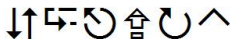


Chosen IV Statistical Analysis for Key Recovery Attacks on Stream Ciphers

Simon Fischer¹, Shahram Khazaei², and Willi Meier¹

¹FHNW and ²EPFL (Switzerland)

ECRYPT

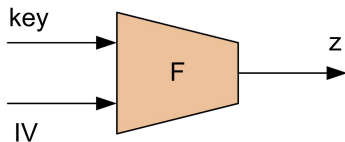


SASC 2008, Lausanne - February 13th

Initialization Function

Initialization function of a stream cipher takes key K and initialization vector V to produce the initial state.

We define the related Boolean function $F(K, V) = z$ where z is the first keystream bit.



Threat model: Adversary can send IV to oracle and gets z according to secret key chosen by the oracle.

Requirements

Security: $F(K, V) = z$ should depend on each key bit and IV bit in a complex way.

Toy example: $F(K, V) = k_1k_2v_1v_3v_6 \oplus k_3k_4v_1 \oplus v_2v_5$.

Efficiency: Because of limited resources, F is often constructed by a number of simple rounds.

For example, Grain-128 and Trivium use Boolean functions of low-degree.

- ▶ Is diffusion sufficient for specified number of rounds?
- ▶ If no, how to mount an attack?

Distinguishing Attacks based on Polynomial Description

Consider $F(K, V)$ with $K = (k_1, \dots, k_n)$ and $V = (v_1, \dots, v_m)$.

Use fixed values for all key bits and for some IV bits:
obtain smaller function $f(v_1, v_2, \dots, v_l)$ with $l < m$.

Adversary can compute the truth table of f with 2^l queries, and also the algebraic normal form (ANF).

Compute the ANF for many different choices of fixed IV bits.
Statistical analysis of ANF's compared to a random function
(each monomial should exist with probability $1/2$ etc.).

[Filiol'02],[Saarinen'06],[O'Neil'07],[Englund-Johansson-Turan'07].

High-degree Monomials

High degree monomials in f are more suspicious to exhibit non-randomness: it will take many clockings before all selected IV bits meet in the same memory cell.

Maximum degree (MD) test in [Englund-Johansson-Turan'07]: check if the maximum degree monomial is produced by f .

The MD test turned out to be a suitable measure for diffusion.

They reported a distinguishing attack on

- ▶ 192/256 rounds in Grain-128
- ▶ 736/1152 rounds in Trivium

Generalization

Make partition $V = (U, W)$ with U as input to f . Focus on single coefficient C of $f(U)$, e.g. maximum degree monomial.

$C = C(K, W)$ depends on the key and remaining IV bits. The values of this Boolean function are derived with the oracle.

Scenarios of attacks (if mixing is not complete):

- ▶ Imbalanced structure of C can be used for dist. attacks.
- ▶ Sometimes, $C(K, W)$ for fixed W does not involve all key bits: key recovery attack on 576/1152 rounds of Trivium [Vielhaber'07].
- ▶ More generally: some key bits in $C(K, W)$ may have only a limited influence. . .

New Scenario

Given $C(K, W)$, find approximation A which depends on subkey of only $t < n$ key bits.

Reduce the search space from 2^n to 2^t . Method of probabilistic neutral key bits [Aumasson-Fischer-Khazaei-Meier-Rechberger'08].

Partition of $K = (L, M)$ with significant key bits L and non-significant key bits M . $A(L, W)$ is defined by replacing non-significant key bits in $C(K, W)$ with fixed values.

Definition: Let γ_i be the bias that complementing the key bit k_i does not change the output of C (neutrality measure).

Significant key bits L : all key bits k_i with $|\gamma_i| < \text{threshold}$.

Key Recovery Attacks

Probabilistic guess and determine: try all possible subkeys L and distinguish correct guess L from incorrect ones.

1. Compute $C(K, W)$ through oracle, for unknown key and N random values of W .
2. For each choice of L do
 - ▶ Compute $A(L, W)$ for the same N values of W .
 - ▶ Check if $C(K, W)$ is equal to $A(L, W)$ for most of the N samples (optimal distinguisher)

Given a candidate subkey L , the entire key is verified by exhaustive search over remaining key part M .

Complexity

Required number of samples N depends on:

- ▶ Correlation ε between $A(L, W)$ and $C(K, W)$ if guessed part L is correct resp. incorrect.
- ▶ The desired levels of p_{fa} and p_{mis} .

Computation of N values of $A(L, W)$ has a cost of $N2^l$.
Repeat for all 2^t guesses of L , hence $N2^{l+t}$.

Set of candidates for subkey L has size $2^t \cdot p_{\text{fa}}$.
Verification of entire key in $2^t p_{\text{fa}} \cdot 2^{n-t} = 2^n p_{\text{fa}}$.

Total complexity: $N2^{l+t} + 2^n p_{\text{fa}}$.

Experimental Results: Trivium

Trivium has internal state of 288 bits.

$n = 80$ key bits, $m = 80$ IV bits, $R = 1152$ rounds.

Example:

- ▶ $F(K, V)$ computes first keystream bit after $r = 672$ rounds.
- ▶ Variable IV part U of $l = 11$ bits, get $f(U)$, focus on coefficient $C(K, W)$ of maximum degree monomial.
- ▶ Compute neutrality measure of key bits (with $|W|=5$).
A set of $t = 29$ key bits ruled out as significant.
- ▶ Leads to approximating function $A(L, W)$ with some correlation to $C(K, W)$.
- ▶ Correct subkey can be detected with time complexity 2^{55} .

Experimental Results: Grain-128

Grain-128 has internal state of 256 bits.

$n = 128$ key bits, $m = 96$ IV bits, $R = 256$ rounds.

Example:

- ▶ Consider $r = 180$ rounds, and use IV part U of $l = 7$ bits.
- ▶ Focus on coefficient $C(K, W)$ of maximum degree monomial.
- ▶ Identify $t = 110$ significant key bits for L and get $A(L, W)$.
- ▶ Can detect subkey in estimated time complexity 2^{124} ,
i.e. improvement factor 2^4 .

Question: How to find weak IV bits?

Conclusions

- ▶ Have applied the recently introduced technique of probabilistic neutral bits.
- ▶ Useful in analysis of initialization of stream ciphers.
- ▶ Contributes to the recent framework of chosen IV statistical analysis.
- ▶ Key recovery with complexity lower than exhaustive key search for simplified versions of two phase 3 eSTREAM candidates.

Thank you!

