Probabilistic Linearization: Internal Differential Collisions in up to 6 Rounds of SHA-3

Zhongyi Zhang[1,2] Chengan Hou[1,2] Meicheng Liu[1,2]

[1] Key Laboratory of Cyberspace Security Defence, Institute of Information Engineering, CAS

[2] School of Cyber Security, University of Chinese Academy of Sciences

2024.08.22

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

Outline

Outline

1. Motivation
    - SHA-3 Hash Function
    - Previous work
    - Our Contribution
2. Overview of the Attack
3. Basic Techniques
4. Results and Summary

Keccak

- NIST SHA-3 hash function competition (2007-2012)
- Designers: Guido Bertoni, Joan Daemen, Michaël Peeters and Gilles Van Assche
- Submitted to SHA-3 competition in 2008
- The winner was announced to be Keccak in 2012
- In 2015, Keccak was standardized by NIST as SHA-3
    - SHA3-224/256/384/512
    - SHAKE128/256 (eXtendable Output Functions, XOFs)

Team Keccak

Guido Bertoni[3], Joan Daemen[2], Seth Hoffert, Michaël Peeters[1], Gilles Van Assche[1] and Ronny Van Keer[1]

[1]STMicroelectronics - [2]Radboud University - [3]Security Pattern

Internal Differential Cryptanalysis

- Developed by Thomas Peyrin (Crypto 2010) to analysis Grøstl.
- **Generalized** by Dinur, Dunkelman and Shamir (FSE 2013) in the analysis of Keccak.
- Improved to **conditional internal differentials** by Zhang, Hou and Liu (EC 2023).

Collision Attacks on Up to 5 Rounds of SHA-3 Using Generalized Internal Differentials

- **Generalized Internal Differential**
- Target Internal Difference Algorithm
- Practical Results:
  - 3-round Keccak-384
  - 3-round Keccak-512
- Theoretical Results:
  - 5-round Keccak-256
  - 4-round Keccak-384

Collision Attacks on Round-Reduced SHA-3 Using Conditional Internal Differentials

- **Conditional Internal Differential**
- Improved Target Internal Difference Algorithm
- Theoretical Results:
  - 5-round SHAKE128/SHA3-224/SHA3-256
  - 4-round SHA3-384/SHA3-512
  - 4/5-round SHAKE256

Collision Attacks on Round-Reduced SHA-3

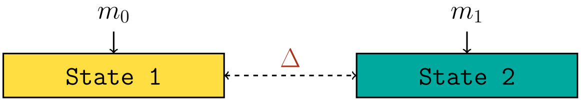
Methods	SHA3-224	SHA3-256	SHA3-384	SHA3-512	SHAKE128	SHAKE256
Differential or SAT-based	2 (practical)	2 (practical)	-	-	-	-
Differential [DDS12]	4 (practical)	4 (practical)	-	-	-	-
Internal Diff. [DDS13]	-	5 (2^{115})	3 (practical) 4 (2^{147})	3 (practical)	-	-
Algebriac Diff. [GLLLQS20]	5 (practical)	5 (practical)	-	-	5 (practical)	-
SAT-based Diff. [HBDM22]	-	-	4 ($2^{59.64}$)	-	-	-
SAT and Quant. [GLST22]	$6^{\dagger}$($2^{96.15}$)	$6^{\dagger}$($2^{102.65}$)	-	-	6 ($2^{123.5}$) $6^{\dagger}$($2^{61.4}$)	-
Internal Diff. [ZHL23]	5 (2^{105})	5 (2^{105})	4 (2^{76})	4 (2^{237})	5 (2^{105})	4 (2^{76}) 5 (2^{185})
Probabilistic Linear. Internal Diff.	5 ($2^{96.67}$)	5 ($2^{96.67}$)	5 ($2^{172.19}$)	4 ($2^{225.29}$)	5 ($2^{96.67}$)	5 ($2^{163.28}$) 6 ($2^{232.29}$)

† Quantum

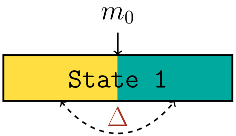
Outline

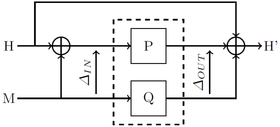
Internal Differential Cryptanalysis

- Standard differential cryptanalysis [BS91]: DES

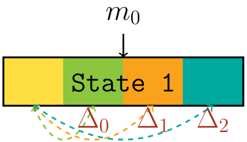


- Internal differential cryptanalysis [Peyrin10]: distinguisher on Grøstl



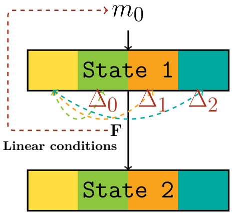


- Generalized internal differential cryptanalysis [DDS13]: collisions in SHA-3



Internal Differential Cryptanalysis

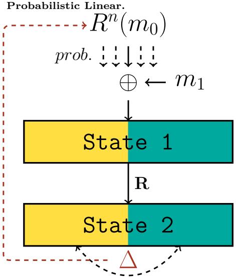
- Conditional internal differential cryptanalysis [ZHL23]: collisions in SHA-3



Find messages conforming 2-round internal differential characteristic by adding linear conditions to the initial state space.

Internal Differential Cryptanalysis

- Probabilistic Linearization (this work): collisions in SHA-3



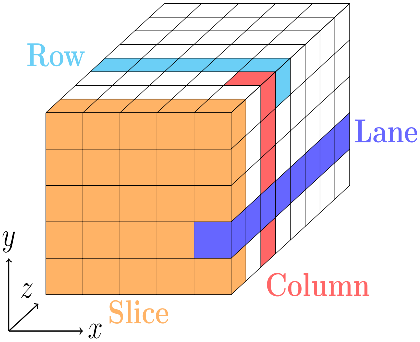
Find the first block message conforming target internal differential characteristic with a certain probability by adding LESS linear conditions to the initial internal difference space.

The Internal State in SHA-3

- 1600 bits: seen as a 5×5 matrix, where each cell is a 64-bit lane in the direction of the z axis $A[x, y], 0 \leq x, y < 5$
- each round R consists of five steps:

$$R = \iota \circ \chi \circ \pi \circ \rho \circ \theta, L \triangleq \pi \circ \rho \circ \theta$$

- χ: the only nonlinear operation, a 5-bit Sbox applies to each row
- ι: cannot be ignored in internal differential cryptanalysis



Symmetric States

- One state has **period** i in the z-axis is called a **symmetric state**

$$A[x][y][(z+i) \mod 64] = A[x][y][z], 0 \leq x, y < 5, 0 \leq z < 64$$

- The fundamental period corresponding to i is gcd (i, 64), i can attain non-trivial values $\{1, 2, 4, 8, 16, 32\}$

Example: A symmetric state with i = 16

```
|2024202420242024|746a746a746a746a|b82eb82eb82eb82e|5642564256425642|6d586d586d586d58|
|0714071407140714|934a934a934a934a|858c858c858c858c|75cb75cb75cb75cb|9e8d9e8d9e8d9e8d|
|6d586d586d586d58|0255025502550255|dd9ddd9ddd9ddd9d|fce0fce0fce0fce0|4a064a064a064a06|
|8482848284828482|3e993e993e993e99|df29df29df29df29|7e547e547e547e54|2013201320132013|
|49ea49ea49ea49ea|f441f441f441f441|e371e371e371e371|c6d9c6d9c6d9c6d9|3541354135413541|
```

Internal Difference Sets

- Given a period i, the set by adding a single **representative state v** to all symmetric states is an **internal difference set** (internal difference)

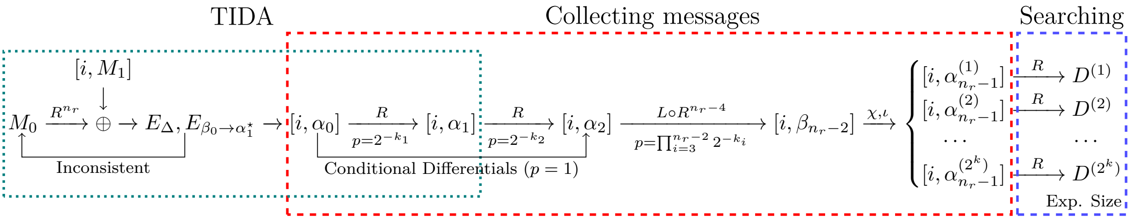
$$[i, v] \triangleq \{v + w \mid w \textit{ is symmetric with period } i\}$$

- The **zero internal difference** is the set of all symmetric states with period i

$$[i, \mathbf{0}] = \{w \mid w \textit{ is symmetric with period } i\}$$

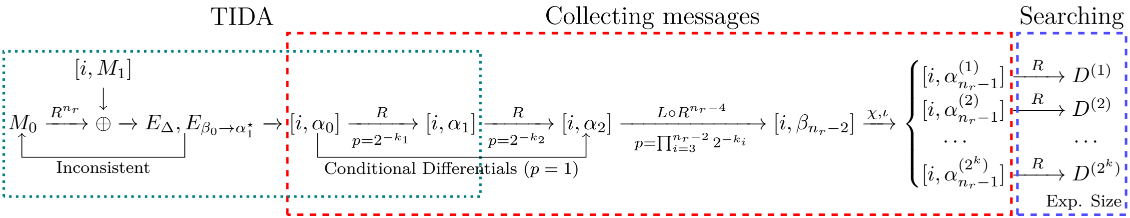
- The action of linear mapping L on any internal difference is equivalent to acting on the representative state

$$L([i, v]) = [i, L(v)]$$



5-round collision attack on SHA-3

- Select M_0 and M_1 by TIDA such that the state enters the internal difference of the second round in a given characteristic
- Calculate the internal difference of M after 4 rounds of round function and store the state into the subset $[i, v_j^{(4)}]$
- Calculate the collision subset of each subset $[i, v_j^{(4)}]$ in turn until one collision is found in a certain collision subset $D^{(j)}$



TIDA

Collecting messages

Searching

$[i, M_1]$

$$M_0 \xrightarrow{R^{n_r}} \oplus \rightarrow E_{\Delta}, E_{\beta_0 \rightarrow \alpha_1^{\star}} \rightarrow [i, \alpha_0] \xrightarrow[p=2^{-k_1}]{R} [i, \alpha_1] \xrightarrow[p=2^{-k_2}]{R} [i, \alpha_2] \xrightarrow[p=\prod_{i=3}^{n_r-2} 2^{-k_i}]{L\circ R^{n_r-4}} [i, \beta_{n_r-2}] \xrightarrow{\chi,\iota} \begin{cases} [i, \alpha_{n_r-1}^{(1)}] \xrightarrow{R} D^{(1)} \\ [i, \alpha_{n_r-1}^{(2)}] \xrightarrow{R} D^{(2)} \\ \cdots \qquad \cdots \\ [i, \alpha_{n_r-1}^{(2^k)}] \xrightarrow{R} D^{(2^k)} \end{cases}$$

Inconsistent

Conditional Differentials $(p = 1)$

Exp. Size

Time complexity $max\{2^{n_1}, 2^{n_2+(k+s)/2}\}$

- Time of TIDA 2^{n_1}.
- Time of obtaining one state passing the internal differential characteristic 2^{n_2}.
- Time of searching collision in the collision subset $2^{(k+s)/2}$.

Outline

1. Motivation
2. Overview of the Attack
3. Basic Techniques
    - Probabilistic Linearization
    - Constructing Internal Differential Characteristics
    - The Expected Size of Collision Subset
4. Results and Summary

TIDA - Connector and Connectivity Problem

- An n_1-round connector of two-block message (M_0, M_1) in a collision attack on n_r-round `SHA-3`:
    - The last $(c+p)$-bit difference input to the first round is fixed;
    - The last $(c+p)$-bit value of the initial state is fixed;
    - The output difference after n_1 round should be equal to the target difference.
- Internal connectivity problem:

$$\Delta(\mathtt{R}(\mathtt{R}^{n_r}(M_0||0^c) \oplus (\overline{M_1}||0^c))) = \alpha_1$$

- TIDA: Transforming internal connectivity problem into a linear system.
- Input difference system: The linear system with respect to the input difference of the first round is the input difference system, regarded as E_Δ. After applying Gaussian elimination to E_Δ, the equations related only to the inner bits are called the inner part of E_Δ, denoted as $E_\mathcal{C}$.

- t-Dimensional Affine Subspace

$$U = \left\{ (x_0, \ldots, x_4) \;\middle|\; \begin{array}{l} \sum_{j=0}^{4} l_j^{(1)} \cdot x_j = q^{(1)}, \\ \vdots \\ \sum_{j=0}^{4} l_j^{(n-t)} \cdot x_j = q^{(n-t)} \end{array} \right\} \triangleq \mathrm{Ker}\left(\sum_{j=0}^{4} l_j^{(1)} \cdot 2^j, q^{(1)} \middle| \ldots \middle| \sum_{j=0}^{4} l_j^{(n-t)} \cdot 2^j, q^{(n-t)} \right)$$

- Difference Density

$$\mathrm{P}(U, \delta_{out}) = \#\{\delta \in U | \delta \to \delta_{out}\} / |U|$$

- Maximum Difference Density Subspace (4-dimensional)

$\delta_{out} = 0x05$

δ_{in}	0x04	0x06	0x07	0x0f	0x11	0x16	0x17	0x19	0x1b	0x1d
Ker(0x01,1)			✓	✓	✓		✓	✓	✓	✓

- When building the input difference system, we select the 4-dimensional affine subspace that contains the most input differences, which is equivalent to adding one linear equation to the system E_Δ for each active Sbox.

Improved Target Internal Difference Algorithm

- Application to 5-round SHA3-384
    - The number of active Sboxes in first χ is 77.
    - TIDA [ZHL23]:
        - $\#E_{\Delta} = 77 * 3 + 83 * 5 = 646$, $\#E_{\mathcal{C}} = 234$
        - Time complexity of obtaining an input difference is 2^{234}
    - This work:
        - $\#E_{\Delta} = 77 * 1 + 83 * 5 = 492$, $\#E_{\mathcal{C}} = 97$
        - Probability $p_1 = 2^{-64.61}$
        - Time complexity of obtaining an input difference is $2^{97+64.61} = 2^{161.61}$

* p_1 is the Probability of the solution of E_{Δ} being the input difference of target difference.

Constructing Internal Differential Characteristics

- **Guideline 1**: The probability of first round differential transition should not be too small.
- **Guideline 2**: The inner part of system E_{Δ} should not have too many equations.

$$\textbf{Guideline 1} \xrightarrow{\text{LESS}} \#AS \xleftarrow{\text{MORE}} \textbf{Guideline 2}$$

- 5-round SHA3-384 internal differential characteristic

$$(\#AS, k_2, k_3, k_4) = (77, 25, 18, 16).$$

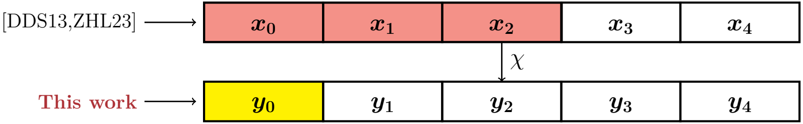
AS is the number of active Sboxes in first χ, the differential transition probability is 2^{-k_i}.

The Expected Size of Collision Subset

The output is the first d bits of the final state

- For $d = 64$, the first output lane depend on the first 3 input lanes
  The size of collision subset is bounded by 2^{56} instead of 2^{3i}

$$y_0 = x_0 \oplus (x_1 + 1)x_2$$



The Expected Size of Collision Subset

The output is the first *d* bits of the final state

collision length	pre size	expected size	collision length	pre size	expected size
1 lane	2^{96}	$\mathbf{2^{56}}$	6 lanes	2^{256}	$\mathbf{2^{216}}$
2 lanes	2^{128}	$\mathbf{2^{98.64}}$	7 lanes	2^{288}	$\mathbf{2^{258.64}}$
3 lanes	2^{160}	$\mathbf{2^{136.58}}$	8 lanes	2^{320}	$\mathbf{2^{296.58}}$
4 lanes	2^{160}	$\mathbf{2^{154.64}}$	9 lanes	2^{320}	$\mathbf{2^{314.64}}$
5 lanes	2^{160}	$\mathbf{2^{160}}$	10 lanes	2^{320}	$\mathbf{2^{320}}$

Table: The expected size for collision length no more than 640 bits

Outline

1. Motivation
2. Overview of the Attack
3. Basic Techniques
4. Results and Summary
    - An Example of the Collision
    - Summary of Attacks on SHA-3

Collision in Keccak[240,160,5,96]

$M_0 =$

```
|6a44|1a51|321-|23f2|66f6|
|41-9|7d57|a1b2|7d8c|1a7a|
|6d-f|c9aa|d211|b134|f229|
|----|----|----|----|----|
|----|----|----|----|----|
```

$$\downarrow R^5$$

$R^5(M_0) =$

```
|169d|-51e|5252|7437|dc49|
|629a|9762|e2--|8eae|5a93|
|67-5|8-dc|c829|f4c4|-aea|
|fbff|bb15|95ee|3b2f|7b41|
|--93|54ba|a9ce|5e4a|4779|
```

$M_1 =$ (left) $= M_1'$ (right)

```
|b9a5|-d74|35f2|51be|f816|  |617d|86ff|df18|fa15|9876|
|9a53|7251|e164|ebe5|482b|  |8a43|e4c7|ca4f|cdc3|482b|
|5--3|14a7|e3cd|-b83|8614|  |683b|2c9f|d2fc|c54d|8c1e|
|----|----|----|----|----|  |----|----|----|----|----|
|----|----|----|----|----|  |----|----|----|----|----|
```

$$\downarrow \oplus R^5(M_0) \qquad \downarrow \oplus R^5(M_0)$$

```
|af38|-86a|67a-|2589|245f|  |77e-|83e1|8d4a|8e22|443f|
|f8c9|e533|-364|654b|12b8|  |e8d9|73a5|284f|436d|12b8|
|37-6|947b|2be4|ff47|8cfe|  |-f3e|ac43|1ad5|3189|86f4|
|fbff|bb15|95ee|3b2f|7b41|  |fbff|bb15|95ee|3b2f|7b41|
|--93|54ba|a9ce|5e4a|4779|  |--93|54ba|a9ce|5e4a|4779|
```

$$\downarrow R^5 \qquad \downarrow R^5$$

```
|5992|37b4|27ce|9981|b9eb|  |5992|37b4|27ce|9981|b9eb|
|e7e5|81a7|eafc|9a8e|6ef8|  |e7e5|3197|37a5|-f1b|25b3|
|e4a9|8b81|8264|187b|e9e9|  |9ed2|fdb-|2baf|4665|a9a9|
|7826|9f-9|a72d|e5bf|3e62|  |ece4|e96d|d75f|--58|7e34|
|-ba6|f6d7|db68|84ce|7744|  |8ba7|-cad|997a|--d1|6af4|
```

- Internal differential characteristic

$$(AS, k_2, k_3, k_4) = (18, 14, 8, 7)$$

- Theoretical time complexity:

$$2^{k_2-9+k_3+(k_4+54)/2-1} = 2^{42.5}$$

- Experimental time complexity: 2^{43}
- Run time: 17 hours (Intel Core i9-13900KF, 32 threads)

Results of Attacks on Reduced SHA-3

- Complexity: $2^{k_3/2} \cdot 2^{s/2}$ (4-round) and $2^{k_3+k_4/2} \cdot 2^{s/2}$ (5-round)

Target	n_r	i	k_1	k_2	k_3	k_4	k_5	Complexity $(\log_2)$
`SHA3-512`	4	32	16	16	170	-	-	225.29
`SHA3-224/256/SHAKE128`	5	32	-	21	18	16	-	96.67
`SHA3-384`	5	32	-	25	18	16	-	170.73
`SHAKE256`	5	32	-	21	18	16	-	163.28
`SHAKE256`	6	32	-	31	25	20	83	232.29

Table: The parameters of characteristics and complexities

Summary and Future Work

- Summary
    - Utilize probabilistic linearization technique to find collisions for up to 6 rounds of all the six SHA-3 functions
    - Present the **first collision attacks** on 5-round **SHA3-384** and 6-round **SHAKE256**
    - and the best collision attack on 4-round **SHA3-512**
- Future work
    - Find **better** internal differential characteristics
    - Apply internal differential analysis to other ciphers

Thank you for your attention!