

Analysis of RIPEMD-160: New Collision Attacks and Finding Characteristics with MILP

Fukang Liu^{1,2}, Gaoli Wang^{3,4}, Santanu Sarkar⁵, Ravi Anand²,
Willi Meier⁶, Yingxin Li³, Takanori Isebe^{2,7}

¹Tokyo Institute of Technology, Tokyo, Japan

²University of Hyogo, Hyogo, Japan

³East China Normal University, Shanghai, China

⁴State Key Laboratory of Cryptology, Beijing, China

⁵Indian Institute of Technology Madras, Chennai, India

⁶FHNW, Windisch, Switzerland

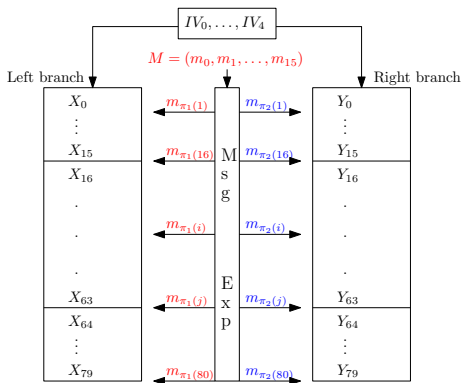
⁷NICT, Tokyo, Japan

Overview

- 1 Background
 - RIPEMD-160
 - Difficulty to Analyze RIPEMD-160
 - Finding Trails
- 2 Our MILP-based Method
 - Problem Analysis
 - Modelling
 - Application
- 3 Summary and Open Questions

RIPEMD-160

- FSE 1996 by Dobbertin et al.
- Strengthens MD5 (double branches, complex round function)
- Bitcoin address (with SHA-256)
- ISO/IEC standard

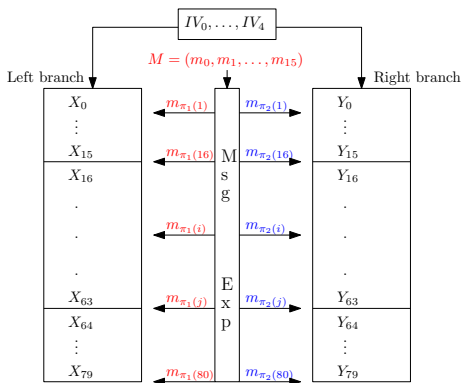


Round Function of RIPEMD-160

Round function (left branch as an example)

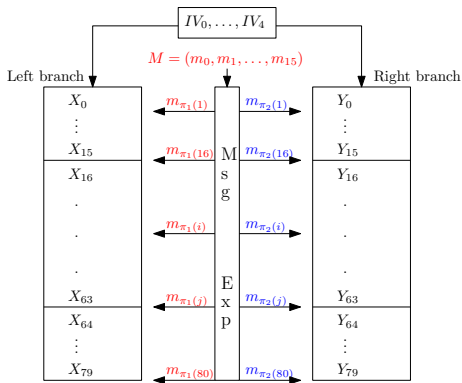
$$Q_i = X_{i-5} \lll 10 \boxplus F_i(X_{i-1}, X_{i-2}, X_{i-3} \lll 10) \boxplus m_{\pi(i)} \boxplus K_i,$$

$$X_i = X_{i-4} \lll 10 \boxplus Q_i \lll s_i.$$



Difficulty to Analyze RIPEMD-160

- Difficulty to analyze RIPEMD-160
 - Finding valid differential trails
 - Finding conforming message pairs
 - Constructing better local collisions



Finding Trails for RIPEMD-160

- Finding valid differential trails
 - Bit conditions by the Boolean functions
 - Extra conditions on Q_i

$$(Q_i \boxplus \alpha_i) \lll s_i = Q_i \lll s_i \boxplus \beta_i,$$

where α_i, β_i are constants derived from the differential trail.

Finding Trails for RIPEMD-160

■ Where are the contradictions?

■ Boolean functions¹:

$$F_{i+1}(X_i, X_{i-1}, X_{i-2} \lll 10), \quad (1)$$

$$F_{i+2}(X_{i+1}, X_i, X_{i-1} \lll 10), \quad (2)$$

$$F_{i+3}(X_{i+2}, X_{i+1}, X_i \lll 10). \quad (3)$$

■ Extra conditions on Q_i :

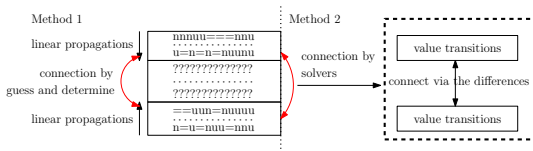
$$(Q_i \boxplus \alpha_i) \lll s_i = Q_i \lll s_i \boxplus \beta_i, \quad (4)$$

$$X_i = Q_i \lll s_i \boxplus X_{i-4} \lll 10 \quad (5)$$

¹e.g., $F(x, y, z) = ONX(x, y, z) = (x \vee \bar{y}) \oplus z$.

Finding Trails for RIPEMD-160

- Finding differential trails: guess-and-determine technique
 - Linearly propagate the difference: sparse part
 - Connect two sparse parts: dense part (**difficult!!!**)

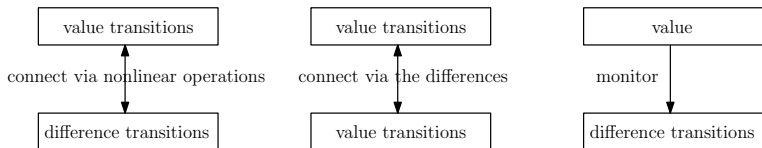


- Not open-sourced (Method 1)
 - 1 But, we can still access the tools by contacting the authors.
 - 2 But, **can we have an alternative choice (maybe a different approach with new potential?)** e.g. MILP/SAT/SMT solvers?
 - 3 I want to have my own tool (personal interest)

Finding Trails for RIPEMD-160

■ Our method (and comparison with other methods)

- left: CRYPTO 2020 (Gimli)
- middle: SAT 2006, CRYPTO 2017 (MD4, SHA-1)
- right: ours



■ Highlight the difference

- Encode the difference transitions into the model
- Remove the expensive value transitions
- A light method (monitoring variables) to detect contradictions

Our Method

■ Analysis of the target problem

- How to describe the following difference transitions²?

$$(\nabla d_i, \nabla d_{i+1}, \nabla d_{i+2}, \nabla d_{i+3}, \nabla d_{i+4}, \nabla m) \rightarrow \nabla d_{i+5}$$

where

$$d_{i+5} = d_{i+1} \lll 10$$

$$\boxplus (F(d_{i+4}, d_{i+3}, d_{i+2} \lll 10) \boxplus (d_i \lll 10) \boxplus m \boxplus c) \lll s.$$

² ∇a : signed difference, δa : modular difference

Our Method

- Equivalent problem (bit-level)
 - How to describe the following difference transitions?

$$(\nabla a_0, \nabla a_1, \nabla a_2, \nabla a_3, \nabla a_4, \nabla m) \rightarrow \nabla a_5$$

where

$$a_5 = a_1 \boxplus (F(a_4, a_3, a_2) \boxplus a_0 \boxplus m \boxplus c) \lll s.$$

Our Method

■ Introduce intermediate variables

$$b_0 = m \boxplus c,$$

$$b_1 = F(a_4, a_3, a_2),$$

$$b_2 = b_0 \boxplus b_1,$$

$$b_3 = b_2 \boxplus a_0,$$

$$b_4 = b_3 \lll s,$$

$$b_5 = a_1 \boxplus b_4,$$

$$a_5 = b_5.$$

- Note 1: basically, we do not need b_0 because $\nabla b_0 = \nabla m$.
- Note 2: m is a free variable

Our Method

■ Main Idea:

- Deterministically compute the signed difference transitions for $b_0 = m \boxplus c$, $b_2 = b_0 \boxplus b_1$ and $b_3 = b_2 \boxplus a_0$. Specifically, for each given $(\nabla x, \nabla y)$, uniquely compute one ∇z such that $\delta z = \delta x \boxplus \delta y$, even though there are many such possible ∇z .
- Compute the signed difference transitions for $b_1 = F(a_4, a_3, a_2)$, where F is a boolean function.
- Compute a possible value of ∇b_5 for $b_4 = b_3 \lll s$ and $b_5 = a_1 \boxplus b_4$.
- Expand ∇b_5 to get all possible ∇a_5 such that the modular differences satisfy $\delta b_5 = \delta a_5$.

Our Method

- The problems to address:
 - Model $\nabla z = \nabla x \boxplus \nabla y$ where we only need to ensure $\delta z = \delta x \boxplus \delta y$.
 - Model $\nabla b_1 = F(\nabla a_4, \nabla a_3, \nabla a_2)$.
 - Model how to expand ∇b_5 to get all possible ∇a_5 such that the modular differences satisfy $\delta b_5 = \delta a_5$.

Describing the Signed Difference

■ Three status: $\{\mathbf{n}, \mathbf{u}, =\}$

■ Method 1: use one ternary variable (inefficient for F^3)

$$v \in \{-1, 0, 1\}$$

■ Method 2: use two binary variables (v, d)

$$(v, d) \in \{(0, 1), (1, 1), (0, 0)\},$$

where $(1, 1)$ is not allowed.

³e.g., $F(x, y, z) = ONX(x, y, z) = (x \vee \bar{y}) \oplus z$.

Modelling the Modular Addition

- Target: $\nabla z = \nabla x \boxplus \nabla y$
 - Use an intermediate variable c to represent the carry
 - No branches

Table: Propagation rules for $(\nabla x[i], \nabla y[i], \nabla c[i]) \rightarrow (\nabla z[i], \nabla c[i + 1])$

[=== \rightarrow ==],	[==n \rightarrow n=],	[==u \rightarrow u=],	[=n= \rightarrow n=],
[=u= \rightarrow u=],	[=nn \rightarrow =n],	[=un \rightarrow ==],	[=nu \rightarrow ==],
[=uu \rightarrow =u],	[n== \rightarrow n=],	[u== \rightarrow u=],	[n=n \rightarrow =n],
[u=n \rightarrow ==],	[n=u \rightarrow ==],	[u=u \rightarrow =u],	[nn= \rightarrow =n],
[nu= \rightarrow ==],	[un= \rightarrow ==],	[uu= \rightarrow =u],	[nnn \rightarrow nn],
[nun \rightarrow n=],	[unn \rightarrow n=],	[nnu \rightarrow n=],	[uun \rightarrow u=],
[unu \rightarrow u=],	[nuu \rightarrow u=],	[uuu \rightarrow uu]	

Modelling the Modular Addition

Table: Propagation rules for $(\nabla x[i], \nabla y[i], \nabla c[i]) \rightarrow (\nabla z[i], \nabla c[i + 1])$

[=== \rightarrow ==],	[==n \rightarrow n=],	[==u \rightarrow u=],	[=n= \rightarrow n=],
[=u= \rightarrow u=],	[=nn \rightarrow =n],	[=un \rightarrow ==],	[=nu \rightarrow ==],
[=uu \rightarrow =u],	[n== \rightarrow n=],	[u== \rightarrow u=],	[n=n \rightarrow =n],
[u=n \rightarrow ==],	[n=u \rightarrow ==],	[u=u \rightarrow =u],	[nn= \rightarrow =n],
[nu= \rightarrow ==],	[un= \rightarrow ==],	[uu= \rightarrow =u],	[nnn \rightarrow nn],
[nun \rightarrow n=],	[unn \rightarrow n=],	[nnu \rightarrow n=],	[uun \rightarrow u=],
[unu \rightarrow u=],	[nuu \rightarrow u=],	[uuu \rightarrow uu]	

Linear inequalities (with LogicFriday):

$$\mathcal{H}_{\text{ADD}} \cdot V_{\text{ADD}}^T \geq C_{\text{ADD}},$$

$$V_{\text{ADD}} = (x_v[i], x_d[i], y_v[i], y_d[i], c_v[i], c_d[i], z_v[i], z_d[i], c_v[i + 1], c_d[i + 1]).$$

Modelling the Modular Addition

$$\mathcal{H}_{\text{ADD}} | \mathcal{C}_{\text{ADD}} = \left[\begin{array}{cccccccccccc|c} 0 & 1 & 0 & 1 & 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & -1 & 0 & -1 & 0 & 1 & 0 & 0 & 0 & -2 \\ 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & -1 & 0 \\ 1 & 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 & -1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & -1 & -1 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & -1 & 1 & 0 & -1 & 0 & 0 & 0 & 0 & -1 \\ 1 & 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 & -1 & -1 & 0 \\ 0 & -1 & 0 & 1 & 0 & -1 & 0 & -1 & 0 & 0 & 0 & -2 \\ 0 & -1 & -1 & 0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 & -2 \\ 0 & 1 & 0 & -1 & 0 & -1 & 0 & -1 & 0 & 0 & 0 & -2 \\ 0 & -1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & -1 & 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 & -1 & 0 & 1 & 0 & 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & -1 & -1 & 0 \\ 1 & -1 & 1 & -1 & 1 & 0 & 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & -1 & 0 & 0 & -1 \\ 1 & 0 & 1 & 0 & 1 & -1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 & -1 & -1 & 0 \\ -1 & 0 & 0 & 0 & -1 & 0 & 1 & 0 & 0 & 1 & -1 & 0 \\ -1 & 0 & -1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & -1 & 0 \\ 0 & 1 & -1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 & -1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 & -1 & -1 & 0 \\ -1 & 0 & 0 & 0 & 0 & 1 & 1 & -1 & 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & -1 & 0 & -1 & 0 & 0 & 0 & 0 & 1 & -2 & 0 \end{array} \right],$$

Modelling the Expansion

- Get all $\nabla\xi$ from ∇z such that $\delta\xi = \delta z$:
 - Use an intermediate variable c to represent the carry
 - Branches!!! (tree structure)

Table: Propagation rules for $(\nabla z[i], \nabla c[i]) \rightarrow (\nabla \xi[i], \nabla c[i + 1])$

$[\text{nn} \rightarrow \text{=n}], [\text{uu} \rightarrow \text{=u}], [\text{nu} \rightarrow \text{==}], [\text{un} \rightarrow \text{==}],$
$[\text{n=} \rightarrow (\text{n=}, \text{un})], [\text{u=} \rightarrow (\text{u=}, \text{nu})],$
$[\text{=n} \rightarrow (\text{n=}, \text{un})], [\text{=u} \rightarrow (\text{u=}, \text{nu})],$
$[\text{==} \rightarrow \text{==}]$

Modelling the Boolean Function

- Example: $\nabla w = ONX(\nabla x, \nabla y, \nabla z)$ where

$$ONX(x, y, z) = (x \vee \bar{y}) \oplus z.$$

- List all possible cases!!!

Table: Valid values of $(\nabla x[i], \nabla y[i], \nabla z[i], \nabla w[i])$

[====],
[==u=], [==uu], [==un], [==n=], [==nn], [==nu],
[=n==], [=n=n], [=n=u], [=u==], [=u=u], [=u=n],
[n==u], [n==n], [u==n], [u==u],
[=nn=], [=uu=], [=nunu], [=nuu], [=unn], [=unu],
[nn=u], [nn==], [nu=u], [nu==], [uu=n], [uu==], [un=n], [un==],
[n=nu], [n=n=], [n=uu], [n=u=], [u=nn], [u=n=], [u=un], [u=u=],
[nnnu], [nnu=], [nun=], [unnn], [uun=], [unu=], [nuuu], [uuun].

Detecting Contradictions in F

- Example: $\nabla w = ONX(\nabla x, \nabla y, \nabla z)$
 - Introduce variables (x, y, z) representing the values into the model
 - Each possible $(\nabla x, \nabla y, \nabla z, \nabla w)$ will impose certain bit conditions on (x, y, z) .
 - List all possible cases.

Table: Valid values of $(\nabla x[i], \nabla y[i], \nabla z[i], \nabla w[i], x[i], y[i], z[i])$

[===, *, *, *],
[==u=, *, 1, *], [==uu, 1, 0, *], [==un, 0, 0, *], [==n=, *, 1, *], [==nn, 1, 0, *], [==nu, 0, 0, *],
[n==, *, *, 0], [n=n, 0, *, 1], [n=u, 1, *, 1], [u==, *, *, 0], [u=u, 0, *, 1], [u=n, 1, *, 1],
[n==u, *, 1, *], [n==u, *, 0, 0], [n==n, *, 0, 1], [u==n, *, 1, *], [u==n, *, 0, 0], [u==u, *, 0, 1],
[nn=, *, *, *], [uu=, *, *, *], [nunn, 0, *, *], [nuu, 1, *, *], [unn, 1, *, *], [unu, 0, *, *],
[nn=u, *, *, 0], [nn==, *, *, 1], [nu=u, *, *, 0], [nu==, *, *, 1], [uu=n, *, *, 0], [uu==, *, *, 1],
[un=n, *, *, 0], [u==, *, *, 1],
[n=nu, *, 1, *], [n=n=, *, 0, *], [n=uu, *, 1, *], [n=u=, *, 0, *], [u=nn, *, 1, *], [u=n=, *, 0, *],
[u=un, *, 1, *], [u=u=, *, 0, *],
[nnnu, *, *, *], [nnu=, *, *, *], [nun=, *, *, *], [unnn, *, *, *], [uun=, *, *, *], [unu=, *, *, *],
[nuuu, *, *, *], [uuun, *, *, *].

Detecting Contradictions in Q

- How to avoid contradictions between a_5 and a_1 ?
 - Modelling the following 3 conditions is sufficient ($\delta q = \delta b_4$)

$$\begin{aligned}q &= a_5 \boxminus a_1, \\(\delta q \boxplus q)[0] &= (\delta b_3 \boxplus q \ggg s)[32 - s], \\(\delta q \boxplus q)[s] &= (\delta b_3 \boxplus q \ggg s)[0].\end{aligned}$$

Other Details

- Many other minor details (see the paper)
 - Model $b_4 = b_3 \lll s$ and $b_5 = a_1 \boxplus b_4$
 - A different way to model the expansion
 - Many optional parameters to control the searching strategies (e.g. where to detect the contradictions)

The Differential Trail for 36-Round RIPEMD-160

Table 10: A partial solution for the 36-round differential characteristic

i	∇X_i	$\pi_r(i)$	i	∇Y_i	$\pi_r(i)$
-5	101001010100101011101011111001000	-5	101001010100101011101011111001000	5	
-4	1110111000100000001110110000011	-4	1110111000100000001110110000011	4	
-3	11111010101100010100111101100010	-3	11111010101100010100111101100010	3	
-2	00011100010001000010011100010010	-2	00011100010001000010011100010010	2	
-1	00111011110101101010010000011111	-1	00111011110101101010010000011111	1	
0	nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn	0	1011100011000001010000010111011	0	
1	m010u0u1m1um0100uuu1um00u0uuu	1	1111110101001110110001010100001	1	
2	1mnu1u0m0m0m10u10uvm01mnu1um1	2	011010010000001010101100101010	2	
3	101nu11111110010mu1110011100011	3	00100111110111101m001101000010m1	3	
4	nnnnnnnn00nnnnnnnnnnnnnnnnnnnnnn	4	010m00001001001m101010101010010	4	
5	1111111010000010000010101100010	5	010010100100011101000m000110010	5	
6	0011000111101101111011010111010	6	10001mnnnnnnnnnnnnnnnn0um101110	6	
7	011010000011101011001111000001	7	0u0m1uum0m010m01mnnm0mnnnnnnnn	7	
8	1001100001011100001001011111011	8	nnnnnnnnnn110u0um0mnnnn1nn0mnnnn	8	
9	0111110001111000110101010100100	9	110010u1000m00u01um01m1010m100m	9	
10	1000010010000000111100110011011	10	u100u01100u0u0111001101u10100111	10	
11	000001010001000101110011011111	11	111m110011110m00110110100m0010	11	
12	101001110010011010101100111110	12	000001010010101110010m011001111	12	
13	1001100000000010001000010001011	13	10001011u1111010m010001u01m010	13	
14	00010101010101010101010101110110	14	01u000011m01m0111001101m010010	14	
15	001000010001111001110010111100	15	101110u110100001010u110010m0010	15	
16	-----	16	11010010100011m010101000001011	16	
17	-----	17	000101u110111011110111101011110	17	
18	-----	18	1010000111000001110011011111110	18	
19	-----	19	0111011100m111001010011m0111111	19	
20	-----	20	010m010011000010100110010100101	20	
21	-----	21	00001110010001011100110110011010	21	
22	-----	22	0001011111010011010u111010100100	22	
23	-----	23	nnnnl-----nnnnnnnnnnnnnnnnnnnnnn	23	
24	-----	24	-----	24	
25	-----	25	-----	25	
26	-----	26	-----	26	
27	-----	27	-----	27	
28	-----	28	-----	28	
29	-----	29	-----	29	
30	-----	30	-----	30	
31	-----	31	-----	31	
32	-----	32	-----	32	
33	-----	33	-----	33	
34	-----	34	-----	34	
35	-----	35	-----	35	
m0	1111101nuu11110101011101110m000	m0	1000100010011011101011100001100		
m1	1110001010001010100001010001010	m1	011010010011101nu110001110001m01		
m2	01110100001111010001110110000001	m2	1001110000111000010010111100101		
m3	0110010101111001001111010101101	m3	00100100001110000011000100111110		
m4	01010101111000110010101001010001	m4	1000011001101010010101001001110		
m5	10000010000110010100000110001110	m5	0010001110110111001111000101001		
m6	00um1011111110w0m11100000100101	m6	1010001011101100110101011101101		
m7	11011100001000100110001010001000	m7	110100010010011101000010001001011		

New Progress

Questions quickly arising:

- Is the technique really useful or efficient? (Other targets like SHA-2?)
- Can it beat ad-hoc dedicated tools? (evidence?)

Our new work⁴:

- The first practical collisions for 40-round RIPEMD-160
- The first practical SFS collisions for 39-round SHA-256 (previous record: 38 rounds at EC 2013 by Mendel et al.)

⁴New Records in Collision Attacks on RIPEMD-160 and SHA-256 (eprint 2023/285)

Summary

- Alternative methods to find trails for the MD-SHA hash family
- Further improve the efficiency?
- More applications? (SHA-256, SHA-512...)

`https://github.com/LFKOKAMI/Find_RIPEMD_Trail`