Linear Structures: Applications to Cryptanalysis of Round-Reduced Keccak

#### Meicheng Liu joint work with Jian Guo and Ling Song

Asiacrypt 2016





#### Outline

#### Introduction SHA-3 hash function

#### Linear Structures

Linear structures of Keccak-f permutation Techniques for keeping 1+1 rounds being linear Techniques for keeping 1+2 rounds being linear

#### Distinguishers

Zero-sum distinguishers on Keccak-f

#### Preimage Attacks

Preimage Attacks on Keccak Setting up linear equations from the output of  $\chi$  Keccak Crunchy Crypto Contest

### Cryptographic hash function

- A cryptographic hash function is a mathematical algorithm that maps data of arbitrary size to a bit string of a fixed size, which is designed to also be one-way function.
- Properties
  - Collision resistance
    - It should be difficult to find a pair of different messages  $m_1$  and  $m_2$  such that  $H(m_1) = H(m_2)$ .
  - Preimage resistance
    - Given an arbitrary *n*-bit value x, it should be difficult to find any message m such that H(m) = x.
  - Second preimage resistance
    - Given message  $m_1$ , it should be difficult to find any different message  $m_2$  such that  $H(m_1) = H(m_2)$ .

#### SHA-3 hash function

- NIST SHA-3 hash function competition (2007–2012)
- Winner: Keccak
  - ▶ The winner was announced to be Keccak in October 2012.
  - Designers: Guido Bertoni, Joan Daemen, Michaël Peeters, and Gilles Van Assche
     Official versions: Keccak-224/256/384/512
     The Keccak web site: http://keccak.noekeon.org/
- In August 2015 NIST announced that SHA-3 had become a hashing standard.
  - SHA3-224/256/384/512
  - SHAKE128/256 (eXtendable Output Functions, XOFs)



Michaël Peeters, Guido Bertoni, Gilles Van Assche and Joan Daemen

#### The Keccak Team

## Specifications of Keccak

- Structure of Keccak
  - Sponge construction



- Keccak-f permutation
  - 1600 bits: a  $5 \times 5$  array of 64-bit lanes
  - 24 rounds
  - each round consists of five steps:

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

•  $\chi$  : the only nonlinear operation

#### SHA-3 hash function

Federal Information Processing Standards (FIPS) 202 instances

Instances			Output	Collision	Preimage
instances	r	C	Length	Resistance	Resistance
SHA3-224	1152	448	224	112	224
SHA3-256	1088	512	256	128	256
SHA3-384	832	768	384	192	384
SHA3-512	576	1024	512	256	512
SHAKE128	1344	256	$\ell$	$\min(\ell/2, 128)$	$\min(\ell, 128)$
SHAKE256	1088	512	$\ell$	$\min(\ell/2, 256)$	$\min(\ell, 256)$

Table: The standard FIPS 202 instances

#### Linear structures of Keccak-f permutation

- Several known attacks are based on the technique of linearizing 1-round Keccak-f
  - Zero-sum distinguishers [AM09]
  - Cube-attack-like cryptanalysis on keyed variants of Keccak [DMP<sup>+</sup>15]
- ▶ We find that 2- and 3-round Keccak-*f* can be linearized

$$\frac{1}{backward} | \frac{1}{forward} | dim \le 512$$

$$\frac{1}{backward} | \frac{2}{forward} | dim \le 194$$

To mount preimage attacks, we often use

$$\frac{1}{\text{backward}} | \frac{1}{\text{forward}} | \text{dim} = 512$$

## Specifications of Keccak

- Structure of Keccak
  - Sponge construction



- Keccak-f permutation
  - 1600 bits: a  $5 \times 5$  array of 64-bit lanes
  - 24 rounds
  - each round consists of five steps:

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

•  $\chi$  : the only nonlinear operation

#### Keccak-f permutation

Internal state A: a  $5 \times 5$  array of 64-bit lanes

$$\begin{aligned} \theta & C[x] = A[x,0] \oplus A[x,1] \oplus A[x,2] \oplus A[x,3] \oplus A[x,4] \\ D[x] &= C[x-1] \oplus (C[x+1] \lll 1) \\ A[x,y] &= A[x,y] \oplus D[x] \end{aligned}$$

$$\rho \ A[x,y] = A[x,y] \lll r[x,y]$$

$$\pi B[y, 2 * x + 3 * y] = A[x, y]$$

$$\chi A[x,y] = B[x,y] \oplus ((\sim B[x+1,y])\&B[x+2,y])$$

- $\iota \ A[0,0] = A[0,0] \oplus RC$
- The constants r[x, y] are the rotation offsets.
- RC[i] are the round constants.
- The only non-linear operation is  $\chi$  step algebraic degree 2

## Techniques for keeping 1+1 rounds being linear

with the degrees of freedom up to 512

Keeping one round forward being linear



Figure: Keeping one round forward being linear with the degrees of freedom up to 512, with yellow bits of degree 1, orange bits of degree at most 1, and the other bits being constants.

- Keeping one round backward being linear
  - linearizing the inverse of  $\chi$  according to its property: restrict the bits of gray lanes to be all ones and the bits of lightgray lanes to be all zeros

#### Linearizing the inverse of $\chi$

The inverse  $\chi^{-1}: b \mapsto a$  has algebraic degree 3, and

$$a_i = b_i \oplus (b_{i+1} \oplus 1) \cdot (b_{i+2} \oplus (b_{i+3} \oplus 1) \cdot b_{i+4}) \tag{1}$$

where  $0 \le i \le 4$  and the indexes are operated on modulo 5. If we impose  $b_3 = 0$  and  $b_4 = 1$ , then we have

$$egin{aligned} &a_0 = b_0 \oplus (b_1 \oplus 1) \cdot (b_2 \oplus 1), \ &a_1 = b_1, \ &a_2 = 1 \oplus b_2 \oplus (b_0 \oplus 1) \cdot b_1, \ &a_3 = 0, \ &a_4 = 1 \oplus (b_0 \oplus 1) \cdot b_1, \end{aligned}$$

and thus all  $a_i$ 's are linear on  $b_0$  and  $b_2$ . That's, for  $b_3 = 0$ ,  $b_4 = 1$  and any fixed  $b_1$ , the algebraic degree of  $\chi^{-1}$  becomes 1.

## Techniques for keeping 1 + 2 rounds being linear with the degrees of freedom up to 194

Keeping two rounds forward being linear



Keeping one round backward being linear

Exploiting the linear structures of Keccak-f

What's a zero-sum distinguisher?

Find a set S such that  $\sum_{x \in S} x = 0$  and  $\sum_{x \in S} f(x) = 0$ .

Known zero-sum distinguisher on Keccak-f permutation



 Our improved zero-sum distinguisher on Keccak-f permutation

$$\frac{m+1}{backward} | \frac{1+n}{forward} | \frac{m+1}{backward} | \frac{2+n}{forward} | \frac{m+1}{forward} | \frac{2}{backward} | \frac{1+n}{backward} | \frac{1+n}{backward}$$

< □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □

Exploiting the linear structures of Keccak-f

What's a zero-sum distinguisher?

Find a set S such that  $\sum_{x \in S} x = 0$  and  $\sum_{x \in S} f(x) = 0$ .

Known zero-sum distinguisher on Keccak-f permutation



 Our improved zero-sum distinguisher on Keccak-f permutation

$$\frac{m+1}{backward} | \frac{1+n}{forward} | \frac{m+1}{backward} | \frac{2+n}{forward} | \frac{2+n}{forward} | \frac{2}{backward} | \frac{2}{backward}$$

► Complexity: 2<sup>1+max(2<sup>n</sup>,3<sup>m</sup>)</sup>

Since deg(χ) = 2 and deg(χ<sup>-1</sup>) = 3, the algebraic degree of n forward Keccak-f rounds is bounded by 2<sup>n</sup>, and m backward rounds by 3<sup>m</sup>.

Exploiting the linear structures of Keccak-f

 Extend the previous zero-sum distinguishers by 2 rounds without increasing the complexities

#R	inv+forw	Best Known	inv+forw	Improved	inv+forw	Further
7	3+4	2 <sup>13</sup> [JN15]	3+4	2 <sup>10</sup>	2+5	2 <sup>9</sup>
8	3+5	2 <sup>18</sup> [АМ09, JN15]	3+5	2 <sup>17</sup>	3+5	2 <sup>10</sup>
9	4+5	2 <sup>33*</sup> [AM09]	4+5	2 <sup>28</sup>	3+6	2 <sup>17</sup>
10	4+6	2 <sup>65</sup> * [АМ09]	4+6	2 <sup>33</sup>	4+6	2 <sup>28</sup>
11	5+6	2 <sup>82*</sup> [AM09]	4+7	2 <sup>65</sup>	4+7	2 <sup>33</sup>
12	5+7	2 <sup>129</sup> [АМ09]	5+7	2 <sup>82</sup>	4+8	2 <sup>65</sup>
13	6+7	2 <sup>244</sup> [амо9]	5+8	2 <sup>129</sup>	5+8	2 <sup>82</sup>
14	6+8	2 <sup>257</sup> [АМ09]	6+8	2 <sup>244</sup>	5+9	2 <sup>129</sup>
15	6+9	2 <sup>513</sup> [АМ09]	6+9	2 <sup>257</sup>		
24	12+12	2 <sup>1575</sup> [BCC11, DL11]				

\*Corrected.

Exploiting the linear structures of Keccak-f

- Practical distinguisher for 11 rounds
- \* The 12-round Keccak-*f* permutations can be distinguished with complexity 2<sup>65</sup> or 2<sup>82</sup>.
  - ► This is of special interests since the 12-round Keccak-*f* permutation variants are used in the CAESAR candidates KEYAK and KETJE.
  - ► Nevertheless, we stress here that this distinguisher does not affect the security of KEYAK or KETJE.

1. linearize two rounds with one round forward and one round backward, and obtain 512 variables such that the first two rounds are linear





1. linearize two rounds with one round forward and one round backward, and obtain 512 variables such that the first two rounds are linear



 to make sure that the state input to the first round corresponds to a legal message, we set up 262 linear equations (256 bits for capacity and 6 bits for padding)

1. linearize two rounds with one round forward and one round backward, and obtain 512 variables such that the first two rounds are linear



 to make sure that the state input to the first round corresponds to a legal message, we set up 262 linear equations (256 bits for capacity and 6 bits for padding)

After the above two steps, there remains 250 free variables such that the bits input to step  $\chi$  of the third round are all linear.

1. linearize two rounds with one round forward and one round backward, and obtain 512 variables such that the first two rounds are linear



 to make sure that the state input to the first round corresponds to a legal message, we set up 262 linear equations (256 bits for capacity and 6 bits for padding)

After the above two steps, there remains 250 free variables such that the bits input to step  $\chi$  of the third round are all linear.

- ▶ Preimage attacks on SHAKE128 with output length 128
  - 3 rounds: set up linear equations by exploiting bilinear structure of χ and guessing some bits input to χ

1. linearize two rounds with one round forward and one round backward, and obtain 512 variables such that the first two rounds are linear



 to make sure that the state input to the first round corresponds to a legal message, we set up 262 linear equations (256 bits for capacity and 6 bits for padding)

After the above two steps, there remains 250 free variables such that the bits input to step  $\chi$  of the third round are all linear.

- ▶ Preimage attacks on SHAKE128 with output length 128
  - 3 rounds: set up linear equations by exploiting bilinear structure of χ and guessing some bits input to χ
  - $\blacktriangleright$  4 rounds: partially linearize the third round, and set up linear equations by bilinear structure of  $\chi$

## Setting up linear equations from the output of $\chi$

Bilinear structure of  $\chi$ 

The algebraic normal form of  $\chi$  mapping 5-bit *a* into 5-bit *b* can be written as  $b_i = a_i \oplus (a_{i+1} \oplus 1) \cdot a_{i+2}$ , and specially we have

$$b_0 = a_0 \oplus (a_1 \oplus 1) \cdot a_2 \tag{2}$$

$$b_1 = a_1 \oplus (a_2 \oplus 1) \cdot a_3 \tag{3}$$

Given two consecutive bits of the output of  $\chi$ , one linear equation on the input bits can be set up. By (3), we have

$$b_1 \cdot a_2 = (a_1 \oplus (a_2 \oplus 1) \cdot a_3) \cdot a_2 = a_1 \cdot a_2 \tag{4}$$

and thus according to (2) we obtain

$$\boldsymbol{b}_0 = \boldsymbol{a}_0 \oplus (\boldsymbol{b}_1 \oplus 1) \cdot \boldsymbol{a}_2. \tag{5}$$

Given three consecutive bits of the output of  $\chi$ , to say  $b_0$ ,  $b_1$  and  $b_2$ , an additional linear equation can be similarly set up:

$$b_1 = a_1 \oplus (b_2 \oplus 1) \cdot a_3. \tag{6}$$

< □ > < @ > < 注 > < 注 > 注目 = の < @ -

# Setting up linear equations from the output of $\chi$ $_{\rm Bilinear\ structure\ of\ }\chi$

The input *a* and output *b* of 5-bit Sbox  $\chi$  satisfy F(a, b) = 0 with

$$F_i(a,b) = b_{i+1} \cdot a_{i+2} + a_i + a_{i+2} + b_i, \ 0 \le i \le 4.$$

Table: Number of Linear Equations on Input Bits Obtained from the Output of 5-bit Sbox  $\chi$ 

#Known consecutive output bits	2	3	4	5
#Linear equations on input bits	1	2	4	5

#### Setting up more linear equations

- 1. The first method is to guess the value of an input bit.
  - guess the value of input bit a<sub>1</sub>
  - obtain the linear equation  $b_0 = a_0 \oplus (a_1 \oplus 1) \cdot a_2$
- 2. The second method is to make use of the probabilistic equation  $b_i = a_i$  with probability 0.75.

< □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □

▶ for 3-round SHAKE128, given a 128-bit hash value *h*:

- ▶ for 3-round SHAKE128, given a 128-bit hash value *h*:
  - ▶ we set up 64 linear equations on the 250 free variables (the first two output bits b<sub>0</sub> and b<sub>1</sub> of 64 Sboxes are known)

 $b_0 = a_0 \oplus (b_1 \oplus 1) \cdot a_2$ 

- ▶ for 3-round SHAKE128, given a 128-bit hash value h:
  - we set up 64 linear equations on the 250 free variables (the first two output bits b<sub>0</sub> and b<sub>1</sub> of 64 Sboxes are known)

$$b_0 = a_0 \oplus (b_1 \oplus 1) \cdot a_2$$

► set up extra 2 × 64 linear equations by guessing 64 bits input to step \(\chi\) of the third round

$$a_2 = c$$
  
 $b_1 = a_1 \oplus (c \oplus 1) \cdot a_3$ 

< □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □

- ▶ for 3-round SHAKE128, given a 128-bit hash value *h*:
  - ▶ we set up 64 linear equations on the 250 free variables (the first two output bits b<sub>0</sub> and b<sub>1</sub> of 64 Sboxes are known)

$$b_0 = a_0 \oplus (b_1 \oplus 1) \cdot a_2$$

► set up extra 2 × 64 linear equations by guessing 64 bits input to step \(\chi\) of the third round

$$a_2 = c$$
  
 $b_1 = a_1 \oplus (c \oplus 1) \cdot a_3$ 

 obtain a linear system of 192 equations on 250 variables, and each solution corresponds to a preimage of h

- ▶ for 3-round SHAKE128, given a 128-bit hash value *h*:
  - ▶ we set up 64 linear equations on the 250 free variables (the first two output bits b<sub>0</sub> and b<sub>1</sub> of 64 Sboxes are known)

$$b_0 = a_0 \oplus (b_1 \oplus 1) \cdot a_2$$

► set up extra 2 × 64 linear equations by guessing 64 bits input to step \(\chi\) of the third round

$$a_2 = c$$
  
 $b_1 = a_1 \oplus (c \oplus 1) \cdot a_3$ 

- obtain a linear system of 192 equations on 250 variables, and each solution corresponds to a preimage of h
- the time complexity of this attack is 1

- ▶ for 3-round SHAKE128, given a 128-bit hash value *h*:
  - ▶ we set up 64 linear equations on the 250 free variables (the first two output bits b<sub>0</sub> and b<sub>1</sub> of 64 Sboxes are known)

$$b_0 = a_0 \oplus (b_1 \oplus 1) \cdot a_2$$

► set up extra 2 × 64 linear equations by guessing 64 bits input to step \(\chi\) of the third round

> $a_2 = c$  $b_1 = a_1 \oplus (c \oplus 1) \cdot a_3$

- obtain a linear system of 192 equations on 250 variables, and each solution corresponds to a preimage of h
- the time complexity of this attack is 1
- similar techniques help us solve two 3-round preimage challenges in Keccak Crunchy Crypto Contest

#### Preimage attacks on Keccak

Exploiting the linear structures of Keccak-f and bilinear structure of  $\chi$ 

#Rounds	Variant	Time	Reference
2	Keccak-224/256	2 <sup>33</sup>	[Naya-PlasenciaRM11]
2	Keccak-224/256	1	Our results
2	Keccak-384/512	$2^{129}/2^{384}$	Our results
3	Keccak[1440, 160, 80]	1	Our results
3	Keccak[640, 160, 80]	2 <sup>7</sup>	Our results
3	SHAKE128	1	Our results
3	Keccak-224/256/384	$2^{97}/2^{192}/2^{322}$	Our results
3	Keccak-512	2 <sup>482</sup>	Our results
3	Keccak-512	2 <sup>506</sup>	[MorawieckiPS13]
4	Keccak[1440, 160, 80]	2 <sup>54</sup>	Our results
4	SHAKE128	2 <sup>106</sup>	Our results
4	Keccak-224/256	$2^{213}/2^{251}$	Our results
4	Keccak-224/256	$2^{221}/2^{252}$	[MorawieckiPS13]
4	Keccak-384/512	$2^{378}/2^{506}$	[MorawieckiPS13]

Keccak team presents challenges for reduced-round Keccak instances, namely Keccak[c = 160, r = b - c] with  $b \ge 200$ :

- The capacity is fixed to 160 bits: this implies a security level of 2<sup>80</sup> against generic collision search.
- The width b of Keccak-f[b] is in {200, 400, 800, 1600}: the width values that support the chosen capacity.
- The number of rounds  $n_r$  ranges from 1 to 12.

For each of these Keccak instances there are two challenges:

- generating a collision in the output truncated to 160 bits;
- generating a preimage of an output truncated to 80 bits.
- ★ The prize for a challenge for  $n_r$  rounds is  $n_r \times 10 \in$ .

A solution for 3-round preimage challenge of width 1600

Challenge: 06 25 a3 46 28 c0 cf e7 6c 75

A solution for 3-round preimage challenge of width 1600

Challenge: 06 25 a3 46 28 c0 cf e7 6c 75

Preimage:

A solution for 3-round preimage challenge of width 800

Challenge: 00 7b b5 c5 99 80 66 0e 02 93

A solution for 3-round preimage challenge of width 800

Challenge: 00 7b b5 c5 99 80 66 0e 02 93

Preimage:

A partially matched solution for 4-round preimage challenge of width 1600

Challenge: 7d aa d8 07 f8 50 6c 9c 02 76

A partially matched solution for 4-round preimage challenge of width 1600

 Challenge:
 7d
 aa
 d8
 07
 f8
 50
 6c
 9c
 02
 76

 Output:
 7d
 aa
 d8
 07
 b0
 50
 6c
 9c
 02
 76

 Message:
 7d
 aa
 d8
 07
 b0
 50
 6c
 9c
 02
 76

34478777714e2563 40cd175047047140 aeb1ccba3d344b91 879550956744910 0909625257444c44 63faab93d1b6a3f5 7aca93b5c0c2afa0 f1b2772194934266 41e5a573d5efc16f 34e0e077bfb4ce43 48bb5cb11aa15738 3ecb466e4aa6fec3 4e3e5449626d5e2d ccec6be24c92d63b fb652d66cc6a4621 356d6bfdd56b1afb d9da9b8c0e366cd3 034ad6fdd9caa885 236ade6960c8edaf 03d6d60e45aeb00e b8132036d4e20f33 8e4a29bbbd2c1cb8 8549b303

A partially matched solution for 4-round preimage challenge of width 1600

 Challenge:
 7d
 aa
 d8
 07
 f8
 50
 6c
 9c
 02
 76

 Output:
 7d
 aa
 d8
 07
 b0
 50
 6c
 9c
 02
 76

 Difference:
 - - -48
 - - - - 

Message:

34d781770fae25d9 4bcdf7304704b1a0 aeb1cc6a3d9a4b9f 879b5b095e744910 09096232b744ac44 63faab93d1b6a3f5 7aca93b5c0c2afa0 f1b2772194934266 41e5a573d5efc16f 34e0e077bfb4ce43 48bb5cb11aa15738 3ecb466e4aa6fec3 4e3e5449626d5e2d ccec6be24c92d63b fb652d66cc6a4621 356d6bfdd56b1afb d9da9b8c0e366cd3 034ad6fdd9caa885 236ade6960c8edaf 03d6d60e45aeb00e b8132036d4e20f33 8e4a29bbbd2c1cb8 8549b303

A solution for 4-round preimage challenge of width 1600

Challenge: 7d aa d8 07 f8 50 6c 9c 02 76

A solution for 4-round preimage challenge of width 1600

Challenge: 7d aa d8 07 f8 50 6c 9c 02 76

Preimage:

e82d8f3276e85543 3cf77a79137cb68c b0d325479f4d33aa 6322817be3f75cdc 1b2d1fc33847eefa 3815737090003e07 f3ae39ce20ca35f1 fe9cf333317e463e 9cb46a02e2c495ce 4dfae61d5770ab3d ea5218e748a57f6b 5cdac47ec1c508be c16d020b

## Summary

- Properties of the nonlinear operation  $\chi$  and its inverse  $\chi^{-1}$
- Linear structures of Keccak-f permutation
- ► Improved zero-sum distinguishers on Keccak-f permutation
- Preimage attacks on Keccak
- Directions of future work
  - how to find linear structures with large space
  - more applications of linear structures