebraic nor statistical.

We are working on such an approach. It is combinatorial in nature. The principle is to rewrite the system to produce a combinatorial equivalent system and to translate the key search by a known combinatorial problem, of a complexity that is reachable in practice (data and time) and lower than the complexity of known attacks in the algebraic or statistical domains. We have named this approach CE (*Combinatorial Equivalence*). Note that the principle of rewriting in the algebraic domain was initiated in [17] for some specific ciphers but does not lead to significant gains in cryptanalysis complexity. The CE cryptanalysis applies both to stream and block ciphers. The rewriting techniques just do not consider the same underlying combinatorial objects. It is also worth mentioning that the CE technique is fully transposable to LFSR-based stream ciphers linear stream ciphers and to ciphertext-only attacks.

We consider real-life cryptanalysis only. We denote it*Effective Cryptanalysis.*

Definition 1.1 (Effective Cryptanalysis)**.** *A cryptanalysis is called effective whenever it can be performed:*

- *In a limited time that allows the cryptanalysis to be played a finite number of times.*
- *On realistic input data sizes that are compatible with the operational reality of the use cases.*

For instance, any attack requiring data without changing keys far beyond the crypto-periods in use in the industrial or governmental world is not effective. Moreover known plaintext attack are realistic as long as the amount of data does not exceed a few kilobits. Any attack whose time complexity is greater than 2^{70} is not effective nowadays. However,

to evaluate the effective security of a system, the size of the required input data N takes precedence over the time complexity T: a system which can be broken with a relatively small value of N despite a high complexity T is in practice considered less secure than a system requiring a much higher value of N even for a significantly lower complexity. It is not so much impossible to increase computing power as it is to have enough input data (at least conceptually) at least if standard cryptographic policies are applied (in particular concerning crypto-periods).

The paper is organized as follows: Section2 presents the specification of the concept cipher *Cipherbent6*.Then in Section 3 the state-of-the-art with respect to stream ciphers cryptanalysis is summarized. The intent is to assess the security of Cipherbent6 with respect to these known techniques. Subsequently, Section 4 summarizes the initial cryptanalysis results and performances we have obtained on Cipherbent6 with respect to the CE cryptanalysis. Finally, Section 5 presents the future work to develop the CE cryptanalysis.

Disclaimer: *this white paper is not basically a research article in the usual sense. The CE technique is not public and is reserved for the industrial and governmental world. This paper is intended to present our results. The proof of our results can be provided on request by a challenge approach (sending an output sequence and returning the key).*

2 Bentcipher6 Specifications

Definition 2.1 (Concept Cipher)**.** *A concept cipher $\mathcal{E}$ is a cryptographic algorithm describing a family $\mathcal{A}$ of cryptographic algorithms of which it captures all the complexity. It then allows the cryptographic security analysis of $\mathcal{A}$ from its representative $\mathcal{E}$. We then can state:*

- *An effective cryptanalysis on $\mathcal{E}$ can be verified by challenge or demonstration, effectively.*
- *Any cryptanalysis that would work on any member of $\mathcal{A}$ would also be efficient on $\mathcal{E}$ with a lower complexity.*
- *A reliable evaluation of the cryptanalysis of systems of the same class can be derived from an effective cryptanalysis of $\mathcal{E}$. As a consequence, any applicable cryptanalysis on $\mathcal{E}$ allows to give an evaluation of the cryptanalysis complexity for the algorithms of $\mathcal{A}$.*
- *Any cryptanalysis of $\mathcal{E}$ is transposable/scalable to any system of class $\mathcal{A}$.*

The concept cipher $\mathcal{E}$ is thus a minimal element of $\mathcal{A}$ ordered by the order relation with respect to the cryptanalysis complexity.

In this paper, we consider a concept cipher called *Bentcipher6*. The corresponding class is that of stream ciphers in which Non Linear Shift Registers (NLFSR) are combined by a non linear Boolean function. Achterbahn Algorithm [8] is a member of this class.

Among many other aspects, concept Cipher $\mathcal{E}$ is worth considering whenever it resists to all known attacks.

As for *Bentcipher6*, the combining function is 6-variable bent function. Its main specifications are the following:

- The combined function is a 6-variable bent function $x_1x_4 \oplus x_2x_6 \oplus x_1x_4$ [13]. It has maximal nonlinearity $\mathcal{NL}(f) = 28$. Its algebraic immunity is not optimal since $\mathcal{AI}(f) = 2 < 3$. However, this non-optimal nature is not expoitable in the case of non-linear feedback shift registers.
- Six nonlinear linear shift registers (NLFSR) of respective length $27, 28, 30, 31, 32, 33$, all having maximum period. These NLFSRs are those used in the Achterbahn stream cipher [8]. Their feedback polynomial are all dense and of high degree. For instance, NLFSR A_{12} has the following Algebraic Normal Form (ANF) for its feedback polynomial:

$$\begin{aligned} A_{12}(x_0, x_1, \ldots, x_{32}) \quad = \quad & x_0 \oplus x_2 \oplus x_7 \oplus x_9 \oplus x_{10} \oplus x_{15} \oplus x_{23} \oplus x_{25} \oplus x_{30} \oplus x_8x_{15} \\ & \oplus x_{12}x_{16} \oplus x_{13}x_{15} \oplus x_{13}x_{25} \oplus x_1x_8x_{14} \oplus x_1x_8x_{18} \oplus x_8x_{12}x_{16} \\ & \oplus x_8x_{14}x_{18} \oplus x_8x_{15}x_{16} \oplus x_8x_{15}x_{17} \oplus x_{15}x_{17}x_{24} \oplus x_1x_8x_{14}x_{17} \\ & \oplus x_1x_8x_{17}x_{18} \oplus x_1x_{14}x_{17}x_{24} \oplus x_1x_{17}x_{18}x_{24} \oplus x_8x_{12}x_{16}x_{17} \oplus \\ & x_8x_{14}x_{17}x_{18} \oplus x_8x_{15}x_{16}x_{17} \oplus x_{12}x_{16}x_{17}x_{24} \oplus x_{14}x_{17}x_{18}x_{24} \oplus \\ & x_{15}x_{16}x_{17}x_{24}. \end{aligned}$$

- The key is the initial content of the 6 registers at time $t = 0$. Thus the key is 181 bit long.

3 Stream Cipher Cryptanalysis State-of-the-Art

While there is a large literature dealing with stream ciphers based on LFSRs and their cryptanalysis, there is quite no result as far as their non linear analogues are concerned.

Stream cipher cryptanalysis techniques are mostly divided in two categories: statistical/-correlation attacks and algebraic attacks. Other variants fall into one or both of these categories.

3.1 Statistical Correlation Attacks

3.1.1 Simple Correlation Attacks

All these attacks rely on a sufficiently high correlation between the sum of a subset of the input variables and the output of the combining Boolean function f [16]. In other words $p = P[f(x) =< x, u >]$ where $< ., . >$ denotes the bitwise scalar product. The Hamming weight of u (which describes the subset of input registers to be taken into account)

determines how many registers have to be considered to go through an exhaustive search simultaneously.

Considering a known plaintext attack (as long as this is operationally acceptable), the required keystream length N depends on the probability on the correlation probability p_i, on the length L of the NLFSR, on the probability of false alarm P_f (*i.e.* the probability that a wrong key is kept as a good candidate) and on the probability of non detection P_m (the good candidate has been rejected and is missing). With $P_m = 10^{-3}$ and $P_f = 2^{-L}$ we then have [1]

$$N < \left(\frac{\sqrt{L} + 3\sqrt{2.p.(1-p)}}{\sqrt{2}(p-0.5)} \right)^2 \quad (1)$$

From Equation1, it follows that the required data size N increases inversely with the probability p and increases with the register length L.

Correlation attacks apply to both LFSR-based and NLFSR-based combining stream ciphers.

3.1.2 Fast Correlation Attacks

Fast correlation attacks [11, 12] rely on the same principle as the correlation attack since they exploit the existence of a correlation between the sum of a subset of the input variables and the output of the combining Boolean function f. But instead of searching through all the possible initializations of target LFSRs, they model LFSR sequences as error-correcting codes on which the combining Boolean function operates as noise of parameter p (Binary Symmetric Channel). The key recovery step consists then in applying a maximum-likelihood decoding algorithm to the linear code defined by the LFSR(s) feedback polynomial(s).

While they are faster than correlation attacks, they require far larger input data (output bits from f or ciphertext bits). In this respect, the effectiveness of fast correlation attacks can be questioned in most operational use-cases. A exhaustive survey on fast correlation attack techniques can be found in [10].

It is worth mentioning that fast correlation attacks regarding NLFSRs-based combiners do not apply. Up to the authors' knowledge, no study has been published on the possible generalization to these combiners.

3.2 Algebraic Attacks

Algebraic attacks [3] and their fast version [4] consist in expressing stream ciphers output bits as equations of low degree where the unknowns are the initialisation bits of the linear shift registers.

The general principle of algebraic attacks is to recover the key (the registers' initialisations) by solving the system of such equations (in a known plaintext context). Generally the number of such equations can be much larger than the number of unknowns. This

makes the resolution of the system less complex. The resolution is performed either by using Groebner bases [15] or by linearizing the system (replacing every monomial of degree greater than 1 by a new unknown variable). The resulting linear system has however too many unknowns (especially is the degree of the equation is beyond 2 or 3) in practice and cannot be solved whenever the algebraic degree of the combining function is large enough. The complexity of these attacks and the required consecutive output bits make them not effective in practice.

These attacks do not apply at the present to NLFSRs combiners.

4 Cipherbent6 Cryptanalysis Results

Cipherbent6 has been designed as a concept cipher for non-linear shift register6based combiners and to test our CE-cryptanalysis. **Cipherbent6 is in no way intended to be a secure encryption system nor usable for real applications except as an evaluation cipher**.

The aim of our work is to provide a cryptanalysis of Cipherbent6 which is as much effective as possible. From a limited size output (known plaintext attack), we intend to recover the 181-bit key.

When considering the existing attacks against stream ciphers presented in Section 3, none of them are applicable except correlation attacks but with an overall complexity of 2^{35} provided that at least 8,000 output bits are known (using Equation 1 and the fact that for a 6-variable bent function $p = 0.5625$). However, our own simulation results show than in practice around 10,000 bits are necessary to recover the 181-bit key uniquely (with a negligible amount of wrong candidates). As for other techniques, their inapplicability comes mostly from the fact that we deal with dense NLFSRs combined with a complex Boolean function.

The cryptanalysis of Cipherbent6 is in two parts:

1. Obtaining a combinatorial equivalent of Cipherbent6 is a **one-time operation** that takes from 24 hours (version 1) to 48 hours (version 2) and has overall complexity of $\mathcal{O}(2^{43})$. At the present time these two first combinatorial equivalents we have produced are not the most optimal ones. Our initial goal was first to find at least one equivalent to validate our approach and second to prove that more optimal equivalents do exist.

2. From a given combinatorial equivalent, the cryptanalysis part requires $N = 2,790$ output bits (version 1) or $N = 1,820$ output bits (version 2). For both versions, it requires a computing time of slightly less 72 hours to retrieve the 181 bits of key. The overall complexity is in $\mathcal{O}(2^{45})$. These preliminary results outperform the known attacks that could be applied (basically correlation attack). This cryptanalysis is not optimized yet and we expect to reduce the cryptanalysis time as well.

Interested people can submit a challenge for us to solve by visiting the *Hope4Sec* website (where the reference source code for Cipherbent6 is provided).

Experiments have been conducted on two AMD Ryzen Threadripper 2990 WX 32-Core Processor x64 (256 Mb of RAM, 36 Tb HDD each).

5 Conclusion et future work

We are currently working to identify more efficient combinatorial equivalents of Cipherbent6. Different candidates are currently under analysis. We expect to reduce the number of the output bits required to less than 1,000 bits in the next step. The ultimate goal/hope is to achieve a number of bits close to the size of the secret key.

The next step is to apply the CE technique to real encryption systems. The first candidate that is closest to Cipherbent6 is the *Achterbahn* algorithm. We expect to find an effective cryptanalysis with around a few Kilobits of output bits (less than 500 Kb).

Finally, our current research deals with the application of CE cryptanalysis to block ciphers.

References

[1] A. Canteaut. Correlation Attack for Stream Ciphers. In: H. van Tilborg, C. A. Henk and S. Jajodia eds, *Encyclopedia of Cryptography and Security*, pp. 261–262, Springer US, 2011.

[2] C. Carlet. *Boolean Functions for Cryptography and Coding Theory*. Cambridge University Press, 2021.

[3] N. Courtois and W. Meier. Algebraic Attacks on Stream Ciphers with Linear Feedback. In: *Proceedings of EUROCRYPT 2003*, Lecture notes in computer science, vol 2656, pp 346–359, 2003.

[4] N. Courtois. Fast Algebraic Attacks on Stream Ciphers With Linear Feedback. In: *Proceedings of CRYPTO 2003*, Lecture notes in computer science, vol 2729, pp. 177–194, 2003.

[5] P. Dąbrowski, G. Labuzek, T. Rachwalik and J. Szmidt. Searching for Nonlinear Feedback Shift Registers with Parallel Computing. *Information Processing Letters*, Vol. 114, Issue 5, pp. 268–272, 2014.

[6] B. A. Davey and H. A. Priestley. *Introduction to Lattices and Order*. Cambridge University Press, 2002.

[7] C. De Cannière, and B. Preneel. "Trivium specifications". eSTREAM submitted papers, 2005. Last retrieved on March 20th, 2023 on `https://www.ecrypt.eu.org/stream/ciphers/trivium/trivium.pdf`

[8] B. M. Gammel, R. Göttfert, O. Kniffler. "ACHTERBAHN-128/80". Achterbahn home page, 2006. Last retrieved on April 25th, 2023 on `http://www.matpack.de/achterbahn/Gammel_Goettfert_Kniffler_Achterbahn-128-80.pdf`

[9] S. W. Golomb. *Shift Register Sequences*. Aegean Park Press, 1981.

[10] F. Jönsson. Some results on fast correlation attacks. PhD thesis, University of Lund, Sweden, 2002.

[11] W. Meier W, O. Staffelbach. Fast correlation attacks on stream ciphers. In: *Advances in cryptology – EUROCRYPT 1988*. Lecture notes in computer science, vol. 330. Springer, pp. 301–314, 1988.

[12] . W. Meier, O. Staffelbach. Fast Correlation Attack on Certain Stream Ciphers. *Journal of Cryptology*, vol. 1, pp. 159–176, 1989.

[13] O. S. Rothaus. On "Bent" Functions. *Journal of Combinatorial Theory (A)*, 20:300-305, 1976.

[14] R. A. Rueppel. *Analysis and Design of Stream Ciphers*. Springer Berlin Heidelberg, 1986.

[15] M. Sala, T. Mora, L. Perret, S. Sakata and C. Traverso. Groebner Bases, Coding and Cryptography. Springer Verlag, 2009.

[16] T. Siegenthaler Decrypting a Class of Stream Ciphers Using Ciphertext Only. *IEEE Trans Comput* C-34(1):81–84, 1985.

[17] G. Yao. *Transformation and Security Analysis of NLFSR-based Stream Ciphers*. PhD Thesis, University of Melbourne, 2021.