MILP Modeling for (Large) S-boxes to Optimize Probability of Differential Characteristics

Ahmed Abdelkhalek[1], **Yu Sasaki**[2], Yosuke Todo[2], Mohamed Tolba[1], and Amr M. Youssef[1]

1:Concordia University, 2: NTT

Talk @ ASK2017, 10 December 2017

Summary

New MILP model for 8-bit S-boxes

- New method to model truncated DDT
- New method to evaluate probability in DDT

Applications

- *SKINNY-128*: the max diff prob reaches 2^{-128} with 14 rounds (prev. 15 rounds)
- *AES-round based Func from FSE2016*: improved the max probability of diff trail

MILP for Differential Cryptanalysis

Mouha et al. at Inscrypt 2011:

Problem of finding optimal differential trail → (convert) → Optimization problem in MILP

Advantage:

Speed of solving MILP has been researched a lot. We can exploit their effort to search for differential propagation trails.

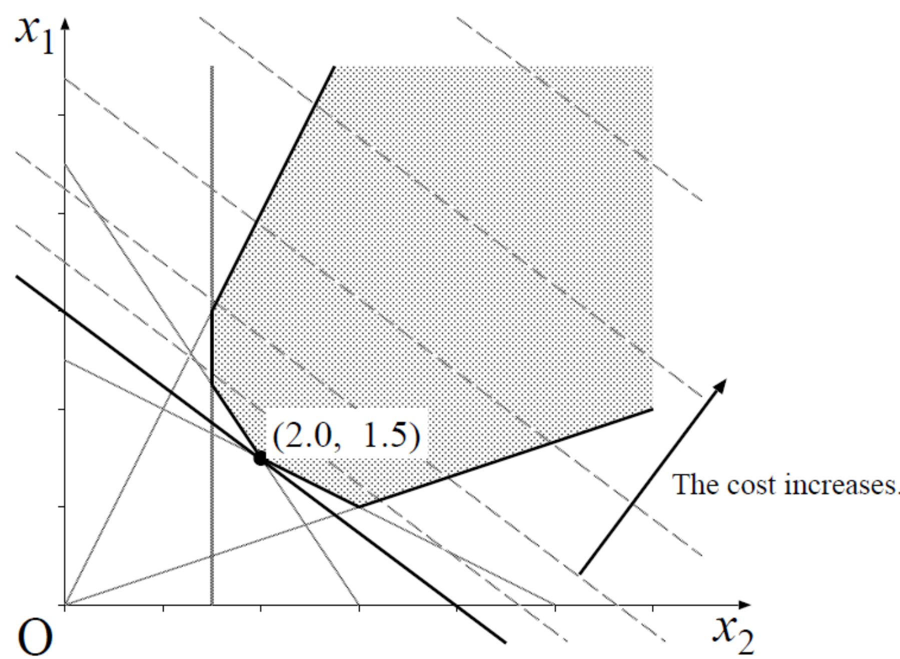
Mixed Integer Linear Programming (MILP)

Optimize objective function within the solution range satisfying all the constraints.

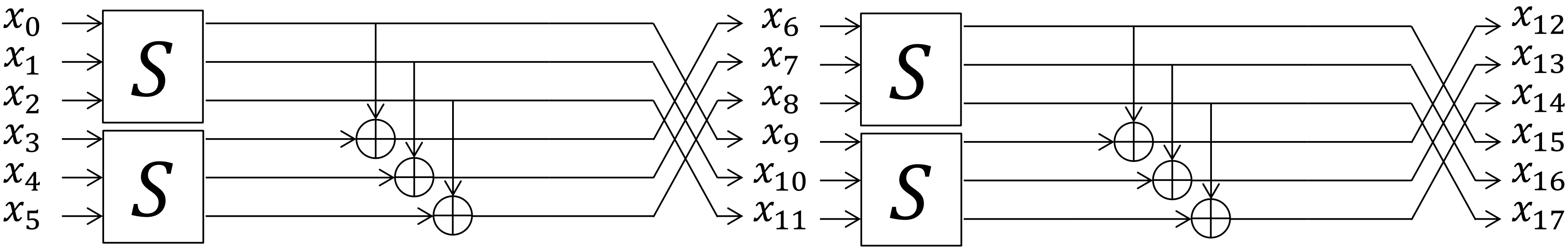
Minimize $50x_1 + 65x_2$

Constraints

$$
\begin{cases}
3x_1 + 2x_2 \geq 9 \\
\frac{1}{15}x_1 + \frac{2}{15}x_2 \geq \frac{1}{3} \\
\frac{1}{6}x_1 \geq \frac{1}{4} \\
x_1 - 3x_2 \leq 0 \\
2x_1 - x_2 \geq 0 \\
x_1 \geq 0 \\
x_2 \geq 0
\end{cases}
$$



MILP Model for 3-Round Toy Cipher



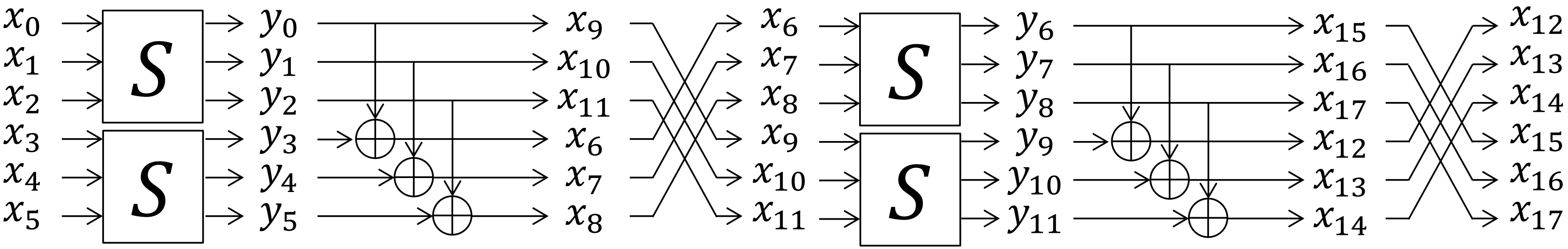
6-bit round function: 3-bit S-box, 3-bit xor, swap

To make the MILP model, define a binary variable $x_i \in \{0,1\}$ for each round;

- $x_i = 0$ denotes the bit i has no difference
- $x_i = 1$ denotes the bit i has difference

Objective Function | Minimize: $x_0 + x_1 + \cdots + x_{6r-1}$

Constraints for Linear Operations

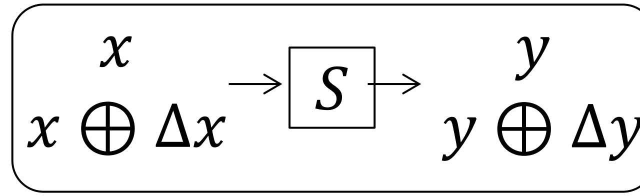


$a \oplus b = c$ can be modeled with 4 inequalities by removing each impossible (a, b, c).

$$(a,b,c) \neq (0,0,1) \Leftarrow \quad a + b - c \geq 0$$
$$(a,b,c) \neq (0,1,0) \Leftarrow \quad a - b + c \geq 0$$
$$(a,b,c) \neq (1,0,0) \Leftarrow -a + b + c \geq 0$$
$$(a,b,c) \neq (1,1,1) \Leftarrow -a - b - c \geq -2$$

Differential Distribution Table (DDT)

We compute the probability that Δx propagates to Δy for each $(\Delta x, \Delta y)$.



Input Difference (Δx)	Output Difference (Δy)							
	0x0	0x1	0x2	0x3	0x4	0x5	0x6	0x7
0x0	1	0	0	0	0	0	0	0
0x1	0	2^{-2}	2^{-2}	0	0	2^{-2}	2^{-2}	0
0x2	0	2^{-2}	2^{-2}	0	0	2^{-2}	2^{-2}	0
0x3	0	0	0	2^{-1}	0	0	0	2^{-1}
0x4	0	0	0	0	2^{-1}	0	0	2^{-1}
0x5	0	2^{-2}	2^{-2}	0	0	2^{-2}	2^{-2}	0
0x6	0	2^{-2}	2^{-2}	0	0	2^{-2}	2^{-2}	0
0x7	0	0	0	2^{-1}	2^{-1}	0	0	0

Truncated DDT ($*$-DDT)

To count the # of active S-boxes, we only care whether each pattern is possible (non-zero probability) or impossible (zero probability). We call it "$*$-DDT".

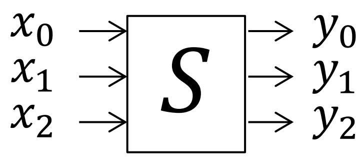
Input Difference (Δx)	Output Difference (Δy)							
	0x0	0x1	0x2	0x3	0x4	0x5	0x6	0x7
0x0	1	0	0	0	0	0	0	0
0x1	0	**1**	**1**	0	0	**1**	**1**	0
0x2	0	**1**	**1**	0	0	**1**	**1**	0
0x3	0	0	0	**1**	0	0	0	**1**
0x4	0	0	0	0	**1**	0	0	**1**
0x5	0	**1**	**1**	0	0	**1**	**1**	0
0x6	0	**1**	**1**	0	0	**1**	**1**	0
0x7	0	0	0	**1**	**1**	0	0	0

Two Methods of Modeling $*$-DDT

	H-representation of convex hull		**Logical condition model (Sun et al.)**	
tool	SAGE Math		N/A	
support alg	greedy	Sub MILP	greedy	Sub MILP
type	heuristic	optimal	heuristic	optimal
coefficients	any integer		{-1, 0, 1}	
#inequ.	small		large	
8-bit S-box	infeasible		?	

Our Focus

Logical Condition Model for S-box



$*$-DDT tells impossible patterns of $(x_2x_1x_0y_2y_1y_0)$.
Each impossible pattern can be removed one inequality.

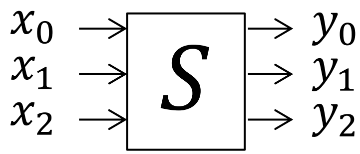
Example: $Pr[(\Delta_i, \Delta_o) = (0x1, 0x2)] = 0$

$$x_2x_1x_0 = 001, \quad y_2y_1y_0 = 010$$

$$x_2 + x_1 - x_0 + y_2 - y_1 + y_0 \geq -1$$

Out of 64 entries of $*$-DDT, about 32 entries are impossible. Each S-box can be modeled with about 32 inequalities.

Reducing the Number of Inequalities



Sun et al. pointed out that several impossible patterns of $(x_2x_1x_0y_2y_1y_0)$ can be removed simultaneously.

Example:

$$Pr[(\Delta_i,\Delta_o) = (0x1{,}0x2)] = Pr[(\Delta_i,\Delta_o) = (0x1{,}0x6)] = 0$$

$$x_2x_1x_0y_2y_1y_0 = 001\mathbf{0}10$$

$$x_2x_1x_0y_2y_1y_0 = 001\mathbf{1}10$$

$$x_2 + x_1 - x_0 - y_1 + y_0 \geq -1$$

Each S-box can be modeled with less than 32 inequalities.

NTT

Two Issues of the Previous S-box Model

1. The number of constraints for each S-box is exponential to the S-box size.
   - 5-bit to 5-bit S-box: feasible
   - 6-bit to 4-bit S-box: feasible
   - 8-bit to 8-bit S-box: infeasible (folklore)

2. Probability of differential transition is ignored.

   An attempt was proposed by Sun et al. in 2014:
   - feasible only up to 4-bit to 4-bit S-box
   - Probability must be 2^{-x} where x is an integer.

New Method to Model $*$-DDT

Core Observation

Minimizing constraints for ∗-DDT

related ↔

Finding the minimum product-of-sum representation of a Boolean function

a well-studied topic!!

NTT

$*$-DDT to Product-of-Sum Representation

- Define a $2n$-bit to 1-bit Boolean function that outputs 1 only when the propagation is possible.
- This can be achieved by listing impossible propagations as a term of product-of-sum or the Conjunctive Normal Form (CNF)
- Indeed, for f to be 1, even a single term must not be 0, i.e. $2n$ variables must avoid impossible patterns.

$$
\begin{aligned}
f(x_2, x_1, x_0, y_2, y_1, y_0) \\
&= (x_2 \vee x_1 \vee x_0 \vee y_2 \vee y_1 \vee \overline{y_0}) \wedge (x_2 \vee x_1 \vee x_0 \vee y_2 \vee \overline{y_1} \vee y_0) \\
&\wedge (x_2 \vee x_1 \vee x_0 \vee y_2 \vee \overline{y_1} \vee \overline{y_0}) \wedge (x_2 \vee x_1 \vee x_0 \vee \overline{y_2} \vee y_1 \vee y_0) \wedge \\
&\cdots \\
&\wedge (\overline{x_2} \vee \overline{x_1} \vee \overline{x_0} \vee \overline{y_2} \vee \overline{y_1} \vee y_0) \wedge (\overline{x_2} \vee \overline{x_1} \vee \overline{x_0} \vee \overline{y_2} \vee \overline{y_1} \vee \overline{y_0})
\end{aligned}
$$

QM, Espresso and LogicFriday

- Finding min. representation of product-of-sum (NP-hard) is well studied in computer science.
- Quine-McCluskey algorithm [Qui52,Qui55,McC56] provides optimal solution and the Espresso algorithm is the heuristic algorithm.
- The freeware called LogicFriday can execute both QM and Espresso.

inequalities to represent $*$-DDT of 8-bit S-boxes

Structure	# non-zero entries	QM	Espresso
AES S-box	33150	-	8302
SKINNY-128 S-box	54067	372	376

Demo

Generating constraints for $*$-DDT of PRESENT S-box by using Logic Friday

Summary for Modeling $*$-DDT

	H-representation of convex hull		Logical condition model (Sun et al.)
tool	SAGE Math		LogicFriday < QM espresso
aux alg	greedy	Sub MILP	no need
type	heuristic	optimal	
coefficients	any integer		{-1, 0, 1}
#inequ.	small		large
8-bit S-box	infeasible		feasible

New Methods to Evaluate Probability

Core Observation

- Separate DDT to multiple tables so that each table contains entries with the same probability.

$$pb\text{-DDT} \begin{cases} 1 & \text{if the entry in DDT has probability } pb \\ 0 & \text{otherwise} \end{cases}$$

- Use conditional constraints (with the big-M method) to activate only a single pb-DDT.

Input Difference (Δx)	Output Difference (Δy)							
	0x0	0x1	0x2	0x3	0x4	0x5	0x6	0x7
0x0	1	0	0	0	0	0	0	0
0x1	0	2^{-2}	2^{-2}	0	0	2^{-2}	2^{-2}	0
0x2	0	2^{-2}	2^{-2}	0	0	2^{-2}	2^{-2}	0
0x3	0	0	0	2^{-1}	0	0	0	2^{-1}
0x4	0	0	0	0	2^{-1}	0	0	2^{-1}
0x5	0	2^{-2}	2^{-2}	0	0	2^{-2}	2^{-2}	0
0x6	0	2^{-2}	2^{-2}	0	0	2^{-2}	2^{-2}	0
0x7	0	0	0	2^{-1}	2^{-1}	0	0	0

DDT

2^{-1}-DDT

Δx	Δy							
	0x0	0x1	0x2	0x3	0x4	0x5	0x6	0x7
0x0	0	0	0	0	0	0	0	0
0x1	0	0	0	0	0	0	0	0
0x2	0	0	0	0	0	0	0	0
0x3	0	0	0	1	0	0	0	1
0x4	0	0	0	0	1	0	0	1
0x5	0	0	0	0	0	0	0	0
0x6	0	0	0	0	0	0	0	0
0x7	0	0	0	1	1	0	0	0

2^{-2}-DDT

Δx	Δy							
	0x0	0x1	0x2	0x3	0x4	0x5	0x6	0x7
0x0	0	0	0	0	0	0	0	0
0x1	0	1	1	0	0	1	1	0
0x2	0	1	1	0	0	1	1	0
0x3	0	0	0	0	0	0	0	0
0x4	0	0	0	0	0	0	0	0
0x5	0	1	1	0	0	1	1	0
0x6	0	1	1	0	0	1	1	0
0x7	0	0	0	0	0	0	0	0

Experimental Data for pb-DDT

Structure		Num. of zero entries	QM	Espresso
AES S-box	2^{-7}	33406	-	8241
	2^{-6}	65281	-	350
SKINNY-128 S-box	2^{-7}	62848	206	208
	2^{-6}	60530	275	283
	$2^{-5.4}$	65472	33	34
	2^{-5}	62708	234	239
	$2^{-4.4}$	65458	42	52
	2^{-4}	64884	147	159
	$2^{-3.7}$	65534	15	15
	$2^{-3.4}$	65518	24	28
	$2^{-3.2}$	65534	15	15
	2^{-3}	65435	62	67
	$2^{-2.7}$	65534	16	16
	$2^{-2.4}$	65532	17	17
	2^{-2}	65513	37	40

Representing Probability of each S-box

Activeness variable

- n_i : 1 if the i-th Sbox is active, 0 otherwise.

Probability Variables

- Q_{i,pb_j}: 1 if the i-th Sbox is active and its differential probability is pb_j, 0 otherwise.

E.g. $Q_{i,2^{-1}}$ and $Q_{i,2^{-2}}$ in the above 3-bit S-box.

The probability when the i-th S-box is active is modeled by

$$\sum_j Q_{i,pb_j} = n_i$$

E.g. $Q_{i,2^{-1}} + Q_{i,2^{-2}} = n_i$

Objective Function

minimize $\sum_{i,j} -\log(pb_j) \times Q_{i,pb_j}$ E.g. $\sum Q_{i,2^{-1}} + 2Q_{i,2^{-2}}$

Activating Inequalities only When Necessary

- We model pb_j-DDT independently for all j.

2^{-1}-DDT

Δx	Δy							
	0x0	0x1	0x2	0x3	0x4	0x5	0x6	0x7
0x0	0	0	0	0	0	0	0	0
0x1	0	0	0	0	0	0	0	0
0x2	0	0	0	0	0	0	0	0
0x3	0	0	0	1	0	0	0	1
0x4	0	0	0	0	1	0	0	1
0x5	0	0	0	0	0	0	0	0
0x6	0	0	0	0	0	0	0	0
0x7	0	0	0	1	1	0	0	0

Inequality to model pb_j-DDT is given by the following form:

$$a_0x_2 + a_1x_1 + a_2x_0 + a_3y_2 + a_4y_1 + a_5y_0 \geq b$$

where, $a_0, a_1, \cdots, a_5 \in \{-1, 0, 1\}$, $b \leq -1$.

- Inequalities to model pb_j-DDT should be meaningful only when $pb_j = 1$.
- big-M method

$$a_0x_2 + a_1x_1 + a_2x_0 + a_3y_2 + a_4y_1 + a_5y_0 + M(1 - Q_{i,pb_j}) \geq b$$

M is a sufficiently big constant.

Summary of Probability Modeling

1. Separate the DDT into pb-DDTs.
2. Add an inequality to represent probability.
3. Model all pb-DDTs with QM or espresso.
4. Add a term for Big-M in each inequality.

Example: actual lp file for SKINNY-128

Applications to SKINNY-128

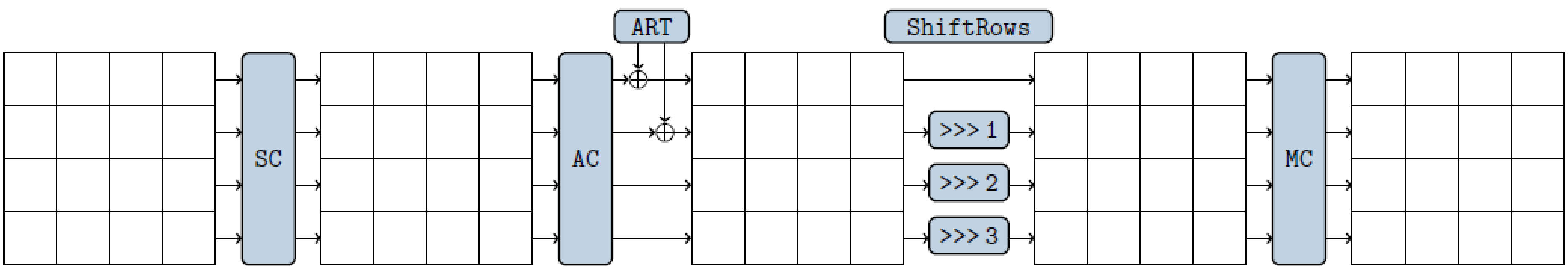
SKINNY

- Proposed at CRYPTO2016 by Beierle et al.
- Tweakable block cipher supporting n-bit block and n-, $2n$-, and $3n$-bit tweakey, where $n \in \{64,128\}$.
- In this talk, we focus our attention on the single-key analysis of SKINNY-128.

SKINNY-128: Round Function

AES-like Round Function

- **SubCells** (SC): Application of an 8-bit Sbox
  Max differential probability of the S-box is 2^{-2}.
- **AddConstants** and **AddRoundTweakey**
- **ShiftRows** (SR): Rotate row i by i bytes to right
- **MixColumns** (MC): Multiply the state by a binary matrix



Previous Bounds

rounds	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
LB (word) [BJK+16]	1	2	5	8	12	16	26	36	41	46	51	55	58	61	66

"LB" denotes lower bound

- Lower bounds can be given by $\#ASbox \times 2^{-2}$.
- Block size is 128 bits. We are targeting differential trails with prob higher than 2^{-128} (64 active S-boxes).
- 15 rounds are secure.

Simple Upper Bounds

rounds	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
LB (word) [BJK+16]	1	2	5	8	12	16	26	36	41	46	51	55	58	61	66
simple UB (bit)	1	2	5	8	12	16	26	36	43	48	52	56	62	68	-

"LB" denotes lower bound and "UB" denotes upper bound.

- We then derived simple upper bounds by assuming all the active S-boxes output the same difference (cancellation by XOR occurs with probability 1)
- Gap exists from 9 rounds to 14 rounds.
- Up to 13 rounds can be attacked simply.
- Is 14-round secure or insecure?

Searching for the Best Diff Trail

- Two-stage strategy by Sun et al.
  1. List up all truncated differentials with word-wise search (fast but may contain contradiction if looked in bit-wise level)
  2. Test the best probability of each truncated diffs.

- The word-wise truncated differential search detect 4 rotation variants. Checking one of them is sufficient.

Cutting-Off Low Probability Transition

Let's consider the 9-round search.

- LB of #ASbox is 41: 2^{-82}
- UB of #ASbox is 43: 2^{-86}

Gap is at most 2^{-4}, thus no need to test the differential propagation with prob 2^{-7} or 2^{-6}.

83% of the non-zero DDT entries propagate with probability 2^{-7} or 2^{-6}. Removing them from the search space has significant impact.

probability	2^{-7}	2^{-6}	$2^{-5.4}$	2^{-5}	$2^{-4.4}$	2^{-4}	$2^{-3.7}$	$2^{-3.4}$	$2^{-3.2}$	2^{-3}	$2^{-2.7}$	$2^{-2.4}$	2^{-2}
DDT value	2	4	6	8	12	16	20	24	28	32	40	48	64
# of entries	2688	5006	64	2828	78	652	2	18	2	101	2	4	23

NTT

Search Results

Rounds	9	10	11	12	13	14
LB	2^{-82}	2^{-92}	2^{-102}	2^{-110}	2^{-116}	2^{-122}
Simple UB	2^{-86}	2^{-96}	2^{-104}	2^{-112}	2^{-124}	2^{-136}
Tight bound	2^{-86}	2^{-96}	2^{-104}	2^{-112}	$\mathbf{2^{-123}}$	$\mathbf{\leq 2^{-128}}$

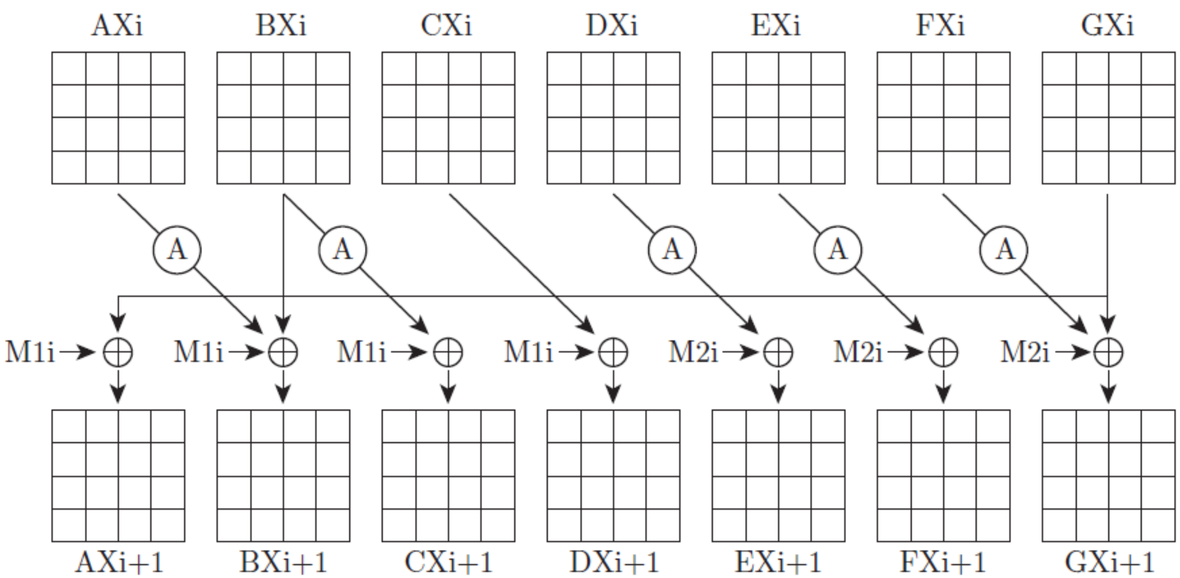
- The cutting-off technique cannot be used for 13 rounds. The experiment took more than 2 weeks.
- All 14-round truncated diffs are the extension of 13-round trail with 3 additional active S-boxes. The maximum prob is $2^{-123-6} = 2^{-129}$.
- Improved diff resistance of SKINNY-128 by 1 round.

Applications to AES-Round Based Function

AES-Round Based Function

- Proposed by Jean and Nikolić at FSE2016.
- many parameters to process multiple AES states
- Lower bound of #active S-boxes is evaluated by MILP. Tightness is unknown. Probability is not evaluated.
- 7 constructions are finally proposed.

C5 construction



NTT

Search Results

C1 construction:

	#Active S-boxes		Probability	
Prev	lower	22	lower	2^{-132}
New	tight	22	tight	2^{-134}

C5 Construction:

	#Active S-boxes		Probability	
Prev	lower	22	lower	2^{-132}
New	lower	24	lower	2^{-144}

Concluding Remarks

Concluding Remarks

New MILP model

- QM and Espresso for modeling $*$-DDT.
- pb-DDT and big-M for evaluating probability.

Applications

- Improved diff resistance of SKINNY-128
- Evaluated prob of AES-round based function.

MILP can be applied to 8-bit Sboxes!!

Thank you for your attention!!