

An attack on hash function HAVAL-128

WANG Xiaoyun¹, FENG Dengguo² & YU Xiuyuan³

1. School of Mathematics and System Sciences, Shandong University, Jinan 250100, China;

2. Institute of Software, Chinese Academy of Sciences, Beijing 100080, China; 3. Mathematics Department, Hangzhou Teachers College, Hangzhou 310012, China

Correspondence should be addressed to Wang Xiaoyun (email: xywang@sdu.edu.cn)

Received October 18, 2004; revised February 2, 2005

Abstract In this paper, we give a fast attack against hash function—HAVAL-128. HAVAL was presented by Y. L. Zheng et al. at Auscrypto'92. It can be processed in 3, 4 or 5 passes, and produces 128, 160, 192, or 224-bit fingerprint. We break the HAVAL with 128-bit fingerprint. The conclusion is that, given any 1024-bit message m , we just make some modifications about m , and the modified message m can collide with another message m' only with the probability of $1/2^7$, where $m' = m + \Delta m$, in which Δm is a fixed difference selected in advance. In addition, two collision examples for HAVAL-128 are given in this paper.

Keywords: hash function, collision, differential attack, differential characteristic.

DOI: 10.1360/122004-107

Hash function directly applies to data integrity, and can be the security guarantee for many cryptosystems and protocols such as signature, group signature, message authentication code, e-cash, bit commitment, coin-flipping, e-voting, etc. According to the structure of the existing hash functions, it can be mainly divided into two kinds: one is based on the cipher blocks, the other is directly constructed. We name the second the standard hash function.

According to the different message processes, the standard hash functions can be divided into two families: MDx family (MD4^[1], MD5^[2], HAVAL^[3], RIPEMD^[4], RIPEMD-160^[5]) and SHA family (SHA-1^[6], SHA-256, 384, 512^[7]). These hashing algorithms reveal the main design technology of the hash functions.

The cryptoanalysis for MDx hash functions has seen much progress. Dobbertin gave an attack for the full MD4^[8] in 1996, which can find a collision with the probability of 2^{-22} . The latest attack on MD4 is a more efficient attack described by Kasselmann in 1997^[9]. As for MD5, den Boer and Bosselaersobber found a kind of pseudo-collision for MD5 which is composed of the same message with two sets of different initial values^[10]. In Eurocrypto'96, Dobbertin presented one collision of MD5 which is made up of two different messages under another set of initial values^[11]. Other analysis results about MDx reduced versions can be seen in refs. [12–16].

For SHA family, Chabaud and Joux^[17] presented a differential attack on SHA-0 which can find a collision with probability 2^{-61} . Joux^[18] announced a 4-block collision for the full SHA-0 in which the complexity of finding the collision is about 2^{51} . Biham and Chen^[19] also gave a near-collision attack on SHA-0, and described their improved results on SHA-0 and SHA-1 in the Rump session^[20].

The purpose of this paper is to break the full HAVAL with 128-bit fingerprint rapidly. Before our break, the latest cryptanalysis is an attack on HAVAL-128 given by Rompay et al.^[21], which can find a collision with about 2^{29} HAVAL-128 computations.

Our attack has the following properties:

1. The attack is a modular differential attack, and all the differential characteristics in the collision are given.
2. The attack gives all the conditions that guarantee all the differential characteristics to occur.
3. The attack is very fast to be performed by computer. Its running time is only about 2^7 times that of HAVAL-128.

1 HAVAL scheme

HAVAL is a hashing algorithm that can compress any length messages in 3, 4 or 5 passes and produce a variable length output —128-bit, 160-bit, 192 or 224-bit fingerprint. The purpose of this paper is to break HAVAL with 3 passes (called HAVAL-128), so we only describe the details of HAVAL-128.

HAVAL-128 employs three passes, H1, H2 and H3. Each pass includes one high non-linear function that performs bit-wise operations on 32-bit words.

$$\begin{aligned}
 f_1(x_6, x_5, x_4, x_3, x_2, x_1, x_0) &= x_1x_4 \oplus x_2x_5 \oplus x_3x_6 \oplus x_0x_1 \oplus x_0, \\
 f_2(x_6, x_5, x_4, x_3, x_2, x_1, x_0) &= x_1x_2x_3 \oplus x_2x_4x_5 \oplus x_1x_2 \oplus x_1x_4 \\
 &\quad \oplus x_2x_6 \oplus x_3x_5 \oplus x_4x_5 \oplus x_0x_2 \oplus x_0, \\
 f_3(x_6, x_5, x_4, x_3, x_2, x_1, x_0) &= x_1x_2x_3 \oplus x_1x_4 \oplus x_2x_5 \oplus x_3x_6 \oplus x_0x_3 \oplus x_0.
 \end{aligned}$$

Description of HAVAL-128. Pad the original message M to make the length of the message be a multiple of 1024 bits. We neglect the padding details^[3] because it has no relation with our attack. For one 1024-bit block m of M , the compressing process is as follows:

1. $a_0 = a, b_0 = b, c_0 = c, d_0 = d, e_0 = e, f_0 = f, g_0 = g, h_0 = h$, (a, b, c, d, e, f, g, h) are the initial input parameters for m . If m is the first 1024-bit message block to be hashed, it is the initial values:

$$\begin{aligned}
 a &= 0x243f6a88, \quad b = 0x85a308d3, \quad c = 0x13198a2e, \quad d = 0x03707344, \\
 e &= 0xa4093822, \quad f = 0x299f31d0, \quad g = 0x082efa98, \quad h = 0xec4e6c89.
 \end{aligned}$$

Otherwise it is the last set of outputs in the former message block.

2. For $j = 1, 2, 3$, for $i = 32(j-1), 32(j-1)+1, 32(j-1)+2, \dots, 32(j-1)+31$

$$\begin{aligned} p_i &= f_j(\phi_j(g_i, f_i, e_i, d_i, c_i, b_i, a_i), \\ r &= (p_i \ggg 7) + (h_i \ggg 11) + m_{ord(j,i)} + k_{j,i}, \\ h_{i+1} &= g_i, \\ g_{i+1} &= f_i, \\ f_{i+1} &= e_i, \\ e_{i+1} &= d_i, \\ d_{i+1} &= c_i, \\ c_{i+1} &= b_i, \\ b_{i+1} &= a_i, \\ a_{i+1} &= r. \end{aligned}$$

3. Let $a = a_{96}, b = b_{96}, c = c_{96}, d = d_{96}, e = e_{96}, f = f_{96}, g = g_{96}, h = h_{96}$.

4. $a = a_{96} + a, b = b_{96} + b, \dots, h = h_{96} + h$. 5. When the message m is the last 1024-bit block, the 128-fingerprint $y_3 * y_2 * y_1 * y_0$ is computed as follows:

$$\begin{aligned} h &= h_3 * h_2 * h_1 * h_0, \\ g &= g_3 * g_2 * g_1 * g_0, \\ f &= f_3 * f_2 * f_1 * f_0, \\ e &= e_3 * e_2 * e_1 * e_0, \\ y_3 &= d + h_3 * g_2 * f_1 * e_0, \\ y_2 &= c + h_2 * g_1 * f_0 * e_3, \\ y_1 &= b + h_1 * g_0 * f_3 * e_2, \\ y_0 &= a + h_0 * g_3 * f_2 * e_1. \end{aligned}$$

The operation in each step employs a constant $k_{j,i}$ (see ref. [3]), where $k_{0,i} = 0$, $0 \leq i \leq 31$. “ \ggg ” represents the circular right shift. “+” is additive operation modular 2^{32} . $x * y$ denotes the concatenation of x and y .

The compressing process for m includes 96 steps. Steps 1–32 belong to pass one, pass two is from 33-th step to 64-th step, and pass three is from 65-th to 96-th step.

The orders of message words in each pass are defined by Table 1.

There are three permutations of variables $\phi_i (i = 1, 2, 3)$ (see Table 2).

2 Some basic conclusions and notations

We first describe some properties of nonlinear functions $f_1 \circ \phi$, $f_2 \circ \phi$ and $f_3 \circ \phi$, which are important for breaking HAVAL-128.

Table 1 The orders of message words

$H_1(ord(1, i))$	00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15
	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
$H_2(ord(2, i))$	05	14	26	18	11	28	07	16	00	23	20	22	01	10	04	08
	30	03	21	09	17	24	29	06	19	12	15	13	02	25	31	27
$H_3(ord(3, i))$	19	09	04	20	28	17	08	22	29	14	25	12	24	30	16	26
	31	15	07	03	01	00	18	27	13	06	21	10	23	11	05	02

Table 2 The ϕ -permutation of variables in three passes

Permutation	x_6	x_5	x_4	x_3	x_2	x_1	x_0
ϕ_1	x_1	x_0	x_3	x_5	x_6	x_2	x_4
ϕ_2	x_4	x_2	x_1	x_0	x_5	x_3	x_6
ϕ_3	x_6	x_1	x_2	x_3	x_4	x_5	x_0

$$f_1(\phi_1(x_6, x_5, x_4, x_3, x_2, x_1, x_0)) = x_2x_3 \oplus x_0x_6 \oplus x_1x_5 \oplus x_2x_4 \oplus x_4,$$

$$f_2(\phi_2(x_6, x_5, x_4, x_3, x_2, x_1, x_0)) = x_0x_3x_5 \oplus x_1x_2x_5 \oplus x_4x_5 \oplus x_0x_2 \oplus x_3x_5 \\ \oplus x_1x_3 \oplus x_1x_2 \oplus x_5x_6 \oplus x_6,$$

$$f_3(\phi_3(x_6, x_5, x_4, x_3, x_2, x_1, x_0)) = x_3x_4x_5 \oplus x_2x_5 \oplus x_1x_4 \oplus x_3x_6 \oplus x_0x_3 \oplus x_0.$$

Proposition 1. Let $y_1 = f_1(\phi_1(x_6, x_5, x_4, x_3, x_2, x_1, x_0))$, $y_{1,i} = f_1(\phi_1(x_6, \dots, x_{i+1}, \neg x_i, x_{i-1}, \dots, x_0))$. Then

1. $y_1 = y_{1,0}$ if and only if $x_6 = 0$.
2. $y_1 = y_{1,1}$ if and only if $x_5 = 0$.
3. $y_1 = y_{1,2}$ if and only if $x_3 \oplus x_4 = 0$.
4. $y_1 = y_{1,3}$ if and only if $x_2 = 0$.
5. $y_1 = y_{1,4}$ if and only if $x_2 = 1$.
6. $y_1 = y_{1,5}$ if and only if $x_1 = 0$.
7. $y_1 = y_{1,6}$ if and only if $x_0 = 0$.

Here, $x_i \in \{0, 1\}$ ($0 \leq i \leq 6$), and $\neg x_i$ is the complement of x_i .

Proof. we just prove (3), others can be proven in the same way.

\implies From $y_1 = y_{1,2}$, we know:

$$x_2x_3 \oplus x_0x_6 \oplus x_1x_5 \oplus x_2x_4 \oplus x_4 = \neg x_2x_3 \oplus x_0x_6 \oplus x_1x_5 \oplus \neg x_2x_4 \oplus x_4.$$

So,

$$x_2x_3 \oplus x_2x_4 = \neg x_2x_3 \oplus \neg x_2x_4.$$

Thus

$$x_3 \oplus x_4 = 0.$$

\Leftarrow From $x_3 \oplus x_4 = 0$, it is easy to prove $y_1 = y_{1,2}$.

Proposition 2. Let $y_2 = f_2(\phi_2(x_6, x_5, x_4, x_3, x_2, x_1, x_0))$, $y_{2,i} = f_2(\phi_2(x_6, \dots, x_{i+1}, \neg x_i, x_{i-1}, \dots, x_0))$. Then

1. $y_2 = y_{2,0}$ if and only if $x_3x_5 \oplus x_2 = 0$.
2. $y_2 = y_{2,1}$ if and only if $x_2x_5 \oplus x_2 \oplus x_3 = 0$.
3. $y_2 = y_{2,2}$ if and only if $x_1x_5 \oplus x_0 \oplus x_1 = 0$.
4. $y_2 = y_{2,3}$ if and only if $x_0x_5 \oplus x_1 \oplus x_5 = 0$.
5. $y_2 = y_{2,4}$ if and only if $x_5 = 0$.
6. $y_2 = y_{2,5}$ if and only if $x_0x_3 \oplus x_1x_2 \oplus x_3 \oplus x_4 \oplus x_6 = 0$.
7. $y_2 = y_{2,6}$ if and only if $x_5 = 1$.

Proposition 3. $y_3 = f_3(\phi_3(x_6, x_5, x_4, x_3, x_2, x_1, x_0))$, $y_{3,i} = f_3(\phi_3(x_6, \dots, x_{i+1}, \neg x_i, x_{i-1}, \dots, x_0))$. Then

1. $y_3 = y_{3,0}$ if and only if $x_3 = 1$.
2. $y_3 = y_{3,1}$ if and only if $x_4 = 0$.
3. $y_3 = y_{3,2}$ if and only if $x_5 = 0$.
4. $y_3 = y_{3,3}$ if and only if $x_0 \oplus x_6 \oplus x_4x_5 = 0$.
5. $y_3 = y_{3,4}$ if and only if $x_1 \oplus x_3x_5 = 0$.
6. $y_3 = y_{3,5}$ if and only if $x_2 \oplus x_3x_4 = 0$.
7. $y_3 = y_{3,6}$ if and only if $x_3 = 0$.

Notations. In order to describe our attack conveniently, we define some notations.

1. $m = (m_0, m_1, \dots, m_{31})$, $m' = (m'_0, m'_1, \dots, m'_{31})$ are two 1024-bit messages.
2. $\Delta m_i = m'_i - m_i$, $\Delta a_i = a'_i - a_i$, \dots , $\Delta h_i = h'_i - h_i$, $\Delta p_i = p'_i - p_i$ denote the modular differences of two variables. These notations are used to describe differential characteristics with \pm symbols in our attack.

It is remarked that the modular difference definition is a bit different from that given by Biham and Shamir^[22], which is defined as the XOR of two variables.

3. $a_i, b_i, c_i, d_i, e_i, f_i, g_i, h_i$ represent the outputs of i -th step.

According to the HAVAL algorithm, we know that $b_i = a_{i-1}$, $c_i = a_{i-2}$, $d_i = a_{i-3}$, $e_i = a_{i-4}$, $f_i = a_{i-5}$, $g_i = a_{i-6}$, $h_i = a_{i-7}$.

4. $x_{i,j}$ denotes the j -th bit of 32-bit word x_i . For example, $a_{i,j}$ is the j -th bit of a_i .

5. $x_i[j]$ is the value by only changing the j -th bit of x_i , so $x_i[j]$ and x_i are only different at the j -th bit.

6. $x_i[j, \dots, j+k]$ is the value by successively changing the j -th, $(j+1)$ -th, \dots $(j+k)$ -th bits of x_i .

7. $i + k$ in 2^{i+k} is additive operation modular 32, $0 \leq i, k \leq 31$.

3 The attack against HAVAL-128

About the security of HAVAL-128, we have the following theorem.

Theorem. There is an attack to find one collision with the running time of about 2^7 HAVAL-128 algorithms.

Proof. The attack is divided into five parts:

1. Select the difference of two messages m, m' as

$$\Delta m = m' - m = (\Delta m_0, \Delta m_1, \dots, \Delta m_{31})$$

such that

$$\Delta m_0 = 2^{10}, \Delta m_{11} = 2^{31}, \Delta m_{18} = 2^3, \Delta m_j = 0, 0 \leq j \leq 31, j \neq 0, 11, 18.$$

By a large amount of cryptanalysis, we find that two messages with such a difference easily consist of one collision.

2. Determine all the differential characteristics such that (m, m') consists of a collision (See Table 3 and Table 4). The collision includes two partial collisions, the internal collision from 1–40 steps and the other internal collision from 86–93 steps.

3. Deduce all the conditions under which the differential characteristics in Table 3 and Table 4 hold.

1) It is clear that $a'_1 = a_1[11]$ in the first step if and only if $a_{1,11} = 0$.

2) From 1 of proposition 1, $a'_2 = a_2$ in the 2nd step if and only if $f_{0,11} = 0$.

3) From 2 of proposition 1, $a'_3 = a_3$ in the 3rd step if and only if $d_{0,11} = 0$.

4) From 3 of proposition 1, $a'_4 = a_4$ in the 4th step if and only if $a_{0,11} = b_{0,11}$ (From the initial values, we know $a_{0,11} = b_{0,11} = 0$).

5) From 4 of proposition 1, $a'_5 = a_5$ in the 5th step if and only if $a_{2,11} = 0$.

6) From 5 of proposition 1, $a'_6 = a_6$ in the 6th step if and only if $a_{3,11} = 1$.

7) We will show the conditions under which the differential characteristics in step 7 hold.

$$(a_6, a_5, a_4, a_3, a_2, a_1[11], a_0, b_0) \longrightarrow (a_7[4, 5, 6, 7, 8], a_6, a_5, a_4, a_3, a_2, a_1[11], a_0).$$

Table 3 Differential characteristics in the partial collision of 1–40 steps

Step	m'_i	Δa_i	The outputs ($a'_i, b'_i, c'_i, d'_i, e'_i, f'_i, g'_i, h'_i$) of $h(m')$
1	m'_0	2^{10}	$a_1[11], a_0, b_0, c_0, d_0, e_0, f_0, g_0$
2	m_1	0	$a_2, a_1[11], a_0, b_0, c_0, d_0, e_0, f_0$
3	m_2	0	$a_3, a_2, a_1[11], a_0, b_0, c_0, d_0, e_0$
4	m_3	0	$a_4, a_3, a_2, a_1[11], a_0, b_0, c_0, d_0$
5	m_4	0	$a_5, a_4, a_3, a_2, a_1[11], a_0, b_0, c_0$
6	m_5	0	$a_6, a_5, a_4, a_3, a_2, a_1[11], a_0, b_0$
7	m_6	2^3	$a_7[4, \dots, 8], a_6, a_5, a_4, a_3, a_2, a_1[11], a_0$
8	m_7	2^{28}	$a_8[29, \dots, 32], a_7[4, \dots, 8], a_6, a_5, a_4, a_3, a_2, a_1[11]$
9	m_8	-2^{21} -2^{22}	$a_9[22, 23], a_8[29, \dots, 32], a_7[4, \dots, 8], a_6, a_5, a_4, a_3, a_2$
10	m_9	0	$a_{10}, a_9[22, 23], a_8[29, \dots, 32], a_7[4, \dots, 8], a_6, a_5, a_4, a_3,$
11	m_{10}	-2^{14}	$a_{11}[15], a_{10}, a_9[22, 23], a_8[29, \dots, 32], a_7[4, \dots, 8], a_6, a_5, a_4$
12	m'_{11}	0	$a_{12}, a_{11}[15], a_{10}, a_9[22, 23], a_8[29, \dots, 32], a_7[4, \dots, 8], a_6, a_5$
13	m_{12}	2^0	$a_{13}[1], a_{12}, a_{11}[15], a_{10}, a_9[22, 23], a_8[29, \dots, 32], a_7[4, \dots, 8], a_6$
14	m_{13}	0	$a_{14}, a_{13}[1], a_{12}, a_{11}[15], a_{10}, a_9[22, 23], a_8[29, \dots, 32], a_7[4, \dots, 8]$
15	m_{14}	0	$a_{15}, a_{14}, a_{13}[1], a_{12}, a_{11}[15], a_{10}, a_9[22, 23], a_8[29, \dots, 32]$
16	m_{15}	2^{17}	$a_{16}[18], a_{15}, a_{14}, a_{13}[1], a_{12}, a_{11}[15], a_{10}, a_9[22, 23]$
17	m_{16}	-2^{11}	$a_{17}[12], a_{16}[18], a_{15}, a_{14}, a_{13}[1], a_{12}, a_{11}[15], a_{10}$
18	m_{17}	0	$a_{18}, a_{17}[12], a_{16}[18], a_{15}, a_{14}, a_{13}[1], a_{12}, a_{11}[15]$
19	m'_{18}	0	$a_{19}, a_{18}, a_{17}[12], a_{16}[18], a_{15}, a_{14}, a_{13}[1], a_{12}$
20	m_{19}	0	$a_{20}, a_{19}, a_{18}, a_{17}[12], a_{16}[18], a_{15}, a_{14}, a_{13}[1]$
21	m_{20}	2^{21}	$a_{21}[22], a_{20}, a_{19}, a_{18}, a_{17}[12], a_{16}[18], a_{15}, a_{14}$
22	m_{21}	0	$a_{22}, a_{21}[22], a_{20}, a_{19}, a_{18}, a_{17}[12], a_{16}[18], a_{15}$
23	m_{22}	0	$a_{23}, a_{22}, a_{21}[22], a_{20}, a_{19}, a_{18}, a_{17}[12], a_{16}[18]$
24	m_{23}	2^6	$a_{24}[7], a_{23}, a_{22}, a_{21}[22], a_{20}, a_{19}, a_{18}, a_{17}[12]$
25	m_{24}	-2^0	$a_{25}[1, 2, 3, 4], a_{24}[7], a_{23}, a_{22}, a_{21}[22], a_{20}, a_{19}, a_{18}$
26	m_{25}	0	$a_{26}, a_{25}[1, 2, 3, 4], a_{24}[7], a_{23}, a_{22}, a_{21}[22], a_{20}, a_{19}$
27	m_{26}	0	$a_{27}, a_{26}, a_{25}[1, 2, 3, 4], a_{24}[7], a_{23}, a_{22}, a_{21}[22], a_{20}$
28	m_{27}	0	$a_{28}, a_{27}, a_{26}, a_{25}[1, 2, 3, 4], a_{24}[7], a_{23}, a_{22}, a_{21}[22]$
29	m_{28}	2^{10}	$a_{29}[11], a_{28}, a_{27}, a_{26}, a_{25}[1, 2, 3, 4], a_{24}[7], a_{23}, a_{22}$
30	m_{29}	0	$a_{30}, a_{29}[11], a_{28}, a_{27}, a_{26}, a_{25}[1, 2, 3, 4], a_{24}[7], a_{23}$
31	m_{30}	0	$a_{31}, a_{30}, a_{29}[11], a_{28}, a_{27}, a_{26}, a_{25}[1, 2, 3, 4], a_{24}[7]$
32	m_{31}	0	$a_{32}, a_{31}, a_{30}, a_{29}[11], a_{28}, a_{27}, a_{26}, a_{25}[1, 2, 3, 4]$
33	m_5	-2^{21}	$a_{33}[22], a_{32}, a_{31}, a_{30}, a_{29}[11], a_{28}, a_{27}, a_{26}$
34	m_{14}	0	$a_{34}, a_{33}[22], a_{32}, a_{31}, a_{30}, a_{29}[11], a_{28}, a_{27}$
35	m_{26}	0	$a_{35}, a_{34}, a_{33}[22], a_{32}, a_{31}, a_{30}, a_{29}[11], a_{28}$
36	m'_{18}	0	$a_{36}, a_{35}, a_{34}, a_{33}[22], a_{32}, a_{31}, a_{30}, a_{29}[11]$
37	m'_{11}	0	$a_{37}, a_{36}, a_{35}, a_{34}, a_{33}[22], a_{32}, a_{31}, a_{30}$
38	m_{28}	0	$a_{38}, a_{37}, a_{36}, a_{35}, a_{34}, a_{33}[22], a_{32}, a_{31}$
39	m_7	0	$a_{39}, a_{38}, a_{37}, a_{36}, a_{35}, a_{34}, a_{33}[22], a_{32}$
40	m_{16}	0	$a_{40}, a_{39}, a_{38}, a_{37}, a_{36}, a_{35}, a_{34}, a_{33}[22]$

It is easy to prove that $a_{5,11} = 1$ and $p_{6,11} = 0$ cause the difference $\Delta p_6 = 2^{10}$, so the difference $\Delta a_7 = 2^3$.

Table 4 Differential characteristics in the partial collision of 86–94 steps

86	m'_0	2^{10}	$a_{86}[11], a_{85}, a_{84}, a_{83}, a_{82}, a_{81}, a_{80}, a_{79}$
87	m'_{18}	0	$a_{87}, a_{86}[11], a_{85}, a_{84}, a_{83}, a_{82}, a_{81}, a_{80}$
88	m_{27}	0	$a_{88}, a_{87}, a_{86}[11], a_{85}, a_{84}, a_{83}, a_{82}, a_{81}$
89	m_{13}	0	$a_{89}, a_{88}, a_{87}, a_{86}[11], a_{85}, a_{84}, a_{83}, a_{82}$
90	m_6	0	$a_{90}, a_{89}, a_{88}, a_{87}, a_{86}[11], a_{85}, a_{84}, a_{83}$
91	m_{21}	0	$a_{91}, a_{90}, a_{89}, a_{88}, a_{87}, a_{86}[11], a_{85}$
92	m_{10}	0	$a_{92}, a_{91}, a_{90}, a_{89}, a_{88}, a_{87}, a_{86}[11], a_{85}$
93	m_{23}	0	$a_{93}, a_{92}, a_{91}, a_{90}, a_{89}, a_{88}, a_{87}, a_{86}[11]$
94	m'_{11}	0	$a_{94}, a_{93}, a_{92}, a_{91}, a_{90}, a_{89}, a_{88}, a_{87}$

Thus $a'_7 = a_7[4, 5, 6, 7, 8]$ if and only if $a_{7,4} = 1, a_{7,5} = 1, a_{7,6} = 1, a_{7,7} = 1, a_{7,8} = 0$.

Because

$$p_{6,11} = a_{4,11}a_{3,11} \oplus a_{6,11}a_{0,11} \oplus a_{5,11}a_{1,11} \oplus a_{4,11}a_{2,11} \oplus a_{2,11},$$

from $p_{6,11} = 0, a_{1,11} = 0, a_{2,11} = 0, a_{3,11} = 1$ and $a_{0,11} = 0$, we know that $a_{4,11} = 0$.

So all the conditions under which the differential characteristic in step 7 is right are as follows:

$$a_{5,11} = 1, a_{7,4} = 1, a_{7,5} = 1, a_{7,6} = 1, a_{7,7} = 1, a_{7,8} = 0, a_{4,11} = 0.$$

8) We can similarly deduce all other conditions which result in other differential characteristics in Table 3 and Table 4. Summing up all the conditions, we obtain table 5. It can be easily verified that these conditions are sufficient for all the differential characteristics in the collision.

4. Modify m so that most of the conditions in Table 5 hold.

(a) Simple modifications

For $j = 0, 1, 2, \dots, 31$, we make some simple modifications for message words to ensure all the conditions in pass 1 hold.

The modification of a_{j+1} : Modify some bits of a_{j+1} in the first pass such that the modified a_{j+1} satisfies all the conditions in $(j+1)$ -th step of Table 5.

For example, a_{18} is modified such that $a_{18,12} = 0, a_{18,18} = 1$ and $a_{18,7} = 0$ by the following method:

$$a_{18} \leftarrow a_{18} \oplus (a_{18,12} \lll 11) \oplus ((a_{18,18} \oplus 1) \lll 17) \oplus (a_{18,7} \lll 6).$$

The modification of m_j :

$$m_j \leftarrow a_{j+1} - (p_j \ggg 7) - (h_j \ggg 11).$$

Table 5 The conditions under which (m, m') is a collision

Step	The conditions of the chaining variable in each step
1	$a_{1,4} = 1, a_{1,5} = 0, a_{1,6} = 0, a_{1,7} = 0, a_{1,8} = 0, a_{1,11} = 0,$
2	$a_{2,11} = 0, a_{2,29} = 1, a_{2,30} = 1, a_{2,31} = 0, a_{2,32} = 0$
3	$a_{3,4} = 0, a_{3,5} = 0, a_{3,6} = 0, a_{3,7} = 1, a_{3,8} = 0, a_{3,11} = 1, a_{3,22} = 0, a_{3,23} = 0,$
4	$a_{4,4} = a_{2,4} + 1, a_{4,7} = 1, a_{4,11} = 0, a_{4,29} = 0, a_{4,30} = 0, a_{4,31} = 0, a_{4,32} = 0,$
5	$a_{5,4} = 1, a_{5,7} = 0, a_{5,8} = 0, a_{5,11} = 1, a_{5,15} = 0, a_{5,22} = 1, a_{5,23} = 0,$ $a_{5,29} = a_{3,29}, a_{5,30} = a_{3,30}$
6	$a_{6,4} = a_{5,4}, a_{6,5} = a_{5,5}, a_{6,6} = a_{5,6}, a_{6,7} = 0, a_{6,8} = 0, a_{6,22} = 0,$ $a_{6,29} = 1, a_{6,30} = 1$
7	$a_{7,1} = 0, a_{7,4} = 1, a_{7,5} = 1, a_{7,6} = 1, a_{7,7} = 1, a_{7,8} = 0, a_{7,11} = 0, a_{7,15} = 0,$ $a_{7,22} = 0, a_{7,29} = 1, a_{7,30} = a_{6,30}, a_{7,31} = a_{6,31}, a_{7,32} = a_{6,32},$
8	$a_{8,4} = 0, a_{8,5} = 0, a_{8,6} = 0, a_{8,7} = 0, a_{8,8} = 0, a_{8,22} = 0, a_{8,23} = a_{7,23}, a_{8,29} = 1,$ $a_{8,30} = 1, a_{8,31} = 1, a_{8,32} = 0,$
9	$a_{9,1} = 0, a_{9,4} = 1, a_{9,5} = 1, a_{9,6} = 1, a_{9,7} = 0, a_{9,8} = 1, a_{9,22} = 1, a_{9,23} = 1,$ $a_{9,29} = 0, a_{9,30} = 0, a_{9,31} = 0, a_{9,32} = 0$
10	$a_{10,7} = a_{7,7}, a_{10,8} = 0, a_{10,15} = a_{9,15}, a_{10,18} = 1, a_{10,22} = 0, a_{10,23} = 0,$ $a_{10,29} = 1, a_{10,30} = 1, a_{10,31} = 1, a_{10,32} = 1$
11	$a_{11,4} = 0, a_{11,5} = 0, a_{11,6} = 0, a_{11,7} = 0, a_{11,8} = 1, a_{11,12} = 0, a_{11,15} = 1,$ $a_{11,22} = 1, a_{11,23} = 1$
12	$a_{12,1} = a_{11,1}, a_{12,15} = 0, a_{12,18} = 0, a_{12,29} = 0, a_{12,30} = 0, a_{12,31} = 0, a_{12,32} = 0$
13	$a_{13,1} = 0, a_{13,4} = 0, a_{13,5} = 0, a_{13,6} = 0, a_{13,7} = 0, a_{13,8} = 0,$ $a_{13,12} = 0, a_{13,15} = 1, a_{13,22} = 0, a_{13,23} = 0$
14	$a_{14,1} = 0, a_{14,18} = 0, a_{14,29} = 0, a_{14,30} = 0, a_{14,31} = 0, a_{14,32} = 1$
15	$a_{15,1} = 1, a_{15,15} = 0, a_{15,18} = 0, a_{15,22} = 0, a_{15,23} = 0$
16	$a_{16,12} = a_{15,12}, a_{16,18} = 0$
17	$a_{17,1} = 0, a_{17,12} = 1, a_{17,15} = 0, a_{17,18} = 0, a_{17,22} = 0$
18	$a_{18,7} = 0, a_{18,12} = 0, a_{18,18} = 1$
19	$a_{19,1} = 0, a_{19,2} = 0, a_{19,3} = 0, a_{19,4} = 0, a_{19,12} = 1$
20	$a_{20,7} = 0, a_{20,18} = 0, a_{20,22} = a_{19,22}$
21	$a_{21,1} = 0, a_{21,2} = 0, a_{21,3} = 0, a_{21,4} = 0, a_{21,12} = 0, a_{21,22} = 0$
22–23	$a_{22,18} = 0, a_{22,22} = 0, a_{23,7} = a_{22,7}, a_{23,11} = 0, a_{23,12} = 0, a_{23,22} = 1$
24	$a_{24,1} = a_{23,1}, a_{24,2} = a_{23,2}, a_{24,3} = a_{23,3}, a_{24,4} = a_{23,4}, a_{24,7} = 0$
25	$a_{25,1} = 0, a_{25,2} = 0, a_{25,3} = 0, a_{25,4} = 1, a_{25,7} = 0, a_{25,11} = 0, a_{25,22} = 0$
26	$a_{26,1} = 0, a_{26,2} = 0, a_{26,3} = 0, a_{26,4} = 0, a_{26,7} = 1$
27	$a_{27,1} = 1, a_{27,2} = 1, a_{27,3} = 1, a_{27,4} = 1, a_{27,11} = 0, a_{27,22} = 0$
28	$a_{28,7} = 0, a_{28,11} = 0, a_{28,22} = 0$
29	$a_{29,1} = 0, a_{29,2} = 0, a_{29,3} = 0, a_{29,4} = 0, a_{29,11} = 0, a_{29,22} = 0$
30	$a_{30,7} = 0, a_{30,11} = 0, a_{30,22} = 1$
31	$a_{31,1} = 0, a_{31,2} = 0, a_{31,3} = 1, a_{31,4} = 0, a_{31,11} = 0, a_{31,22} = 0$
32–34	$a_{32,11} = 0, a_{32,22} = 0, a_{33,11} = 1, a_{33,22} = 1, a_{34,22} = 1$
35–37	$a_{35,11} = a_{34,11} + 1, a_{35,22} = 0, a_{36,22} = 1, a_{37,22} = 1$
82–89	$a_{82,11} = 1, a_{83,11} = 0, a_{84,11} = 0, a_{85,11} = 1, a_{86,11} = 0, a_{87,11} = 0, a_{89,11} = 0$

After the above modifications, the modified m and m' satisfy all the conditions in pass 1. So (m, m') is a collision with the probability of 2^{14} .

(b) Advanced modifications

If $a_{33,11} = 0$, we modify successively $m_5, m_6, m_7, m_8, m_9, m_{10}, m_{11}, m_{12}, m_{13}$ according to Table 6 ($j = 11$).

We can easily know that this modification only causes the change of 11-th bit for a_6 in pass 1. After the modification, $a_{33,11}$ changes from 0 to 1, and all the conditions in the first pass (Table 5) remain unchanged.

If $a_{33,22} = 0$, we modify $m_5, m_6, m_7, m_8, m_9, m_{10}, m_{11}, m_{12}, m_{13}$ successively by Table 6 ($j = 22$).

Table 6 The modification of m_5, \dots, m_{123} for correcting $a_{34,j}, j = 11, 22$

Step	m_k	The modified m_k	
6	m_5	$m_5 \leftarrow m_5 + 2^{j-1}$	$a_6[j], a_5, a_4, a_3, a_2, a_1, a_0, b_0$
7	m_6	$m_6 \leftarrow a_7 - (p_6 \ggg 7) - (h_6 \ggg 11)$	$a_7, a_6[j], a_5, a_4, a_3, a_2, a_1, a_0$
8	m_7	$m_7 \leftarrow a_8 - (p_7 \ggg 7) - (h_7 \ggg 11)$	$a_8, a_7, a_6[j], a_5, a_4, a_3, a_2, a_1$
9	m_8	$m_8 \leftarrow a_9 - (p_8 \ggg 7) - (h_8 \ggg 11)$	$a_9, a_8, a_7, a_6[j], a_5, a_4, a_3, a_2$
10	m_9	$m_9 \leftarrow a_{10} - (p_9 \ggg 7) - (h_9 \ggg 11)$	$a_{10}, a_9, a_8, a_7, a_6[j], a_5, a_4, a_3$
11	m_{10}	$m_{10} \leftarrow a_{11} - (p_{10} \ggg 7) - (h_{10} \ggg 11)$	$a_{11}, a_{10}, a_9, a_8, a_7, a_6[j], a_5, a_4$
12	m_{11}	$m_{11} \leftarrow a_{12} - (p_{11} \ggg 7) - (h_{11} \ggg 11)$	$a_{12}, a_{11}, a_{10}, a_9, a_8, a_7, a_6[j], a_5$
13	m_{12}	$m_{12} \leftarrow a_{13} - (p_{12} \ggg 7) - (h_{12} \ggg 11)$	$a_{13}, a_{12}, a_{11}, a_{10}, a_9, a_8, a_7, a_6[j]$
14	m_{13}	$m_{13} \leftarrow a_{14} - (p_{13} \ggg 7) - (h_{13} \ggg 11)$	$a_{14}, a_{13}, a_{12}, a_{11}, a_{10}, a_9, a_8, a_7$

It is clear that, after the above modifications, (m, m') consists of a collision with the probability of $1/2^{12}$. More complicated modifications can make all the conditions in passes 1 and 2 hold. So, the collision probability can be improved to $1/2^7$.

At the end of this section, we give two collisions which are only different in the final word (see Table 7). Such a collision structure reveals that it is very easy to find collisions of HAVAL-128, and we can find many collisions with various structures such as collisions different only in a few words, one word, or especially one byte, etc.

It is remarked that the words m_0, m_1, \dots, m_{31} are located from right to left, and in each word, the least significant bit is in the left.

4 Conclusion

Summing up the above descriptions, we know that for any randomly selected message m we can easily get a collision with an efficient probability by modifying m . So, HAVAL-128 is not secure.

In this paper, we described only one kind of collision with a fixed message difference.

Table 7 Two pairs of collision for HAVAL-128

	6377448b	d9e59f18	f2aa3cbb	d6cb92ba	ee544a44	879fa576	1ca34633	76ca5d4f
M_0	a67a8a42	8d3adc8b	b6e3d814	5630998d	86ea5dcd	a739ae7b	54fd8e32	acbb2b36
	38183c9a	b67a9289	c47299b2	27039ee5	dd555e14	839018d8	aabbd9c9	d78fc632
	fff4b3a7	40000096	7f466aac	ffffbc0	5f4016d2	5f4016d0	12e2b0	f4307f87
	6377488b	d9e59f18	f2aa3cbb	d6cb92ba	ee544a44	879fa576	1ca34633	76ca5d4f
M'_0	a67a8a42	8d3adc8b	b6e3d814	d630998d	86ea5dcd	a739ae7b	54fd8e32	acbb2b36
	38183c9a	b67a9289	c47299ba	27039ee5	dd555e14	839018d8	aabbd9c9	d78fc632
	fff4b3a7	40000096	7f466aac	ffffbc0	5f4016d2	5f4016d0	12e2b0	f4307f87
h_0	95b5621c	ca62817a	a48dacd8	6d2b54bf				
	6377448b	d9e59f18	f2aa3cbb	d6cb92ba	ee544a44	879fa576	1ca34633	76ca5d4f
M_1	a67a8a42	8d3adc8b	b6e3d814	5630998d	86ea5dcd	a739ae7b	54fd8e32	acbb2b36
	38183c9a	b67a9289	c47299b2	27039ee5	dd555e14	839018d8	aabbd9c9	d78fc632
	fff4b3a7	40000096	7f466aac	ffffbc0	5f4016d2	5f4016d0	12e2b0	f5b16963
	6377488b	d9e59f18	f2aa3cbb	d6cb92ba	ee544a44	879fa576	1ca34633	76ca5d4f
M'_1	a67a8a42	8d3adc8b	b6e3d814	d630998d	86ea5dcd	a739ae7b	54fd8e32	acbb2b36
	38183c9a	b67a9289	c47299ba	27039ee5	dd555e14	839018d8	aabbd9c9	d78fc632
	fff4b3a7	40000096	7f466aac	ffffbc0	5f4016d2	5f4016d0	12e2b0	f5b16963
h_1	b0e99492	d64eb647	5149ef30	4293733c				

In fact, we can find many other kinds of collisions for HAVAL-128. For example, if initial values satisfy that,

$$f_{0,i} = 0, d_{0,i} = 0, a_{0,i} = b_{0,i},$$

we can use the similar attack to get many collisions (m, m') with the following differences:

$\Delta m_0 = 2^{i-1}$, $\Delta m_{11} = 2^{i-12}$, $\Delta m_{18} = 2^{i-8}$, $\Delta m_j = 0$, $0 \leq j \leq 31$, $j \neq 0, 11, 18$, where $0 \leq i \leq 31$. The success probability to find one collision is also about $1/2^7$.

If any of $f_{0,i} = 0$, $d_{0,i} = 0$, $a_{0,i} = b_{0,i}$ does not hold, we can find a 2048-bit collision $(m_0 * m_1, m_0 * m'_1)$. The rough attack is that, we first find a random 1024-bit message m_0 such that the last set of outputs satisfy $f_{96,i} = 0$, $d_{96,i} = 0$ and $a_{96,i} = b_{96,i}$, then we can find another message block collision (m_1, m'_1) by using the above last set of outputs as the initial values. A real collision $(m_0 * m_1, m_0 * m'_1)$ for HAVAL-128 with two blocks is found.

In addition, our attack is available to attack HAVAL-160.

Acknowledgements We would like to thank Jim Phalen and Deb Phalen for their helpful language modification during the writing of this paper. This work was supported by the National Natural Science Foundation of China (Grant No. 90304009), and the "973 Project" (Grant No. G19990358).

References

1. Rivest, R. L., The MD4 message digest algorithm, Advances in Cryptology, Crypto'90, 1991, LNCS 537: 303–311.

2. Rivest, R. L., The MD5 message-digest algorithm, Request for Comments (RFC 1320), 1992.
3. Zheng, Y., Pieprzyk, J., Seberry, J., HAVAL—A one-way hashing algorithm with variable length of output, *Advances in Cryptology, Auscrypto'92*, LNCS 718: 83–104.
4. RIPE, Integrity primitives for secure information systems, Final report of RACE integrity primitives evaluation (RIPE-RACE 1040), LNCS 1007, 1995.
5. Dobbertin, H., Bosselaers, A., Preneel, B., RIPMEMD-160: A strengthened version of RIPMMD, *Fast Software Encryption*, 1996, LNCS 1039: 71–82.
6. FIPS 180-1, Secure hash standard, NIST, US Department of Commerce, Washington D. C.: Springer-Verlag, 1996.
7. FIPS 180-2, Secure hash standard, <http://csrc.nist.gov/publications/>, 2002.
8. Dobbertin, H., Cryptanalysis of MD4, *Fast Software Encryption*, 1996, LNCS 1039: 53–69.
9. Kasselmann, P., A fast attack on the MD4 hash function, in *Proceedings of the 1997 South African Symposium on Communications and Signal Processing (COMSIG'97)*, 1997, 147–150.
10. Boer, B. den, Bosselaers, A., Collisions for the compression function of MD5, *Advances in Cryptology, Eurocrypt'93*, 1994, LNCS 765: 293–304.
11. Dobbertin, H., Cryptanalysis of MD5 compress, *Advances in Cryptology, Eurocrypt'96, Rump Session*, 1996.
12. Boer, B. den, Bosselaers, A., An attack on the last two rounds of MD4, *Advances in Cryptology, Crypto'91*, 1992, LNCS 576: 194–203.
13. Dobbertin, H., RIPEMD with two round compress function is not collision-free, *J. Cryptology*, 1997, 10(1): 51–70.
14. Her, Y. S., Sakurai, K., Kim, S. H., Attack for finding collision in reduced versions of 3-pass and 4-pass HAVAL, in *Proceedings of International Conference on Computers, Communications and Systems (2003ICCCS)*, CE-15: 75–78.
15. Kasselmann, P. R., Penzhorn, W. T., Cryptanalysis of reduced version of HAVAL, *Electronic Letters*, 2000, 36(1): 30–31.
16. Park, S., Sung, S. H., Chee, S. et al., On the security of reduced versions of 3-pass HAVAL, *Proceedings of ACISP*, 2002, 406–419.
17. Chabaud, F., Joux, A., Differential collisions in SHA-0, *Advances in Cryptology, Crypto'98*, 1998, LNCS 1462: 56–71.
18. Joux, A., Collisions for SHA-0, *Rump Session of Crypto'04*.
19. Biham, E., Chen, R., Near collision for SHA-0, *Advances in Cryptology, Crypto'04*, 2004, LNCS 3152: 290–305.
20. Biham, E., Chen, R., New results on SHA-0 and SHA-1, *Rump Session of Crypto'04*.
21. Rompay, B. V., Biryukov, A., Preneel, B. et al., Cryptanalysis of 3-pass HAVAL, *Asiacrypt'2003*, 2003, LNCS 2894: 228–245.
22. Biham, E., Shamir, A., *Differential cryptanalysis of the data encryption standard*, Berlin: Springer-Verlag, 1993.