Cryptanalysis of Round-Reduced KECCAK Using Non-linear Structures

Mahesh Sreekumar Rajasree(✉)

Center for Cybersecurity, Indian Institute of Technology Kanpur, Kanpur, India
mahesr@cse.iitk.ac.in

Abstract. In this paper, we present new preimage attacks on KECCAK-384 and KECCAK-512 for 2, 3 and 4 rounds. The attacks are based on non-linear structures (structures that contain quadratic terms). These structures were studied by Guo et al. [13] and Li et al. [18,19] to give preimage attacks on round reduced KECCAK. We carefully construct non-linear structures such that the quadratic terms are not spread across the whole state. This allows us to create more linear equations between the variables and hash values, leading to better preimage attacks. As a result, we present the best theoretical preimage attack on KECCAK-384 and KECCAK-512 for 2 and 3-rounds and also KECCAK-384 for 4-rounds.

Keywords: KECCAK · SHA-3 · Hash function · Cryptanalysis · Preimage attack

1 Introduction

Cryptographic hash functions are widely used in modern cryptography such as in digital signatures, message integrity and authentication. The U.S. National Institute of Standards and Technology (NIST) announced the "NIST hash function competition" for the Secure Hash Algorithm-3 (SHA-3) in 2006. They received 64 proposals from around the world. Among these, KECCAK designed by Bertoni, Daemen, Peeters, and Van Assche [4] became one of the candidates for SHA-3. It won the competition in October 2012 and was standardized as a "Secure Hash Algorithm 3" [12].

The KECCAK hash family is based on the sponge construction [5]. Its design was made public in 2008 and since then, it has received intense security analysis. In 2016, Guo et al. [13] formalised the idea of linear structures and gave practical preimage attacks for 2 rounds KECCAK-224/256. They also gave better preimage attacks for KECCAK-384/512, all variants of 3-rounds KECCAK as well as preimage attacks for 4-rounds KECCAK-224/256. Li et al. [19] improved the complexity of preimage attack for 3-rounds KECCAK-256 by using a new type of structure called cross-linear structure. The best-known attacks for 3 and 4 rounds KECCAK-224/256 are given by Li et al. [18] using a new technique called allocating approach, which consists of two phases - Precomputation phase

F. Hao et al. (Eds.): INDOCRYPT 2019, LNCS 11898, pp. 175–192, 2019.
https://doi.org/10.1007/978-3-030-35423-7_9

and Online phase. They gave the first practical preimage attack for 3-rounds KECCAK-224. Theoretical preimage attacks for higher rounds on KECCAK are considered in [2,7,20]. Apart from the attacks mentioned above, there are several other attacks against KECCAK such as preimage attacks in [16,17,21,22], collision attacks in [8–10,15,23] and distinguishers in [1,6,11,13,14].

Table 1. Summary of preimage attacks

Rounds	Instances	Complexity	References
1	224	Practical	[17]
	256		
	384		
	512		
2	224	Practical	[13]
	256	Practical	
	384	2^{129}	
	512	2^{384}	
2	384	Time 2^{89}	[16]
		Space 2^{87}	
2	384	$\mathbf{2^{113}}$	Subsection 3.2
	512	$\mathbf{2^{321}}$	Subsection 3.1
3	224	2^{38}	[18]
	256	2^{81}	
3	384	2^{322}	[13]
	512	2^{482}	
3	384	$\mathbf{2^{321}}$	Subsection 3.4
	512	$\mathbf{2^{475}}$	Subsection 3.5
4	224	2^{207}	[18]
	256	2^{239}	
4	384	2^{378}	[20]
	512	2^{506}	
4	384	$\mathbf{2^{371}}$	Subsection 3.6

Our Contributions: In this paper, we give the best theoretical preimage attacks for KECCAK-384 for $2, 3, 4$ rounds and KECCAK-512 for $2, 3$ rounds. This is achieved by carefully constructing non-linear structures such that the quadratic terms are not spread throughout the whole state and the number of free variables in the system of equations is more. Table 1 summaries the best theoretical preimage attacks up to four rounds and our contributions. The space complexity is most of the attacks is constant unless it is explicitly mentioned.

Organization: The rest of the paper contains the following sections. In Sect. 2, we will give a brief description about KECCAK, some preliminaries and notations that are used throughout the paper and useful observations about KECCAK. Section 3 contains detailed description of all our preimage attacks. Finally, we conclude in Sect. 4.

2 Structure of KECCAK

KECCAK hash function is based on sponge construction [5] which uses a padding function *pad*, a bitrate parameter r and a permutation function f as shown in Fig. 1.

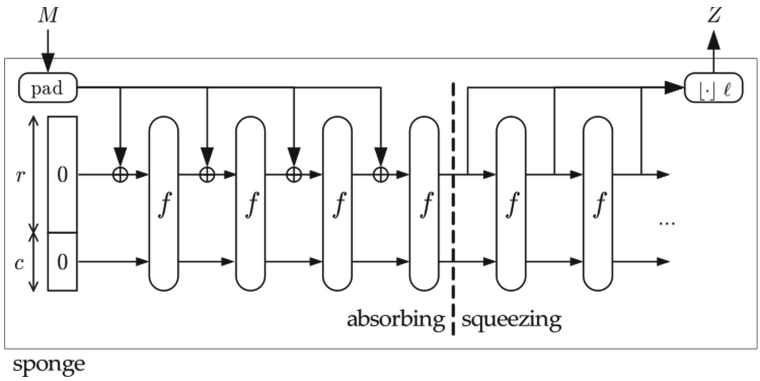

Fig. 1. Sponge function [5]

2.1 Sponge Construction

As shown in Fig. 1, the sponge construction consists of two phases - absorbing and squeezing. It first applies the padding function *pad* on the input string M which produces M' whose length is a multiple of r. In the absorbing phase, M' is split into blocks of r bits namely $m_1, m_2, ...m_k$. The initial state (IV) is a b bit string containing all 0. Here $b = r + c$ where c is called the capacity. The first r bits of IV is XORed with first block m_1 and is given as input to f. The output is XORed with the next message block m_2 and then is given as input to f again. This process is continued till all the message blocks have been absorbed.

The squeezing phase extracts the required output, which can be of any length. Let ℓ be the required output length. If $\ell \leq r$, then the first ℓ bits of the output of absorbing phase is the output of the sponge construction. Whereas, if $\ell > r$, then more blocks of r bits are extracted by repeatedly applying f on the output of the absorbing phase. This process is repeated enough number of times until

we have extracted at least ℓ bits. The final output of the sponge construction is the first ℓ bits that have been extracted.

In the KECCAK hash family, the permutation function f is a KECCAK-$f[b]$ permutation, and the *pad* function appends 10^*1 to input M. KECCAK-f is a specialization of KECCAK-p permutation.

$$\text{KECCAK-}f[b] = \text{KECCAK-p}[b, 12 + 2\gamma]$$

where $\gamma = log_2(b/25)$.

The official version of KECCAK have $r = 1600 - c$ and $c = 2\ell$ where $\ell \in \{224, 256, 384, 512\}$ called KECCAK-224, KECCAK-256, KECCAK-384 and KECCAK-512.

2.2 KECCAK-p Permutation

KECCAK-p permutation is denoted by KECCAK-p$[b, n_r]$, where $b \in \{25, 50, 100, 200, 400, 800, 1600\}$ is the length of the input string and n_r is the number of rounds of the internal transformation. The parameter b is also called the width of the permutation. The b bit input string can be represented as a $5 \times 5 \times w$ 3-dimensional array known as state as shown in Fig. 2. A lane in a state S is denoted by $S[x, y]$ which is the substring $S[x, y, 0]|S[x, y, 1]| \ldots |S[x, y, w-1]$ where w is equal to $b/25$ and "|" is the concatenation function.

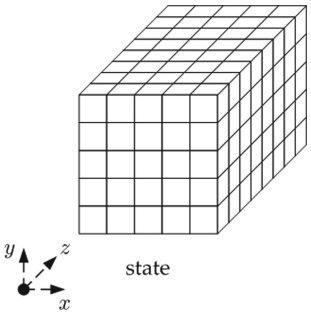

Fig. 2. KECCAK state [3]

In each round, the state S goes through 5 step mappings θ, ρ, π, χ and ι, i.e. $Round(S, i_r) = \iota(\chi(\pi(\rho(\theta(S)))), i_r)$ where i_r is the round index. Except for χ, rest of the step mappings are linear. In the following, S' is the state after applying the corresponding step mapping to S, "$\oplus$" denotes bitwise XOR and "$\cdot$" denotes bitwise AND.

1. θ: The θ step XOR's $S[x, y, z]$ with parities of its neighbouring columns in the following manner.

$$S'[x, y, z] = S[x, y, z] \oplus P[(x + 1) \ mod \ 5][(z - 1) \ mod \ 64] \\ \oplus P[(x - 1) \ mod \ 5][z]$$

where $P[x][z]$ is the parity of a column, i.e.,

$$P[x][z] = \bigoplus_{i=0}^{4} S[x, i, z]$$

2. ρ: The ρ step simply rotates each lane by a predefined value given in the table below, i.e.

$$S'[x, y] = S[x, y] << r[x][y]$$

where $<<$ means bitwise rotation towards MSB of the 64-bit word.

4	18	2	61	56	14
3	41	45	15	21	8
2	3	10	43	25	39
1	36	44	6	55	20
0	0	1	62	28	27
$y\backslash x$	0	1	2	3	4

3. π: The π step interchanges the lanes of the state S.

$$S'[y, 2x + 3y] = S[x, y]$$

4. χ: The χ step is the only non-linear operation among the 5 step mappings due to the quadratic term.

$$S'[x, y, z] = S[x, y, z] \oplus ((S[(x + 1) \ mod \ 5, y, z] \oplus 1) \cdot S[(x + 2) \ mod \ 5, y, z])$$

5. ι: The ι step is the only step that depends on the round number.

$$S'[0, 0] = S[0, 0] \oplus RC_i$$

where RC_i is a constant which depends on i where i is the round number.

2.3 Preliminaries and Notations

In this paper, we will be using the following observations made by Guo et al. [13]. The χ step mapping is a row dependent operation. Let a_0, a_1, a_2, a_3, a_4 be the 5 input bits to the χ operation and b_0, b_1, b_2, b_3, b_4 be the 5 output bits.

Observation 1. *Let d_0, d_1, d_2, d_3, d_4 be the elements of a column. Then, the parity of column can be fixed to a constant c by choosing for any $i \in \{0, 1, 2, 3, 4\}$*

$$d_i = c \oplus \left(\bigoplus_{j=1}^{j=4} d_{i+j} \right)$$

Observation 2. *If the output of χ for an entire row is known, i.e. $\chi([a_0, a_1, a_2, a_3, a_4]) = [b_0, b_1, b_2, b_3, b_4]$, then we have*

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

Observation 3. *If we are given two consecutive bits b_i, b_{i+1} of the output of χ, we can set up the following linear equation on the input bits.*

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

In the rest of the paper, all the message variables and hash values are represented in the form of lanes (array) of length 64, and we will use + symbol in place of $\oplus$. For a state A, $A[x, y]$ denotes a lane where $0 \leq x, y \leq 4$. In all the equations, the value inside the brackets '()' indicates the offset by which the lane is shifted. For example, $A[x, y](k)$ denotes lane $A[x, y]$ rotated by an offset of k. Every operation between two lanes is bitwise.

3 Our Preimage Attacks

In this section, we present the preimage attacks for round reduced KECCAK. In [13], the authors try to set up linear equations between message bits (variables) and hash bits by controlling the diffusion due to θ and χ from producing any non-linear terms. Observation 1 is used to manage the diffusion due to θ. Lanes are fixed to constant to prevent χ from creating any non-linear terms. Furthermore, for KECCAK-384/512, the first row of the hash digest can be inverted due to Observation 2.

In most cases, the number of linear equations between the variables and hash values is strictly less than the hash length. Therefore, they repeat the whole procedure enough number of times by appropriately changing the constants in the system of linear equations. This gives a successful preimage attack. In [18,19], similar techniques are used to restrict χ from producing many non-linear terms. Here, we allow χ to produce non-linear terms, but at the same time, we control the number of non-linear terms in the state.

3.1 Preimage Attack on 2 Rounds KECCAK-512

In this subsection, we describe our preimage attack for 2-rounds KECCAK-512. The best-known attack for this variant of KECCAK is by Guo et al. [13] with a complexity of 2^{384}. Their preimage attack is based on a linear structure by keeping four lanes as variables. We give two preimage attacks using six lanes as variables. In the first preimage attack, we keep the lanes in column 1, 3, 4 as variables and get an attack of complexity 2^{337} which can be improved to 2^{321}. However, the second preimage attack chooses a different set of lanes as variables and also has complexity of 2^{321}.

Preimage Attack with Complexity 2^{337}: In Fig. 3, we set the lanes in column 1, 3 and 4 as variables, and the rest of the lanes are set to some constant. Therefore, we have $6 \times 64 = 384$ variables. To avoid the propagation by θ in the first round, we use Observation 1, i.e., $\bigoplus_{j=0}^{4} A[i,j] = \alpha_i, \forall i \in [0,2,3]$ where α_i is some constant and hence include $3 \times 64 = 192$ linear constraints to the system. Also, since the hash length is 512, we can invert the first row of the hash value due to Observation 2.

Observe that after the application of the χ operation in the first round, state (4) contains a lane with quadratic terms. Due to the θ of the second round, these will get propagated only to the neighbouring columns. Hence, majority of the lanes in the state (5) contains only linear terms. But, while equating state (6) and state (7), we are only able to obtain $2 \times 64 = 128$ linear equations between the hash values and the variables. Observe that we have set up only 320 linear equations but have 384 variables.

Applying the techniques used in [13], we can linearize the quadratic term and use them to create more linear equations between hash value and the variables. Notice that in state (5), there is atmost one quadratic term in each polynomial. This is because the state before the application of θ in the second round has only one lane containing polynomials with only one quadratic term. More precisely, $A[4,4]$ of state (4) contains a polynomial of the form $p_1 + \overline{p_2}.p_3$ where p_i's are linear polynomials. This non-linear polynomial can be linearlized by adding one more linear equation to the system, say $p_3 = \beta$ where β is a constant. Therefore, if we linearize one quadratic term in state (4), we will be able to linearize 11 quadratic terms in state (5). But, only 3 out of the 11 linearized terms can be equated to the values in state (7). Therefore, we can set up an additional 64 linear equations of which $3\lfloor 64/4 \rfloor = 48$ equations are between message bits and hash values. But, we need to include one more linear equation for the last message bit to be 1 to satisfy the padding condition of KECCAK. Therefore, we have a system of linear equation in 384 variables and 384 equations. Since, we have $128 + 48 - 1 = 175$ linear equations between hash values and variables, we get a valid preimage with probability $1/2^{337}$.

To get a successful preimage attack, we must repeat the above procedure for at least 2^{337} times where the system of linear equations are different each time. Observe that there is enough degrees of freedom to perform this, i.e. 192 bits from $A[1,0], A[1,1]$ and $A[4,0]$ and 192 bits from α_i for $i \in [0,2,3]$ which sums up to 384 bits. Therefore, we have a preimage attack for 2-rounds KECCAK-512 with complexity of 2^{337}.

Improved Analysis: In the previous analysis, by equating state (6) and (7), we were able to obtain 128 linear equations between the hash values and variables. Let us now focus on the second χ operation on the second row of state (6). Observe that the second and fourth lanes of second row in state (6) are linear whereas we know the values of the first three consecutive lanes of the output of the second χ operation. Using Observation 3, we can set up an additional 64 linear equations which sums up to $128 + 64 - 1$ linear equations between the

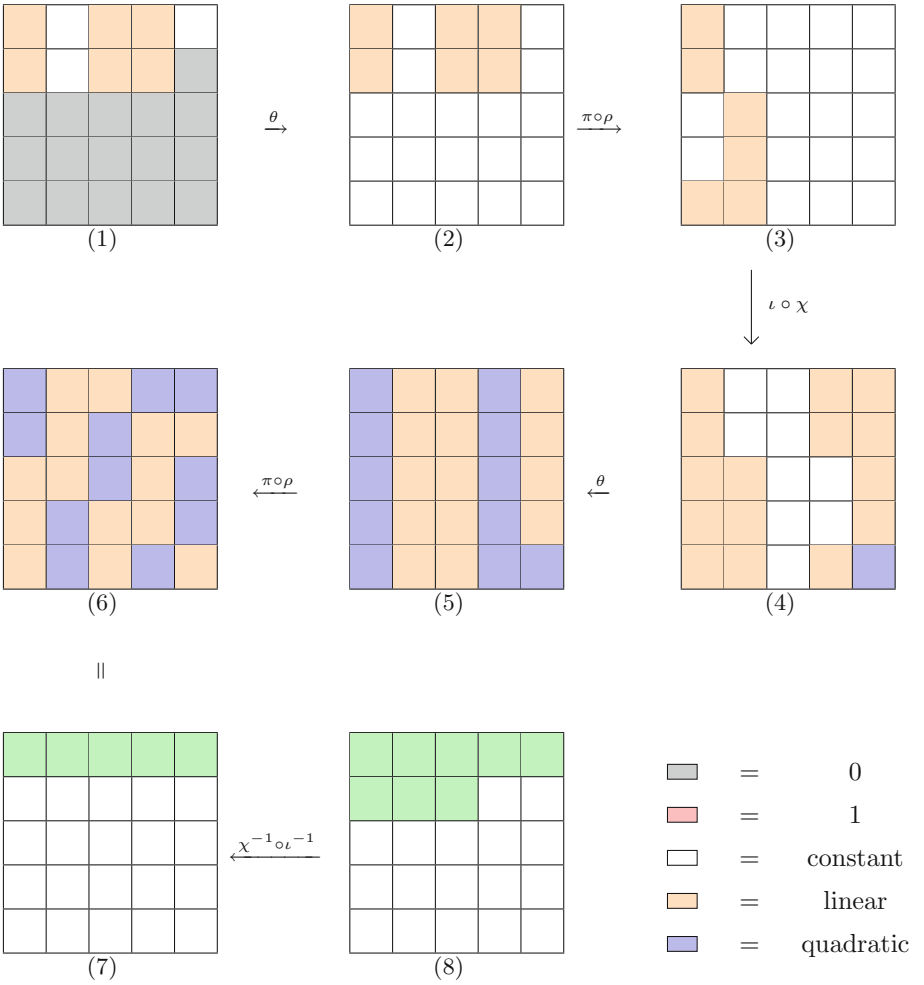

Fig. 3. Preimage attack on 2-round KECCAK-512

hash value and variables. Therefore, we have a preimage attack for 2-rounds KECCAK-512 with complexity of 2^{321}.

By choosing a different set of lanes as variables, we have another preimage attack with complexity 2^{321}. In Fig. 4, columns $1, 2$ and 4 are set as variables and the rest are set to constant. We also set $\bigoplus_{j=0}^{4} A[i,j] = \alpha_i, \forall i \in [0,1,3]$ where α_i is some constant, thus adding 192 linear equations to the system. Observe that in this case, we can set up $3 \times 64 - 1$ linear equations between the hash values and the variables. We must also include one more linear constraint for the last bit of message to be 1 to satisfy the padding condition for KECCAK. Therefore, we have a system of linear equation in 384 variables and 384 equations.

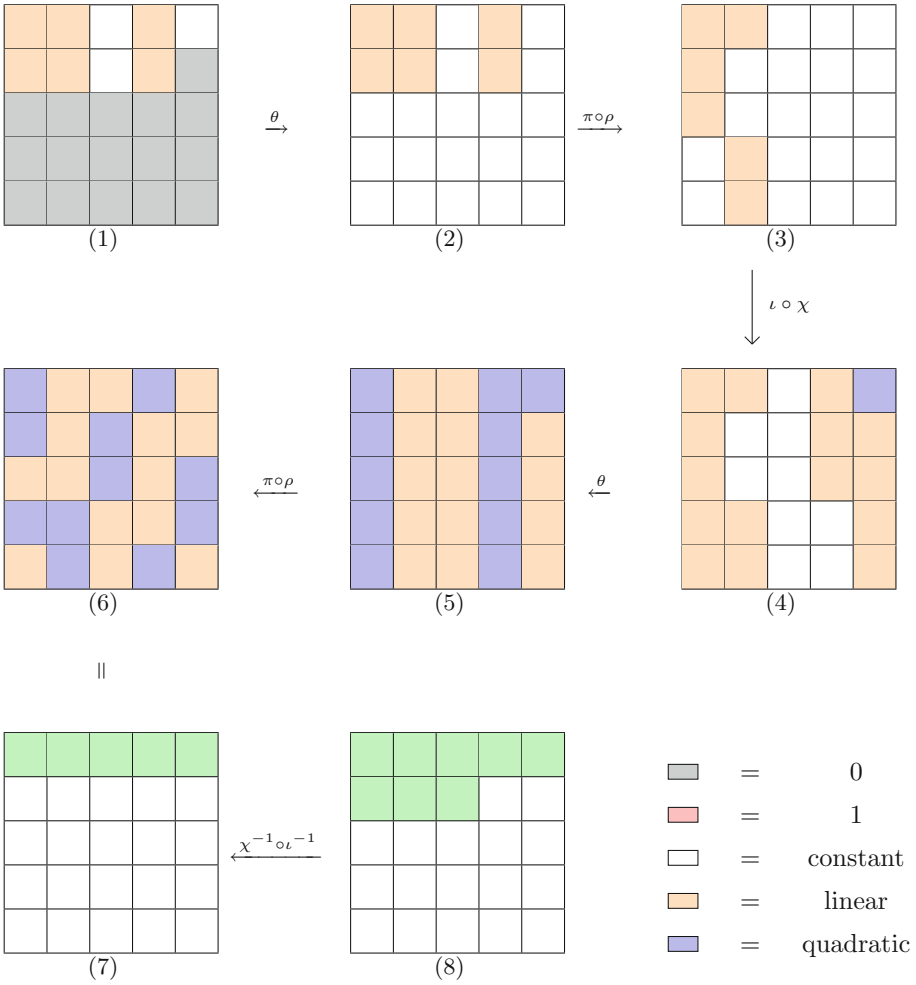

Fig. 4. Better preimage attack on 2-round KECCAK-512

Since we are able to set up only 191 linear equations between the hash values and the variables, we get a valid preimage with probability $1/2^{321}$. Observe that there is enough degrees of freedom to repeat this procedure for 2^{321} due 192 bits from $A[2,0], A[2,1]$ and $A[4,0]$ and 192 bits from α_i for $i \in [0,1,3]$ which sums up to 384 bits. Therefore, we have a preimage attack for 2-rounds KECCAK-512 with complexity of 2^{321}.

3.2 Preimage Attack on 2 Rounds KECCAK-384

The preimage attack given by Guo et al. [13] for 2 rounds KECCAK-384 has a complexity of 2^{129} by constructing a linear structure with 6×64 variables. In our

attack, we use 8×64 variables as shown in Fig. 5. In-order to avoid propagation by θ in first round, we add the following 3×64 linear constraints into the system, $\bigoplus_{j=0}^{4} A[i,j] = \alpha_i, \forall i \in [0,2,3]$ where α_i is some constant.

By equating state (5) and state (6), we get $2 \times 64 = 128$ linear equations between variables and hash values. Observe that we have only set up 320 linear equations but have $8 \times 64 = 512$ variables. Applying the linearization technique used in Subsect. 3.1, we can set up an additional 3×64 linear equations of which $3\lfloor(3 \times 64)/4)\rfloor = 144$ equations are between message bits and hash values. After satisfying the padding rule, we have a complexity gain over brute force of $2^{128+144-1} = 2^{271}$ and hence a preimage attack of complexity $2^{384-271} = 2^{113}$. Observe that we have enough degrees of freedom to repeat this procedure for

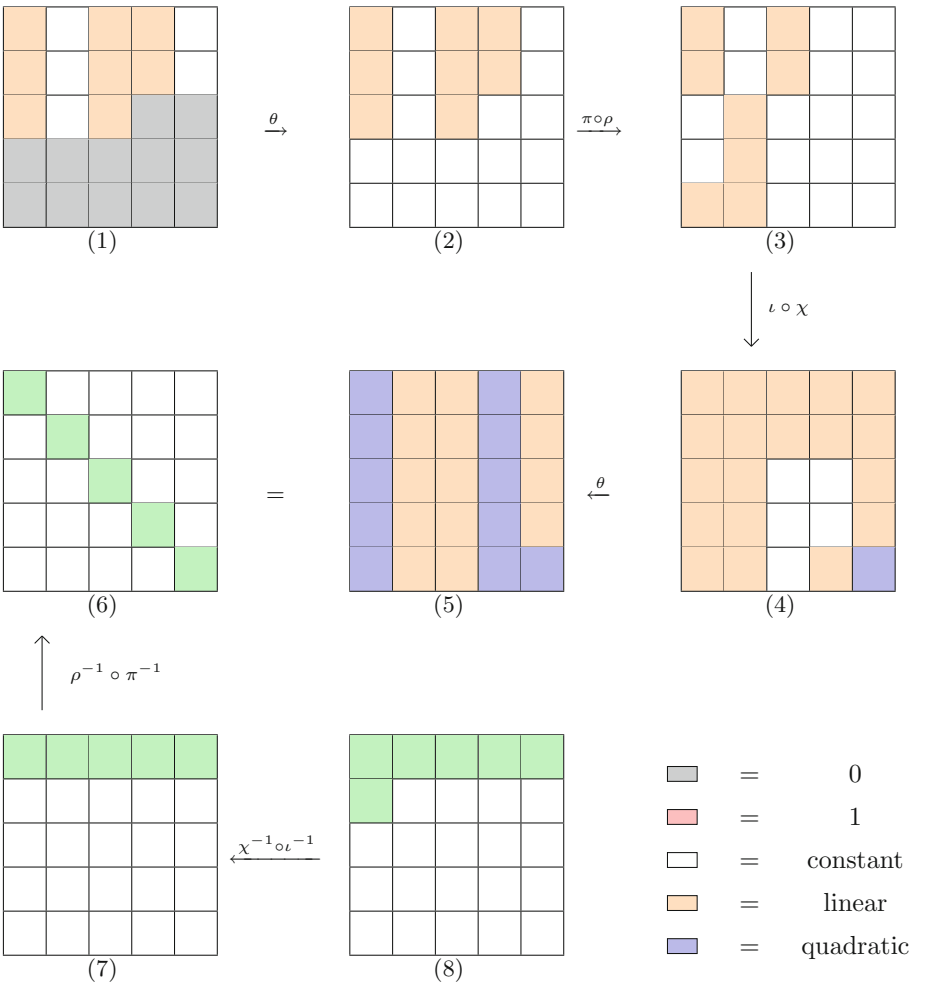

Fig. 5. Preimage attack on 2-round KECCAK-384

2^{113} times. Note that this result cannot be compared with the preimage attack given by Kumar et al. [16] because their attack has a space complexity of 2^{87}.

3.3 Preimage Attack for Higher Rounds

In the previous subsections, we were able to get better preimage attack due to the fact the states are not filled with quadratic terms. If we were to find a similar attack for 3-rounds, we need to keep the following guidelines in mind.

1. The state after the application of second θ must be sparse of lanes with linear terms and comprised mostly of lanes with constant terms. This is because it would lead to a state with lesser quadratic terms after the application of χ of the second round.
2. Even if the propagation due to the θ in the third round cannot be restricted, the state before the application of the third θ must contain all its quadratic terms either in a single column or in two columns adjacent to each other. This would lead to a state with at least one column containing linear terms only after the application of θ.

3.4 Preimage Attack on 3 Rounds KECCAK-384

The following is our attack on KECCAK-384 for 3-rounds which uses two message blocks as shown in Fig. 6. The first message block is chosen in such a way that after the application of 3 round KECCAK on this block, we get a state such that $A[1,3] = A[3,3] = 0$ and $A[1,4] = A[4,4] = 1$ where A is state (2) as shown in Fig. 6. The first message block can be found by randomly choosing $2^{4\times 64}$ message block and expecting one of them to give the required output. This works because the output of a hash function is random and therefore the complexity for brute force preimage attack is $1/2^l$ where l is the number of bits in the hash digest. The same technique has been used in [18] subsection 4.3.

The second message block contains $6 \times 64 = 384$ variables. We want to keep the columns 2, 4 and 5 unchanged after the application of first θ. For this, we first set $\bigoplus_{j=0}^{4} A[i,j] = \alpha_i$, for $i \in [0,2]$ and then set up equation between column 1 and column 3 so that column 2 does not get affected after the application of first θ. This means that the α_i's are dependent. Similarly, c_2 and c_3 can be set according to α_i's such that column 4 and 5 do not get affected after the first θ. Therefore, we have 2×64 linear equations in our system. c_1 can be fixed to some randomly chosen value.

To avoid propagation after second θ, we set up 3×64 linear equations to make the column parties equal to some constant β_i. Observe that after the application of the second χ, there are two lanes with quadratic terms in state (8). But after the application of the third θ, the fourth column will contain only linear terms. By equating state (9) and state (10), we can set up 63 linear equations between message bits and hash values. Also, we have one more equation to keep the last message bit equal to 1. Therefore, we have a preimage attack with a time

of complexity $2^{384-63} = 2^{321}$ because computing the first message block has a complexity of 2^{256}.

Note that there are enough degrees of freedom due to the 256 bits from α_i's and the β_i's, 64 bits from c_1 and enough bits from the first message block.

3.5 Preimage Attack on 3 Rounds KEECAK-512

We use two message blocks and $4 \times 64 = 256$ variables for this attack as shown in Fig. 7. The first message block is used so that we get enough degree of freedom to launch a preimage attack. Observe that after the application of θ in first round, we require certain lanes to be 1/0 in state (4). To achieve this, we first set $A[1,0] \oplus A[1,1] = \alpha_1$ where α_1 is some constant. Then, we set up 64 linear equations of the form $\bigoplus_{i=0}^{4} (A[1,i] \oplus A[3,i](1) = e_2 + 1)$. Observe that due to this constraint, after the application of first θ, we will get $A[2,0] = A[2,4] = 1$ and $A[2,1] = 0$ where A is state (4). Similarly, by fixing x_6 and x_2 appropriately, we can get the required state (4).

To avoid propagation due to the θ in second round, we add only 64 linear equations to the system to make the parity of the first columns in state (6) as a constant. Observe that after the application of θ of the third round, the lanes in the first two columns will contain only one quadratic term. So, if we linearize one quadratic term in $A[2,4]$ of state (9), then we have linearized five polynomials in column 2 of state (10). Similarly, if we linearize one quadratic term in $A[4,2]$ of state (9), then we have linearized five polynomial in column 1 of state (10).

But, out of these 6 linearized polynomials, only one can be used to create a linear equation between message bits and hash value by equating state (10) and state (11). Therefore, we have $\lfloor 64/2 \rfloor = 32$ linear equations between message bits and hash value and hence obtained a preimage of complexity $2^{512-32+1} = 2^{481}$. Due to the first message block, we have enough degree of freedom.

Improved Analysis. Observe that if we carefully linearize one quadratic term from $A[2,4]$ and one from $A[4,2]$ of state (9), we also linearize one more polynomial in column 4 of state (10), i.e. we have also linearized a polynomial in the lane $A[3,3]$. Therefore, now we have $3\lfloor \frac{64}{5} \rfloor + 2 = 3 \times 12 + 2 = 38$. Therefore, we have an improved preimage attack of complexity $2^{512-38+1} = 2^{475}$.

3.6 Preimage Attack on 4 Rounds KECCAK-384

This attack requires two message blocks and $6 \times 64 = 384$ variables as shown in Fig. 8. As done in Subsect. 3.4, the first message block is found by trying randomly many message blocks so that after the application of 4-rounds and XORing the second message block, we get state (2). Observe that in state (2), there are two lanes with entries c and $\overline{c}$. We also require state (2) to satisfy one more equation.

$$d(-1) + \overline{b}(-2) + (g(-1) + (\overline{c} + a + b)(-2))(-2) + (a + b)(1) = \overline{k} \qquad (1)$$

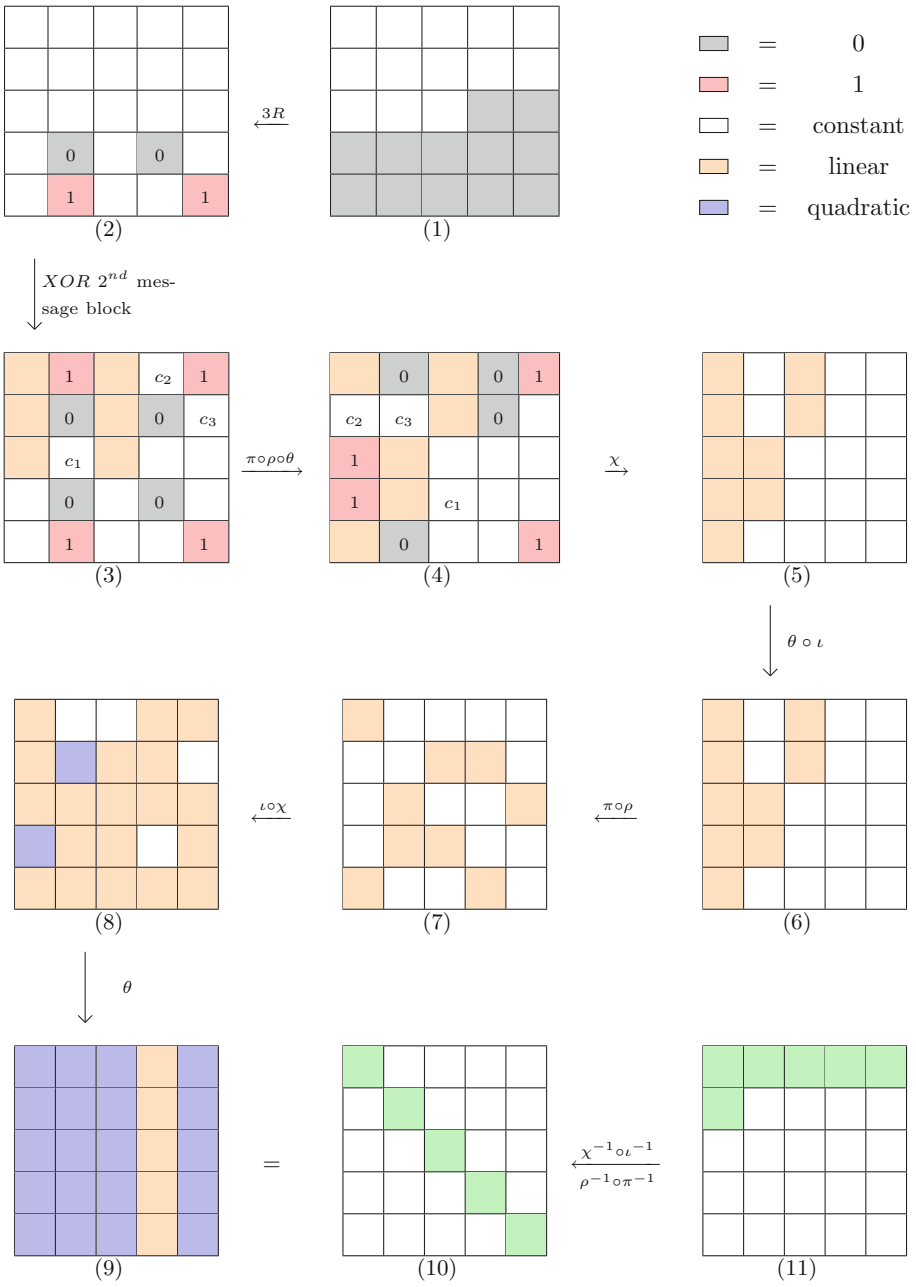

Fig. 6. Preimage attack on 3-round KECCAK-384

Therefore, we would require a complexity of 2^{128} to find the appropriate first message block. We will use the following strategy to obtain state (3). We include

$A[0,0] = A[0,2]$ to the system of linear equations, fix $x_1 = 0$ and randomly assign value to x_7 whereas we fix $x_2 = \overline{c}, x_3 = d, x_5 = g$. Since we require state (3) after the application of θ, we have the following equations.

$$(a+b) + (A[2,0] + A[2,2] + e)(1) = \overline{c} \tag{2}$$

$$(A[2,0] + A[2,2] + e) + (x_6 + x_7 + i + j + k)(1) = g \tag{3}$$

$$(x_6 + x_7 + i + j + k) + (A[1,0] + A[1,2] + c)(1) = \overline{b} \tag{4}$$

$$(A[1,0] + A[1,2] + c) + (x_4 + f + h)(1) = d \tag{5}$$

$$(x_4 + f + h) + (a+b)(1) = \overline{k} \tag{6}$$

Therefore, we add Eqs. (7) and (9) to the system of equations and fix x_6 and x_4 according to Eqs. (8) and (10). Observe that due to the following equations, all equations from (2)–(6) are satisfied, particularly, Eq. (6) is satisfied due to Eq. (1).

$$A[2,0] + A[2,2] = (\overline{c} + a + b)(-1) + e \tag{7}$$

$$x_6 = g(-1) + (\overline{c} + a + b)(-2) + x_7 + i + j + k \tag{8}$$

$$A[1,0] + A[1,2] = \overline{b}(-1) + (g(-1) + (\overline{c} + a + b)(-2))(-1) + c \tag{9}$$

$$\begin{aligned} x_4 &= d(-1) + f + h + (\overline{b} + x_6 + x_7 + i + j + k)(-2) \\ &= d(-1) + f + h + \overline{b}(-2) + (g(-1) + (\overline{c} + a + b)(-2))(-2) \end{aligned} \tag{10}$$

Also, we include 2×64 linear equations for restricting the propagation due to θ in the second round. Observe that each polynomial in the state (9) has 11 quadratic terms. In [13] subsection 6.3, Guo et al. gave a technique that carefully linearizes the quadratic terms such that if the number of free variables is t, we can construct $2\lfloor (t-5)/8 \rfloor$ linear equations between hash values and the variables. Let A denotes state (8), B denotes the state after χ of third round and C denotes the state after θ of fourth round. From the definition of χ and θ and neglecting ι step for the sake of simplicity,

$$B[x,y,z] = A[x,y,z] \oplus (A[x+1,y,z] \oplus 1) \cdot A[x+2,y,z]$$

$$C[x,y,z] = B[x,y,z] \oplus \bigoplus_{y'=0}^{4} B[x-1,y',z] \oplus \bigoplus_{y'=0}^{4} B[x+1,y',z-1]$$

We can linearize $B[x-1,y,z]$ and $B[x,y,z]$ by guessing the value of $A[x+1,y,z]$ for $0 \le y \le 4$. Similarly, we can linearize $B[x+1,y,z-1]$ and $B[x+2,y,z-1]$ by guessing the value of $A[x+3,y,z-1]$ for $0 \le y \le 4$. This helps us in linearizing $C[x,y,z]$, but observe that

$$C[x+1,y+1,z] = B[x+1,y+1,z] \oplus \bigoplus_{y'=0}^{4} B[x,y',z] \oplus \bigoplus_{y'=0}^{4} B[x+2,y',z-1]$$

which contain a quadratic part in $B[x+1,y+1,z]$. By linearizing this term, we set up 13 linear equations of which two equations are between message bits and hash

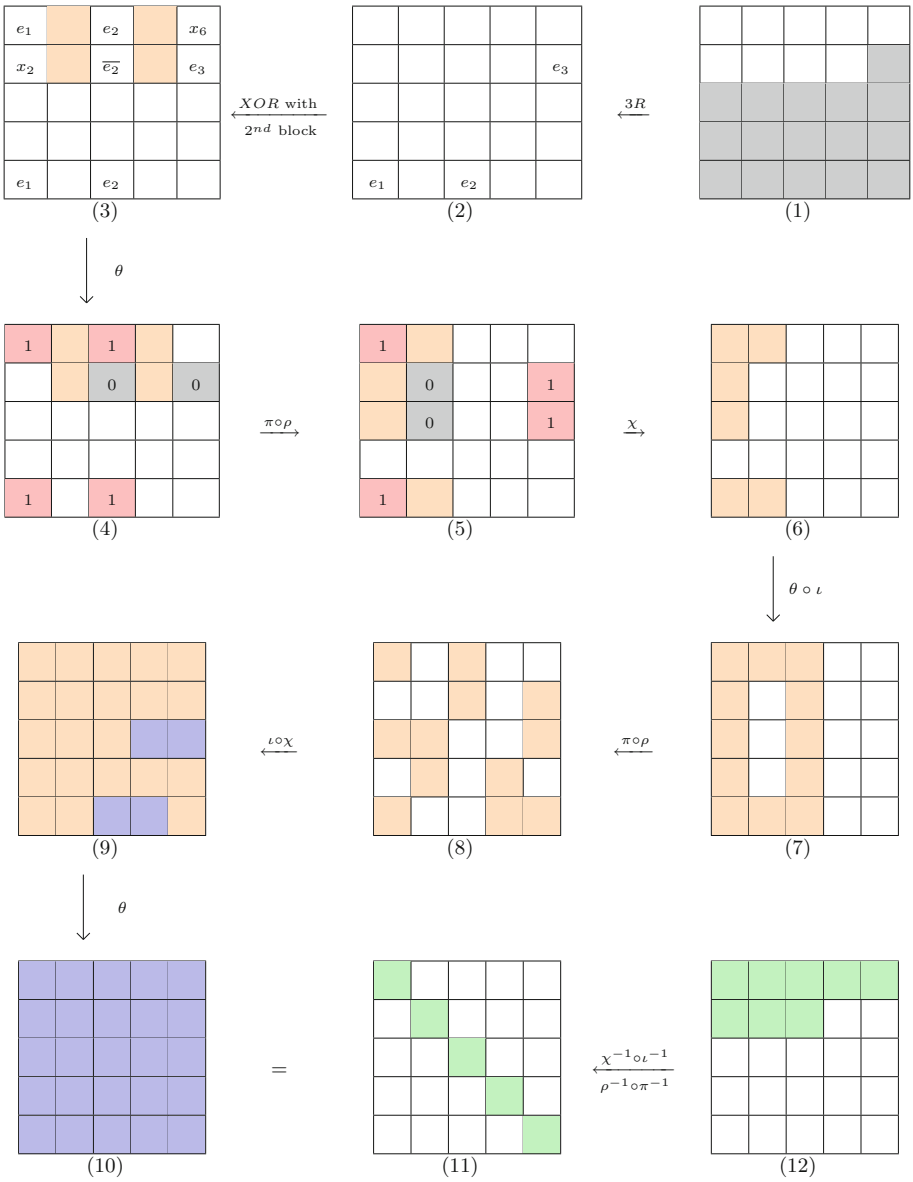

Fig. 7. Preimage attack on 3-round KECCAK-512

values. Similarly, by carefully observing $C[x+2, y+2, z-1]$ and $C[x+3, y+3, z-1]$ and linearizing them, we can set up another 8 linear equations of which two equations are between message bits and hash values. For more details, refer [13]. In our case, the number of free variable $t = 64$ and therefore, we can set up 14 linear equations between message bits and hash values. Observe that we have

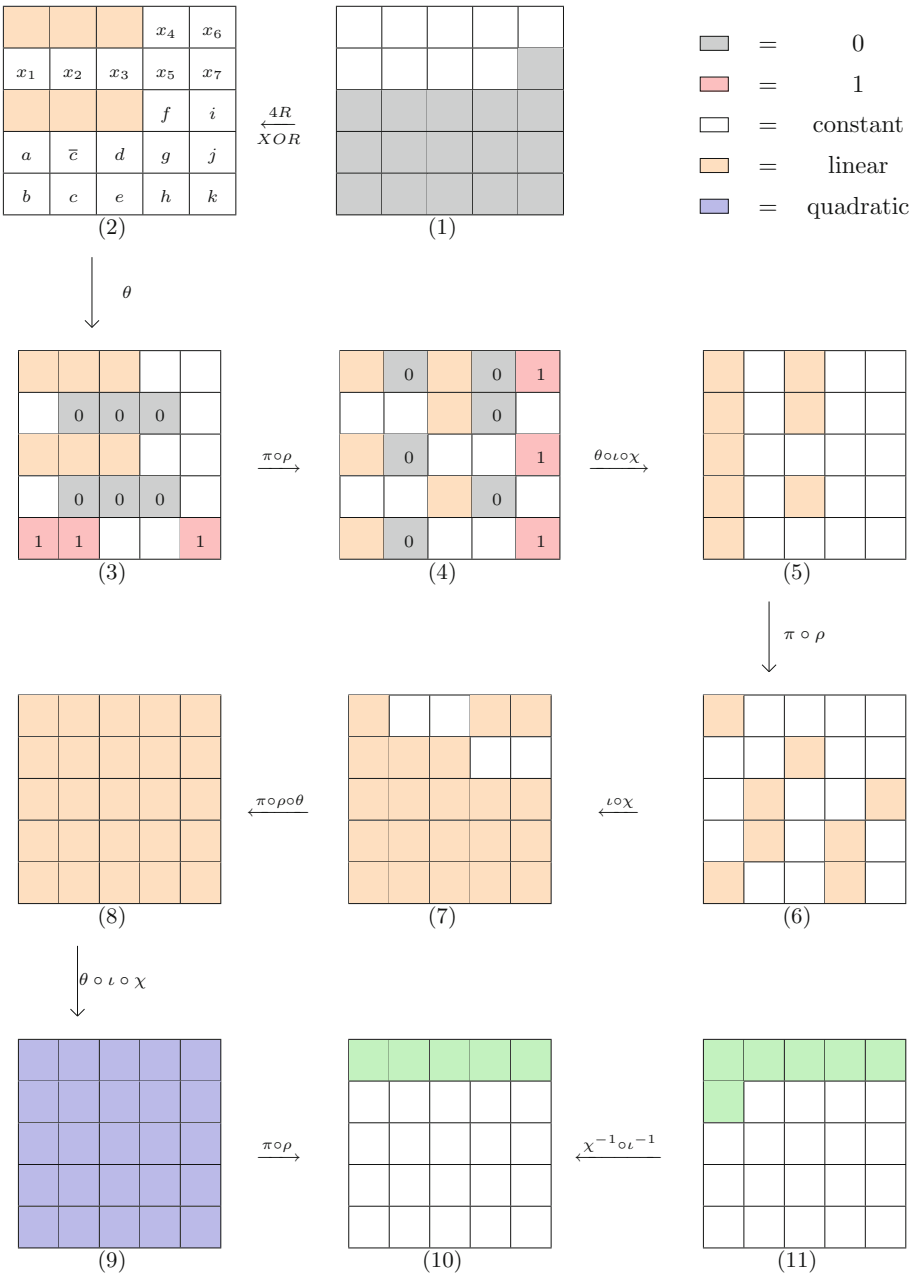

Fig. 8. Preimage attack on 4-round KECCAK-384

enough degree of freedom due to x_7, the parity of the two columns of the second θ and rest from the first message block. Therefore, the complexity of our attack is 2^{371}.

4 Conclusion

In this paper, we give the best theoretical preimage attacks on 2, 3 rounds KECCAK-512 and 2, 3, 4 rounds KECCAK 384 by studying non-linear structures carefully. It would be interesting to see whether non-linear structures along with other techniques can be used to find better preimage attacks for higher rounds.

Acknowledgement. We would like to thank Rajendra Kumar for valuable discussions and anonymous reviewers of INDOCRYPT 2019 for their helpful comments.

References

1. Aumasson, J.P., Meier, W.: Zero-sum distinguishers for reduced Keccak-f and for the core functions of Luffa and Hamsi. In: Rump Session of Cryptographic Hardware and Embedded Systems-CHES 2009, p. 67 (2009)
2. Bernstein, D.J.: Second preimages for 6 (7?(8??)) rounds of Keccak. NIST mailing list (2010)
3. Bertoni, G., Daemen, J., Peeters, M., Assche, G.: The Keccak reference (2011). http://keccak.noekeon.org/keccak-reference-3.0.pdf
4. Bertoni, G., Daemen, J., Peeters, M., Van Assche, G.: Keccak specifications. Submission to NIST (round 2), pp. 320–337 (2009)
5. Bertoni, G., Daemen, J., Peeters, M., Van Assche, G.: Cryptographic sponges (2011). http://sponge.noekeon.org
6. Boura, C., Canteaut, A., De Cannière, C.: Higher-order differential properties of KECCAK and *Luffa*. In: Joux, A. (ed.) FSE 2011. LNCS, vol. 6733, pp. 252–269. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-21702-9_15
7. Chang, D., Kumar, A., Morawiecki, P., Sanadhya, S.K.: 1st and 2nd preimage attacks on 7, 8 and 9 rounds of Keccak-224,256,384,512. In: SHA-3 workshop, August 2014
8. Dinur, I., Dunkelman, O., Shamir, A.: New attacks on Keccak-224 and Keccak-256. In: Canteaut, A. (ed.) FSE 2012. LNCS, vol. 7549, pp. 442–461. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-34047-5_25
9. Dinur, I., Dunkelman, O., Shamir, A.: Collision attacks on up to 5 rounds of SHA-3 using generalized internal differentials. In: Moriai, S. (ed.) FSE 2013. LNCS, vol. 8424, pp. 219–240. Springer, Heidelberg (2014). https://doi.org/10.1007/978-3-662-43933-3_12
10. Dinur, I., Dunkelman, O., Shamir, A.: Improved practical attacks on round-reduced Keccak. J. Cryptol. **27**(2), 183–209 (2014)
11. Duc, A., Guo, J., Peyrin, T., Wei, L.: Unaligned rebound attack: application to Keccak. In: Canteaut, A. (ed.) FSE 2012. LNCS, vol. 7549, pp. 402–421. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-34047-5_23
12. Dworkin, M.J.: SHA-3 standard: permutation-based hash and extendable-output functions. Technical report (2015)

13. Guo, J., Liu, M., Song, L.: Linear structures: applications to cryptanalysis of round-reduced KECCAK. In: Cheon, J.H., Takagi, T. (eds.) ASIACRYPT 2016. LNCS, vol. 10031, pp. 249–274. Springer, Heidelberg (2016). https://doi.org/10.1007/978-3-662-53887-6_9
14. Jean, J., Nikolić, I.: Internal differential boomerangs: practical analysis of the round-reduced Keccak-f permutation. In: Leander, G. (ed.) FSE 2015. LNCS, vol. 9054, pp. 537–556. Springer, Heidelberg (2015). https://doi.org/10.1007/978-3-662-48116-5_26
15. Kölbl, S., Mendel, F., Nad, T., Schläffer, M.: Differential cryptanalysis of Keccak variants. In: Stam, M. (ed.) IMACC 2013. LNCS, vol. 8308, pp. 141–157. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-45239-0_9
16. Kumar, R., Mittal, N., Singh, S.: Cryptanalysis of 2 round KECCAK-384. In: Chakraborty, D., Iwata, T. (eds.) INDOCRYPT 2018. LNCS, vol. 11356, pp. 120–133. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-05378-9_7
17. Kumar, R., Rajasree, M.S., AlKhzaimi, H.: Cryptanalysis of 1-round KECCAK. In: Joux, A., Nitaj, A., Rachidi, T. (eds.) AFRICACRYPT 2018. LNCS, vol. 10831, pp. 124–137. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-89339-6_8
18. Li, T., Sun, Y.: Preimage attacks on round-reduced KECCAK-224/256 via an allocating approach. In: Ishai, Y., Rijmen, V. (eds.) EUROCRYPT 2019. LNCS, vol. 11478, pp. 556–584. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-17659-4_19
19. Li, T., Sun, Y., Liao, M., Wang, D.: Preimage attacks on the round-reduced Keccak with cross-linear structures. IACR Trans. Symmetric Cryptol. 39–57 (2017)
20. Morawiecki, P., Pieprzyk, J., Srebrny, M.: Rotational cryptanalysis of round-reduced KECCAK. In: Moriai, S. (ed.) FSE 2013. LNCS, vol. 8424, pp. 241–262. Springer, Heidelberg (2014). https://doi.org/10.1007/978-3-662-43933-3_13
21. Morawiecki, P., Srebrny, M.: A SAT-based preimage analysis of reduced Keccak hash functions. Inf. Process. Lett. **113**(10–11), 392–397 (2013)
22. Naya-Plasencia, M., Röck, A., Meier, W.: Practical analysis of reduced-round KECCAK. In: Bernstein, D.J., Chatterjee, S. (eds.) INDOCRYPT 2011. LNCS, vol. 7107, pp. 236–254. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-25578-6_18
23. Song, L., Liao, G., Guo, J.: Non-full sbox linearization: applications to collision attacks on round-reduced KECCAK. In: Katz, J., Shacham, H. (eds.) CRYPTO 2017. LNCS, vol. 10402, pp. 428–451. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-63715-0_15