

Massive Superpoly Recovery with a Meet-in-the-middle Framework Improved Cube Attacks on Trivium and Kreyvium

Jiahui He¹, Kai Hu^{1,2}, Hao Lei¹, Meiqin Wang¹

¹Shandong University

²Nanyang Technological University

Eurocrypt 2024 @ Zurich

May 27, 2024

- 1 Cube Attack and Superpoly Recovery
- 2 Recalling Core Monomial Prediction (CMP)
- 3 Refinement of CMP Theory
 - SR Problem
 - Perfectly Accurate CMP Theory
 - Equivalence between CMP and MP
 - Meet-in-the-middle Framework
- 4 Application to Superpoly Recovery
- 5 Key Recovery from Superpoly

- ▶ Proposed by Dinur and Shamir at EUROCRYPT 2009 ¹
- ▶ Any output bit of a symmetric cipher can be expressed as an Boolean function f of \mathbf{k} and \mathbf{v} . The cube attack focuses on the coefficient of a monomial $t_I = \prod_{i \in I} v[i]$.
 - I is called cube indices and $v[i]$'s ($i \in I$) are called cube variables.
 - $f(\mathbf{k}, \mathbf{v}) = p(\mathbf{k}, \mathbf{v}) \cdot t_I + q(\mathbf{k}, \mathbf{v})$,
 - $p(\mathbf{k}, \mathbf{v})$ is called the superpoly of t_I and can be computed as $p(\mathbf{k}, \mathbf{v}) = \sum_{\mathbf{v} \in C_I} f(\mathbf{k}, \mathbf{v})$
- ▶ Cube attack
 - Offline phase: Fix non-cube variables to constants (usually $\mathbf{0}$) and recover the expression of the superpoly
 - Online phase: Calculate the value of the superpoly and establish an equation with respect to \mathbf{k} . The attacker can extract some information about \mathbf{k} by solving the equation.

¹I. Dinur, A. Shamir: Cube Attacks on Tweakable Black Box Polynomials. EUROCRYPT 2009

Previous Techniques for Superpoly Recovery

▶ Experimental methods

- Only linear or quadratic superpolies are targeted

▶ Division Property

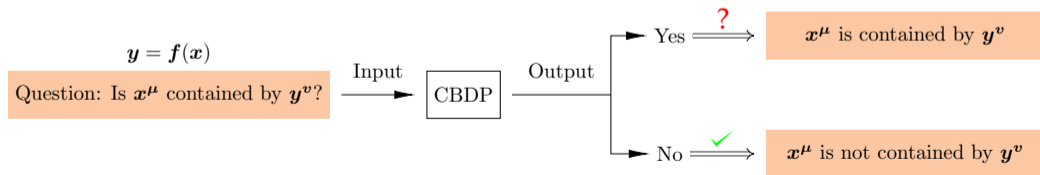
- Todo et al. proposed the division property at EUROCRYPT 2015
- Todo et al. introduced the division property into the cube attack at CRYPTO 2017
- Wang et al. recovered the exact superpoly using three-subset division property at ASIACRYPT 2019
- Hao et al. proposed the three-subset division property without the unknown subset (3SDPwoU) and utilized the Gurobi solver to recover the exact superpolies at EUROCRYPT 2020

▶ Methods based on monomial prediction

- Hu et al. reinterpreted the 3SDPwoU from the perspective of monomial propagation at ASIACRYPT 2020
- Hu et al. proposed a nested framework of monomial prediction to accelerate the recovery of superpoly at ASIACRYPT 2021
- He et al. proposed the core monomial prediction technique at ASIACRYPT 2022 as a scaled-down version of monomial prediction

Conventional Bit-based Division Property (CBDP)

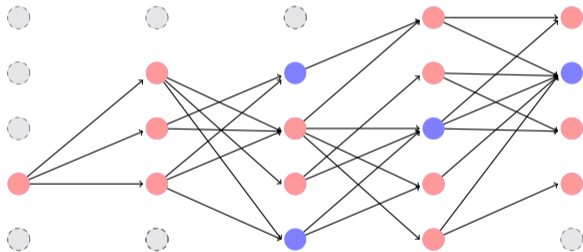
- ▶ CBDP was originally proposed as by Todo et al. at EUROCRYPT 2015 as a bit-level generalization of the integral property.
- ▶ At ASIACRYPT 2016, Xiang et al. introduced the concept of the division trail.
- ▶ At CRYPTO 2017, Todo et al. showed that CBDP can achieve perfect accuracy in claiming that a specific set of monomials of the input is not contained by the ANF of a output bit. However, CBDP can produce false positives.



$${}^1x^\mu = \prod_{u[i]=1} x[i]$$

Monomial Prediction (MP)

- ▶ For a composite Boolean function $\mathbf{y} = \mathbf{f}(\mathbf{x}) = \mathbf{f}^r \circ \dots \circ \mathbf{f}^0(\mathbf{x})$, the monomial prediction/three-subset division property without unknown subset (3SDPwoU) technique can predict if $\mathbf{x}^u \rightarrow \mathbf{y}^v$, by counting the number of monomial/division trails.



- ▶ Let $f(\mathbf{k}, \mathbf{v}) = p(\mathbf{k}, \mathbf{v}) \cdot t_I + q(\mathbf{k}, \mathbf{v})$. To recover p after fixing non-cube variables to $\mathbf{0}$, use the monomial prediction technique to find all possible $g(\mathbf{k})$ that satisfy $g(\mathbf{k}) \cdot t_I \rightarrow f$, then $p = \sum g(\mathbf{k})$.

Target cipher

- 1 Given an r -round cipher as

$$f(\mathbf{k}, \mathbf{v}) = f^{r-1} \circ f^{r-2} \circ \dots \circ f^0,$$

where f^i is the round function at round i with \mathbf{x}^i and \mathbf{x}^{i+1} being the input and output, respectively.

- 2 The initial state \mathbf{x}^0 is loaded with \mathbf{k} , \mathbf{v} , constant 0 bits and constant 1 bits. The output bit z is the sum of monomials of \mathbf{x}^r .
- 3 Choose cube indices I and set the non-cube variables to constants, we want to compute the superpoly of t_I in z .

¹Sometimes we use $\pi(\mathbf{x}^r, \mathbf{t}^r)$ to represent $\mathbf{x}^{r\mathbf{t}^r}$ for the sake of clarity

Initial Core Monomial Prediction (CMP)

Flag technique

For each bit b of a round state \mathbf{x}^i , we express it a polynomial g of cube variables and \mathbf{k} , and assign it an additional flag $b.F \in \{1_c, 0_c, \delta\}$.

- ▶ 0_c means g is constant 0.
- ▶ δ means g involves at least one cube variable.
- ▶ 1_c means g is non-zero and does not involve cube variables.

Example

Let $(k[0], k[1], v[0], v[1], v[2])$ be the input of f and cube indices $I = \{0, 1\}$. We set $v[2] = 0$.

- ▶ $g = k[0]k[1] + 1 \Rightarrow b.F = 1_c$
- ▶ $g = 0 \Rightarrow b.F = 0_c$
- ▶ $g = k[0]v[0] + k[1] \Rightarrow b.F = \delta$
- ▶ $g = k[1]v[2] + k[0] \Rightarrow b.F = 1_c$

Operations of flags

We define $=$, \oplus and \times operations for the elements of $\{1_c, 0_c, \delta\}$, corresponding to the basic operations COPY, XOR and AND, respectively:

- ▶ The $=$ operations sets one element equal to another element.
- ▶ The \oplus operations follows the rules:
 - $1_c \oplus 1_c = 1_c$
 - $0_c \oplus x = x \oplus 0_c = x$ for arbitrary $x \in \{1_c, 0_c, \delta\}$
 - $\delta \oplus x = x \oplus \delta = \delta$ for arbitrary $x \in \{1_c, 0_c, \delta\}$
- ▶ The \times operation follows the rules:
 - $1_c \times x = x \times 1_c = x$ for arbitrary $x \in \{1_c, 0_c, \delta\}$
 - $0_c \times x = x \times 0_c = 0_c$ for arbitrary $x \in \{1_c, 0_c, \delta\}$
 - $\delta \times \delta = \delta$

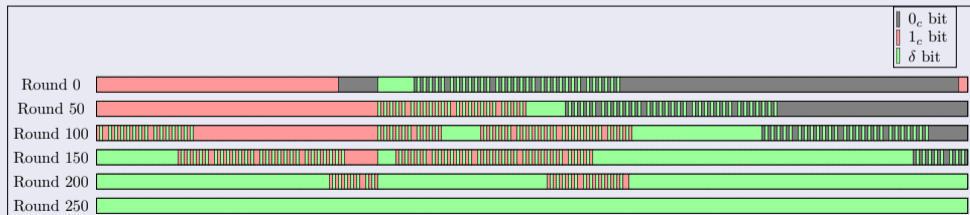
Remark: The flags can be computed round by round quickly without the help of the MILP solver.

Initial Core Monomial Prediction (CMP)

Example

This is the distribution of flags at different rounds of TRIVIUM for the cube indices

$$I = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 13, 15, 17, 19, 21, 24, 26, 28, 30, 32, 34, 36, 39, 41, 43, 45, 47, 49, 51, 54, 56, 58, 60, 62, 64, 66, 69, 71, 73, 75, 77, 79\}.$$



Initial Core Monomial Prediction (CMP)

Flag masks

For i -th round state \mathbf{x}^i , we use three binary vectors $\mathbf{M}^{i,\delta}$, $\mathbf{M}^{i,1_c}$, $\mathbf{M}^{i,0_c}$ to indicate the δ , 1_c , 0_c bits of \mathbf{x}^i , respectively.

$\mathbf{M}^{i,\delta}[j] = 1$ if and if only $x^i[j]$ is a δ bit.

$\mathbf{M}^{i,1_c}[j] = 1$ if and if only $x^i[j]$ is a 1_c bit.

$\mathbf{M}^{i,0_c}[j] = 1$ if and if only $x^i[j]$ is a 0_c bit.

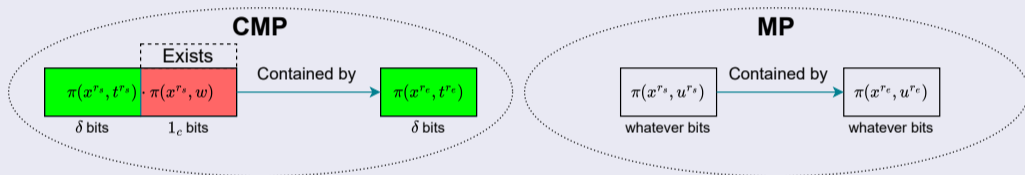
Example

Let $\mathbf{x}^i = (x[0], x[1], x[2], x[3])$ with $x[0].F = 0_c$, $x[1].F = x[2].F = \delta$, $x[3].F = 1_c$. Then, $\mathbf{M}^{i,\delta} = (0, 1, 1, 0)$, $\mathbf{M}^{i,1_c} = (0, 0, 0, 1)$, $\mathbf{M}^{i,0_c} = (1, 0, 0, 0)$. Obviously, $\mathbf{M}^{i,0_c} = \neg(\mathbf{M}^{i,\delta} \vee \mathbf{M}^{i,1_c})$.

Initial Core Monomial Prediction (CMP)

Definition

If $\pi(\mathbf{x}^{r_s}, \mathbf{t}^{r_s})$ can propagate to $\pi(\mathbf{x}^{r_e}, \mathbf{t}^{r_e})$ by CMP, we denote it by $\pi(\mathbf{x}^{r_s}, \mathbf{t}^{r_s}) \xrightarrow{\mathcal{C}} \pi(\mathbf{x}^{r_e}, \mathbf{t}^{r_e})$.



Core monomial trail

The connection of transitions $\pi(\mathbf{x}^{r_s}, \mathbf{t}^{r_s}) \xrightarrow{\mathcal{C}} \pi(\mathbf{x}^{r_s+1}, \mathbf{t}^{r_s+1}) \xrightarrow{\mathcal{C}} \dots \xrightarrow{\mathcal{C}} \pi(\mathbf{x}^{r_e}, \mathbf{t}^{r_e})$ is called an r' -round core monomial trail, where $r' = r_e - r_s$.

Initial Core Monomial Prediction (CMP)

General propagation rules and models

For a specific round function \mathbf{x}^i , we list all pairs $(\mathbf{t}^i, \mathbf{t}^{i+1})$ that satisfy $\pi(\mathbf{x}^i, \mathbf{t}^i) \xrightarrow{\mathcal{C}} \pi(\mathbf{x}^{i+1}, \mathbf{t}^{i+1})$, then we establish certain rules such that for any pair $(\mathbf{t}^i, \mathbf{t}^{i+1})$, the transition from \mathbf{t}^i to \mathbf{t}^{i+1} fulfills the rules. These rules can be encoded as linear inequalities suitable for MILP modeling.

Rules for basic operations and S-box

Using the general propagation rules and models, we can derive the propagation rules and models for basic operations COPY, AND, XOR and single S-box. This allowing us to construct an MILP model to capture the propagation of CMP through any cipher.

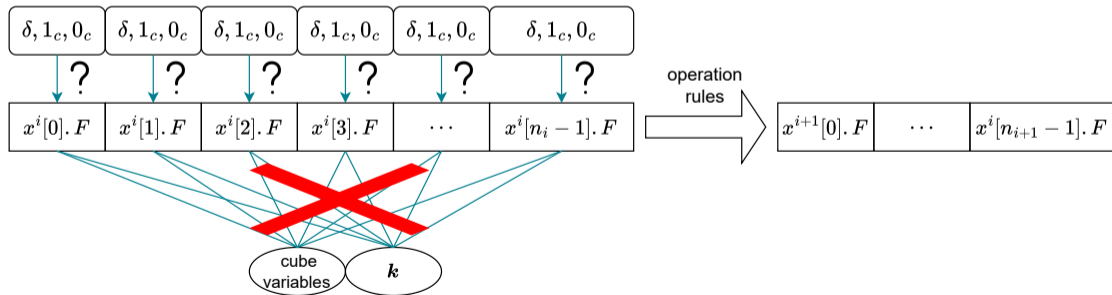
$$\mathbf{f}^3 \quad \mathbf{f}^2 \quad \mathbf{f}^1 \quad \mathbf{f}^0$$

$$\mathbf{f} = \text{XOR} \circ \text{AND} \circ \text{COPY} \circ \text{S-box}$$

Indeterminate Flags

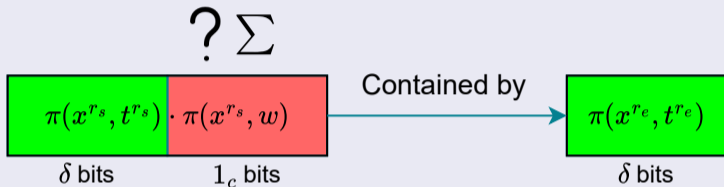
Indeterminate flag masks

The flags are no longer determined based on cube variables and k . They become variables that can take δ , 1_c or 0_c now, but operation rules still hold. If the flags of x^i are determined, then the flags of any round state after round i are determined. Flag masks become variables as well.

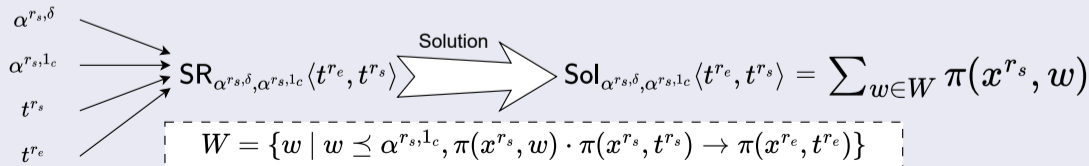


SR problem

SR problem



Instance of SR problem



Assigning different values to the flag masks can lead to different propagation of CMP.

Relation between SR problem and CMP

$\text{Sol}_{\alpha^{r_s, \delta}, \alpha^{r_s, 1c}} \langle \mathbf{t}^{r_e}, \mathbf{t}^{r_s} \rangle$ is not 0 if and only if $\pi(\mathbf{x}^{r_s}, \mathbf{t}^{r_s}) \xrightarrow{C} \pi(\mathbf{x}^{r_e}, \mathbf{t}^{r_e})$.

Reducing superpoly recovery to solving SR problem

Let $M^{0, \delta}$, $M^{0, 1c}$ and $M^{0, 0c}$ take the particular values $\gamma^{0, \delta}$, $\gamma^{0, 1c}$ and $\gamma^{0, 0c}$ respectively, where

$\gamma^{0, \delta}[j] = 1$ if and only if $x^0[j]$ is loaded with a cube variable.

$\gamma^{0, 1c}[j] = 1$ if and only if $x^0[j]$ is loaded with a constant 1 or a secret variable.

$\gamma^{0, 0c} = \neg(\gamma^{0, \delta} \vee \gamma^{0, 1c})$.

We compute the values of $M^{r, \delta}$, $M^{r, 1c}$, $M^{r, 0c}$ as $\gamma^{r, \delta}$, $\gamma^{r, 1c}$, $\gamma^{r, 0c}$. Then computing $\text{Coe} \langle z, t_I \rangle$ can be reduced to solving instances of the form $\text{SR}_{\gamma^{0, \delta}, \gamma^{0, 1c}} \langle \mathbf{t}^r, \mathbf{t}^0 \rangle$.

Solve SR Problem Based on CMP Theory

Contribution of core monomial trail

$$\text{Contr}\langle \ell \rangle = \text{Sol}\langle t^{r_s+1}, t^{r_s} \rangle \times \text{Sol}\langle t^{r_s+2}, t^{r_s+1} \rangle \times \dots \times \text{Sol}\langle t^{r_e-1}, t^{r_e-2} \rangle \times \text{Sol}\langle t^{r_e}, t^{r_e-1} \rangle$$

$$\ell = \boxed{\pi(x^{r_s}, t^{r_s})} \xrightarrow{c} \boxed{\pi(x^{r_s+1}, t^{r_s+1})} \xrightarrow{c} \boxed{\dots} \xrightarrow{c} \boxed{\pi(x^{r_e-1}, t^{r_e-1})} \xrightarrow{c} \boxed{\pi(x^{r_e}, t^{r_e})}$$

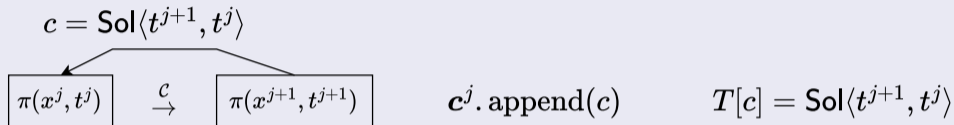
Solving SR problem by core monomial trails

Let $\mathbf{M}^{r_s, \delta}$, $\mathbf{M}^{r_s, 1_c}$ take arbitrary values $\alpha^{r_s, \delta}$, $\alpha^{r_s, 1_c}$ and $\mathbf{M}^{r_s, 0_c} = \neg(\alpha^{r_s, \delta} \vee \alpha^{r_s, 1_c})$. Given an instance $\text{SR}\langle t^{r_e}, t^{r_s} \rangle$, if $\pi(x^{r_s}, t^{r_s}) \stackrel{c}{\not\sim} \pi(x^{r_e}, t^{r_e})$, the solution of this instance is 0; otherwise, we can calculate the solution of this instance by

$$\text{Sol}\langle t^{r_e}, t^{r_s} \rangle = \sum_{\ell \in \pi(x^{r_s}, t^{r_s}) \stackrel{c}{\rightsquigarrow} \pi(x^{r_e}, t^{r_e})} \text{Expr}\langle \text{Contr}\langle \ell \rangle, x^{r_s} \rangle.$$

Equivalence Between CMP and MP

Iterative variable substitution algorithm



One-to-one correspondence between core monomial trail and monomial trail

$$\begin{aligned} \pi(\mathbf{x}^{r_s}, \mathbf{t}^{r_s}) &\xrightarrow{\mathcal{C}} \pi(\mathbf{x}^{r_{s+1}}, \mathbf{t}^{r_{s+1}}) \xrightarrow{\mathcal{C}} \dots \xrightarrow{\mathcal{C}} \pi(\mathbf{x}^{r_e}, \mathbf{t}^{r_e}) \Leftrightarrow \\ &\pi(\mathbf{c}^{r_s} \parallel \dots \parallel \mathbf{c}^{r_e-1} \parallel \mathbf{x}^{r_s}, \mathbf{w}^{r_s} \parallel \dots \parallel \mathbf{w}^{r_e-1} \parallel \mathbf{t}^{r_s}) \\ &\rightarrow \pi(\mathbf{c}^{r_s+1} \parallel \dots \parallel \mathbf{c}^{r_e-1} \parallel \mathbf{x}^{r_{s+1}}, \mathbf{w}^{r_{s+1}} \parallel \dots \parallel \mathbf{w}^{r_e-1} \parallel \mathbf{t}^{r_{s+1}}) \rightarrow \dots \\ &\rightarrow \pi(\mathbf{c}^{r_e-1} \parallel \mathbf{x}^{r_e-1}, \mathbf{w}^{r_e-1} \parallel \mathbf{t}^{r_e-1}) \\ &\rightarrow \pi(\mathbf{x}^{r_e}, \mathbf{t}^{r_e}). \end{aligned}$$

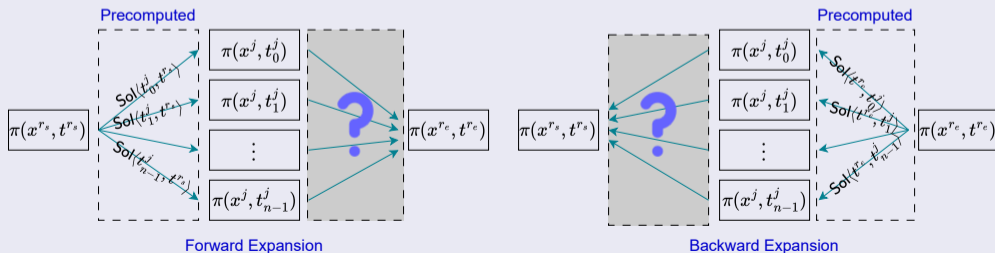
Forward and Backward Expansion

Divide-and-conquer

Given an instance $\text{SR}_{\alpha^{r_s, \delta}, \alpha^{r_s, 1_c}} \langle \mathbf{t}^{r_e}, \mathbf{t}^{r_s} \rangle$, we have

$$\text{Sol} \langle \mathbf{t}^{r_e}, \mathbf{t}^{r_s} \rangle = \sum_{\mathbf{t}^j} \text{Expr} \langle \text{Sol} \langle \mathbf{t}^{r_e}, \mathbf{t}^j \rangle, \mathbf{x}^{r_s} \rangle \cdot \text{Sol} \langle \mathbf{t}^j, \mathbf{t}^{r_s} \rangle,$$

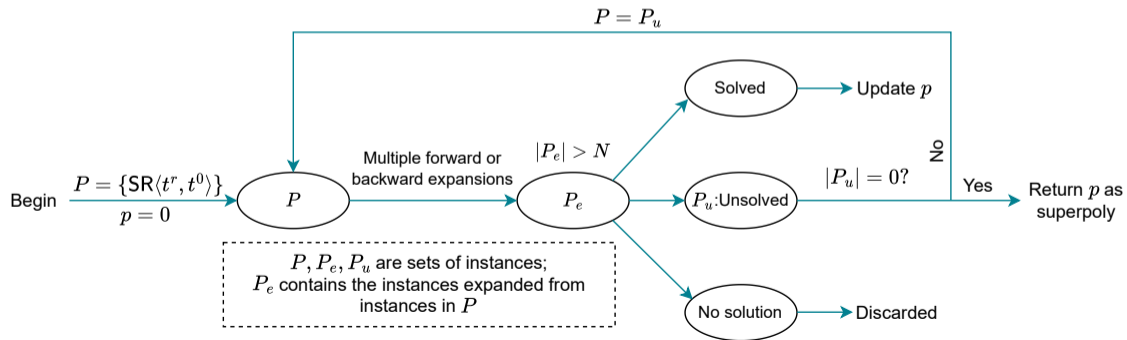
where the summation is over all \mathbf{t}^j 's that satisfy $\pi(\mathbf{x}^{r_s}, \mathbf{t}^{r_s}) \xrightarrow{C} \pi(\mathbf{x}^j, \mathbf{t}^j)$ and $\pi(\mathbf{x}^j, \mathbf{t}^j) \xrightarrow{C} \pi(\mathbf{x}^{r_e}, \mathbf{t}^{r_e})$.



Meet-in-the-middle (MITM) Framework

Applying forward and backward expansion interchangeably and recursively

Both the forward expansion and backward expansion divide an instance into multiple simpler instances. If they are applied interchangeably and recursively, we can develop a meet-in-the-middle framework.



New Results of Superpoly Recovery

- Summary of our new results of the superpoly recovery. #Cube means the number of cubes whose superpolies are recovered.

| Cipher | Rounds | #Cube | Cube size | Overall time cost | Balancedness | #Involved key bits |
|---------------|--------|-------|-----------|-------------------|--------------|--------------------|
| TRIVIUM | 849 | 2 | 44 | 76 hours | 0.50 | 80 |
| | 850 | 1 | 44 | 81 hours | 0.50 | 80 |
| | 851 | 1 | 44 | 600 hours | 0.50 | 80 |
| GRAIN-128AEAD | 192 | 1 | 94 | 9 days | 0.49 | 128 |
| Kreyvium | 896 | 1 | 123 | 180 hours | 0.50 | 30 |
| | 896 | 1 | 124 | 58 hours | 0 | 0 |
| | 897 | 1 | 124 | 67 hours | 0.50 | 60 |
| | 898 | 1 | 124 | 182 hours | 0.50 | 75 |
| | 899 | 1 | 124 | 272 hours | 0.50 | 121 |

- The balancedness is estimated with experiments on 2^{15} randomly-generated keys
- Conducted by Gurobi Solver (version 9.1.2) on a work station with $2 \times$ AMD EPYC 7302 16-core (32 siblings) Processor 3.3 GHz, (totally 64 threads), 256G RAM, and Ubuntu 20.10.

Key Recovery Attack

Based on memory-efficient variant of Möbius transform proposed by Dinur at EUROCRYPT 2021. For the equation $p(\mathbf{k}) = a$ established in the online phase of cube attack:

- 1 Guess m key bits.
- 2 Simplify p to a polynomial p' of unguessed key bits.
- 3 Perform Möbius transform on p' .
- 4 Exhaustive search.

| Cipher | Rounds | Type | Time | Memory (bits) | Reference |
|----------|--------|------|--------------|---------------|----------------------------|
| Trivium | 849 | Cube | 2^{79} | 2^{42} | This Work |
| | 850 | Cube | 2^{79} | 2^{46} | This Work |
| | 851 | Cube | 2^{79} | 2^{49} | This Work |
| Kreyvium | 896 | Cube | 2^{127} | Practical | This Work |
| | 897 | Cube | 2^{127} | Practical | This Work |
| | 898 | Cube | 2^{127} | Practical | This Work |
| | 899 | Cube | 2^{127} | Practical | This Work |
| | 900 | Cube | $2^{127.58}$ | Practical | Previous Work ² |

²Hao Fan, Yonglin Hao, Qingju Wang, Xinxin Gong, and Lin Jiao: Key filtering in cube attacks from the implementation aspect. CANS 2023

- ▶ Refine the CMP theory to be perfectly accurate.
- ▶ Superpoly recovery and cube attacks for up to 851 rounds of TRIVIUM and 899 rounds of Kreyvium.

Thanks for your attention!

hejiahui2020@mail.sdu.edu.cn