SuperBall: A New Approach for MILP Modelings of Boolean Functions

Ting Li[1], Yao Sun [1,2]

1. SKLOIS,Institute of Information Engineering, CAS
2.School of Cyber Security, University of Chinese Academy of Sciences

Mar. 2023

Outline

Backgrounds

MILP has been widely used in cryptanalysis:

- Differential attacks
- Impossible differential attacks
- Cube attacks
- ...

Backgrounds

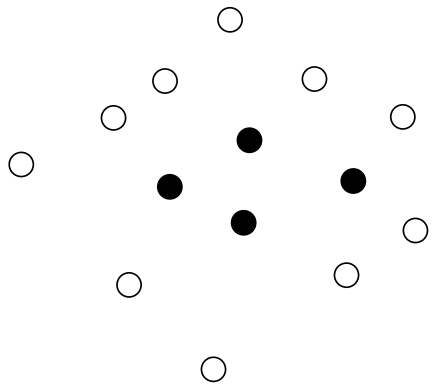
- DDT
- Linear layer
- Division property
- ……

→ Boolean function $F: \mathbb{F}_2^n \to \mathbb{F}_2$ → MILP Models

↓ Support of F: $\{v \in \mathbb{F}_2^n | F(v) = 1\}$

↓ Solve

= Solution



Backgrounds

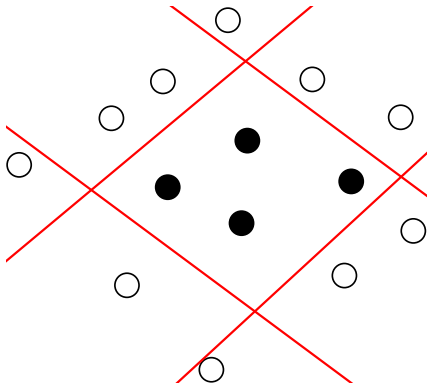
- DDT
- Linear layer
- Division property
- ……

→ Boolean function $F: \mathbb{F}_2^n \to \mathbb{F}_2$ → MILP Models

↓ Support of F: $\{v \in \mathbb{F}_2^n | F(v) = 1\}$

↓ Solve

= Solution



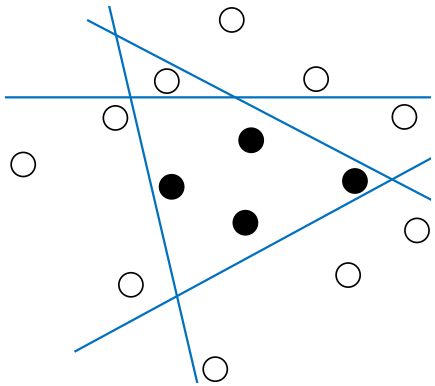
Backgrounds

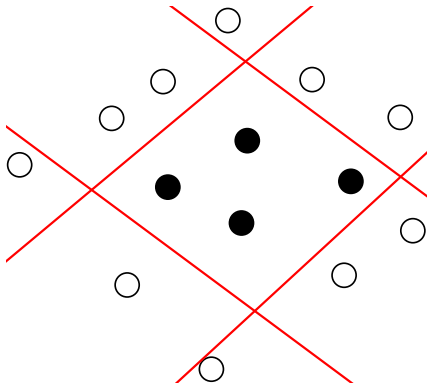
- DDT
- Linear layer
- Division property
- ……

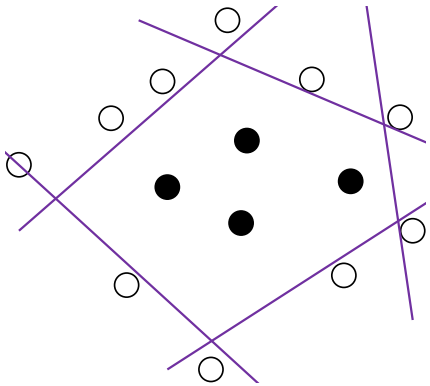
→ Boolean function $F: \mathbb{F}_2^n \to \mathbb{F}_2$ → MILP Models

↓ Support of F: $\{v \in \mathbb{F}_2^n | F(v) = 1\}$ = Solution

Solve







Motivations

Problem:
Which type of MILP model could achieve the best efficiency?

Resolution:

Step 1: Generating various inequalities for Boolean functions

Step 2: Construct various types of MILP models

Step 3: Compare the efficiency of MILP models

Contributions

1. Propose a novel approach to generate various inequalities.
2. Improve the Sasaki-Todo algorithm to construct the MILP model.
3. Find a type of model that has better efficiency.

Generating Inequalities of Boolean Functions

Previous works:

- H-representation of the convex hull (Sage Software) [SHW+14b]
- Logical condition modeling [SHW+14b]
- Product-of-sum representation of Boolean functions(Logic Friday Software) [AST+17]
- Improved logical condition modeling [BC20]

Generating Inequalities of Boolean Functions

Main idea 1: method of undetermined coefficients

Example 1

Given $S = \{1000, 1010, 0110, 1110, 1001, 0101, 1101\} \subset \mathbb{F}_2^4$, how to compute a polynomial $f = a_0x_0 + a_1x_1 + a_2x_2 + a_3x_3 + b$ s.t. $S \subset \mathrm{Sol}(f \geq 0)$?

$$1000\text{: } x_0 = 1, x_1 = x_2 = x_3 = 0.$$

$$\begin{cases} f(1000) = & a_0 + b & \geq 0, \\ f(1010) = & a_0 + a_2 + b & \geq 0, \\ f(0110) = & a_1 + a_2 + b & \geq 0, \\ f(1110) = & a_0 + a_1 + a_2 + b & \geq 0, \\ f(1001) = & a_0 + a_3 + b & \geq 0, \\ f(0101) = & a_1 + a_3 + b & \geq 0, \\ f(1101) = & a_0 + a_1 + a_3 + b & \geq 0. \end{cases} \tag{1}$$

Generating Inequalities of Boolean Functions

Main idea 1: method of undetermined coefficients

Example 2

Given $S = \{1000, 1010, 0110, 1110, 1001, 0101, 1101\} \subset \mathbb{F}_2^4$, how to compute a polynomial $f = a_0x_0 + a_1x_1 + a_2x_2 + a_3x_3 + b$ s.t. $S \subset \mathrm{Sol}(f \geq 0)$ and $\mathrm{Sol}(f < 0)$ is as large as possible?

$$
\begin{array}{lrll}
? & f(0000) = & b & < 0, \\
? & f(0100) = & a_1 + b & < 0, \\
? & f(1100) = & a_0 + a_1 + b & < 0, \\
? & f(0010) = & a_2 + b & < 0, \\
? & f(0001) = & a_3 + b & < 0, \\
? & f(0011) = & a_2 + a_3 + b & < 0, \\
? & f(1011) = & a_0 + a_2 + a_3 + b & < 0, \\
? & f(0111) = & a_1 + a_2 + a_3 + b & < 0, \\
? & f(1111) = & a_0 + a_1 + a_2 + a_3 + b & < 0.
\end{array}
\tag{2}
$$

Generating Inequalities of Boolean Functions

Main idea 1: method of undetermined coefficients

Example 2

Given $S = \{1000, 1010, 0110, 1110, 1001, 0101, 1101\} \subset \mathbb{F}_2^4$, how to compute a polynomial $f = a_0x_0 + a_1x_1 + a_2x_2 + a_3x_3 + b$ s.t. $S \subset \mathrm{Sol}(f \geq 0)$ and $\mathrm{Sol}(f < 0)$ is as large as possible?

$$f(v) - \gamma \cdot (1 - y_v) < 0, v \notin S. \tag{3}$$

$$y_v = \begin{cases} 0 \Rightarrow & f(v) < \gamma, \\ 1 \Rightarrow & f(v) < 0. \end{cases}$$

$$|\mathrm{Sol}(f < 0)| = \sum_{v \notin S} y_v$$

Generating Inequalities of Boolean Functions

Main idea 1: method of undetermined coefficients

Example 2

Given $S = \{1000, 1010, 0110, 1110, 1001, 0101, 1101\} \subset \mathbb{F}_2^4$, how to compute a polynomial $f = a_0x_0 + a_1x_1 + a_2x_2 + a_3x_3 + b$ s.t. $S \subset \mathrm{Sol}(f \geq 0)$ and $\mathrm{Sol}(f < 0)$ is as large as possible?

$$\text{Maximize} : \sum_{v \notin S} y_v,$$

s.t.

$$\begin{cases} f(v) \geq 0, & v \in S, \\ f(v) - \gamma \cdot (1 - y_v) < 0, & v \notin S. \end{cases}$$

$S \subset \mathbb{F}_2^n$, # Constraints = 2^n. If n is large, the system is hard to solve.

Generating Inequalities of Boolean Functions

Main idea 2: divide-and-conquer strategy

Example 2

Given $S = \{1000, 1010, 0110, 1110, 1001, 0101, 1101\} \subset \mathbb{F}_2^4$, how to compute a polynomial $f = a_0x_0 + a_1x_1 + a_2x_2 + a_3x_3 + b$ s.t. $S \subset \mathrm{Sol}(f \geq 0)$ and $\mathrm{Sol}(f < 0)$ is as large as possible?

Pattern 1: $a_2 \geq 0$, $a_3 \geq 0$,
Pattern 2: $a_2 < 0$, $a_3 \geq 0$,
Pattern 3: $a_2 \geq 0$, $a_3 < 0$,
Pattern 4: $a_2 < 0$, $a_3 < 0$.

Generating Inequalities of Boolean Functions

Main idea 2: divide-and-conquer strategy

Example 2

Given $S = \{1000, 1010, 0110, 1110, 1001, 0101, 1101\} \subset \mathbb{F}_2^4$, how to compute a polynomial $f = a_0x_0 + a_1x_1 + a_2x_2 + a_3x_3 + b$ s.t. $S \subset \mathrm{Sol}(f \geq 0)$ and $\mathrm{Sol}(f < 0)$ is as large as possible?

Pattern 1: $a_2 \geq 0$, $a_3 \geq 0$

$$\begin{cases} f(1000) = & a_0 + b & \geq 0, \\ f(1010) = & a_0 + a_2 + b & \geq 0, \rightarrow \textit{useless} \\ f(0110) = & a_1 + a_2 + b & \geq 0, \\ f(1110) = & a_0 + a_1 + a_2 + b & \geq 0, \\ f(1001) = & a_0 + a_3 + b & \geq 0, \rightarrow \textit{useless} \\ f(0101) = & a_1 + a_3 + b & \geq 0, \\ f(1101) = & a_0 + a_1 + a_3 + b & \geq 0. \end{cases} \tag{4}$$

$B = \{1000, 0110, 1110, 0101, 1101\} \subset S$

Generating Inequalities of Boolean Functions

Main idea 2: divide-and-conquer strategy

Example 2

Given $S = \{1000, 1010, 0110, 1110, 1001, 0101, 1101\} \subset \mathbb{F}_2^4$, how to compute a polynomial $f = a_0x_0 + a_1x_1 + a_2x_2 + a_3x_3 + b$ s.t. $S \subset \mathrm{Sol}(f \geq 0)$ and $\mathrm{Sol}(f < 0)$ is as large as possible?

Pattern 1: $a_2 \geq 0$, $a_3 \geq 0$

$$
\begin{array}{lll}
? & f(0000) = & b & < 0, \\
? & f(0100) = & a_1 + b & < 0, \\
? & f(1100) = & a_0 + a_1 + b & < 0, \\
? & f(0010) = & a_2 + b & < 0, \\
? & f(0001) = & a_3 + b & < 0, \\
? & f(0011) = & a_2 + a_3 + b & < 0, \\
\times & f(1011) = & a_0 + a_2 + a_3 + b & < 0, \\
\times & f(0111) = & a_1 + a_2 + a_3 + b & < 0, \\
\times & f(1111) = & a_0 + a_1 + a_2 + a_3 + b & < 0.
\end{array}
\tag{5}
$$

$R = \{0000, 0100, 1100, 0010, 0001, 0011\} \subset \mathbb{F}_2^4 \setminus S$

Generating Inequalities of Boolean Functions

Main idea 2: divide-and-conquer strategy

Example 2

Given $S = \{1000, 1010, 0110, 1110, 1001, 0101, 1101\} \subset \mathbb{F}_2^4$, how to compute a polynomial $f = a_0x_0 + a_1x_1 + a_2x_2 + a_3x_3 + b$ s.t. $S \subset \mathrm{Sol}(f \geq 0)$ and $\mathrm{Sol}(f < 0)$ is as large as possible?

Pattern 1: $a_2 \geq 0$, $a_3 \geq 0$

$$\text{Maximize} : \sum_{v \in R} y_v,$$

$$s.t. \begin{cases} f(v) \geq 0, & v \in B, \\ f(v) - \gamma \cdot (1 - y_v) < 0, & v \in R. \end{cases}$$

$$\#\ \text{Constraints} = |B| + |R| = 5 + 6 = 11 < 2^4.$$

Generating Inequalities of Boolean Functions

Generate various inequalities:

- Compute a polynomial s.t. $\mathrm{Sol}(f = 0)$ is as large as possible:

$$\text{Maximize} : \sum_{v \in R} z_v,$$

$$s.t. \begin{cases} f(v) \geq 0, & v \in B, \\ f(v) - \gamma \cdot (1 - z_v) \leq 0, & v \in S. \end{cases}$$

- Compute the sparsest polynomial: $a_j \geq 0$:

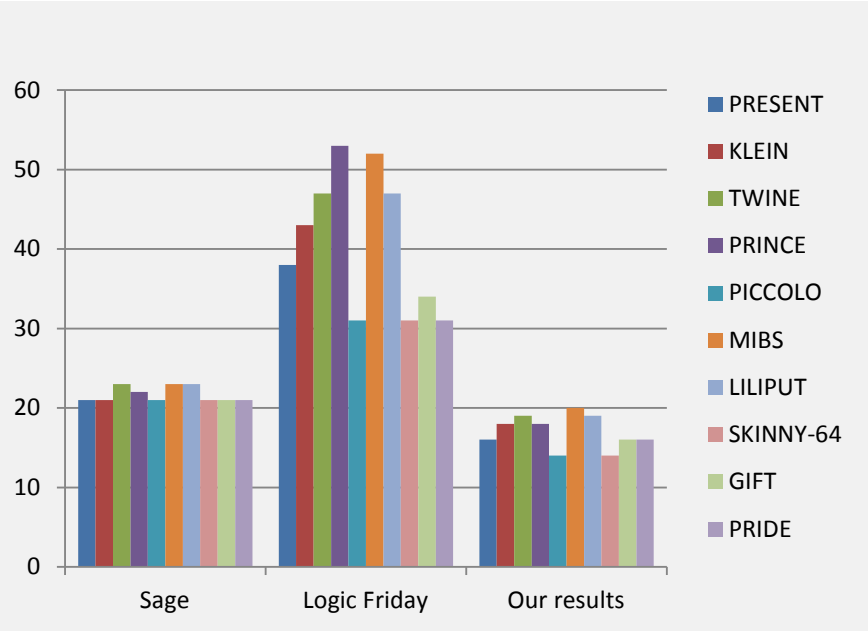
$$\text{Maximize} : \sum w_j,$$

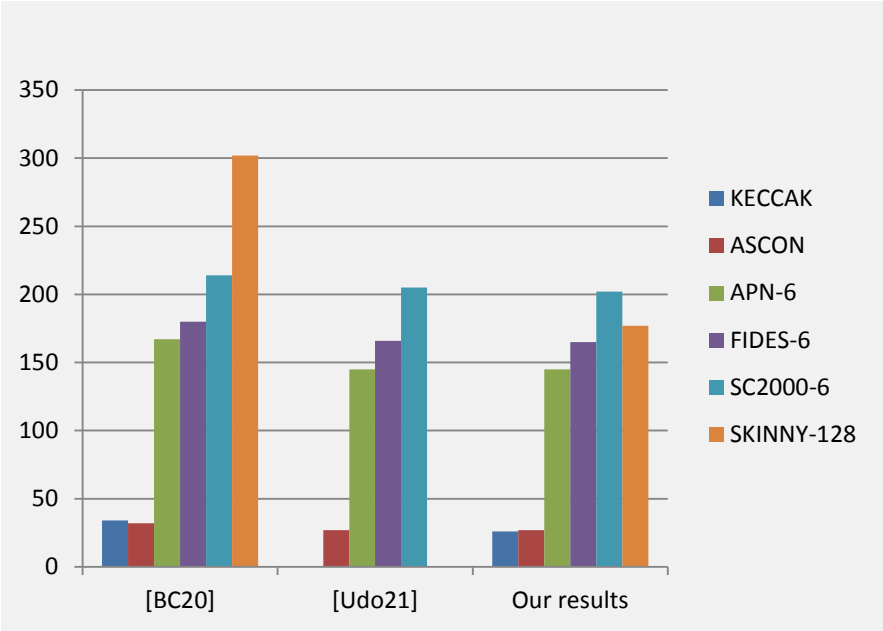
$$s.t. \begin{cases} f(v) \geq 0, & v \in B, \\ a_j \geq 1 - w_j \text{ and } a_j \leq R \cdot (1 - w_j). \end{cases}$$

Generating Inequalities of Boolean Functions

The size of the minimal model could reflect the diversity of the set of candidate inequalities.

Comparisons of the minimal sizes of models. Boolean functions are deduced from Sboxes.





Constructing an MILP Model

Sasaki-Todo method [ST17a]

$S = \{v_4, v_5, v_6, v_7\}$.

	d_1	d_2	d_3	d_4	d_5	d_6
	$f_1 \geq 0$	$f_2 \geq 0$	$f_3 \geq 0$	$f_4 \geq 0$	$f_5 \geq 0$	$f_6 \geq 0$
v_0	0	1	0	1	1	0
v_1	1	0	1	0	0	0
v_2	0	0	1	1	1	0
v_3	1	0	0	0	0	1
v_4	0	0	0	0	0	0
v_5	0	0	0	0	0	0
v_6	0	0	0	0	0	0
v_7	0	0	0	0	0	0

$Minimize: \sum d_i$

$d_2 + d_4 + d_5 \geq 1$

$d_1 + d_3 \geq 1$

$d_3 + d_4 + d_5 \geq 1$

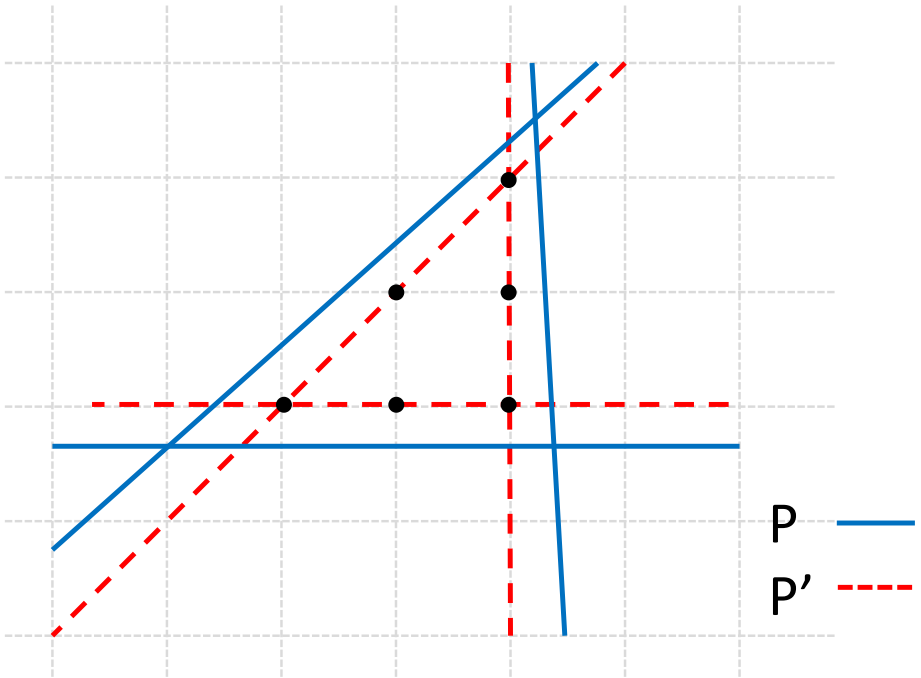
$d_1 + d_6 \geq 1$

$0: f_i(v_j) \geq 0; \quad 1: f_i(v_j) < 0;$

The model with the minimal size dose not always lead to the best efficiency.

Constructing an MILP Model

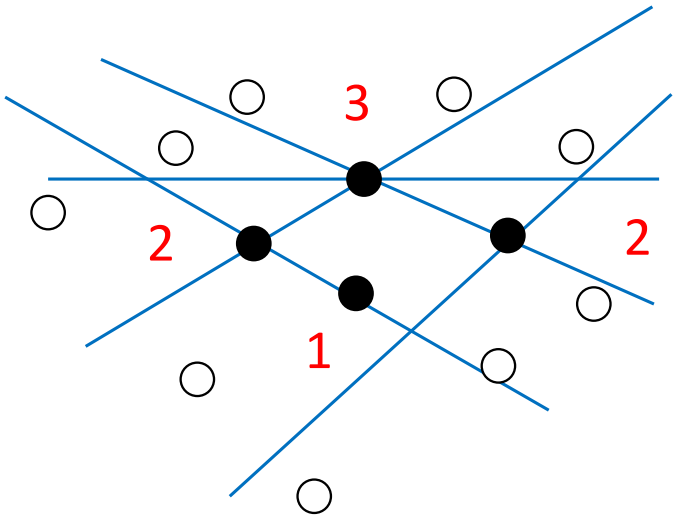
An efficient model should balance size and strength. [Vie15]
Strength : search space explored for optimal integer solution.



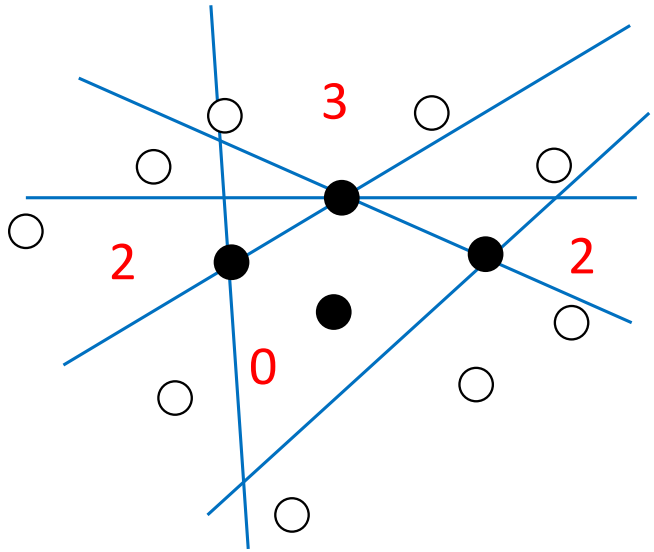
Constructing an MILP Model

Strength is hard to calculate.

- Approximate strength (A.S.) :
  $\min_{v \in B} |\{f_i \mid v \in \mathrm{Sol}(f_i = 0), 0 \le i \le m\}|$
- Cover rate (C.Rate) : $|(\cup_{0 \le i \le m} \mathrm{Sol}(f_i = 0)) \cap B| / |B|$.



A.S. = 1
C.Rate = 1

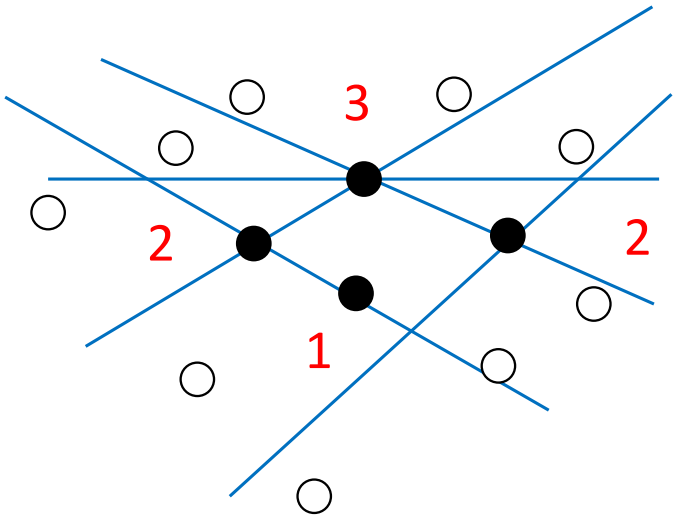


A.S. = 0
C.Rate = 0.75

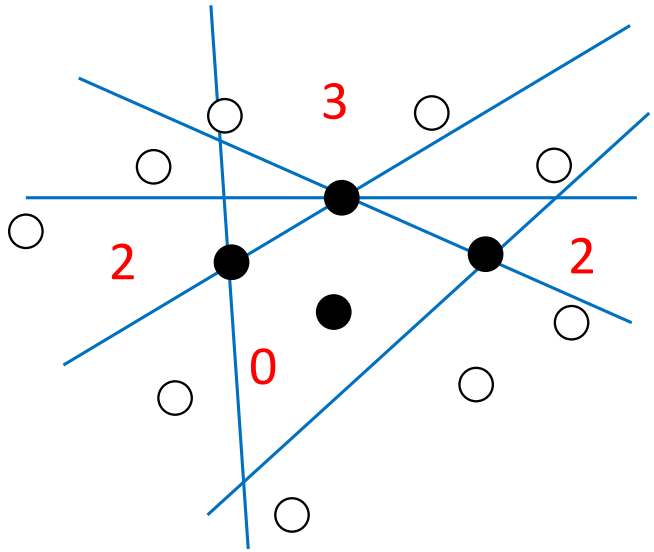
Constructing an MILP Model

Strength is hard to calculate.

- Approximate strength (A.S.) :
  $\min_{v \in B} |\{f_i \mid v \in \mathrm{Sol}(f_i = 0), 0 \leq i \leq m\}|$
- Cover rate (C.Rate) : $|(\cup_{0 \leq i \leq m} \mathrm{Sol}(f_i = 0)) \cap B| / |B|$.



A.S. = 1
C.Rate = 1



A.S. = 0
C.Rate = 0.75

Constructing an MILP Model

The improved Sasaki-Todo method computes the model with a minimal size among the models with an approximate strength being at least s.

$S = \{v_4, v_5, v_6, v_7\}$.

	d_1	d_2	d_3	d_4	d_5	d_6	
	$f_1 \geq 0$	$f_2 \geq 0$	$f_3 \geq 0$	$f_4 \geq 0$	$f_5 \geq 0$	$f_6 \geq 0$	$Minimize: \sum d_i$
v_0	0	1	0	1	1	0	$d_2 + d_4 + d_5 \geq 1$
v_1	1	0	1	0	0	0	$d_1 + d_3 \geq 1$
v_2	0	0	1	1	1	0	$d_3 + d_4 + d_5 \geq 1$
v_3	1	0	0	0	0	1	$d_1 + d_6 \geq 1$
v_4	$0^\star$	$\bar{0}$	$0^\star$	$0^\star$	$0^\star$	$\bar{0}$	$d_2 + d_6 \geq s$
v_5	$0^\star$	$0^\star$	$\bar{0}$	$0^\star$	$0^\star$	$\bar{0}$	$d_3 + d_6 \geq s$
v_6	$0^\star$	$0^\star$	$0^\star$	$\bar{0}$	$\bar{0}$	$0^\star$	$d_4 + d_5 \geq s$
v_7	$\bar{0}$	$\bar{0}$	$0^\star$	$0^\star$	$0^\star$	$0^\star$	$d_1 + d_2 \geq s$

$1: f_i(v_j) < 0;\ \bar{0}: f_i(v_j) = 0; 0^\star: f_i(v_j) > 0;$

Constructing an MILP Model

When $s = 0$, this method computes a model with the minimal size, and the cover rate of this model is maximal in this size.

$S = \{v_4, v_5, v_6, v_7\}$.

	d_1	d_2	d_3	d_4	d_5	d_6
	$f_1 \geq 0$	$f_2 \geq 0$	$f_3 \geq 0$	$f_4 \geq 0$	$f_5 \geq 0$	$f_6 \geq 0$
v_0	0	1	0	1	1	0
v_1	1	0	1	0	0	0
v_2	0	0	1	1	1	0
v_3	1	0	0	0	0	1
v_4	$0^\star$	$0^\star$	$0^\star$	$0^\star$	$0^\star$	$0^\star$
v_5	$0^\star$	$0^\star$	$\bar{0}$	$0^\star$	$0^\star$	$\bar{0}$
v_6	$0^\star$	$0^\star$	$0^\star$	$\bar{0}$	$\bar{0}$	$0^\star$
v_7	$\bar{0}$	$\bar{0}$	$0^\star$	$0^\star$	$0^\star$	$0^\star$

Minimize: $8 \cdot \sum d_i - \sum c_i$

$d_2 + d_4 + d_5 \geq 1$

$d_1 + d_3 \geq 1$

$d_3 + d_4 + d_5 \geq 1$

$d_1 + d_6 \geq 1$

$d_3 + d_6 \geq c_5$

$d_4 + d_5 \geq c_6$

$d_1 + d_2 \geq c_7$

$1: f_i(v_j) < 0;\ \bar{0}: f_i(v_j) = 0; 0^\star: f_i(v_j) > 0;$

Experiments

Models:

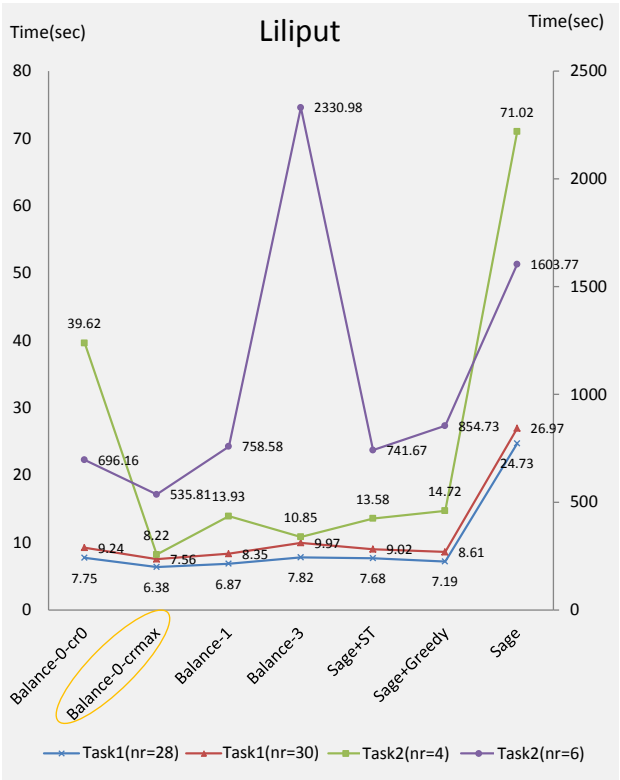
- Balance-*s* (This paper).
- Sage+ST(Sasaki-Todo algorithm) [ST17a]
- Sage+Greedy [SHW+14a]
- Sage
- Espresso [AST+17]

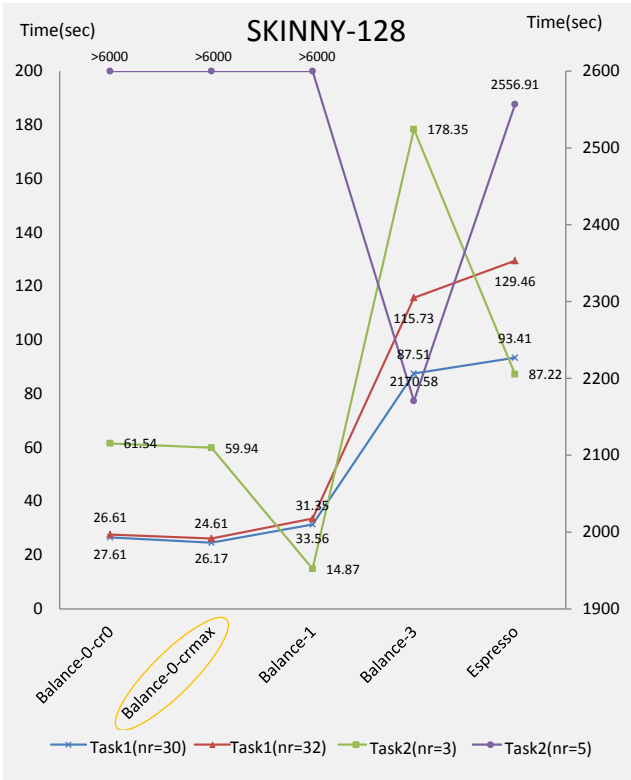
	LILIPUT			SKINNY		
	Size	A.S.	C.Rate	Size	A.S.	C.Rate
Balance-0-cr0	19	0	0	177	0	0
Balance-0-crmax	19	0	93/106	177	0	6420/11469
Balance-1	22	1	1	189	1	1
Balance-3	34	3	1	367	3	1
Sage+ST	23	0	101/106	-	-	-
Sage+Greedy	26	0	102/106	-	-	-
Sage	324	12	1	-	-	-
Espresso	-	-	-	377	4	1

Experiments

Two traditional search tasks:

1. Verification of differential pairs(Liliput:nr =28/30, SKINNY-128:nr=30/32).
2. Finding the minimal number of active Sboxes(Liliput:nr = 4/6, SKINNY-128:nr=3/5).





Thanks! Any Questions ?