Key-Recovery Attacks on Keccak-Based Constructions

Ling Song

Joint work with Jian Guo, Danping Shi and San Ling

NANYANG TECHNOLOGICAL UNIVERSITY

中国科学院信息工程研究所
INSTITUTE OF INFORMATION ENGINEERING,CAS

10 October, 2018 @ Milano, Italy

Outlines

Outline

KECCAK

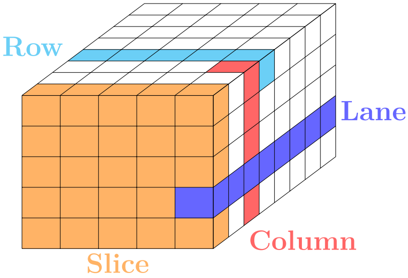
- Permutation-based hash function
  - Designed by Guido Bertoni, Joan Daemen, Michaël Peeters and Gilles Van Assche
  - Selected as SHA-3 standard
  - Underlying permutation: KECCAK-$p[1600, 24]$
- KECCAK under keyed modes: KMAC, KECCAK-MAC
- Its relatives
  - Authenticated encryption: KEYAK, KETJE
  - Pseudorandom function: KRAVATTE
  - Permutation: XOODOO

KECCAK-$p[b, n_r]$ Permutation

- b bits: seen as a 5×5 array of $\frac{b}{25}$-bit lanes, $A[x, y]$
- n_r rounds
- each round R consists of five steps:

$$R = \iota \circ \chi \circ \pi \circ \rho \circ \theta$$

- χ : S-box on each row
- π, ρ: change the position of state bits



http://www.iacr.org/authors/tikz/

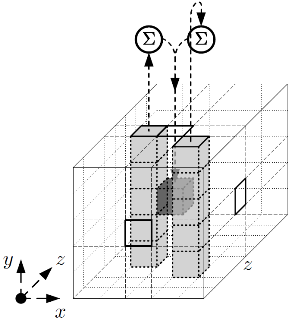
KECCAK-p Round Function: θ

θ step: adding two columns to the current bit

$$C[x] = A[x,0] \oplus A[x,1] \oplus A[x,2] \oplus A[x,3] \oplus A[x,4]$$

$$D[x] = C[x-1] \oplus (C[x+1] \lll 1)$$

$$A[x,y] = A[x,y] \oplus D[x]$$

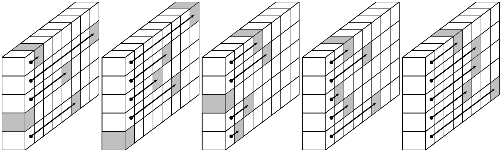


http://keccak.noekeon.org/

- The Column Parity kernel
  - If $C[x] = 0, 0 \leq x < 5$, then the state A is in the CP kernel.

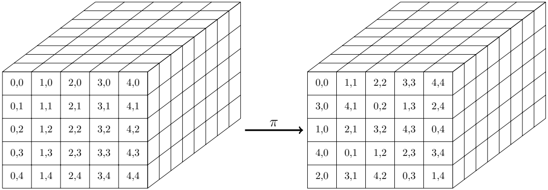
KECCAK-*p* Round Function: ρ, π

ρ step: lane level rotations, $A[x, y] = A[x, y] \lll r[x, y]$



http://keccak.noekeon.org/

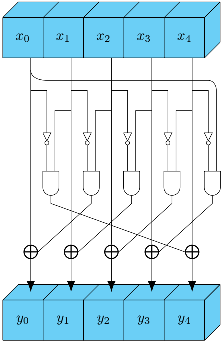
π step: permutation on lanes, $A[y, 2 * x + 3 * y] = A[x, y]$



KECCAK-p Round Function: χ

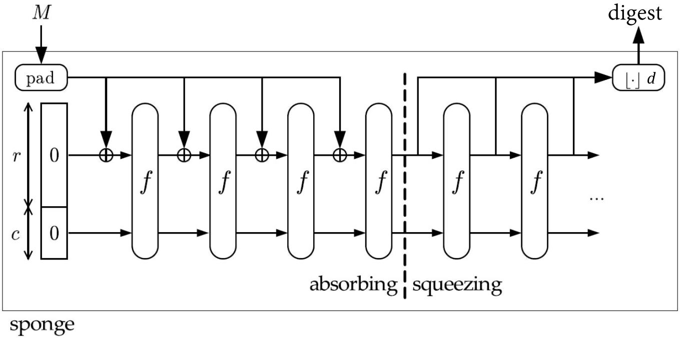
χ step: 5-bit S-boxes, nonlinear operation on rows

$$y_0 = x_0 + (x_1 + 1) \cdot x_2,$$
$$y_1 = x_1 + (x_2 + 1) \cdot x_3,$$
$$y_2 = x_2 + (x_3 + 1) \cdot x_4,$$
$$y_3 = x_3 + (x_4 + 1) \cdot x_0,$$
$$y_4 = x_4 + (x_0 + 1) \cdot x_1.$$



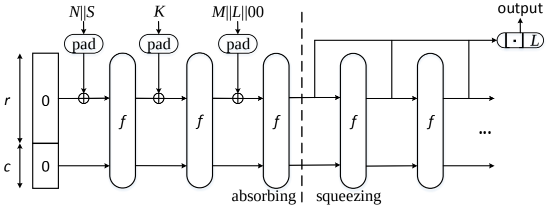
- Nonlinear term: product of two adjacent bits in a row.
- The algebraic degree of n rounds is 2^n.

KECCAK: KECCAK-$p[1600, 24] + $Sponge

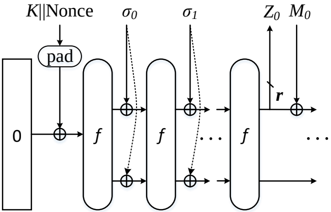


- Sponge construction [BDPV11]
  - b-bit permutation f
  - Two parameters: bitrate r, capacity c, and $b = r + c$.
- KECCAK-MAC
  - Take $K||M$ as input

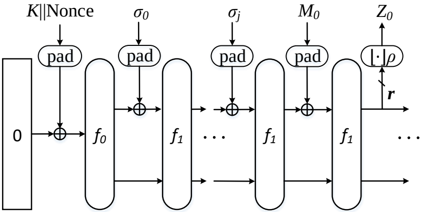
Keyed KECCAK Constructions



KMAC



KEYAK



KETJE

Key Recovery Attacks

Intuition: $deg(\chi) = 2$. Consider algebraic cryptanalsis, in paticular, cube attacks.

Key Recovery Attacks

Intuition: $deg(\chi) = 2$. Consider algebraic cryptanalsis, in paticular, cube attacks.

Contributions

- Mixed Integer Linear Programming models for searching two types of cube attacks
- Best key recovery attacks on round-reduced KMAC, KEYAK, KETJE and KECCAK-MAC so far
- Solve the open problem of "Full State Keyed Duplex (Sponge)"

Key Recovery Attacks

Intuition: $deg(\chi) = 2$. Consider algebraic cryptanalsis, in paticular, cube attacks.

Contributions

- Mixed Integer Linear Programming models for searching two types of cube attacks
- Best key recovery attacks on round-reduced KMAC, KEYAK, KETJE and KECCAK-MAC so far
- Solve the open problem of "Full State Keyed Duplex (Sponge)"

"Whether these attacks can still be extended to more rounds by exploiting full-state absorbing remains an open question".

— *the* KEYAK *designers*

Key Recovery Attacks

Intuition: $deg(\chi) = 2$. Consider algebraic cryptanalsis, in paticular, cube attacks.

Contributions

- Mixed Integer Linear Programming models for searching two types of cube attacks
- Best key recovery attacks on round-reduced KMAC, KEYAK, KETJE and KECCAK-MAC so far
- Solve the open problem of "Full State Keyed Duplex (Sponge)"

Ling Song, Jian Guo: *Cube-Attack-Like Cryptanalysis of Round-Reduced Keccak Using MILP*. IACR Transactions on Symmetric Cryptology, 2018(3), 182-214.

Ling Song, Jian Guo, Danping Shi, San Ling: *New MILP Modeling: Improved Conditional Cube Attacks on Keccak-based Constructions*. To appear in ASIACRYPT 2018

Outline

Cube Attacks [DS09] (Higher Order Differential Cryptanalysis)

- Given a Boolean polynomial $f(k_0, \dots, k_{n-1}, v_0, \dots, v_{m-1})$ and a monomial $t_I = v_{i_1} \dots v_{i_d}$, $I = \{v_{i_1}, \dots, v_{i_d}\}$, f can be written as

$$f(k_0, \dots, k_{n-1}, v_0, \dots, v_{m-1}) = t_I \cdot p_{S_I} + q$$

  - q contains terms that are not divisible by t_I
  - p_{S_I} is called the superpoly of I in f
  - $v_{i_1}, \dots, v_{i_d}$ are called cube variables. d is the dimension.
- The the cube sum is exactly

$$\sum_{(v_{i_1}, \dots, v_{i_d}) \in C_I} f(k_0, \dots, k_{n-1}, v_0, \dots, v_{m-1}) = p_{S_I}$$

- Cube attacks: p_{S_I} is a linear polynomial in key bits.
- Cube testers: distinguish p_{S_I} from a random function.
- If $deg(f) < d$, $p_{S_I} = 0$

Cube-Attack-Like Cryptanalysis [DMP+15]

Renamed auxCube

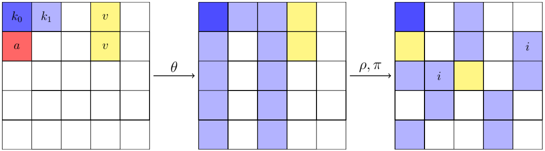
Idea: do not recover the exact linear p_{S_I} but try to limit the number (n_i) of key bits involved in p_{S_I} using n_a auxiliary variables.

Preprocessing phase Build a lookup table. The complexity is 2^{n_i+d}.

n_i key bits	Cube sum
00...00	01011...
00...01	11010...
...	...
11...11	10110...

Online phase Guess the value of n_a auxiliary variables and then query the cipher to obtain the cube sum; look up the table to recover the n_i key bits. The complexity is 2^{n_a+d}.

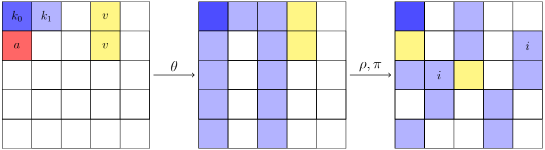
auxCube On KECCAK



$$d = 64,\ n_a = 64,\ n_i = 64,$$

The algebraic degree of n rounds is 2^n. Linearize the first round by avoiding adjacent cube variables.

auxCube On KECCAK



$$d = 64,\ n_a = 64,\ n_i = 64,$$

The algebraic degree of n rounds is 2^n. Linearize the first round by avoiding adjacent cube variables.

Task of the MILP Model

1. Find 2^{n-1}-dimensional cubes where n is as large as possible; (attack more rounds).
2. Find balanced attacks where n_i and n_a are close and as small as possible. (low complexity).

Conditional Cube Testers of Keccak [HWX+17]

Renamed conCube

conCube

- Linearize the first round.
- There exist p cube variables that are not multiplied with any cube variable even in the second round under certain *conditions*.

Type I conCube

- $p = 1$.
- Given such a cube with $d = 2^{n-1}$, $p_{S_I} = 0$ for n-round KECCAK if the conditions are met.

Type II conCube

- $p = d$.
- Given such a cube with $d = 2^{n-2} + 1$, $p_{S_I} = 0$ for n-round KECCAK if the conditions are met.

ConCube on KECCAK

If the conditions involve the key, the conditional cube can be used to recover the key.

Task of the MILP Model

1. Find Type I (II) cubes with dimension 2^{n-1} $(2^{n-2}+1)$ where n is as large as possible; (attack more rounds).
2. The number of conditions is minimized. (low complexity).

Outline

Mixed Integer Linear Programming

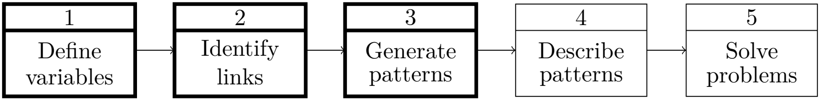
- An MILP problem is of the form

$$\begin{aligned} \texttt{min} \quad & c^T x \\ Ax &\geq b \\ x &\geq 0 \\ x &\in \mathbb{Z} \end{aligned}$$

- Solvers
  - Gurobi, CPLEX, SCIP, ...
- Application to cryptanalysis since Mouha et al.'s pioneering work [MWGP11]

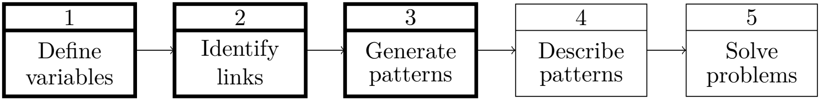
MILP-based Cryptanalysis

1. Define variables which are mostly binary for the crypto problem.
2. Identify links between the variables
3. Generate all valid patterns for the variables
4. Describe valid patterns with inequalities
5. Solve the MILP problem



MILP-based Cryptanalysis

1. Define variables which are mostly binary for the crypto problem.
2. Identify links between the variables
3. Generate all valid patterns for the variables
4. Describe valid patterns with inequalities
5. Solve the MILP problem



Example: construct an MILP model for searching Type II conCubes (for FKD)

1. Modeling the first χ
2. Modeling the activeness of column sums

Modeling the First χ

1. Define variables

Let $a[x][y][z]$ be the state:

$$a \xrightarrow{\pi\circ\rho\circ\theta} \mathbf{b} \xrightarrow{\chi} \mathbf{c} \xrightarrow{\pi\circ\rho\circ\theta} \mathbf{d} \xrightarrow{\chi} e$$

$A[x][y][z] = 1$ if $a[x][y][z]$ is active, *i.e.*, containing cube variables:

$$A \xrightarrow{\pi\circ\rho\circ\theta} \mathbf{B} \xrightarrow{\chi} \mathbf{C} \xrightarrow{\pi\circ\rho\circ\theta} \mathbf{D} \xrightarrow{\chi} E$$

$V[x][y][z] = 1$ indicates that bit $b[x][y][z]$ is constrained to the value of $H[x][y][z]$.

Modeling the First χ

2. Identify links: propagation of variables through χ

Observation

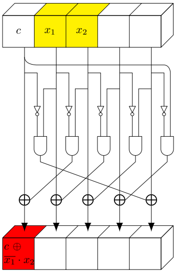
1. Linearize χ by avoiding adjacent variables in the input.
2. Bit 1 (0) on the left (right) of the variable helps to restrict the diffusion of variables through χ, while an unknown constant diffuses the variable in an uncertain way.

Modeling the First χ

2. Identify links: propagation of variables through χ

Observation

1. Linearize χ by avoiding adjacent variables in the input.
2. Bit 1 (0) on the left (right) of the variable helps to restrict the diffusion of variables through χ, while an unknown constant diffuses the variable in an uncertain way.

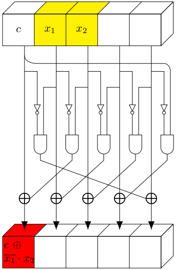


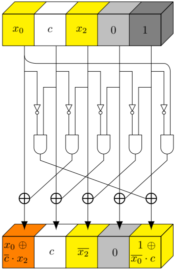
Modeling the First χ

2. Identify links: propagation of variables through χ

Observation

1. Linearize χ by avoiding adjacent variables in the input.
2. Bit 1 (0) on the left (right) of the variable helps to restrict the diffusion of variables through χ, while an unknown constant diffuses the variable in an uncertain way.





Modeling the First χ

3. Generate valid patterns

$$c[x] = b[x] + (b[x+1] + 1) \cdot b[x+2]^1$$

$b[x]$	$b[x+1]$	$b[x+2]$	$c[x]$

[1] Omit coordinates $[y][z]$.

Modeling the First χ

3. Generate valid patterns

$$c[x] = b[x] + (b[x+1] + 1) \cdot b[x+2]$$ [1]

$b[x]$	$b[x+1]$	$b[x+2]$	$c[x]$
cst	cst	cst	cst

[1] Omit coordinates $[y][z]$.

Modeling the First χ

3. Generate valid patterns

$$c[x] = b[x] + (b[x+1] + 1) \cdot b[x+2]^1$$

$b[x]$	$b[x+1]$	$b[x+2]$	$c[x]$
cst	cst	cst	cst
var	cst	*	var

[1] Omit coordinates $[y][z]$.

Modeling the First χ

3. Generate valid patterns

$$c[x] = b[x] + (b[x+1] + 1) \cdot b[x+2]^{1}$$

$b[x]$	$b[x+1]$	$b[x+2]$	$c[x]$
cst	cst	cst	cst
var	cst	*	var
cst	cst	var	var (deg $\leq$ 1)

[1] Omit coordinates $[y][z]$.

Modeling the First χ

3. Generate valid patterns

$$c[x] = b[x] + (b[x+1] + 1) \cdot b[x+2]^{1}$$

$b[x]$	$b[x+1]$	$b[x+2]$	$c[x]$
cst	cst	cst	cst
var	cst	*	var
cst	cst	var	var (deg $\leq$ 1)
cst	1	var	cst

[1] Omit coordinates $[y][z]$.

Modeling the First χ

3. Generate valid patterns

$$c[x] = b[x] + (b[x+1]+1) \cdot b[x+2]^1$$

$b[x]$	$b[x+1]$	$b[x+2]$	$c[x]$
cst	cst	cst	cst
var	cst	*	var
cst	cst	var	var (deg ≤ 1)
cst	1	var	cst
⋮	⋮	⋮	⋮

[1] Omit coordinates $[y][z]$.

Modeling the First χ

3. Generate valid patterns

$$B[x] = \begin{cases} 0, & b[x] \text{ is a constant;} \\ 1, & b[x] \text{ is a var.} \end{cases} \qquad V[x] = \begin{cases} 0, & \text{no condidtion on } b[x]; \\ 1, & b[x] \text{ is restricted to } 0/1. \end{cases}$$

Modeling the First χ

3. Generate valid patterns

$$B[x] = \begin{cases} 0, & b[x] \text{ is a constant;} \\ 1, & b[x] \text{ is a var.} \end{cases} \qquad V[x] = \begin{cases} 0, & \text{no condidtion on } b[x]; \\ 1, & b[x] \text{ is restricted to } 0/1. \end{cases}$$

Table: Diffusion of variables through χ

$B[x]$	$B[x+1]$	$B[x+2]$	$V[x+1]$	$V[x+2]$	$H[x+1]$	$H[x+2]$	$C[x]$
0	0	0	*	*	*	*	0
1	0	0	*	*	*	*	1
0	0	1	0	0	*	*	1
0	0	1	1	0	1	*	0
0	0	1	1	0	0	*	1
0	1	0	0	0	*	*	1
0	1	0	0	1	*	0	0
0	1	0	0	1	*	1	1
1	0	1	0	0	*	*	1
1	0	1	1	0	*	*	1

Modeling the First χ

4. Describe valid patterns with inequality

By generating the convex hull of the set of patterns [SHW+14], we get

$$
\begin{aligned}
-B[x] - B[x+1] &\geq -1 \\
-B[x] + C[x] &\geq 0 \\
-B[x+2] - V[x+2] &\geq -1 \\
-B[x+1] - V[x+1] &\geq -1 \\
-B[x] - B[x+1] - H[x+2] + C[x] &\geq -1 \\
B[x] - V[x+1] - H[x+1] - C[x] &\geq -2 \\
B[x] - V[x+2] + H[x+2] - C[x] &\geq -1 \\
B[x] + B[x+1] + B[x+2] - C[x] &\geq 0 \\
-B[x+1] - B[x+2] + V[x+1] + V[x+2] + C[x] &\geq 0 \\
-B[x+1] - B[x+2] + V[x+2] + H[x+1] + C[x] &\geq 0
\end{aligned}
$$

Modeling the Activeness of Column Sums

1. Define variables

For the state

- $a \xrightarrow{\pi\circ\rho\circ\theta} b \xrightarrow{\chi} c \xrightarrow{\pi\circ\rho\circ\theta} d \xrightarrow{\chi} e$
- Column sums before χ: $g_1[x][z] = \sum_y b[x][y][z]$
- Column sums after χ: $g_2[x][z] = \sum_y c[x][y][z]$

Variables for the activeness

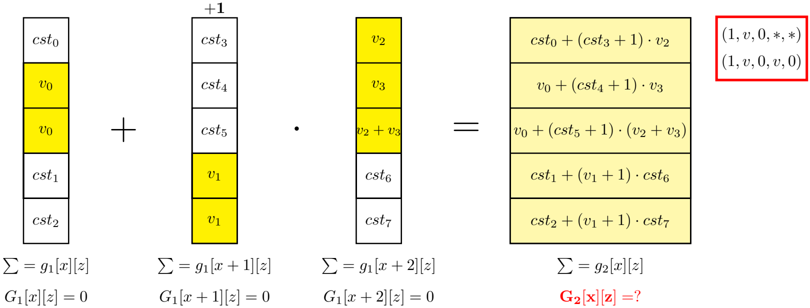
- $G_1[x][z] = 1$ if $g_1[x][z]$ is active.
- $G_2[x][z] = 1$ if $g_2[x][z]$ is active.

In which case $G_2[x][z]=0$?

Modeling the Activeness of Column Sums

2. Identify links for $G_2[x][z]$

$$b[x] + (b[x+1]+1) \cdot b[x+2] = c[x]$$

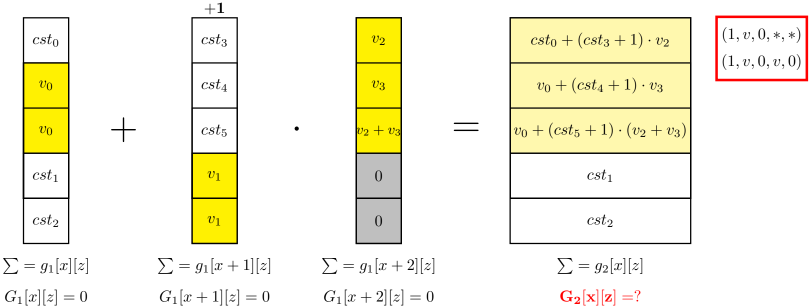


Cond1: $G_1[x][z]$ must be 0.

Modeling the Activeness of Column Sums

2. Identify links for $G_2[x][z]$

$$b[x] + (b[x+1]+1) \cdot b[x+2] = c[x]$$

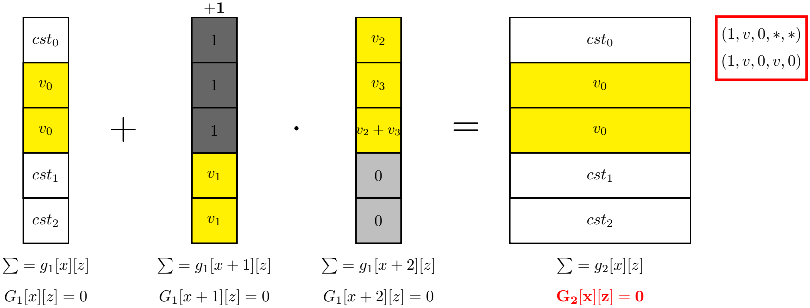


Cond2: No variable in column $(x+1, z)$ of b propagates to column (x, z) of c.

Modeling the Activeness of Column Sums

2. Identify links for $G_2[x][z]$

$$b[x] + (b[x+1]+1)\cdot b[x+2] = c[x]$$

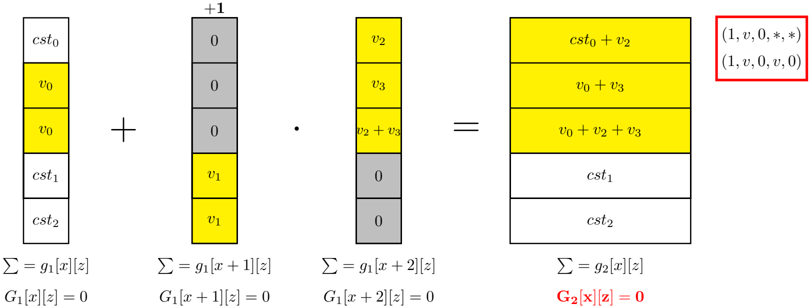


Cond3.1: No variable in column $(x+2)$ of b propagates to column (x, z) of c.

Modeling the Activeness of Column Sums

2. Identify links for $G_2[x][z]$

$$b[x] + (b[x+1]+1)\cdot b[x+2] = c[x]$$



Cond3.2: All the variables in column $(x+2)$ of b propagate to column (x,z) of c and $G_1[x+2][z]=0$.

Modeling the Activeness of Column Sums

2. Identify links for $G_2[x][z]$

Condition for $G_2[x][z] = 0$

Cond1 $\bigwedge$ Cond2 $\bigwedge$ (Cond3.1 $\bigvee$ Cond3.2)

Modeling the Activeness of Column Sums

2. Identify links for $G_2[x][z]$

Condition for $G_2[x][z] = 0$

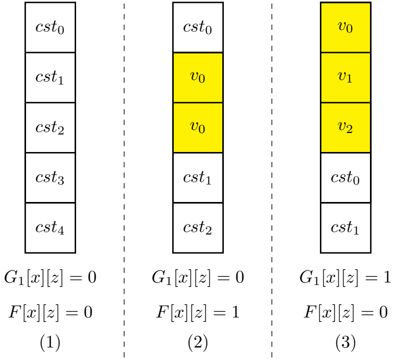
Cond1 $\bigwedge$ Cond2 $\bigwedge$ (Cond3.1 $\bigvee$ Cond3.2)

$\Rightarrow$ Model each part individually.

Model for Cond1

$G_1[x][z]$ together with $F[x][z]$ describe a column before χ.

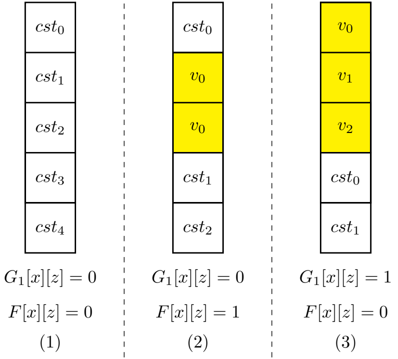
1. The column is not active, *i.e.*, there is no variable;
2. The column is active and the column sum is active;
3. The column is active and the column sum is inactive.



Model for Cond1

$G_1[x][z]$ together with $F[x][z]$ describe a column before χ.

1. The column is not active, *i.e.*, there is no variable;
2. The column is active and the column sum is active;
3. The column is active and the column sum is inactive.



The patterns of $B[x][y][z], y = 0, \cdots, 4$ and $F[x][z]$, $G_1[x][z]$ fall into a set of 58 discrete points in $\mathbb{R}^7$.

Model for Cond1

Table: Inequalities modeling the activeness of a column

$$
\begin{aligned}
-F[x][z] - G_1[x][z] &\geq -1 \\
-B[x][0][z] + F[x][z] + G_1[x][z] &\geq 0 \\
-B[x][1][z] + F[x][z] + G_1[x][z] &\geq 0 \\
-B[x][2][z] + F[x][z] + G_1[x][z] &\geq 0 \\
-B[x][3][z] + F[x][z] + G_1[x][z] &\geq 0 \\
-B[x][4][z] + F[x][z] + G_1[x][z] &\geq 0 \\
\sum_{y} B[x][y][z] - 2F[x][z] - G_1[x][z] &\geq 0
\end{aligned}
$$

Modeling the Activeness of Column Sums

2. Identify links for $G_2[x][z]$

Condition for $G_2[x][z] = 0$

Cond1 $\bigwedge$ **Cond2** $\bigwedge$ (Cond3.1 $\bigvee$ Cond3.2)

$\Rightarrow$ Model each part individually.

Model for Cond2

Variables

- Cond2 $\leftrightarrow M[x][z] = 0$
- $P[x][y][z] = 1$ if the variable at $(x+1, y, z)$ is propagated to (x, y, z) uncertainly.

Inequalities

$$M[x][z] - P[x][y][z] \geq 0, y = 0, \cdots, 4.$$

$$\sum_y P[x][y][z] - M[x][z] \geq 0.$$

$P[x]$	$B[x+1]$	$V[x+2]$	inequalities
0	0	*	$-P[x] + B[x+1] \geq 0$
1	1	0	$-P[x] - V[x+2] \geq -1$
0	1	1	$P[x] - B[x+1] + V[x+2] \geq 0$

Modeling the Activeness of Column Sums

2. Identify links for $G_2[x][z]$

Condition for $G_2[x][z] = 0$

Cond1 $\bigwedge$ Cond2 $\bigwedge$ **(Cond3.1** $\bigvee$ **Cond3.2)**

$\Rightarrow$ See the paper.

The Full Model

Objective

$$min \quad \sum V[x][y][z]$$

Linear constraints

- Dimension

$$\sum B[x][y][z] - \sum F[x][z] = 2^{n-2} + 1$$

- Other inequalities

Outline

Results of Key Recovery Attacks

- First analytical results on KMAC
- Improve the attack against Lake Keyak (128) from 6 to 8 rounds in the NR setting, and attack 9 rounds if the key size is 256 bits.
- Solve the FKD open problem

Target	$\lvert K \rvert$	c	Rounds	Time	Reference	Type
KMAC128	128	256	7/24	2^{76}	this	conCube
KMAC256	256	512	9/24	2^{147}	this	

Target	$\lvert K \rvert$	NR	Rounds	Time	Reference	Type
Lake KEYAK	128	Yes	6/12	2^{37}	[DMP+15]	cube
	128	No	8/12	2^{74}	[HWX+17]	conCube
	128	Yes	8/12	$2^{71.01}$	this	conCube
	256	Yes	9/14	$2^{137.05}$	this	
River KEYAK	128	Yes	8/12	2^{77}	this	
FKD[1600]	128	No	9/-	2^{90}	this	

NR: nonce-respected

Attack complexity improvements on KETJE

Target	$\|K\|$	Rounds	T	M	Reference	Type
KETJE Major	128	7/13	2^{83}	-	[LBD+17]	conCube
	128	7/13	$2^{71.24}$	-	this	
KETJE Minor	128	7/13	2^{81}	-	[LBD+17]	
	128	7/13	$2^{73.03}$	-	this	
KETJE Sr V1	128	7/13	2^{115}	2^{50}	[DMP+15]	auxCube
	128	7/13	2^{91}	-	this	conCube
KETJE Sr V2	128	7/13	$2^{113.58}$	2^{48}	[DLWQ17]	auxCube
	128	7/13	2^{99}	2^{33}	this	
KETJE Jr V1	96	5/13	2^{56}	2^{38}	[DLWQ17]	
	96	5/13	$2^{36.86}$	2^{18}	this	
	72	6/13	$2^{68.04}$	2^{34}	this	
KETJE Jr V2	96	5/13	$2^{50.32}$	2^{32}	[DLWQ17]	
	96	5/13	$2^{34.91}$	2^{15}	this	
	80	6/13	$2^{59.17}$	2^{25}	this	
XOODOO	128	6/-	2^{89}	2^{55}	this	

Attacks on Keccak-MAC

Target	$\|K\|$	c	Rounds	Time	Reference	Type
Keccak-MAC	128	256/512	7/24	2^{72}	[HWX+17]	conCube
		768	7/24	2^{75}	[LBD+17]	
		1024	6/24	$2^{58.3}$		
		1024	6/24	2^{40}	this	
		1024	7/24	2^{111}	this	auxCube

Comparison of auxCube and conCube

	auxCube	conCube
Model	1 round, simple	2 rounds, complex
Degree of freedom	When DF is small, e.g. KETJE	When DF is large, e.g. FKD
Fully unknown internal state	No	Yes, e.g. KMAC, FKD
Memory	Non-negligible	Negligible

Conclusion

1. Two MILP models for searching cubes for KECCAK.
2. First attacks on KMAC and XOODOO, and improved attacks on KEYAK and KETJE.
3. Solve the FKD open problem.
4. The security of Keccak-based constructions is far from being threatened.

Conclusion

1. Two MILP models for searching cubes for Keccak.
2. First attacks on KMAC and Xoodoo, and improved attacks on Keyak and Ketje.
3. Solve the FKD open problem.
4. The security of Keccak-based constructions is far from being threatened.

Thank you for your attention!