Preimage Attacks on up to 5 Rounds of SHA-3 Using Internal Differentials

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

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

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

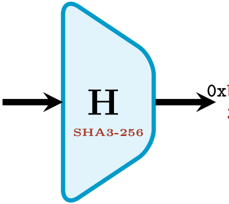
2025.05.08

Outline

Outline

Hash Function

On dirait que le soleil a soif. Une averse est un verre d'eau; une pluie est tout de suite bue. Le matin tout ruisselait, l'après-midi tout poudroi.



`0xbe9d17d323bc17d7935e91bc6b7086d0`
`3c8dffc9fbc29b41d5e4bcb7f4f81a92`

- A cryptographic hash function is a mathematical algorithm that maps an **arbitrary length input** (the message M) to a **fixed length d-bit output**.
- Security properties
    - Pre-image resistance
      - Given digest h, it should be difficult to find message m such that $H(m) = h$
    - Second pre-image resistance
    - Collision resistance

Keccak

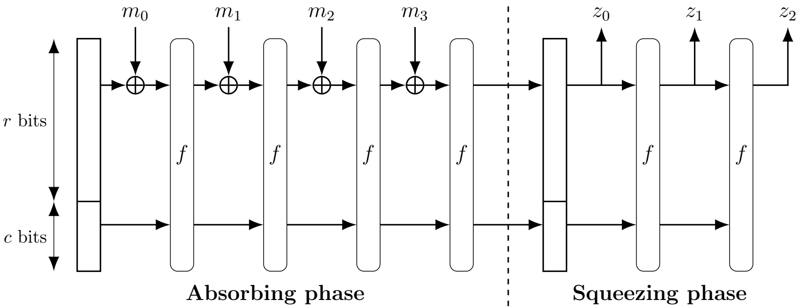
- NIST SHA-3 hash function competition (2007-2012)
- Designers: Guido Bertoni, Joan Daemen, Michaël Peeters and Gilles Van Assche
- Submitted to SHA-3 competition in 2008
- The winner was announced to be Keccak in 2012
- In 2015, Keccak was standardized by NIST as SHA-3
  - SHA3-224/256/384/512
  - SHAKE128/256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

Structure of Sponge construction



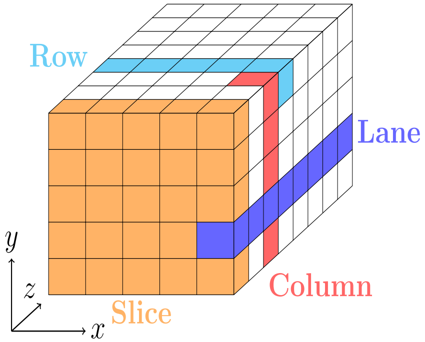
- b-bit permutation Keccak-f, f contains 24 rounds
- Two parameters: bitrate r and capacity c, $b = r + c$
- The message is padded and then split into r-bit blocks

The Internal State

- 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



Preimage Attacks on Round-Reduced SHA-3

- [MPS13]: Preimage attacks under rotational cryptanalysis
  First preimage attack on 4-round Keccak
- [GLS16]: A 2-round linear structure of Keccak round function
  One of the major cryptanalytic tools for security evaluation of SHA-3
- [Dinur21]: A polynomial method-based algorithm for solving equation systems
  Achieved the best 4-round attack on Keccak-384 and Keccak-512
- [Bernstein10] and [CKMS14]: Algebraic techniques to speed up a brute-force (second) preimage search for up to 9 rounds of Keccak

* The best known preimage attacks with more than a tiny advantage over a brute-force search currently only reach 4 rounds

Functions	SHA3-224	SHA3-256	SHA3-384	SHA3-512	SHAKE128	SHAKE256
Security Strengths in Bits	224	256	384	512	$\min(d, 128)$	$\min(d, 256)$

Preimage Attacks on Round-Reduced SHA-3

Methods	SHA3-224	SHA3-256	SHA3-384	SHA3-512	SHAKE128	SHAKE256
Rotational [MPS13]	4 (2^{221})	4 (2^{252})	4 (2^{378})	4 (2^{506})	-	-
Linear Structure [GLS16]	4 (2^{213})	4 (2^{251})	3 (2^{322})	3 (2^{482})	4 (2^{106})	4 (2^{251})
Linear Structure [LS19]	4 (2^{207})	4 (2^{239})	-	-	-	4 (2^{239})
Linear Structure [HLY21]	4 (2^{192})	4 (2^{218})	-	-	-	-
Solving Poly. [Dinur21]	4 ($2^{217\dagger}$)	4 ($2^{246\dagger}$)	4 ($2^{374\dagger}$)	4 ($2^{502\dagger}$)	-	-
MITM [QHD+23]	-	-	-	4 ($2^{504.58}$)	-	-
Squeeze-MITM This work	4 ($2^{135.5}$) 5 ($2^{216.03\dagger}$)	4 ($2^{151.5}$) 5 ($2^{254.33\dagger}$)	4 ($2^{277.8}$)	3 ($2^{504.2}$)	4 ($2^{81.5}$) 5 ($2^{100.5}$)	4 ($2^{151.5}$) 5 ($2^{254.33\dagger}$)

† Bit operations

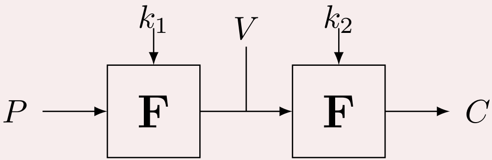
Outline

Meet-in-the-Middle (MITM) Attack [DH77]

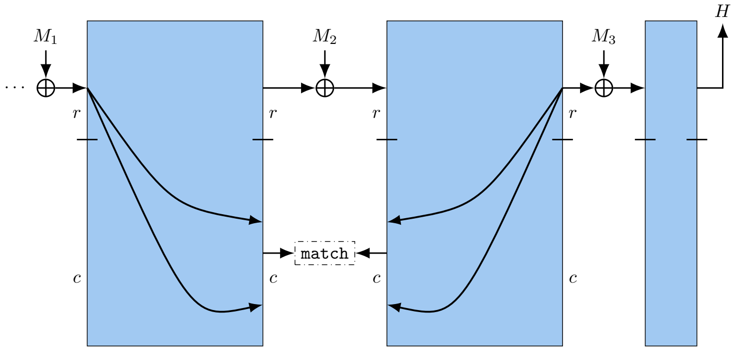
- Proposed by Diffie and Hellman in 1977
- An efficient exhaustive search way based on the birthday attack

Example: Double Encryption

- $C = E_K(P) = F_{K_2}(F_{K_1}(P)), K = K_1||K_2$
- The time complexity: $2^{|K_1|+|K_2|} \rightarrow 2^{|K_1|+|K_2|-n}$
- Meet in the middle: Store $F_{K_1}(P)$ in V and match $F_{K_1}(P) = F_{K_2}^{-1}(C)$



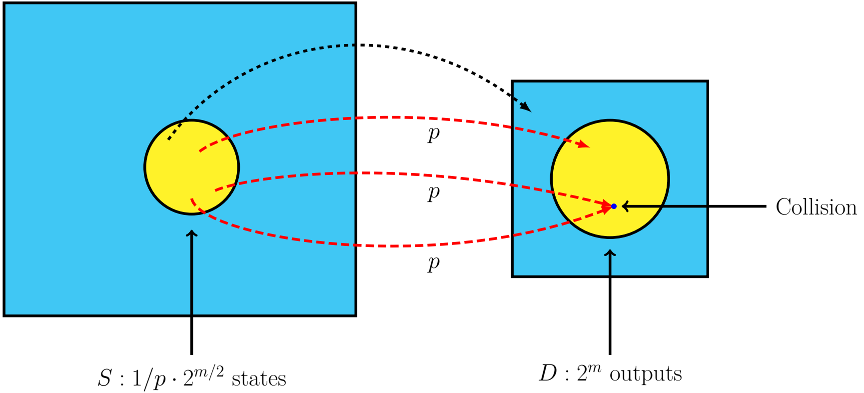
MITM Preimage Attacks on Sponge Constructions [BDPV2008]



- $H = f(M_3 + f(M_2 + f(M_1)))$, multiple blocks of message $M_1||M_2||M_3$
- Meet in the middle: Store $f(M_1)$ in V, match $f(M_1)$ and $f^{-1}(f^{-1}(H) + M_3)$ in the capacity part
- Set the value of the middle block $M_2 = f(M_1) + f^{-1}(f^{-1}(H) + M_3)$

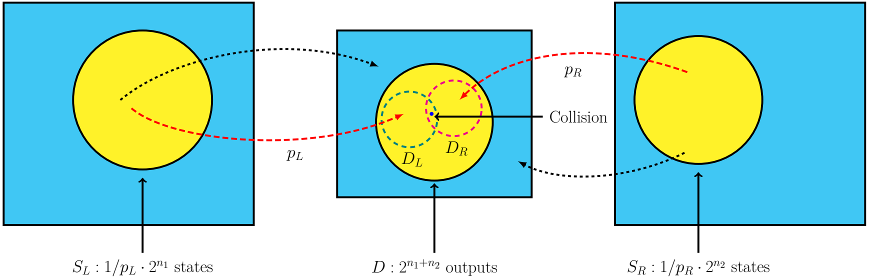
Squeeze Attack [DDS 2013]

- Proposed by Dinur, Dunkelman and Shamir in 2013 to perform collision attacks
- An input subset S is mapped with probability p to the output subset D
- The time complexity of the attack is $1/p \cdot \sqrt{|D|}$

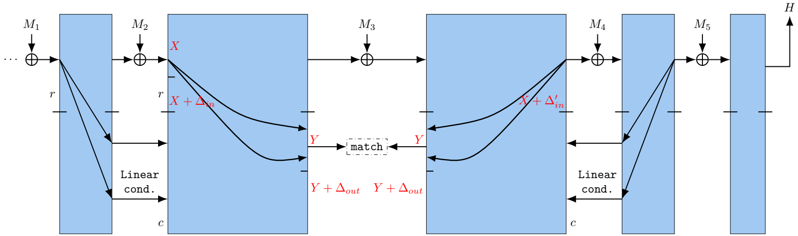


Squeeze Meet-in-the-Middle Attack (This work)

- An input subset $S_L(S_R)$ is mapped with probability $p_L(p_R)$ to the output subset $D_L(D_R)$
- Use squeeze attack to find a collision in the middle subset D, $|D| = |D_L| \cdot |D_R|$
- The time complexity of the attack is $1/p_L \cdot |D_L| + 1/p_R \cdot |D_R|$



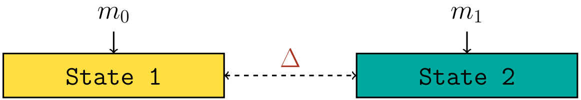
Squeeze MITM Preimage Attacks on Sponge Constructions



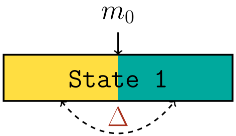
- $H = f(M_5 + f(M_4 + f(M_3 + f(M_2 + f(M_1)))))$
- Meet in the middle: Store $f(M_2 + f(M_1))$ in V, match $f(M_2 + f(M_1))$ and $f^{-1}(f^{-1}(f^{-1}(H) + M_5) + M_4)$ in the capacity part
- Set the value of $M_3 = f(M_2 + f(M_1)) + f^{-1}(f^{-1}(f^{-1}(H) + M_5) + M_4)$
- Preset the internal difference to launch the squeeze attack

Internal Differential Cryptanalysis

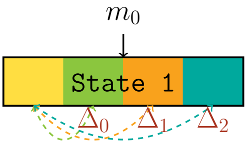
- Standard differential cryptanalysis [BS91]:



- Internal differential cryptanalysis [Peyrin10]:



- Generalized internal differential cryptanalysis [DDS13]:



Symmetric States

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

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

- 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

2025202520252025	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)

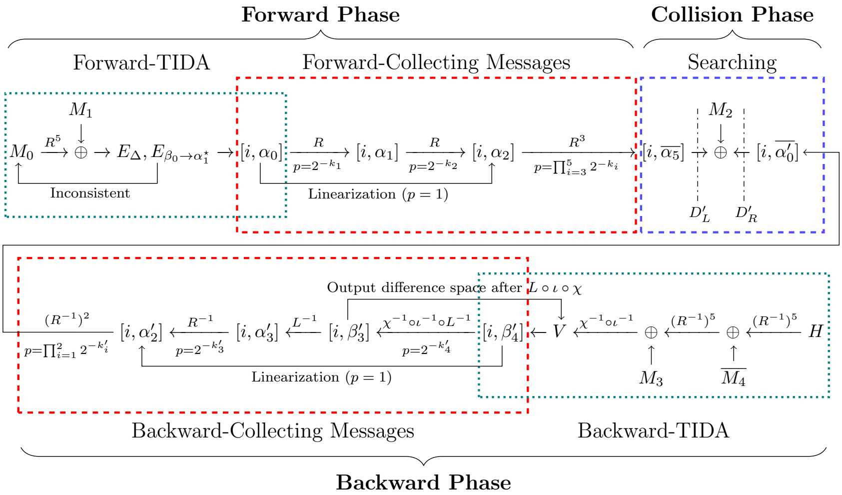
$$[i, v] \triangleq \{v + w \mid w \textit{ is symmetric with period } i\}$$

- 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 preimage attack on SHA-3

- Search for $f(f(M_0) \oplus M_1)$ and $f^{-1}(f^{-1}(f^{-1}(H) \oplus M_4) \oplus M_3)$ with the same capacity.

Outline

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:

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

$$\Delta(\mathtt{R}^{-1}(\mathtt{R}^{-n_r}(M_0||0^c) \oplus (\overline{M_1}||0^c))) = \alpha_1'$$

- Forward-TIDA: Transforming internal connectivity problem into linear systems.
- Backward-TIDA: Transforming internal connectivity problem into lookup tables.

Backward Target Internal Difference Algorithm

Definition (Value-Difference Distribution Table)

Set Sbox $\mathtt{S} : \mathbb{F}_2^5 \to \mathbb{F}_2^5$. Given any input difference δ_{in} and $2t$-bit value (y, y'), $\delta_{in} \in \mathbb{F}_2^5, y, y' \in \mathbb{F}_2^t$, the entry $\mathtt{VDDT}(\alpha, y, y')$ in the value-difference distribution table records the number of elements in the set $\{(x, x') \in \mathbb{F}_2^{2\times(5-t)} | \mathtt{S}(x||y) + \mathtt{S}(x'||y') = \delta_{in}\}$.

- In the attack, $\mathtt{S}$ in the definition is replaced by $\mathtt{S}^{-1}$ in the backward phase.
- Taking SHA3-384 as an example,
  - Randomly select the fifth block M_4 to get the output difference that matches the last two planes in the backward internal difference characteristic.
  - Calculating the probability of a complete match by check the VDDT with 2t-bit value of each Sbox in the third plane, where $t = 2$.

Inverse Sbox Linearization

Observation

Given $\delta_{in}, \delta_{out} \in \mathbb{F}_2^5$, denote the set $V = \{y = (y_0, ..., y_4) : \mathtt{S}^{-1}(y) + \mathtt{S}^{-1}(y + \delta_{out}) = \delta_{in}\}$ and $\mathtt{S}^{-1}(V) = \{\mathtt{S}^{-1}(y) : y \in V\}$, we have

a. if $\mathtt{DDT}(\delta_{in}, \delta_{out}) = 2$ or 4, then V is an inverse-linearizable affine subspace.

b. if $\mathtt{DDT}(\delta_{in}, \delta_{out}) = 8$, then there are two 2-dimensional inverse-linearizable subspaces $W_i \subset V$, $i = 0, 1$, such that $W_0 \cup W_1 = V$. And there are two independent linear conditions L_1 and L_2, all elements in V satisfy L_1, and 6 elements satisfy L_2.

δ_{out}	V	L_1	L_2
0x01	0x14,0x15,0x11,0x10,0x1a,0x1b,0x1d,0x1c	$y_4 = 1$	$y_1 = 0$
0x11	0x06,0x17,0x02,0x13,0x08,0x19,0x0e,0x1f	$y_4 + y_0 = 0$	$y_1 = 1$
0x09	0x00,0x09,0x05,0x0c,0x0a,0x03,0x0d,0x04	$y_4 = 0$	$y_1 = 0$
0x19	0x12,0x0b,0x16,0x0f,0x18,0x01,0x1e,0x07	$y_4 + y_0 = 1$	$y_1 = 1$

Inverse Sbox Linearization

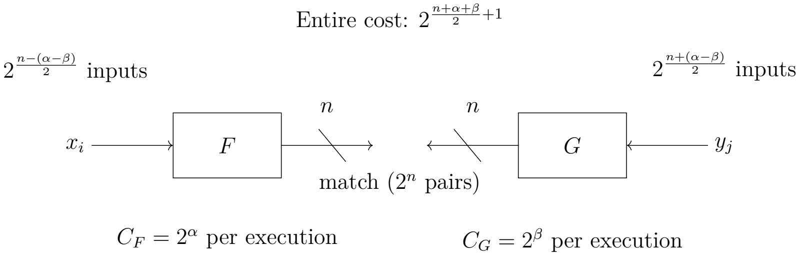
Definition (Inverse-Linearizable affine subspace)

Inverse-Linearizable affine subspaces are affine output subspaces on which the inverse of Sbox substitution is equivalent to a linear transformation. If V is a inverse-linearizable affine subspaces of an Sbox operation $\mathtt{S}(\cdot)$, $\forall y \in V$, $\mathtt{S}^{-1}(y) = A \cdot y + b$, where A is a matrix and b is a constant vector.

$$\mathtt{S}^{-1}(y) = \begin{pmatrix} (y_3y_4 + y_2 + y_4)\overline{y_1} + y_0 \\ (y_4y_0 + y_3 + y_0)\overline{y_2} + y_1 \\ (y_0y_1 + y_4 + y_1)\overline{y_3} + y_2 \\ (y_1y_2 + y_0 + y_2)\overline{y_4} + y_3 \\ (y_2y_3 + y_1 + y_3)\overline{y_0} + y_4 \end{pmatrix} \xrightarrow{\begin{pmatrix} y_0 = 0 \\ y_2 = 0 \\ y_1 = y_3 + y_4 \end{pmatrix}} \begin{pmatrix} 0 \\ y_4 \\ 0 \\ y_3 \\ 0 \end{pmatrix}.$$

* Add at least 3 linear conditions

Unbalanced MITM Attacks [Sasaki2014]



1. For $j = 1, ..., 2^{(\alpha-\beta)/2}$, compute $G(y_j)$, and store the result in a list L where the data is indexed by $G(y_j)$.
2. Set For $j = 1, ..., 2^{(\alpha+\beta)/2}$, compute $F(x_i)$, and search for a match with $F(y_i)$ in the list L.

* Adjust the number of messages required in the forward and backward phases to minimize the overall complexity of the attack.

Outline

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

Results of Attacks on Reduced SHA-3

- Complexity: $2^{1+(k_3+k_4+k_5+k_1'+k_2')/2} \cdot 2^{c/4}$

<table>
<tr><th>Target</th><th>n_r</th><th>k_2</th><th>k_3</th><th>k_4</th><th>k_5</th><th>k_1'</th><th>k_2'</th><th>k_3'</th><th>k_4'</th><th>Compl. ($\log_2$)</th></tr>
<tr><td>SHA3-512</td><td>3</td><td>7</td><td>3</td><td>-</td><td>-</td><td>4</td><td>64</td><td>-</td><td>-</td><td>504.2</td></tr>
<tr><td>SHAKE128</td><td>4</td><td>24</td><td>22</td><td>0</td><td>-</td><td>11</td><td>12</td><td>239</td><td>-</td><td>81.5</td></tr>
<tr><td>SHA3-224</td><td>4</td><td>24</td><td>22</td><td>0</td><td>-</td><td>11</td><td>12</td><td>239</td><td>-</td><td>135.5</td></tr>
<tr><td>SHA3-256/SHAKE256</td><td>4</td><td>24</td><td>22</td><td>0</td><td>-</td><td>11</td><td>12</td><td>239</td><td>-</td><td>151.5</td></tr>
<tr><td>SHA3-384</td><td>4</td><td>25</td><td>18</td><td>5</td><td>-</td><td>33</td><td>48</td><td>210</td><td>-</td><td>277.8</td></tr>
<tr><td>SHAKE128</td><td>5</td><td>155</td><td>34</td><td>15</td><td>11</td><td>7</td><td>4</td><td>98</td><td>399</td><td>100.5</td></tr>
<tr><td>SHA3-224</td><td>5</td><td>158</td><td>30</td><td colspan="2">17.64</td><td>13</td><td>8</td><td>88</td><td>384</td><td>216.03</td></tr>
<tr><td>SHA3-256/SHAKE256</td><td>5</td><td>158</td><td>30</td><td colspan="2">17.64</td><td>13</td><td>8</td><td>88</td><td>384</td><td>254.33</td></tr>
</table>

Table: The parameters of characteristics and complexities

Preimage in Keccak[704,96,4,800]

M_0

```
|4477e7e-|f34d-eb2|4c-f5416|32f84625|856e5ca4|
|14d6ca2b|84c5ee95|89447339|df8afe-2|25ba93-9|
|2-3a416a|4c269ac9|d683d724|54736fee|83914e3c|
|3a971-95|ea84e-71|5b3e9e24|e615-cc3|3897c99c|
|cf799c75|1db9a6f3|
```

$\downarrow R^4$

```
|e36dcbea|155-55-2|2765-778|3c229-3e|7-1ef815|
|-55785da|f629fffa|4c6ea-9f|e6-e-5a1|aa27-45d|
|c-58f97a|526d-b49|f77238-1|7dfc43-4|ee87875f|
|45366383|d36febcb|6d125fb7|54d9fc8c|f75de98b|
|2-6fabaf|-d8a11b8|b-a-b-e-|-77f398b|5b1a8--3|
```

$\oplus \leftarrow M_1$

```
|fdc352de|5462c379|2c41b666|f3f-6-b7|c757cc58|
|6-619d38|5acc46e6|d4e36-89|7b-fe876|2578b266|
|-b-f1561|97fb3844|47ce3c6d|-e255445|8db-7b6f|
|cd5a6481|3d391e53|152da7-2|7968a323|9-e88-9-|
|88eba33b|ed19e12b|
```

$\downarrow R^4$

```
|e6e7e66d|9fed9eed|c253c253|8-57815c|f98bf9--|
|78657835|38-73847|ebcf6bdf|3b6f3b6f|8ce1-ca1|
|e91ce91c|6a996399|92789258|db58d358|75a87588|
|f514fd14|2-cf28cf|263-2e3-|a359ab59|789f7-9f|
|47394739|faecfaec|1-631-63|5a9-5a9-|4b244b24|
```

$\oplus \leftarrow M_2$

```
|ce-1ce8b|2-232123|d427d4-7|d642d759|3c6b3cf-|
|7ea87ef8|79157955|71dff1ef|16af16af|22f-a2a-|
|5bfe5bfe|ebdbe2db|c-16c-16|e315eb15|fbedfbdd|
|899a819a|af6fa76f|43724b52|a837a-37|769a7e8a|
|fe3efe3e|fb13fb13|
```

$\downarrow R^4$

```
|aa85eb77|c3c45cec|5-2d6964|35e9cd47|87e3c87a|
|7de--4-b|6531-c35|3628b-19|9e2175dc|9f29f151|
|718d-7b8|2a58171-|5176bfa4|f1db32b7|8d8f1697|
|1bd5-5-1|82d33e74|66bc5a64|84cc6583|63-21898|
|b646e981|-6c2-af2|27b2f468|8ebbcb5d|6be494d2|
```

$\oplus \leftarrow M_3$

```
|eb9b9334|fcb75626|cb165a43|7783963b|48c118-6|
|-9596525|c147b74e|d61c643e|8bac495e|-4a6c49f|
|1fdae537|45a8f-d3|dae1b176|eb19c7b9|3639663-|
|-e2a75b8|1dc2ea53|b3949b35|1999d84-|3718d67a|
|46e4b83b|b7b17-bd|
```

$\downarrow R^4$

```
|8e2186fd|116355-3|f9745742|5a429ab3|68da87f7|
|9efa48aa|de64d61e|d611f-19|91c38af8|ffb7d44c|
|d14721f-|69-7-af6|37-6fcc2|fdaa4269|2a457aed|
|adfad8ff|7e8-9aaf|d8-11a56|d-5b42af|6fbfdeb-|
|-6d24425|c53571dc|5987d-64|d87dca25|22-917d4|
```

$\oplus \leftarrow \overline{M_4}$

```
|695-14e9|2a748ec6|ffffffff|ffffffff|ffffffff|
|ffffffff|ffffffff|ffffffff|ffffffff|ffffffff|
|ffffffff|ffffffff|ffffffff|ffffffff|ffffffff|
|ffffffff|ffffffff|ffffffff|ffffffff|ffffffff|
|ffffffff|ffffffff|
```

$\downarrow R^4$

H

```
|ffffffff|ffffffff|ffffffff|ffffffff|ffffffff|
|ffffffff|ffffffff|ffffffff|ffffffff|ffffffff|
|ffffffff|ffffffff|ffffffff|ffffffff|ffffffff|
|ffffffff|ffffffff|ffffffff|ffffffff|ffffffff|
|ffffffff|ffffffff|ffffffff|ffffffff|ffffffff|
```

- Digest is a full 1-bit string
- Internal differential characteristic

$$(k_2, k_3, \overline{k_4}, k'_1, k'_2, k'_3) = (27, 12.5, 0, 6, 8, 125)$$

- Theoretical time complexity:

$$2^{1+(k_3+\overline{k_4}+k'_1)/2+96/4} = 2^{34.25}$$

- Experimental time complexity: 2^{35}

Summary and Future Work

- Summary
  - Utilize squeeze MITM to find preimage for up to 5 rounds of all the six SHA-3 functions
  - Present the **first preimage attacks** on 5-round **SHA3-224/SHA3-256/SHAKE128/SHAKE256**
  - and the **best preimage attacks** on 4-round **SHA3-224/SHA3-256/SHA3-384/SHAKE128/SHAKE256**
- Future work
  - Find **better** internal differential characteristics
  - Apply internal differential analysis to other ciphers

Thank you for your attention!