ISSN 1000-9825, CODEN RUXUEW
Journal of Software, Vol.20, No.6, June 2009, pp.1617−1624
doi: 10.3724/SP.J.1001.2009.03349

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel/Fax: +86-10-62562563

MD5 碰撞攻击中的充要条件集*

陈士伟+, 金晨辉

(信息工程大学 电子技术学院,河南 郑州 450004)

Set of Necessary and Sufficient Conditions in Collision Attacks on MD5

CHEN Shi-Wei+, JIN Chen-Hui

(Institute of Electronic Technology, University of Information Engineering, Zhengzhou 450004, China)

+ Corresponding author: E-mail: chenshiwei1012@sohu.com

Chen SW, Jin CH. Set of necessary and sufficient conditions in collision attacks on MD5. *Journal of Software*, 2009,20(6):1617−1624. http://www.jos.org.cn/1000-9825/3349.htm

Abstract: By analyzing the properties of the nonlinear functions used in MD5 and the differences in terms of XOR and subtraction modulo 2^{32}, this paper proves that some sufficient conditions presented by Liang Jie and Lai Xuejia are also necessary to guarantee the differential path from the 23rd step to the 62nd step and give a set of necessary and sufficient conditions to guarantee the output differences of the last two steps. Then, according to the set of necessary and sufficient conditions this paper presents an improved collision attack algorithm on MD5. Finally, it analyzes the average computational complexity of the attack algorithm which is 0.718 7 times of that of the previous collision attack algorithms and proves the efficiency of the improved algorithm by computer simulations.

Key words: MD5; a set of necessary and sufficient conditions; collision attack; differential path

摘 要: 通过分析MD5中非线性函数的性质以及模2^{32}减差分和异或差分的性质,证明了Liang Jie和Lai Xuejia给出的产生MD5碰撞的充分条件集中的条件是保证第23~62步的差分路径满足的充要条件,给出了保证第63、64步的输出差分满足的充要条件集.利用得到的充要条件集,提出了对MD5的改进的碰撞攻击算法,该算法的平均计算复杂度约为已有碰撞攻击算法的0.718 7倍,并通过实验对该算法的改进效果进行了验证.

关键词: MD5;充要条件集;碰撞攻击;差分路径

中图法分类号: TP309 **文献标识码**: A

杂凑函数是将任意长度的消息变换为固定长度二元字符串的一类函数,变换的结果称为杂凑值.MD5是重要的杂凑算法之一,被广泛应用于数字签名和许多密码协议中.1993年,den Boer和Bosselaers[1]提出了MD5的伪碰撞消息,Dobbertin[2]在1996年提出了对MD5的半自由起始碰撞攻击.在2004年的欧洲密码年会上,王小云等人[3]公布了MD5的碰撞对.之后,越来越多的密码学者开始关注于杂凑函数的分析研究.Hawkes等人[4]依据文献[3]给出的碰撞对对碰撞攻击算法进行了猜想,但没有发现其中的一个关键技巧,即消息修改技术.在2005年的

* Supported by the Science Fund for Distinguished Young Scholars of He'nan Province of China under Grant No.0312001800 (河南省杰出青年科学基金)

Received 2007-10-24; Accepted 2008-04-02

欧洲密码年会上,王小云等人[5]首次提出碰撞攻击算法的主要思路,即寻找一条以较大概率产生MD5 碰撞的差分路径,构造保证差分路径成立的充分条件集,然后利用所提出的消息修改技术,使充分条件集中尽可能多的条件一定得到满足,最后通过随机检验的方法使余下的充分条件得到满足.

近年来,许多密码学者对MD5 碰撞攻击算法中的充分条件集以及消息修改技术进行了研究.通过大量的实验,Yajima和Shimoyama[6]发现王小云等人[5]给出的充分条件集并不总能产生碰撞,并对其中的一些条件作了修正.Nakano等人[7]指出,文献[5]的充分条件集中有 16 个冗余条件,并解释了其中一些条件冗余的原因.Sasaki等人[8]系统地给出了构造产生MD5 碰撞的充分条件集的方法.此外,Liang和Lai[9]增加了若干新条件,以使循环左移和模 2^{32}减运算可以交换,从而得到了总能产生MD5 碰撞的充分条件集,进而提出了快速碰撞攻击算法.与此同时,Wang等人[10]研究了同样的问题,并提出了不同的碰撞攻击算法.Stevens[11]优化了产生MD5 碰撞的充分条件集,并提出了新的寻找第 1 块消息的算法.在消息修改技术方面,文献[9–14]提出了改进的消息修改技术,使更多的充分条件得到满足.其中,Sasaki等人[13]提出的改进的消息修改技术可以使文献[5]中第 1 次迭代的第 1~22 步的所有充分条件一定成立.故当通过随机检测的方法使余下的充分条件得到满足时,碰撞攻击算法的计算复杂度由余下的充分条件的个数决定.然而我们发现,余下的某些条件并不是保证差分路径满足的充要条件,这将导致差分路径成立的概率降低,从而增加了算法的计算复杂度.文献[15]提出了对文献[4]的改进方法,崔国华等人[16]结合现有的碰撞分析结论,为杂凑算法的改进提供了相应的思路.

本文通过分析MD5 中非线性函数的性质及模 2^{32}减差分和异或差分的性质,给出了保证第 23~64 步的差分路径成立的充要条件集.利用所得到的充要条件集,本文提出了改进的碰撞攻击算法,该算法的平均计算复杂度约为已有碰撞攻击算法的计算复杂度的 0.718 7 倍,并通过实验对算法的改进效果进行了验证.

1 MD5 算法的简述及符号表示

MD5 是将任意长度的消息转化为 128-比特杂凑值的一个函数,它以 512-比特分组来处理输入消息,每一分组又被划分为 16 个 32-比特消息子分组.MD5 算法的压缩函数包括 4 轮,每一轮包括 16 步.链接变量 a,b,c 和 d 的初始值为

$$a_0 = 0x67452301,\ b_0 = 0xefcdab89,\ c_0 = 0x98badcfe,\ d_0 = 0x10325476.$$

在每一轮中,链接变量 a,b,c 和 d 按下面的顺序更新:

$$a = b + ((a + f(b,c,d) + m + s) <<< k),\ d = a + ((d + f(a,b,c) + m + s) <<< k),$$
$$c = d + ((c + f(d,a,b) + m + s) <<< k),\ b = c + ((b + f(c,d,a) + m + s) <<< k),$$

其中,+是模 2^{32}加法,s和k是给定的常数,m是 32-比特消息子分组,$x<<<k$表示x循环左移k-比特,f是轮函数.f在每一轮的表示如下:

$$\text{第1轮}: f = F(x,y,z) = (x \wedge y) \vee [(\neg x) \wedge z], \quad \text{第2轮}: f = G(x,y,z) = (x \wedge z) \vee [y \wedge (\neg z)],$$
$$\text{第3轮}: f = H(x,y,z) = x \oplus y \oplus z, \quad \text{第4轮}: f = I(x,y,z) = y \oplus [x \vee (\neg z)].$$

对于 $x = \sum_{i=1}^{n} x_i 2^{i-1} \in Z/(2^n),\ x_i \in \{0,1\}$,称$x_i$为$x$的第$i$比特.本文将使用以下符号表示:

$M_0(M_0')$:第 1 块 512-比特输入消息; $M_1(M_1')$:第 2 块 512-比特输入消息;

a_i,d_i,c_i,b_i:输入消息为(M_0,M_1)时,第 $4i-3$ 步~第 $4i$步的输出值($0 \le i \le 16$);

a_i',d_i',c_i',b_i':输入消息为 (M_0',M_1') 时,第 $4i-3$ 步~第 $4i$ 步的输出值($0 \le i \le 16$);

$a_{i,j},d_{i,j},c_{i,j},b_{i,j}$:$a_i,d_i,c_i,b_i$的第$j$-比特($0 \le i \le 16$); $a_{i,j}',d_{i,j}',c_{i,j}',b_{i,j}'$: a_i',d_i',c_i',b_i' 的第j-比特($0 \le i \le 16$);

$H_i(H_i')$:经i次迭代后的结果($i \ge 0$);m_i:第i个 32-比特消息子分组($0 \le i \le 15$);

x_H,x_L:$x = x_H 2^{n-k} + x_L$,x_H为x的高k比特的值,x_L为x的低$n-k$比特的值;

$\Delta X,\delta X$:$\Delta X = (X' - X) \bmod 2^{32}$,$\delta X = X' \oplus X$($X$可以取$a_i,b_i,c_i,d_i$或$\phi_i$);

$|A|$:集合 A 中元素的个数.

2 保证差分路径成立的充要条件集

首先回顾王小云等人[5]给出的产生MD5 碰撞的差分路径,然后分析MD5 算法中非线性函数的性质及模 2^{32} 减差分和异或差分的性质,进而给出保证第 23~64 步的差分路径成立的充要条件集.

2.1 王小云等人提出的产生MD5碰撞的差分路径

王小云等人[5]首次提出的产生MD5 碰撞的差分路径表示为

$$\Delta H_0 = 0 \xrightarrow{(M_0, M_0')} \Delta H_1 \xrightarrow{(M_1, M_1')} \Delta H = 0,$$

其中,

$$M_0' - M_0 = (0,0,0,0,2^{31},0,0,0,0,0,0,2^{15},0,0,2^{31},0),$$

$$M_1' - M_1 = (0,0,0,0,2^{31},0,0,0,0,0,0,-2^{15},0,0,2^{31},0),$$

$$\Delta H_1 = (2^{31}, 2^{31}+2^{25}, 2^{31}+2^{25}, 2^{31}+2^{25}) = (\Delta a_{16}, \Delta d_{16}, \Delta c_{16}, \Delta b_{16}).$$

此外,可以从文献[5]的表 3 和表 5 中得到各步的链接变量在产生 MD5 碰撞的差分路径中所应该满足的输出差分.

2.2 保证差分路径成立的充要条件集

2.2.1 MD5 中的非线性函数的性质

MD5 的压缩函数包括 4 轮,每一轮使用不同的非线性函数.下面将对后 3 轮中非线性函数的一些性质进行分析,并给出 3 个引理.

引理 1. 设$x,y,z \in /Z(2^{32}), G(x,y,z)=(x \wedge z) \vee (y \wedge (\neg z)), \Delta z=2^{31}$,则

(1) 若$\Delta x=\Delta y=2^{31}$,则$\Delta G(x,y,z)=2^{31}$的充要条件是$x_{32}=y_{32}$;

(2) 若$\Delta x=0, \Delta y=2^{31}$,则$\Delta G(x,y,z)=0$ 的充要条件是$x_{32} \oplus y_{32} \oplus z_{32}=0$;

(3) 若$\Delta x=\Delta y=0$,则$\Delta G(x,y,z)=2^{31}$的充要条件是$x_{32}=y_{32} \oplus 1$.

引理 2. 设$x,y,z \in /Z(2^{32}), H(x,y,z)=x \oplus y \oplus z, \Delta x=\Delta y \in \{0,2^{31}\}$,则$\Delta H(x,y,z)=2^{31}$的充要条件是$\Delta z=2^{31}$.

引理 3. 设$x,y,z \in /Z(2^{32}), \Delta x=\Delta y=\Delta z=2^{31}, I(x,y,z)=y \oplus (x \vee (\neg z))$,则:

(1) $\Delta I(x,y,z)=2^{31}$的充要条件是$x_{32}=z_{32}$;

(2) $\Delta I(x,y,z)=2^{31}$的充要条件是$z_{32}=x_{32} \oplus 1$.

2.2.2 保证第 23~64 步的输出差分成立的充要条件集

王小云等人[5]给出了产生MD5 碰撞的差分路径以及保证差分路径成立的充分条件集.随后,Liang Jie和Lai Xuejia[9]增加了若干新条件,从而得到了总能产生碰撞的充分条件集.利用上面得到的非线性函数的性质,可以证明文献[9]的充分条件集中的条件是保证第 23~62 步的输出差分成立的充要条件.

命题1. 文献[9]给出的产生MD5 碰撞的充分条件集中的条件是保证第 1 次迭代中第 23~62 步的输出差分成立的充要条件.

证明:从文献[5]的表 3 中可以得到第 1 次迭代中各步的链接变量所应该满足的输出差分.

在第 23 步的运算中,由于$c_6=d_6+[(G(d_6,a_6,b_5)+c_5+m_{15}+0xd8a1e681) <<<14]$且$\Delta d_6=\Delta a_6=\Delta b_5=2^{31}, \Delta c_5=2^{31}+2^{17}, \Delta m_{15}=0$,记$\Sigma_{23}=G(d_6,a_6,b_5)+c_5+m_{15}+0xd8a1e681$,则$\Delta c_6=0$ 的充要条件是

$$[(\Sigma_{23}+\Delta G(d_6,a_6,b_5)+2^{31}+2^{17}) <<<14] - (\Sigma_{23} <<<14) = 2^{31}.$$

上式等价于$\Delta G(d_6,a_6,b_5)=2^{31}$且$[(\Sigma_{23}+2^{17}) <<<14] - (\Sigma_{23} <<<14)=2^{31}$.再由引理 1 中的(1)得知,$\Delta G(d_6,a_6,b_5)=2^{31}$的充要条件是$d_{6,32}=a_{6,32}$,且$[(\Sigma_{23}+2^{17}) <<<14] - (\Sigma_{23} <<<14)=2^{31}$的充要条件[10]为$(\Sigma_{23})_L<2^{17}$且$(\Sigma_{23})_H<2^{14}$.易知$(\Sigma_{35})_H<2^{14}$成立,且$(\Sigma_{23})_L<2^{17}$成立的充要条件是$\Sigma_{23,18}=0$,故$\Delta c_6=0$ 的充要条件是$\Sigma_{23,18}=0$ 且$d_{6,32}=a_{6,32}$成立.类似地,利用所得到的非线性函数的性质,可以证明文献[9]的充分条件集中的其他条件是保证第 24~62 步的输出差分满足的充要条件.命题 1 得证. □

下面将推导出最后两步的输出差分$\Delta c_{16}=2^{31}+2^{25}$且$\Delta b_{16}=2^{31}+2^{25}$满足的充要条件.先给出两个引理:

引理 4. 设 $x_i,y_i,\alpha_i,\beta_i\in\{0,1\}$, $h(x_i,y_i)=x_i\vee(y_i\oplus 1)$,则

$$h(x_i\oplus\alpha_i,y_i\oplus\beta_i)\oplus h(x_i,y_i)=\begin{cases}y_i\alpha_i, & \text{若}\beta_i=0\\ y_i\alpha_i\oplus x_i\oplus\alpha_i\oplus 1, & \text{若}\beta_i=1\end{cases}.$$

引理 5. 设$X',X\in Z(2^{32})$,$\Delta X=X'-X$,则$\Delta X=2^{31}+2^{25}$的充要条件是对任意的 $0\le i\le 25$ 均有 $X'_i=X_i$ 且下面的(1)和(2)中有一个成立:

(1) $X'_{32}=X_{32}$,且对满足 $0\le t\le 5$ 的 t 均有 $X'_{26+t}=0,X_{26+t}=1$;

(2) $X'_{32}\oplus X_{32}=1$,存在 $q(0\le q\le 5)$使得 $X'_{26+q}=1,X_{26+q}=0$,并对满足 $0\le t<q$ 的 t 均有 $X'_{26+t}=0,X_{26+t}=1$,且对满足 $q<t\le 5$ 的 t 均有 $X'_{26+t}=X_{26+t}$.

利用引理 4 和引理 5,下面将给出保证$\Delta c_{16}=2^{31}+2^{25}$和$\Delta b_{16}=2^{31}+2^{25}$成立的充要条件集.

定理 1. 设 $c_{15},b_{15},a_{16},d_{16},c_{16},b_{16}\in Z/(2^{32})$,$\Delta c_{15}=\Delta b_{15}=\Delta a_{16}=2^{31}$,$\Delta d_{16}=2^{31}+2^{25}$,$\Delta m_2=0$ 且$\Delta m_9=0$,则$\Delta c_{16}=2^{31}+2^{25}$且$\Delta b_{16}=2^{31}+2^{25}$成立的充要条件是下面 4 个结论中至少 1 个成立:

(1) $d_{16,32}=b_{15,32},c_{16,32}=a_{16,32}$,且存在$q,p(0\le q\le p\le 5)$使得对满足 $0\le t<q$的t均有$d_{16,26+t}=1$ 且$d_{16,26+q}=0$,且对满足$0\le t<p$的t均有$c_{16,26+t}=1$ 和$c_{16,26+p}=0$,且对满足 $0\le t\le q$的t均有$b_{15,26+t}=0$,同时有

$$a_{16,26+i}=\begin{cases}1, & \text{若}\ 0\le i\le q\\ 0, & \text{若}\ q<i\le p\end{cases}.$$

(2) $d_{16,32}=b_{15,32}$,且存在$q(0\le q\le 5)$使得对满足 $0\le t<q$的t均有$d_{16,26+t}=1$ 和$d_{16,26+q}=0$,并对满足 $0\le t\le 6$ 的t均有$c_{16,26+t}=1$,且对满足 $0\le t\le q$的t均有$b_{15,26+t}=0$,同时有

$$a_{16,26+i}=\begin{cases}1, & \text{若}\ 0\le i\le q\\ 0, & \text{若}\ q<i\le 5\end{cases}.$$

(3) $c_{16,32}=a_{16,32}=\oplus 1$,且对满足 $0\le t\le 6$ 的t均有$d_{16,26+t}=1$,对满足 $0\le t<5$ 的t均有$c_{16,26+t}=1$ 和$c_{16,31}=0$,且对满足$0\le i\le 5$ 的i均有$b_{15,26+i}=0$ 和$a_{16,26+i}=1$.

(4) $d_{16,32}=1,c_{16,32}=0$,且对满足 $0\le t\le 5$ 的t均有$d_{16,26+t}=1$ 和$c_{16,26+t}=1$,且对满足 $0\le i\le 5$ 的i均有$b_{15,26+i}=0$ 和$a_{16,26+i}=1$.

证明:首先给出$\Delta c_{16}=2^{31}+2^{25}$成立的充要条件.在MD5 算法中第 1 次迭代的第 63 步运算中,有

$$c_{16}=d_{16}+\{[I(d_{16},a_{16},b_{15})+c_{15}+m_2+0x2ad7d2bb]<<<15\},$$

其中,$I(d_{16},a_{16},b_{15})=a_{16}\oplus[d_{16}\vee(\neg b_{15})]$.

由于$\Delta d_{16}=2^{31}+2^{25}$,$\Delta a_{16}=\Delta b_{15}=\Delta c_{15}=2^{31}$且$\Delta m_2=0$,故$\Delta c_{16}=2^{31}+2^{25}$等价于$\Delta I(d_{16},a_{16},b_{15})=2^{31}$,即$\delta I(d_{16},a_{16},b_{15})=2^{31}$.又由$\Delta b_{15}=\Delta a_{16}=2^{31}$知$\delta b_{15}=\delta a_{16}=2^{31}$,故$\Delta c_{16}=2^{31}+2^{25}$的充要条件是

$$(\delta d_{16}\,\&\,b_{15})\oplus[(d_{16}\oplus\delta d_{16})\,\&\,2^{31}]=2^{31}\tag{1}$$

由引理 5 可知,$\Delta d_{16}=2^{31}+2^{25}$可以分为两种情况,从而细化$\Delta c_{16}=2^{31}+2^{25}$的充要条件.

情况 1:$\delta d_{16}\bmod 2^{25}=0$, $d'_{16,32}\oplus d_{16,32}=1$,存在$q(0\le q\le 5)$使得 $d'_{16,26+q}=1$,$d_{16,26+q}=0$,并对任意的 $0\le t<q$均有 $d'_{16,26+t}=0$,$d_{16,26+t}=1$,同时,对任意的$q<t\le 5$ 均有 $d'_{16,26+t}=d_{16,26+t}$,则 $\delta d_{16}=d'_{16}\oplus d_{16}=2^{31}\oplus\overset{q}{\underset{t=0}{\oplus}}2^{25+t}$,故此时公式(1)等价于$\left(\overset{q}{\underset{t=0}{\oplus}}2^{25+t}\,\&\,b_{15}\right)\oplus[(d_{16}\oplus b_{15})\,\&\,2^{31}]=0$,即$d_{16,32}=b_{15,32}$,且对任意的 $0\le t\le q$均有$b_{15,26+t}=0$.

情况 2:$\delta d_{16}\bmod 2^{25}=0$, $d'_{16,32}=d_{16,32}$,且对任意的 $0\le t\le 5$ 均有 $d'_{16,26+t}=0$,$d_{16,26+t}=1$,则 $\delta d_{16}=d'_{16}\oplus d_{16}=\overset{5}{\underset{t=0}{\oplus}}2^{25+t}$,故此时公式(1)等价于 $(d_{16}\,\&\,2^{31})\oplus\left[\left(\overset{5}{\underset{t=0}{\oplus}}2^{25+t}\right)\&\,b_{15}\right]=2^{31}$,即$d_{16,32}=1$,且对满足 $0\le t\le 5$ 的t均有$b_{15,26+t}=0$.

下面在$\Delta c_{16}=2^{31}+2^{25}$的条件下,给出$\Delta b_{16}=2^{31}+2^{25}$的充要条件.在MD5 算法第 1 次迭代的第 64 步运算中,

$$b_{16}=c_{16}+\{[I(c_{16},d_{16},a_{16})+b_{15}+m_9+0xeb86d391]<<<21\},$$

其中,$I(c_{16},d_{16},a_{16})=d_{16}\oplus[c_{16}\vee(\neg a_{16})]$.

由于$\Delta c_{16}=2^{31}+2^{25}$,$\Delta d_{16}=2^{31}+2^{25}$,$\Delta a_{16}=2^{31}+2^{25}$且$\Delta m_9=0$,故$\Delta b_{16}=2^{31}+2^{25}$的充要条件是$\Delta I(c_{16},d_{16},a_{16})=2^{31}$,即

$$\delta I(c_{16},d_{16},a_{16})=(\delta c_{16}\,\&\,a_{16})\oplus(\delta a_{16}\,\&\,c_{16})\oplus(\delta c_{16}\,\&\,\delta a_{16})\oplus\delta d_{16}\oplus\delta a_{16}=2^{31}.$$

再由$\Delta a_{16}=2^{31}$知$\delta a_{16}=2^{31}$,故$\Delta b_{16}=2^{31}+2^{25}$的充要条件是

$$(\delta c_{16}\,\&\,a_{16})\oplus[2^{31}\,\&\,(c_{16}\oplus\delta c_{16})]\oplus\delta d_{16}=0 \tag{2}$$

由引理 5 知,$\Delta d_{16}=2^{31}+2^{25}$和$\Delta c_{16}=2^{31}+2^{25}$都可分为两种情况,下面结合这 4 种情况,将公式(2)具体化.

情况I:$\delta d_{16}\bmod 2^{25}=0$, $d'_{16,32}\oplus d_{16,32}=1$,并存在$q(0\le q\le 5)$使得 $d'_{16,26+q}=1$,$d_{16,26+q}=0$,且对满足 $0\le t<q$的t均有 $d'_{16,26+t}=0$,$d_{16,26+t}=1$.同时,对满足$q<t\le 5$ 的t均有 $d'_{16,26+t}=d_{16,26+t}$;$\delta c_{16}\bmod 2^{25}=0$, $c'_{16,32}\oplus c_{16,32}=1$,并存在$p(0\le p\le 5)$使得 $c'_{16,26+p}=1$,$c_{16,26+p}=0$,且对满足 $0\le t<p$的t均有 $c'_{16,26+t}=0$,$c_{16,26+t}=1$,同时,对满足$p<t\le 5$ 的t均有 $c'_{16,26+t}=c_{16,26+t}$,则 $\delta d_{16}=d'_{16}\oplus d_{16}=2^{31}\oplus\mathop{\oplus}_{t=0}^{q}2^{25+t}$ 且 $\delta c_{16}=c'_{16}\oplus c_{16}=2^{31}\oplus\mathop{\oplus}_{t=0}^{p}2^{25+t}$,故此时公式(2)等价于 $[2^{31}\,\&\,(a_{16}\oplus c_{16})]\oplus\left(\mathop{\oplus}_{t=0}^{p}2^{25+t}\,\&\,a_{16}\right)\oplus\mathop{\oplus}_{t=0}^{q}2^{25+t}=0$,即$a_{16,32}=c_{16,32}$,$q\le p$且

$$a_{16,26+i}=\begin{cases}1, & \text{若 } 0\le i\le q\\ 0, & \text{若 } q<i\le p\end{cases}.$$

情况II:$\delta d_{16}\bmod 2^{25}=0$, $d'_{16,32}\oplus d_{16,32}=1$,并存在$q(0\le q\le 5)$使得 $d'_{16,26+q}=1$,$d_{16,26+q}=0$,且对满足 $0\le t<q$的t均有 $d'_{16,26+t}=0$,$d_{16,26+t}=1$.同时,对满足$q<t\le 5$ 的t均有 $d'_{16,26+t}=d_{16,26+t}$;$\delta c_{16}\bmod 2^{25}=0$, $c'_{16,32}=c_{16,32}$,且对满足 $0\le t\le 5$ 的t均有 $c'_{16,26+t}=0$,$c_{16,26+t}=1$,则 $\delta d_{16}=d'_{16}\oplus d_{16}=2^{31}\oplus\mathop{\oplus}_{t=0}^{q}2^{25+t}$ 且 $\delta c_{16}=c'_{16}\oplus c_{16}=\mathop{\oplus}_{t=0}^{5}2^{25+t}$,故此时公式(2)等价于 $(2^{31}\,\&\,c_{16})\oplus 2^{31}\oplus\left(\mathop{\oplus}_{t=0}^{5}2^{25+t}\,\&\,a_{16}\right)\oplus\mathop{\oplus}_{t=0}^{q}2^{25+t}=0$,即$c_{16,32}=1$ 且

$$a_{16,26+i}=\begin{cases}1, & \text{若 } 0\le i\le q\\ 0, & \text{若 } q<i\le 5\end{cases}.$$

情况III:$\delta d_{16}\bmod 2^{25}=0$, $d'_{16,32}=d_{16,32}$,且对满足 $0\le t\le 5$ 的t均有 $d'_{16,26+t}=0$,$d_{16,26+t}=1$;$\delta c_{16}\bmod 2^{25}=0$, $c'_{16,32}\oplus c_{16,32}=1$,并存在$p(0\le p\le 5)$使得 $c'_{16,26+p}=1$,$c_{16,26+p}=0$,且对满足 $0\le t<p$的t均有 $c'_{16,26+t}=0$,$c_{16,26+t}=1$,同时,对满足$p<t\le 5$ 的t均有 $c'_{16,26+t}=c_{16,26+t}$,则 $\delta d_{16}=d'_{16}\oplus d_{16}=\mathop{\oplus}_{t=0}^{5}2^{25+t}$ 且 $\delta c_{16}=c'_{16}\oplus c_{16}=2^{31}\oplus\mathop{\oplus}_{t=0}^{p}2^{25+t}$,故此时公式(2)等价于 $[2^{31}\,\&\,(a_{16}\oplus c_{16})]\oplus\left(\mathop{\oplus}_{t=0}^{p}2^{25+t}\,\&\,a_{16}\right)\oplus\mathop{\oplus}_{t=0}^{5}2^{25+t}=2^{31}$,即$a_{16,32}\ne c_{16,32}$,$p=5$,且对满足 $0\le i\le 5$ 的i均有$a_{16,26+i}=1$.

情况IV:$\delta d_{16}\bmod 2^{25}=0$, $d'_{16,32}=d_{16,32}$,且对满足 $0\le t\le 5$ 的t均有 $d'_{16,26+t}=0$,$d_{16,26+t}=1$;$\delta c_{16}\bmod 2^{25}=0$, $c'_{16,32}=c_{16,32}$,且对满足 $0\le t\le 5$ 的t均有$c'_{16,26+t}=0$,$c_{16,26+t}=1$,则 $\delta d_{16}=d'_{16}\oplus d_{16}=\mathop{\oplus}_{t=0}^{5}2^{25+t}$ 且 $\delta c_{16}=c'_{16}\oplus c_{16}=\mathop{\oplus}_{t=0}^{5}2^{25+t}$,故此时公式(2)等价于 $(2^{31}\,\&\,c_{16})\oplus\left(\mathop{\oplus}_{t=0}^{5}2^{25+t}\,\&\,a_{16}\right)\oplus\mathop{\oplus}_{t=0}^{5}2^{25+t}=0$,即$c_{16,32}=0$,且对满足 $0\le i\le 5$ 的i均有$a_{16,26+i}=1$.

最后,将情况 1 分别与情况I和情况II结合,可得定理 1 中的(1)和定理 1 中的(2),将情况 2 分别与情况III和情况IV结合,即可得定理 1 中的(3)和定理 1 中的(4),故$\Delta c_{16}=2^{31}+2^{25}$且$\Delta b_{16}=2^{31}+2^{25}$成立的充要条件是定理 1 中的(1)~定理 1 中的(4)中的至少 1 个成立. □

注 1:

(1) 定理 1 的 4 个结论即为$\Delta c_{16}=2^{31}+2^{25}$且$\Delta b_{16}=2^{31}+2^{25}$同时成立的充要条件集.由于 4 个结论都包括条件$b_{15,26}=0$ 和$a_{16,26}=1$,故将去掉$b_{15,26}=0$ 和$a_{16,26}=1$ 后的 4 个集合分别记为*Set*1,*Set*2,*Set*3 和*Set*4;

(2) 如果$q=p=0$,则定理 1 之(1)为文献[5,9−14]的碰撞攻击中所用的保证Δc_{16}和Δb_{16}满足的充分条件集.由q的不同取值可知,*Set*2 包括 6 个充分条件集;

(3) 由于*Set*3 包含 24 个充分条件,故*Set*3 中的条件全部满足的概率为 $1/2^{24}$;类似地,*Set*4 中的条件全部满足的概率也为 $1/2^{24}$.

3 改进的碰撞攻击算法及计算复杂度分析

已有的碰撞攻击算法[5,9−14]通过检验链接变量是否满足充分条件来判断产生碰撞的差分路径是否成立.但是,根据上面得到的充要条件集可知,这样将遗漏一些可以使差分路径成立的充分条件集,使得碰撞攻击算法继续运行的概率减小.故本文提出的改进碰撞攻击算法将直接检验第 23~64 步的输出差分是否成立,即利用保证各步输出差分成立的充要条件集进行检验.这样就增大了差分路径成立的概率,从而降低了算法的计算复杂度.

3.1 改进的碰撞攻击算法

寻找第 1 块消息的改进碰撞攻击算法如下:

步骤 1.随机选择 16 个 32-比特消息子分组,计算第 1~16 步的输出值;

步骤 2.利用单一的消息修改技术,使第 1 轮的链接变量满足文献[9]中的充分条件及多重消息修改过程中所增加的额外条件[13];

步骤 3.随机选择a_5,使其满足文献[9]中的充分条件及多重消息修改过程中所增加的额外条件[13];

步骤 4.利用文献[13]中提出的多重消息修改技术,使第 18~22 步的链接变量满足充分条件;

步骤 5.计算第 23~64 步的输出值,检验各步的输出差分是否满足差分路径,若有一个不成立,则返回至步骤 3;若都成立,则计算第 2 次迭代的初始值,若初始值的任何一个充分条件不得到满足,则返回至步骤 3;否则,输出 16 个 32-比特消息子分组.

同样地,可以给出寻找第 2 块碰撞消息的充要条件集,并提出改进的碰撞攻击算法,不再详细列出.

3.2 改进的碰撞攻击算法与现有算法的计算复杂度比较

由于Δc_{16}和Δb_{16}是每次迭代中最后两步的输出差分,且利用充要条件集进行判断可使保证Δc_{16}和Δb_{16}成立的概率增大,故第 3.1 节提出的碰撞攻击算法可以降低所有已有碰撞算法[5,9−14]的计算复杂度.记Pr_{block1}为第 1 次迭代中保证Δc_{16}和Δb_{16}成立的概率, Pr'_{block1} 为已有碰撞攻击算法的第 1 次迭代中保证Δc_{16}和Δb_{16}成立的概率, Pr_1为$Set1$ 中至少 1 个集合满足的概率,Pr_2为$Set2$ 中至少 1 个集合满足的概率.下面将对本文提出的碰撞攻击算法与已有算法的计算复杂度进行比较.

定理 2. $Pr_1\approx 0.347692$, $Pr'_{block1}=0.25$.

证明:设 $0\le q\le p\le 5$,记

$$C_1^{(p)}=\{(c_{16,26},c_{16,27},...,c_{16,31}):(c_{16,26},c_{16,27},...,c_{16,25+p},c_{16,26+p})=(0,0,...,0,1)\},$$

$$D_1^{(q)}=\{(d_{16,26},d_{16,27},...,d_{16,31}):(d_{16,26},d_{16,27},...,d_{16,25+q},d_{16,26+q})=(0,0,...,0,1)\},$$

$$A_1^{(qp)}=\{(a_{16,27},...,a_{16,31}):(a_{16,27},...,a_{16,26+q})=(1,...,1)\ \text{且}\ (a_{16,27+q},...,a_{16,26+p})=(0,...,0)\},$$

$$B_1^{(q)}=\{(b_{15,27},...,b_{15,31}):(b_{15,27},...,b_{15,26+q})=(0,...,0)\},$$

$$\Omega_1^{(qp)}=\{(a,b,c,d):a\in A_1^{(qp)},b\in B_1^{(q)},d\in D_1^{(q)},c\in C_1^{(p)}\},$$

则 $Set1=\bigcup_{0\le q\le p\le 5}\Omega_1^{(qp)}$,故 $Pr_1=2^{-22}\times\left|\bigcup_{0\le q\le p\le 5}\Omega_1^{(qp)}\right|=2^{-22}\times\sum_{0\le q\le p\le 5}2^{2(10-q-p)}\approx 0.347692$.

如果$q=p=0$,则$Set1$ 即为现有算法[5,9−14]给出的保证Δc_{16}和Δb_{16}满足的充分条件集,故 $Pr'_{block1}=0.25$. □

类似于定理 2 的证明,可证得定理 3.

定理 3. $Pr_2\approx 1.627604\times 10^{-4}$.

由于各充分条件集是独立的,故由定理 2、定理 3 和注 1 之(3)可知,$Pr_{block1}=Pr_1+Pr_2+2\times 2^{-24}\approx 0.34785$.

记$C_{block1}(M_0)$为第 1 次迭代中找到使得除条件$d_{16,26}=0$ 和$c_{16,26}=0$ 之外的所有充分条件满足的一对消息所需的平均计算复杂度.设每次搜索是相互独立的,且经t步搜索后Δc_{16}和Δb_{16}得到满足,则Δc_{16}和Δb_{16}得到满足的概率为$(1-Pr_{block1})^{t-1}\times Pr_{block1}$,其数学期望为 $1/Pr_{block1}$.故本文提出的改进碰撞攻击算法的计算复杂度为$C_{block1}(M_0)/0.34785$,而已有的碰撞攻击算法的计算复杂度为$C_{block1}(M_0)/0.25$.尽管文献[5,9−14]中各个碰撞攻击算法的$C_{block1}(M_0)$有可能不同,但由于这些算法都利用充分条件来检验输出差分是否成立,故本文提出的寻找第

1 块碰撞消息的改进碰撞攻击算法的平均计算复杂度为已有算法[5,9−14]计算复杂度的 0.25/0.34785≈0.7187 倍.

最后,本文以文献[13]中的碰撞攻击算法为例,在Pentium4 2.5GHZCPU的PC机上进行实验,在输入的随机消息相同的条件下,本文比较了文献[13]的利用充分条件集的碰撞攻击算法和本文提出的利用充要条件集的碰撞攻击算法的运行时间,从而对上述算法的改进效果进行了验证.表 1 给出了一个实例,其中利用充要条件集得到的碰撞消息的第 1 个 512 比特块(M_0)不满足文献[5,9−14]中关于第 63、64 步的充分条件集.

Table 1 An example of the comparison between the results of two kinds of collision attack algorithms

表 1 两种碰撞攻击算法运行结果的比较的例子

Results of the collision attack using sufficient conditions	The first block message (M_0)	The second block message (M_1)	Hash value
	31b0fc1d c3d1ab74 4d87d8ed 38faec60 6545ea7e 4d3669e9 005738ea 42ef6e35 05632dd5 00543c01 7d9ae8d7 93aa1562 424e9edf 8b2b1a0d dd433de6 04c8dbcd	3a1d861d bbb27be8 0595af57 4619560f 45ca187e 4db3bf3b 55eb2912 bf7159d3 81731f9d 4bc03dba af1bb88e c026f089 e1df98d7 839bfa73 e529d578 e9250451	2ab9fda7 fea13c30 dd6d4400 0c9dc129
Runtime	125 min	157min	
Results of our improved collision attack	The first block message (M_0)	The second block message (M_1)	Hash value
	339cec1d 43fd0292 039d4b3c 93012c82 6fee0a7d a1ccc205 005738ea 42ef6e35 05632dd5 ffd43c01 7d9ee8df 93aa156a 424e9edf 8b2b1a15 dcc53de6 04cccccd	3919861d 37f7fdc1 c57ca62f f2366617 4741f68d 4d345d4b f7aaa91a bf6de7cf 8177207d 4ba03d79 a15bb896 68a4f46b e1fd3ca9 8b241759 776a57c3 f19aad55	87a56b1b 26d8580e a53e3525 14060c53
Runtime	95min	65min	
Ratio of runtime	95/125=0.76	65/157=0.41	

4 结束语

本文通过分析MD5 中的非线性函数的性质以及模 2^{32}减差分和异或差分的性质,证明了Liang Jie和Lai Xuejia给出的产生MD5碰撞的充分条件集中的条件是保证第23~62步的输出差分满足的充要条件,给出了保证最后两步的输出差分成立的充要条件.利用得到的充要条件集,本文提出了改进的碰撞攻击算法,该算法的平均计算复杂度大约为已有的碰撞攻击算法的计算复杂度的 0.718 7 倍,并通过实验对改进效果进行了验证.

References:

[1] den Boer B, Bosselaers A. Collisions for the compression function of MD5. In: Proc. of the Advances in Cryptology, Eurocrypt 1993. Springer-Verlag, 1994.

[2] Dobbertin H. Cryptanalysis of MD5 compress. In: Proc. of the Presented at the Rump Session of Eurocrypt 1996. 1996.

[3] Wang XY, Feng DG, Lai XJ, Yu HB. Collisions for hash functions MD4, MD5, HAVAL-128 and RIPEMD. Report, 2004/199, Cryptology ePrint Archive, 2004.

[4] Hawkes P, Paddon M, Rose GG. Musings on the Wang *et al*. MD5 collision. Report, 2004/264, Cryptology ePrint Archive, 2004.

[5] Wang XY, Yu HB. How to break MD5 and other hash functions. In: Proc. of the Eurocrypt 2005. LNCS 3494, Berlin: Springer-Verlag, 2005. 19−35.

[6] Yajima J, Shimoyama T. Wang's sufficient conditions of MD5 are not sufficient. Report, 2005/263, Cryptology ePrint Archive, 2005.

[7] Nakano Y, Kuwakado H, Morii M. Redundancy of the Wang-Yu sufficient conditions. Report, 2006/406, Cryptology ePrint Archive, 2006.

[8] Sasaki Y, Naito Y, Yajima J, Shimoyama T, Kunihiro N, Ohta K. How to construct sufficient condition in searching collisions of MD5. Report, 2006/074, Cryptology ePrint Archive, 2006.

[9] Liang J, Lai XJ. Improved collision attack on hash function MD5. Report, 2005/425, Cryptology ePrint Archive, 2005.

[10] Wang ZY, Zhang HG, Qin ZP, Meng QS. A fast attack on the MD5 hash function. Journal of Shanghai Jiaotong University, 2006, E-11(2):140−145.

[11] Stevens M. Fast collision attack on MD5. Report, 2006/104, Cryptology ePrint Archive, 2006.

[12] Klima V. Finding MD5 collisions on a notebook PC using multi-message modifications. Report, 2005/102, Cryptology ePrint Archive, 2005.

[13] Sasaki Y, Naito Y, Kunihiro N, Ohta K. Improved collision attack on MD5. Report, 2005/400, Cryptology ePrint Archive, 2005.

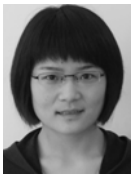
[14] Klima V. Tunnels in hash functions: MD5 collisions within a minute. Report, 2006/105, Cryptology ePrint Archive, 2006.

[15] Gregory H. Further musings on the Wang *et al*. MD5 collision: Improvements and corrections on the work of Hawkes, Paddon, and Rose. Report, 2007/375, Cryptology ePrint Archive, 2007.

[16] Cui GH, Zhou RH, Su L. Research on the analysis of the MD5 resistibility. Computer Engineering and Science, 2007,29(1):45−48 (in Chinese with English abstract).

附中文参考文献:

[16] 崔国华,周荣华,粟栗.关于 MD5 强度分析的研究.计算机工程与科学,2007,29(1):45−48.



陈士伟(1983－),女,河南南阳人,硕士生,主要研究领域为密码学.

金晨辉(1965－),男,博士,教授,博士生导师,主要研究领域为密码学,信息安全.