# Algebraic Analysis of the Initialization Phase Grain Version 1 using a Degree-Oriented Interpolation Technique

Andreas Schaffhauser
FernUniversität in Hagen
Dept. of Mathematics and Computer Science
58084 Hagen, Germany
andreas.schaffhauser@fernuni-hagen.de

29th Crypto Day, 6/7 September 2018

We present a degree-oriented interpolation technique to express the keystream bits $z_0$ to $z_{39}$ of the keystream generator Grain Version 1 (Grain v1) in algebraic normal form (ANF). Because of the fact that the resulting algebraic expressions just consist of secret key bits, a nonlinear system of equations can be established, whose solution is the secret key itself. This report shows one basic way to prepare an algebraic attack against a stream cipher.

## References

(2008). The eSTREAM Project Website. HTML. URL `http://www.ecrypt.eu.org/stream/project.html`.

Martin Hell, Thomas Johannson & Willi Meier (2007). Grain: a stream cipher for constrained environments. *International Journal of Wireless and Mobile Computing* **2**, 86–93. URL `https://dx.doi.org/10.1504/IJWMC.2007.013798`.

## 1 Introduction

An algebraic attack against a stream cipher is possible, if there are any relations between the key/state bits and the resulting keystream bits. Many recent stream ciphers are based on the same principle: starting from a secret key and an initial vector (IV), non-/linear feedback shift registers (NLFSR/LFSR) are initialized with them. Outgoing from that state of the registers, bits are taken as input of a filter function to calculate a keystream bit $z_i$. This keystream bit is used for encrypting one of the plaintext bits $p_i$ with an XOR-Operation, resulting in the cipher bit $c_i$, simply expressed as $c_i = p_i \oplus z_i$. Thereby every $z_i$ contains information about secret key bits $k_i$, it is possible to express them in ANF, consisting only secret key bits.

**crypto day matters 29** (2018)
Beiträge vom 29. Krypto-Tag, 06.-07.09.2018, Bosch, Renningen.
doi:10.18420/cdm-2018-29-02

# 2 Methodology

## 2.1 Grain v1

The used cipher for our investigations is Grain v1. This cipher was submitted to the eSTREAM Project by Hell, Johannson & Meier (2007). The cipher consists of an 80 bits NLFSR and an 80 bits LFSR. 5 Bits of these 160 state bits get used for the filter function $h(x)$. For the calculation of a keystream bit $z_i$ a sum of certain NLFSR bits and the result of the filter function $h(x)$ is added with an XOR-Operation. For the exact definitions, please refer to the specification.

The 80 key bits of the secret key are loaded into the NLFSR registers. The 64 bits of the IV are loaded into LFSR registers. Figure 1 shows the start state of the cipher.

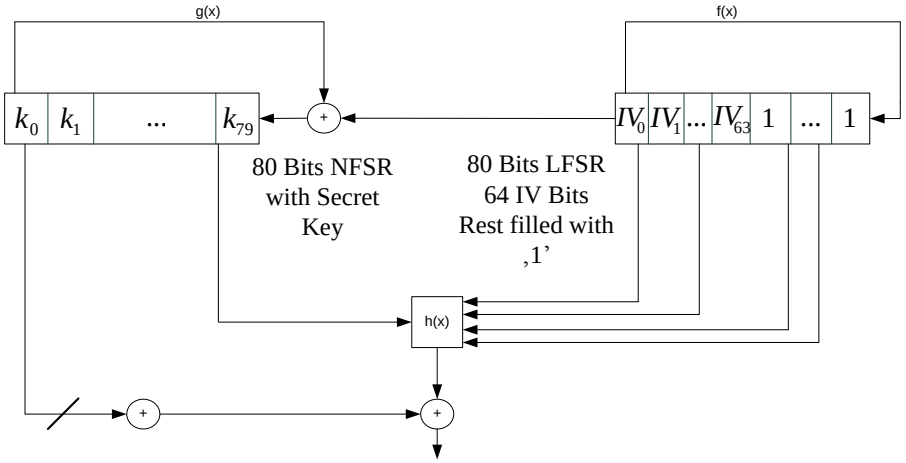


Figure 1: Start State Grain v1

During the initialization phase 160 keystream bits ($z_0 - z_{159}$) are fed back to increase the degree of confusion into the both shift registers again.

## 2.2 Key Bit Dependencies

Starting from the state, seen in Figure 1, keystream bits get calculated. The Grain v1 implementation is extened that each keystream bit $z_i$ not only has its value, but also the key bits influencing them. As a result, a list is written containing the set of responsible key bits for each keystream bit, for example:

$$z_0 : \{1,\ 2,\ 4,\ 10,\ 31,\ 43,\ 56\} \tag{1}$$

This means the key bits $k_1$, $k_2$, $k_4$, $k_{10}$, $k_{31}$, $k_{43}$ and $k_{56}$ have influence of the ANF of $z_0$.

### 2.3 Degree-Oriented Interpolation

With the help of the list with key bit dependecies, it is possible to interpolate the ANF for the respective $z_i$. We assume every possible value combination of the key bit set and calculate the respective $z_i$. But we start to look for the coefficients of the unknown ANF at degree 0. Then we search for the coefficients of degree 1, then degree 2, etc. This procedure stops on our target degree $d$. By omitting all coefficients greater than target degree $d$, we omit coefficients influecing the ANF in $1/2^{D+1}, 1/2^{D+2}, ...$ cases. If a keystream bit depends on $n$ key bits, the omitting changes the calculation effort from $\mathcal{O}(2^n)$ to $\mathcal{O}(n^d)$. The structuring of the interpolation order keeps the effort as low as possible and allows to find well hitting algebraic expressions for keystream bits having a huge set of key bit dependencies. In addition, an addressing scheme of the key bit dependencies and address distribution across multiple processors can generate massive parallelization that expands the boundaries of findable algebraic expressions. To decide whether a coefficient of a higher degree is present in the unknown ANF, the estimated value for the respective $z_i$ must be calculated. For this purpose all already found coefficients have to stored in a common data structure. Due the matter of fact that with increasing $z_i$ and increasing degree the number of coefficients rises also, a data structure with a time complexity $\mathcal{O}(1)$ for searching an already found coefficient is recommended (e.g. a hash map).

## 3 Results and Future Work

The properties of the found algebraic expressions can be seen in Table 1.

| $z_i$ | $\sum$ Coefficients | Hit Quality | Alg. Deg. |
|---|---|---|---|
| $z_0 - z_{15}$ | 7 | 100% | 1 |
| $z_{16}$ | 15 | 100% | 2 |
| $z_{17}$ | 233 | 100% | 7 |
| $z_{18} - z_{23}$ | 38 | 100% | 6 |
| $z_{24} - z_{28}$ | 60 | 100% | 6 |
| $z_{29} - z_{31}$ | 59 | 100% | 6 |
| $z_{32}$ | 67 | 100% | 6 |
| $z_{33}$ | 283 | 100% | 7 |
| $z_{34}$ | 57020 | 91% | 9 |
| $z_{35}$ | 28614 | 90% | 8 |
| $z_{36}$ | 25700 | 90% | 8 |
| $z_{37-38}$ | 265267 | 89% | 9 |
| $z_{39}$ | 262610 | 93% | 9 |

Table 1: $z_i$ ANF Properties

The future work will focus on solving this nonlinear system of equations.