How to Break MD5 and other hash functions

Xiaoyun Wang and Hongbo Yu (China)

Presented by: Saar Benodiz May 2012

Outline

- Introduction
- Description of MD5
- Differential Attack for Hash Functions
- Message Modification
- Generate Messages
- Summary

Introduction

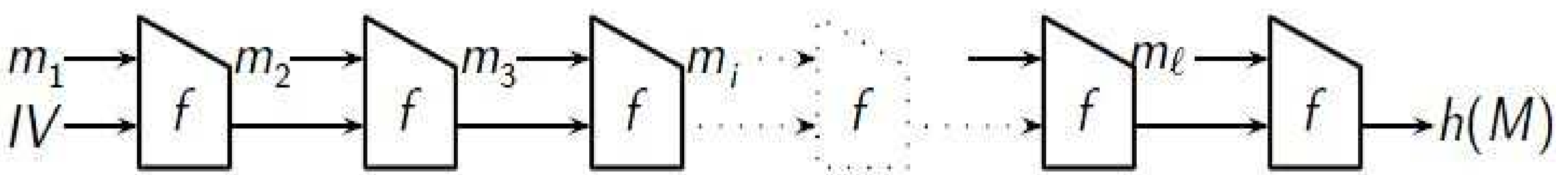
- MD5 was designed in 1992 as an improvement of MD4.
- In this lecture we present a new powerful attack on MD5 which allows us to find collisions efficiently.
- We used this attack to find collision of MD5 in about 15 minutes up to an hour computation time.

Introduction

- The attack is a differential attack, which unlike most differential attack, does not use the exclusive-or as a measure of difference, but instead uses also modular integer subtraction as the measure.
- An application of this attack to MD4 can find collision in less than a fraction of a second.
- This attack is also applicable to other hash functions, such as RIPEMD and HAVAL.

Description of MD5

- Take messages of size up to 2^{64} and outputs 128 bit.
- A message is padded so the length is a multiple of 512.
- Each 512 bit block is compressed individually



- (Merkle-Damgard)
- IV is 4 word each 32-bit a,b,c,d . The output of each f is a,b,c,d for next level.

Description of MD5

- Let $h_{i-1} = (a_0, b_0, c_0, d_0)$
- Let M_i message block be $M_i = (w_0, w_1, \ldots, w_{15})$
- For i=0 to 63

$a_{i+1} = d_i$

$d_{i+1} = c_i$

$c_{i+1} = b_i$

$b_{i+1} = b_i + (a_{i+} F_i(b_i, c_i, d_i) + w_{g(i)} + k_i) <<< s_i$

All additions are modulo 2^{32}

Description of MD5

- For each f there are 4 rounds and each round has 16 steps
- For fixed i, 4 consecutive steps will yield

$a_{i+4} = b_i + ((a_i + F_i(b_i, c_i, d_i) + w_{g(i)} + k_i) <<< s_i)$

$d_{i+4} = a_i + ((d_i + F_{i+1}(a_i, b_i, c_i) + w_{g(i+1)} + k_{i+1}) <<< s_{i+1})$

$c_{i+4} = d_i + ((c_i + F_{i+2}(d_i, a_i, b_i) + w_{g(i+2)} + k_{i+2}) <<< s_{i+2})$

$b_{i+4} = c_i + ((b_i + F_{i+3}(c_i, d_i, a_i) + w_{g(i+3)} + k_{i+3}) <<< s_{i+3})$

Description of MD5

- Each round, a different message word is used, a different round constant is used, and a different function and rotations, this provides non-linearity.

$F_i(X,Y,Z)=(X \wedge Y) \vee (\sim X \wedge Z)$ $0 \leq i \leq 15$

$F_i(X,Y,Z)=(X \wedge Z) \vee (Y \wedge \sim Z)$ $16 \leq i \leq 31$

$F_i(X,Y,Z)=X \oplus Y \oplus Z$ $32 \leq i \leq 47$

$F_i(X,Y,Z)=Y \oplus (X \vee \sim Z)$ $48 \leq i \leq 63$

Ki is constant that based on sin

Finding Collisions on MD5

- MD5 has a 128 bit hash so a brute force attack to find a collision requires at most 2^{128} applications of MD5 and 2^{64} by the birthday paradox
- In 1993, B. den Boer and A. Bosselaers found collision of the same message with two different sets of initial values.
- In 1996 H. Dobbertin presented collision of two different block with chosen IV

Finding Collisions on MD5

- Xiaoyun Wang and Hongbo Yu show an attack that requires $2^{39} + 2^{32}$ MD5 operations
- This attack takes at most an hour and 5 minutes on a IBM P690 (supercomputer)
- we want to find a pair (M_0, M_1) and (M_0', M_1') such that:

$$\begin{aligned}
(a, b, c, d) &= \mathrm{MD5}(a_0, b_0, c_0, d_0, M_0), \\
(a', b', c', d') &= \mathrm{MD5}(a_0, b_0, c_0, d_0, M_0'), \\
\mathrm{MD5}(a, b, c, d, M_1) &= \mathrm{MD5}(a', b', c', d', M_1'),
\end{aligned}$$

Differential Attack for Hash Functions

- The attack uses two types of differentials
- XOR differential: $\Delta X = X \oplus X'$
- Modular differential: $\Delta X = X' - X \bmod 2^{32}$ $(2^{31} - 2^{31})$

- The combination of both kinds of differences give us more information than each of them keep by itself.

Differential Attack for Hash Functions

- For example When X'-X= 2^6 the xor diffrents can have many possibilities.

1. One-bit difference in bit 7, i.e., 0x00000040. In this case means that bit 7 in X' is 1 and bit 7 in X is 0.

   X' = 0100 0000

   X = 0000 0000

2. Two-bit difference, in which a different carry is transferred from bit 7 to bit 8, i.e., 0x000000C0.

   X' = 1000 0000

   X = 0100 0000

Differential Attack for Hash Functions

- Xor difference is marked by the list of active bits with their relative sign ,For example, the difference -2^6 [7,8,9,...22,-23] All bits of X from bit 7 to bit 22 are 0, and bit 23 is 1, while all bits of X` from bit 7 to bit 22 are 1, and bit 23 is 0.

- For $M=(m_0,\ldots,m_{n-1})$ and $M'=(m'_0,\ldots m'_{n-1})$ the full hash differential is:

  $\Delta H_0 \rightarrow \Delta H_1 \rightarrow \ldots \rightarrow \Delta H_{n=} \Delta H$

If M and M' are a collision pair $\Delta H=0$

Round Differential

- Provided that the hash function as 4 rounds and each round as 16 step we can represent each function as:
- $\Delta H_i \rightarrow \Delta H_{i+1}$:
- $\Delta H_i \overset{P_1}{-->} \Delta R_{i+1,1} \overset{P_2}{-->} \Delta R_{i+1,2} \overset{P_3}{-->} \Delta R_{i+1,3} \overset{P_4}{-->} \Delta R_{i+1,4} = \Delta H_{i+1}$
- And each round as:
- $\Delta R_{j-1} \overset{P_{j,1}}{-->} \Delta x_1 \overset{P_{j,2}}{-->} \ldots \overset{P_{j,16}}{-->} \Delta x_{16} = \Delta R_j$
- The probability P of $\Delta H_i \rightarrow \Delta H_{i+1}$ is:

$$P \geq \prod_{l=1}^{4} P_j \ and \ P_j \geq \prod_{t=1}^{16} P_{jt}$$

Round Differential

- Each of these differentials has a probabilistic relationship with the next.
- Ideally, we'd like to be able to set up 2 messages where we can guarantee with probability 1 that $\Delta H=0$
- This can be assured by modifying M so the first round differential will be what you want
- More modifications will improve the probability for the second, third and fourth round differentials
- ΔM_0 has been picked to improve this as well

Differential Attack on MD5

- Find $M=(M_0,M_1\ )$ and $M'=(M'_0,M'_1)$
- $\Delta M_0=M'_0-M_0=(0,0,0,0,2^{31},0,0,0,0,0,0,2^{15},0,0,2^{31},0)$
- $\Delta M_1=M'_1-M_1=(0,0,0,0,2^{31},0,0,0,0,0,0,-2^{15},0,0,2^{31},0)$
- $\Delta H_1=(2^{31},2^{31}+2^{25},2^{31}+2^{25},2^{31}+2^{25})$

- ΔM_0 has been picked to improve the probability that the round differentials will hold(ΔH_1).
- M'_0 differ in the 5th, 12th and 15th words only
- Same for M_1 and M'_1.
- ΔM_1 has been selected not only to ensure both 3-4 round differentials will hold, but also to produce output difference that can be cancelled with the output difference ΔH_1

Sufficient Conditions

Step

Chaining Variable for M_0

Message Word for M_0

Shift Rotation

Message Word Difference

Chaining Variable Difference

Chaining Variable for M_0'

Table 3. The Differential Characteristics in the First Iteration Differential

Step	The output in i-th step for M_0	w_i	s_i	Δw_i	The output difference in i-th step	The output in i-th step for M_0'
4	b_1	m_3	22			
5	a_2	m_4	7	2^{31}	-2^{6}	$a_2\,[7,\ldots,22,-23]$
6	d_2	m_5	12		$-2^{6}+2^{23}+2^{31}$	$d_2\,[-7,24,32]$
7	c_2	m_6	17		$-1-2^{6}+2^{23}-2^{27}$	$c_2\,[7,8,9,10,11,-12,-24,-25,-26,$ $27,28,29,30,31,32,1,2,3,4,5,-6]$
8	b_2	m_7	22		$1-2^{15}-2^{17}-2^{23}$	$b_2\,[1,16,-17,18,19,20,-21,-24]$,
9	a_3	m_8	7		$1-2^{6}+2^{31}$	$a_3\,[-1,2,7,8,-9,-32]$
10	d_3	m_9	12		$2^{12}+2^{31}$	$d_3\,[-13,14,32]$
11	c_3	m_{10}	17		$2^{30}+2^{31}$	$c_3\,[31,32]$
12	b_3	m_{11}	22	2^{15}	$-2^{7}-2^{13}+2^{31}$	$b_3\,[8,-9,14,\ldots,19,-20,32]$
13	a_4	m_{12}	7		$2^{24}+2^{31}$	$a_4\,[-25,26,32]$
14	d_4	m_{13}	12		2^{31}	$d_4\,[32]$
15	c_4	m_{14}	17	2^{31}	$2^{3}-2^{15}+2^{31}$	$c_4\,[4,-16,32]$
16	b_4	m_{15}	22		$-2^{29}+2^{31}$	$b_4\,[-30,32]$
17	a_5	m_1	5		2^{31}	$a_5\,[32]$
18	d_5	m_6	9		2^{31}	$d_5\,[32]$
19	c_5	m_{11}	14	2^{15}	$2^{17}+2^{31}$	$c_5\,[18,32]$
20	b_5	m_0	20		2^{31}	$b_5\,[32]$
21	a_6	m_5	5		2^{31}	$a_6\,[32]$
22	d_6	m_{10}	9		2^{31}	$d_6\,[32]$
23	c_6	m_{15}	14			c_6
24	b_6	m_4	20	2^{31}		b_6
25	a_7	m_9	5			a_7
26	d_7	m_{14}	9	2^{31}		d_7
27	c_7	m_3	14			c_7
...	...	...	...	...	...	...
34	d_9	m_8	11			d_9
35	c_9	m_{11}	16	2^{15}	2^{31}	$c_9\,[*32]$
36	b_9	m_{14}	23	2^{31}	2^{31}	$b_9\,[*32]$
37	a_{10}	m_1	4		2^{31}	$a_{10}\,[*32]$
38	d_{10}	m_4	11	2^{31}	2^{31}	$d_{10}\,[*32]$
39	c_{10}	m_7	16		2^{31}	$c_{10}\,[*32]$
...	...	...	...	...	...	...
45	a_{12}	m_9	4		2^{31}	$a_{12}\,[*32]$
46	d_{12}	m_{12}	11		2^{31}	$d_{12}\,[32]$
47	c_{12}	m_{15}	16		2^{31}	$c_{12}\,[32]$
48	b_{12}	m_2	23		2^{31}	$b_{12}\,[32]$
49	a_{13}	m_0	6		2^{31}	$a_{13}\,[32]$
50	d_{13}	m_7	10		2^{31}	$d_{13}\,[-32]$
51	c_{13}	m_{14}	15	2^{31}	2^{31}	$c_{13}\,[32]$
52	b_{13}	m_5	21		2^{31}	$b_{13}\,[-32]$
...	...	...	...	...	...	...
58	d_{15}	m_{15}	10		2^{31}	$d_{15}\,[-32]$
59	c_{15}	m_6	15		2^{31}	$c_{15}\,[32]$
60	b_{15}	m_{13}	21		2^{31}	$b_{15}\,[32]$
61	$aa_0 = a_{16} + a_0$	m_4	6	2^{31}	2^{31}	$aa_0' = aa_0[32]$
62	$dd_0 = d_{16} + d_0$	m_{11}	10	2^{15}	2^{31}	$dd_0' = dd_0[26,32]$
63	$cc_0 = c_{16} + c_0$	m_2	15		2^{31}	$cc_0' = cc_0[-26,27,32]$
64	$bb_0 = b_{16} + b_0$	m_9	21		2^{31}	$bb_0' = bb_0[26,-32]$

Sufficient Conditions

- Derive a set of sufficient conditions that guarantee the differential characteristic in Step 8 of MD5 (Table 3) to hold:
- The differential characteristic in Step 8 of MD5 is:

$$(\Delta c_2, \Delta d_2, \Delta a_2, \Delta b_1) \longrightarrow \Delta b_2.$$

- Each chaining variable satisfies one of the following equations.
- a_I, b_I, c_I, d_I respectibely denote the outputs of the (4i-3)-th, (4i-2)-th,(4i-1)-th and 4i-th steps for compressing M wherere 1>=i<=16. $a`_I$, $b`_I$, $c`_I$, $d`_I$ are defined similarly

$$\begin{aligned}
b_1' &= b_1 \\
a_2' &= a_2[7, ..., 22, -23] \\
d_2' &= d_2[-7, 24, 32] \\
c_2' &= c_2[7, 8, 9, 10, 11, -12, -24, -25, -26, 27, 28, 29, 30, 31, 32, 1, 2, 3, 4, 5, -6] \\
b_2' &= b_2[1, 16, -17, 18, 19, 20, -21, -24]
\end{aligned}$$

Sufficient Conditions

- According to the operations in the 8-th step, we have

$$b_2 = c_2 + ((b_1 + F(c_2, d_2, a_2) + m_7 + t_7) \lll 22$$

$$b_2' = c_2' + ((b_1 + F(c_2', d_2', a_2') + m_7' + t_7) \lll 22$$

$$\phi_7 = F(c_2, d_2, a_2) = (c_2 \wedge d_2) \vee (\neg c_2 \wedge a_2)$$

In the above operations, c_2 occurs twice in the right hand side of the equation. In order to distinguish the two, let c_2^F denote the c_2 inside F, and c_2^{NF} denote the c_2 outside F.

The derivation is based on the following two facts:

1. Since $\Delta b_1 = 0$ and $\Delta m_7 = 0$, we know that $\Delta b_2 = \Delta c_2^{NF} + (\Delta\phi_7 \lll 22)$.
2. Fix one or two of the variables in F so that F is reduced to a single variable.

By the similar method, we can derive a set of sufficient conditions (see Table 4 and Table 6) which guarantee all the differential characteristics in the collision differential to hold.

Sufficient Conditions

Table 4

c_1	$c_{1,7}=0,\ c_{1,12}=0,\ c_{1,20}=0$
b_1	$b_{1,7}=0,\ b_{1,8}=c_{1,8},\ b_{1,9}=c_{1,9},\ b_{1,10}=c_{1,10},\ b_{1,11}=c_{1,11},\ b_{1,12}=1,\ b_{1,13}=c_{1,13},$ $b_{1,14}=c_{1,14},\ b_{1,15}=c_{1,15},\ b_{1,16}=c_{1,16},\ b_{1,17}=c_{1,17},\ b_{1,18}=c_{1,18}, b_{1,19}=c_{1,19},$ $b_{1,20}=1,\ b_{1,21}=c_{1,21},\ b_{1,22}=c_{1,22},\ b_{1,23}=c_{1,23},\ b_{1,24}=0,\ b_{1,32}=1$
a_2	$a_{2,1}=1,\ a_{2,3}=1,\ a_{2,6}=1,\ a_{2,7}=0,\ a_{2,8}=0,\ a_{2,9}=0,\ a_{2,10}=0,\ a_{2,11}=0,$ $a_{2,12}=0,\ a_{2,13}=0,\ a_{2,14}=0,\ a_{2,15}=0,\ a_{2,16}=0,\ a_{2,17}=0,\ a_{2,18}=0,\ a_{2,19}=0,$ $a_{2,20}=0,\ a_{2,21}=0,\ a_{2,22}=0,\ a_{2,23}=1,\ a_{2,24}=0,\ a_{2,26}=0,\ a_{2,28}=1,\ a_{2,32}=1$
d_2	$d_{2,1}=1,\ d_{2,2}=a_{2,2},\ d_{2,3}=0,\ d_{2,4}=a_{2,4},\ d_{2,5}=a_{2,5},\ d_{2,6}=0,\ d_{2,7}=1,\ d_{2,8}=0,$ $d_{2,9}=0,\ d_{2,10}=0,\ d_{2,11}=1,\ d_{2,12}=1,\ d_{2,13}=1,\ d_{2,14}=1,\ d_{2,15}=0,\ d_{2,16}=1,$ $d_{2,17}=1,\ d_{2,18}=1,\ d_{2,19}=1,\ d_{2,20}=1,\ d_{2,21}=1,\ d_{2,22}=1,\ d_{2,23}=1,\ d_{2,24}=0,$ $d_{2,25}=a_{2,25},\ d_{2,26}=1,\ d_{2,27}=a_{2,27},\ d_{2,28}=0,\ d_{2,29}=a_{2,29},\ d_{2,30}=a_{2,30},$ $d_{2,31}=a_{2,31},\ d_{2,32}=0$
c_2	$c_{2,1}=0,\ c_{2,2}=0,\ c_{2,3}=0,\ c_{2,4}=0,\ c_{2,5}=0,\ c_{2,6}=1,\ c_{2,7}=0,\ c_{2,8}=0,\ c_{2,9}=0,$ $c_{2,10}=0,\ c_{2,11}=0,\ c_{2,12}=1,\ c_{2,13}=1,\ c_{2,14}=1,\ c_{2,15}=1,\ c_{2,16}=1,\ c_{2,17}=0,$ $c_{2,18}=1,\ c_{2,19}=1,\ c_{2,20}=1,\ c_{2,21}=1,\ c_{2,22}=1,\ c_{2,23}=1,\ c_{2,24}=1,\ c_{2,25}=1,$ $c_{2,26}=1,\ c_{2,27}=0,\ c_{2,28}=0,\ c_{2,29}=0,\ c_{2,30}=0,\ c_{2,31}=0,\ c_{2,32}=0$
b_2	$b_{2,1}=0,\ b_{2,2}=0,\ b_{2,3}=0,\ b_{2,4}=0,\ b_{2,5}=0,\ b_{2,6}=0,\ b_{2,7}=1,\ b_{2,8}=0,\ b_{2,9}=1,$ $b_{2,10}=0,\ b_{2,11}=1,\ b_{2,12}=0,\ b_{2,14}=0,\ b_{2,16}=0,\ b_{2,17}=1,\ b_{2,18}=0,\ b_{2,19}=0,$ $b_{2,20}=0,\ b_{2,21}=1,\ b_{2,24}=1,\ b_{2,25}=1,\ b_{2,26}=0,\ b_{2,27}=0,\ b_{2,28}=0,\ b_{2,29}=0,$ $b_{2,30}=0,\ b_{2,31}=0,\ b_{2,32}=0$
a_3	$a_{3,1}=1,\ a_{3,2}=0,\ a_{3,3}=1,\ a_{3,4}=1,\ a_{3,5}=1,\ a_{3,6}=1,\ a_{3,7}=0,\ a_{3,8}=0,\ a_{3,9}=1,$ $a_{3,10}=1,\ a_{3,11}=1,\ a_{3,12}=1,\ a_{3,13}=b_{2,13},\ a_{3,14}=1,\ a_{3,16}=0,\ a_{3,17}=0,\ a_{3,18}=0,$ $a_{3,19}=0,\ a_{3,20}=0,\ a_{3,21}=1,\ a_{3,25}=1,\ a_{3,26}=1,\ a_{3,27}=0,\ a_{3,28}=1,\ a_{3,29}=1,$ $a_{3,30}=1,\ a_{3,31}=1,\ a_{3,32}=1$
d_3	$d_{3,1}=0,\ d_{3,2}=0,\ d_{3,7}=1,\ d_{3,8}=0,\ d_{3,9}=0,\ d_{3,13}=1,\ d_{3,14}=0,\ d_{3,16}=1,$ $d_{3,17}=1,\ d_{3,18}=1,\ d_{3,19}=1,\ d_{3,20}=1,\ d_{3,21}=1,\ d_{3,24}=0,\ d_{3,31}=1,\ d_{3,32}=0$

Message Modification

- It is easy to modify M_0 such that the conditions of round 1 in Table 4 hold with probability 1
- For example We want $c_{1,7}=0$, $c_{1,12}=0$, $c_{1,20}=0$ So we modify m_2 as follows.

$$c_1^{new} \leftarrow c_1^{old} - c_{1,7}^{old} \cdot 2^6 - c_{1,12}^{old} \cdot 2^{11} - c_{1,20}^{old} \cdot 2^{19}$$

$$m_2^{new} \leftarrow ((c_1^{new} - c_1^{old}) \ggg 17) + m_2^{old}.$$

Message Modification

- By modifying each message word of message m_0, all the conditions in round 1 of Table 4 hold (first 16 step). The first iterations differential hold with probability 2^{-43} .

- The same modification is applied to m_1 ,After modification, the second iterations differential hold with probability 2^{-37} .

Multi-Message Modification

- It is even possible to fulfill a part of the conditions of the first 32 steps by a multi-message modification.
- For example, $a_{5,32} = 1$, we correct it into $a_{5,32} = 0$ by modifying m_1, m_2, m_3, m_4, m_5 such that the modification generates a partial collision from 2-6 steps, and remains that all the conditions in round 1 hold.
- Some other conditions can be corrected by the similar modification technique.

Message Modification

- By our modification, 37 conditions in round 2-4 are undetermined in the table 4, and 30 conditions in round 2-4 are undetermined in the table 6.
- So the first iteration differential hold with probability 2^{-37}.
- The second iteration differential hold with probability 2^{-30}.

Generate $M_{0+}M`_0$

- Select random message M_0
- Modify M_0 so it meets the conditions
- $M_0' = M_0 + \Delta M_0$
- This will result in ΔH_1 with probability 2^{-37}
- Test the messages on MD5.
- This doesn't require more then 2^{39} MD5 operations

Generate $M_{1+}M`_1$

- Select random message M_1
- Modify M_1 so it meets the conditions
- $M_1' = M_1 + \Delta M_1$
- Use ΔH_1 as IV , The probability that $\Delta H = 0$ is 2^{-30}
- Test if the messages lead to a collision.
- This doesn't require more then 2^{32} MD5 operations

Creating More Collisions

- To select another message M_0 is only to change the last two words from the previous selected message $M_{0.}$

- it is easy to find many second blocks M_1 , $M`_1$ which lead to collisions.

Summary

- This paper described a powerful attack against hash functions, and in particular showed that finding a collision of MD5 is easily feasible.

- This attack is also able to break efficiently other hash functions, such as HAVAL-128, MD4, RIPEMD, and SHA-0.

References

- How To Break MD5 and Other Hash Functions – Xiaoyun Wang and Hongbo Yu
- www.cs.virginia.edu/cs588/lectures/md5-collisions.ppt