第42卷 第4期 山东大学学报(理学版) 2007年4月
Vol.42 No.4 JOURNAL OF SHANDONG UNIVERSITY Apr. 2007

文章编号:1671-9352(2007)04-0001-05

MD4算法分析

黎 琳

(山东大学 数学与系统科学学院,山东 济南 250100)

摘要:采用比特追踪法对MD4进行攻击,利用差分特性,找到近似碰撞路线,使得给定一个消息 m,可以以高概率找到另一消息 m' 产生碰撞,并保持较低的Hamming重量.

关键词:Hash函数;近似碰撞;差分特征;Hamming重量

中图分类号:TP309 **文献标识码**:A

Cryptanalysis of MD4

LI Lin

(School of Math. and System Sci., Shandong Univ., Jinan 250100, Shandong, China)

Abstract: Bit flipping and differential characteristics are used to attack MD4. A near collision path can be found. Given a message m, a message m' can be found with high probability and low hamming weight to get a collision.

Key words: Hash function; near collision; differential characteristics; Hamming weight

0 引言

Hash函数是信息安全领域重要的研究课题.它不仅可以用于数字签名方案,还可以用于验证信息来源的真实性及信息数据的完整性.它可将任意长度的消息压缩到固定长度的消息摘要.通常,标准Hash函数主要分为两大类:MDx系列,包括MD4,MD5,HAVAL,RIPEMD,RIPEMD-128等和SHA系列包括SHA-1, SHA-256等[1~5].

MD4算法是Rivest于1990年提出的Hash函数算法.其最初的设计目的是抵抗碰撞攻击和第二原根攻击,但已有的攻击表明未达到其设计目标.1996年,H.Dobbertin提出了对MD4算法以概率 2^{-22} 找到碰撞的成功攻击[6].1998年,H.Dobbertin表明MD4算法的前两圈不是单向的,这意味着对于MD4的前两圈可能存在原根和第二原根[7].近期,王小云等提出了一种新的对于Hash函数MD4和RIPEMD的碰撞攻击方法,并应用于MD5、HAVAL-128、SHA-0和SHA-1,取得较大轰动[8~12].

本文采用王小云等提出的比特追踪的方法对MD4的进行攻击,利用差分特性,找到近似碰撞路线,使得给定一个消息 m,可以以高概率找到另一消息 m' 产生碰撞,并保持较低的Hamming重量.

本文首先在第一部分详细描述了HMAC和MD4算法,并给出了文中用到的一些符号的定义;第二部分给出了对MD4算法的分析方法,最后我们对全文进行了总结.

1 MD4算法

1.1 MD4算法

收稿日期:2007-03-26

基金项目:国家自然科学基金重点资助项目(90604036);国家杰出青年基金资助项目(60525201);国家973计划资助项目(2007CB807902)

作者简介:黎 琳(1978-),女,博士研究生,主要研究方向对称密码的分析与设计.

MD4算法是Rivest于1990年提出的Hash函数算法,通过3圈操作可以将任意长度的消息压缩成128位的Hash函数值.算法每圈都包含一个对32比特字进行比特运算的高度非线性圈函数,分别为:

$$F(X,Y,X)=(X\wedge Y)\vee(\neg X\wedge Z),$$
$$G(X,Y,Z)=(X\wedge Y)\vee(X\wedge Z)\vee(Y\wedge Z),$$
$$H(X,Y,Z)=X\oplus Y\oplus Z.$$

其中,X,Y,Z是32比特的字.运算符$\oplus,\wedge,\vee,\neg$分别表示异或,与,或和补运算.

算法中每圈运算都包括16步相同的操作.每步中都有变量a,b,c,d,且变量的值将不断得到更新.

$$\phi_0(a,b,c,d,m_k,s)=((a+F(b,c,d)+m_k)\bmod 2^{32})<<<s,$$
$$\phi_1(a,b,c,d,m_k,s)=((a+G(b,c,d)+m_k+0x5a827999)\bmod 2^{32})<<<s,$$
$$\phi_2(a,b,c,d,m_k,s)=((a+H(b,c,d)+m_k+0x6ed9edba1)\bmod 2^{32})<<<s.$$

其中,m_k是32位明文分组,$<<<s$表示循环左移s位,+表示模2^{32}的加法运算.

MD4的初始值为:

$$(a,b,c,d)=(0x67452301,0xefcdab89,0x98badcfe,0x10325476).$$

对于消息$\bar{M}$的一个512比特的消息分组M,$M=(m_0,m_1,\cdots m_{15})$,运算过程如下:

(1) 消息分组M的输入值为(aa,bb,cc,dd).如果M是第一个512比特被压缩的分组,则其初始值为(aa,bb,cc,dd),否则其初始值是前一个消息分组压缩后的输出值.

(2) 完成以下48步运算(3圈)

For $j=1,2,3$,
  For $i=0,1,2,3$,
    $a=\phi_j(a,b,c,d,\omega_{j,4i},s_{j,4i})$,
    $d=\phi_j(a,b,c,d,\omega_{j,4i+1},s_{j,4i+1})$,
    $c=\phi_j(a,b,c,d,\omega_{j,4i+2},s_{j,4i+2})$,
    $b=\phi_j(a,b,c,d,\omega_{j,4i+3},s_{j,4i+3})$.

其中,$s_{j,4i+k}(k=0,1,2,3)$是常量.每圈中消息字的顺序及移位值见参考文献[1].

(3) 将链接变量a,b,c,d分别加入输入链接变量产生当前消息分组的最终链接变量.

$$aa=(a+aa)\bmod 2^{32},$$
$$bb=(b+bb)\bmod 2^{32},$$
$$cc=(c+cc)\bmod 2^{32},$$
$$dd=(d+dd)\bmod 2^{32}.$$

如果M是最后一个消息分组,则$H(\bar{M})=aa\,|\,bb\,|\,cc\,|\,dd$是消息$\bar{M}$的Hash函数值.否则以$(aa,bb,cc,dd)$作为输入值,对下一个512比特的消息分组重复以上过程.

1.2 符号说明

为了便于说明,我们定义以下符号.

(1) $M=(m_0,m_1,\cdots m_{15})$与$M'=(m'_0,m'_1,\cdots m'_{15})$是两个512比特的消息分组.

(2) a_i,d_i,c_i,b_i分别表示消息分组M第$4i-3,4i-2,4i-1,4i$步的输出,其中$1\le i\le 16$.

(3) a'_i,d'_i,c'_i,b'_i分别表示消息分组M'第$4i-3,4i-2,4i-1,4i$步的输出,其中$1\le i\le 16$.

(4) $\Delta m_i=m'_i-m_i$表示两个消息字的模差分.这些差分可正可负,用于描述带的差分特征.

(5) $a_{i,j},b_{i,j},c_{i,j},d_{i,j}$分别表示$a_i,b_i,c_i,d_i$第$j$比特的值,其中,第1比特表示最低比特位,第32比特表示最高比特位.

(6) $x_i[j],x_i[-j]$是x_i只改变第j比特后的值.$x_i[j]$的值表示将x_i的第j比特从0变到1;$x_i[-j]$表示x_i的第j比特从1变到0.

(7) $x_i[\pm j_1,\pm j_2,\cdots\pm j_l]$表示连续改变$x_i$的第$j_1,j_2,\cdots j_l$比特后的值.

2　对MD4算法的攻击

利用比特追踪法对任意512比特的消息 m,选取 $m' = m + \Delta m = (m'_0, m'_1, \cdots m'_{15})$.如果每步某个输入变量仅改变一个或少数几个比特,则该步输出值可能不发生变化或仅改变一个或少数几个比特.根据这些改变的差分特性及非线性函数性质,我们可以找出近似碰撞路线及满足路线的充分条件.

2.1　选择明文差分

定义两个明文 m 和 m' 的关分 Δm:

$$m = (m_0, m_1, \cdots m_{15}),$$

$$m' = (m'_0, m'_1, \cdots m'_{15}),$$

令 $\Delta m = m' - m = (\Delta m_0, \Delta m_1, \cdots \Delta m_{15})$,经过分析,我们选择明文差分满足:

$$\Delta m_7 = 2^{21},$$

就可以找到近似碰撞.

2.2　确定差分特性

表1表示差分特性.表中1~7列分别表示步数,消息变量,选择的消息的顺序,移位值,明文差分,输出差分,输出变量值.

表1　碰撞的差分特征

Table 1　Differential characteristics of the collision differential for MD4

Step	Output for M	m_i	s_i	Δm_i	Output difference	Output for M'
1	a_1	m_0	3			a_1
2	d_1	m_1	7			d_1
3	c_1	m_2	11			c_1
4	b_1	m_3	19			b_1
5	a_2	m_4	3			a_2
6	d_2	m_5	7			d_2
7	c_2	m_6	11			c_2
8	b_2	m_7	19	2^{21}	2^8	$b_2[9]$
9	a_3	m_8	3			a_3
10	d_3	m_9	7			d_3
11	c_3	m_{10}	11			c_3
12	b_3	m_{11}	19		2^{27}	$b_3[-28,-29,-30,-31,32]$
13	a_4	m_{12}	3			a_4
14	d_4	m_{13}	7		-2^6	$d_4[-7]$
15	c_4	m_{14}	11			c_4
16	b_4	m_{15}	19		2^{14}	$b_4[15]$
17	a_5	m_0	3			a_5
18	d_5	m_4	5		-2^{11}	$d_5[12,13,14,-15]$
19	c_5	m_8	9			c_5
20	b_5	m_{12}	13			b_5
21	a_6	m_1	3			a_6
22	d_6	m_5	5		-2^{16}	$d_6[-17]$
23	c_6	m_9	9			c_6
24	b_6	m_{13}	13			b_6
25	a_7	m_2	3			a_7
26	d_7	m_6	5		-2^{21}	$d_7[-22]$
27	c_7	m_{10}	9			c_7
28	b_7	m_{14}	13			b_7
29	a_8	m_3	3			a_8

续表

Step	Output for M	m_i	s_i	Δm_i	Output difference	Output for M'
30	d_8	m_7	5	2^{21}		d_8
31	c_8	m_{11}	9			c_8
32	b_9	m_{15}	13			b_9
…	…	…	…	…	…	…
43	c_{11}	m_5	11			c_{11}
44	b_{11}	m_{13}	15			b_{11}
45	a_{12}	m_3	3			a_{12}
46	d_{12}	m_{11}	9			d_{12}
47	c_{12}	m_7	11	2^{21}	2^{32}	$c_{12}[1]$
48	b_{12}	m_{15}	15		2^{15}	$b_{12}[16]$

2.3 确定满足差分特征的充分条件

根据差分特征和非线性函数的性质,我们能得到满足表1中差分特征的充分条件表.从表2,我们可以得到近似碰撞概率分别为 2^{-56}.

表2 碰撞的充分条件

Table 2 Sufficient conditions for collisions of MD4

Step	Output for M	
1	a_1	
2	d_1	
3	c_1	
4	b_1	
5	a_2	
6	d_2	
7	c_2	$c_{2,9} = d_{2,9}$
8	b_2	$b_{2,9} = 0$
9	a_3	$a_{3,9} = 0$
10	d_3	$d_{3,9} = 1$
11	c_3	$c_{3,28} = d_{3,28}, c_{3,29} = d_{3,29}, c_{3,30} = d_{3,30}, c_{3,31} = d_{3,31}, c_{3,32} = d_{3,32}$
12	b_3	$b_{3,28} = 1, b_{3,29} = 1, b_{3,30} = 1, b_{3,31} = 1, b_{3,32} = 0, a_{4,7} = b_{3,7}$
13	a_4	$a_{4,28} = 0, a_{4,29} = 0, a_{4,30} = 0, a_{4,31} = 0, a_{4,32} = 1$
14	d_4	$d_{4,7} = 1, d_{4,28} = 1, d_{4,29} = 1, d_{4,30} = 1, d_{4,31} = 1, d_{4,32} = 1$
15	c_4	$c_{4,7} = 0, c_{4,15} = d_{4,15}$
16	b_4	$b_{4,15} = 0, b_{4,7} = c_{4,7}$
17	a_5	$a_{5,11} = c_{4,11}, a_{5,12} = b_{4,12}, a_{5,13} = b_{4,13}, a_{5,14} = b_{4,14}$
18	d_5	$d_{5,12} = 0, d_{5,13} = 0, d_{5,14} = 0, d_{5,15} = 1$
19	c_5	$c_{5,12} = a_{5,12}, c_{5,13} = a_{5,13}, c_{5,14} = a_{5,14}, c_{5,15} = a_{5,15} + 1$
20	b_5	$b_{5,12} = c_{5,12}, b_{5,13} = c_{5,13}, b_{5,14} = c_{5,14}, b_{5,15} = c_{5,15}$
21	a_6	$a_{6,17} = b_{5,17}$
22	d_6	$d_{6,17} = 1$
23	c_6	$c_{6,17} = a_{6,17}$
24	b_6	$b_{6,17} = c_{6,17}$
25	a_7	$a_{7,22} = b_{6,22}$
26	d_7	$d_{7,22} = 1$
27	c_7	$c_{7,22} = a_{7,22}$
28	b_7	$b_{7,22} = c_{7,22}$
29	a_8	
30	d_8	
31	c_8	

续表

Step	Output for M	
32	b_9	
…	…	…
43	c_{11}	
44	b_{11}	
45	a_{12}	
46	d_{12}	
47	c_{12}	$c_{12,1}=0$
48	b_{12}	$b_{12,17}=0$

根据表1和表2,我们给定一消息 m,可以找到另一消息 m',产生碰撞,其概率为 2^{-56},Hamming 重量为2.

3 结论

本文采用比特追踪法对 MD4 进行攻击,利用差分特性,找到近似碰撞路线,使得给定一个消息 m,可以以高概率找到另一消息 m' 产生碰撞,其概率为 2^{-56},Hamming 重量为 2.通过给定的路线跟条件,可能存在两个明文分组间的较好的第二原根攻击.也可用于秘密前缀的 HMAC-MD4 的密钥恢复攻击.

参考文献:

[1] Rivest R L. The MD4 message digest algorithm[A]. Advances in Cryptology, Crypto'90[C]. Berlin: Springer-Verlag, 1990.

[2] Rivest R L. The MD5 message digest algorithm[S]. Request for Comments (RFC 1320), Internet Activities Board, Internet Privacy Task Force, 1992.

[3] Zheng Y, Pieprzyk J, Seberry J. Haval-A one-way hashing algorithm with variable length of output[A]. Advances in Cryptology, Auscrypto'92 Processings [C]. New York: Springer-Verlag, 1992. 83 ~ 104.

[4] Dobbertin H, Bosselaers A, Preneel B. RIPEMD-160: A strengthened version of RIPEMD[A]. Fast Software Encryption, LNCS 1039 [C]. Berlin: Springer-Verlag, 1996. 71-82.

[5] Dobbertin H. RIPEMD with two round compress function is not collision-free[J]. Journal of Cryptology, 1997, 10:51 ~ 69.

[6] Dobbertin H. Cryptanalysis of MD4[A]. Fast Software Encryption, LNCS 1039[C]. Berlin: Springer-Verlag, 1996.

[7] Dobbertin H. The first two rounds of MD4 are not one-way[A]. Fast Software Encryption, LNCS 1372[C]. Berlin: Springer-Verlag, 1998. 284 ~ 292.

[8] Wang X Y, Yu H B. How to break MD5 and other hash functions[A]. Eurocrypt'05, LNCS 3621[C]. Berlin: Springer-Verlag, 2005. 19 ~ 35.

[9] Wang X Y, Lai X J, Feng D G, et al. Cryptanalysis of the hash functions mD4 and RIPEMD[A]. Advances in Cryptology-Eurocrypt 05, LNCS3494[C]. Berlin: Springer-Verlag, 2005. 1 ~ 18.

[10] Wang X Y, Feng D G, Yu X Y. An attack on HAVAL function Haval-128[J]. Science in China Ser F Information Sciences, 2005, 48 (5):1 ~ 12.

[11] Wang X Y, Yu H B, Lisa Y. Efficient collision search attacks on SHA-0[A]. Crypto'05, LNCS 3621[C]. Berlin: Springer-Verlag, 2005. 1 ~ 16.

[12] Wang X Y, Lisa Y, Yu H B. Finding collisions on the full SHA-1[A]. Crypto'05, LNCS 3621[C]. Berlin: Springer-Verlag, 2005. 17 ~ 36.

(编辑:李晓红)