产生 MD5 碰撞的新的充分条件集

陈士伟 金晨辉
(信息工程大学电子技术学院 郑州 450004)

摘 要：Wang Xiaoyun 等(2005)给出了 MD5 能产生碰撞的一个充分条件集，并首次成功对 MD5 进行了碰撞攻击。Yuto Nakano 等(2006)指出上述充分条件集中有 16 个条件是冗余的，并给出了其中 14 个条件冗余的原因。Liang Jie 和 Lai Xuejia(2005)指出 Wang Xiaoyun 等给出的充分条件集并非总能产生碰撞，并增加新的条件使之总能产生碰撞，同时提出了一个新的碰撞攻击算法。本文证明了 Yuto Nakano 等给出的 16 个冗余条件中有两个并不冗余，且 Liang Jie 和 Lai Xuejia 增加的新条件中有两个是冗余的，指出 Liang Jie 和 Lai Xuejia 的碰撞攻击算法在消息修改时忽视了被修改条件之间的制约性，因而未必总能产生碰撞，本文对此进行了修正，给出新的充分条件集，并通过实验验证了该充分条件集总能产生碰撞。

关键词：保密通信；MD5；碰撞攻击；充分条件集；冗余性；制约性
中图分类号：TN918.1 **文献标识码**：A **文章编号**：1009-5896(2009)03-0740-05

A New Set of Sufficient Conditions for MD5 Collisions

Chen Shi-wei Jin Chen-hui
(*Institute of Electronic Technology*, *the University of Information Engineering*, *Zhengzhou* 450004, *China*)

Abstract: Wang *et al.* (2005) proposed a collision attack on MD5 and gave a set of sufficient conditions to yield a MD5 collision. Yuto Nakano *et al.* (2006) pointed out that there were 16 redundant conditions in Wang *et al.*'s set of sufficient conditions and explained why 14 out of them were redundant. This paper will propose that two of them are not redundant actually and present two new redundant conditions in the set of sufficient conditions presented by Liang Jie and Lai Xuejia in 2005. Additionally, it will show that there is a mistake in Liang Jie and Lai Xuejia's collision attack algorithm for the second-block message because they do not consider the dependence in the sufficient conditions, and correct the mistake. Finally, a new set of sufficient conditions is obtained and it could always yield a MD5 collision according to computer simulations.

Key words: Secret communication; MD5; Collision attack; A set of sufficient conditions; Redundancy; Dependence

1 引言

杂凑函数是一类重要的密码算法，被广泛使用于数字签名、数据完整性、电子货币等等。一个安全的杂凑函数应该满足单向性和无碰撞性。目前，MD5 和 SHA-1 是广泛使用的两大杂凑算法。

2004 年，Wang Xiaoyun 等[1]首次公布了 MD5 算法的碰撞对，但并没有公开产生碰撞的攻击算法。随后，Philip Hawkes 等[2]利用文献[1]给出的碰撞对对攻击算法进行了猜测。在 2005 年的欧洲密码年会上，Wang Xiaoyun 等[3]首次公布了对 MD5 的碰撞攻击算法。该碰撞攻击的主要思想是首先寻找能产生碰撞的差分路径，然后构造使差分路径满足的充分条件集，并利用消息修改技术使尽可能多的条件以概率 1 成立，从而降低碰撞攻击的计算复杂性。随后，Yuto Nakano 等[4]指出文献[3]给出的充分条件集中有 16 个条件是冗余的，并给出了其中 14 个条件冗余的原因。通过大量的试验，Jun Yajima 和 Takeshi Shimoyama [5]指出文献[3]给出的充分条件集并非总能产生碰撞，并对其做出了修正。Liang Jie 和 Lai Xuejia [6]指出 Jun Yajima 和 Takeshi Shimoyama 给出的充分条件集也不总能产生碰撞，并增加新的条件使之总能产生碰撞。与此同时，Wang Zhangyi 等[7]也对此问题进行了研究，并增加了相同的新条件。随后，Yu Sasaki 等[8]提出了构造充分条件集的一般方法，Klima[9]提出了新的多重消息修改技术，Black 等[10]研究了 MD5 的碰撞攻击算法并提出了改进的方法。

循环左移和模 2^{32} 减法运算是 MD5 碰撞攻击用到的两个基本运算。本文首先给出循环左移运算与模 2^{32} 减法可交换的充要条件，并据此证明了文献[6]新增加的条件中有两个是冗余的。此外，本文还证明了 Yu Nakano 等[4]给出的 16 个冗余条件中的两个并不冗余，并指出 Liang Jie 和 Lai Xuejia 的碰撞攻击算法在消息修改时忽视了条件之间的制约性，因而未必总能产生碰撞。本文对上述问题进行了修正，给出了

2007-9-26 收到，2008-06-27 改回
河南省杰出青年科学基金(0312001800)资助课题

新的充分条件集，并通过实验验证了该充分条件集总能产生碰撞。

2　MD5 算法简介

MD5 是将任意长度的消息变成 128-bit 杂凑值的一个杂凑函数，它以 512-bit 分组来处理输入消息，每一分组又划分为 16 个 32-bit 子分组。MD5 的压缩函数包括 4 轮，每一轮有 16 步。在每一次迭代中，链接变量按照以下运算顺序更新：

$$a = b + [(a + f(b,c,d) + m + \text{const}) <<< k]$$
$$d = a + [(d + f(a,b,c) + m + \text{const}) <<< k]$$
$$c = d + [(c + f(d,a,b) + m + \text{const}) <<< k]$$
$$b = c + [(b + f(c,d,a) + m + \text{const}) <<< k]$$

其中‘+’是模2^{32}加法， const 和 k 是给定的常数， m 是32-bit 的消息子分组。‘$x <<< k$’ 表示 x 循环左移 k 比特。f 是每轮所用的非线性函数，表示如下：

第1轮： $f = F(x,y,z) = (x \wedge y) \vee [(\neg x) \wedge y]$

第2轮： $f = G(x,y,z) = (x \wedge z) \vee [y \wedge (\neg z)]$

第3轮： $f = H(x,y,z) = x \oplus y \oplus z$

第4轮： $f = I(x,y,z) = y \oplus [x \vee (\neg z)]$

对于 $x = \sum_{i=1}^{n} x_i 2^{i-1} \in Z/(2^n)$，$x_i \in \{0,1\}$，我们称 x_i 为 x 的第 i 比特。本文使用的符号表示如下：

$M_0(M_0')$： 第1块512-bit消息分组； $M_1(M_1')$： 第2块512-bit消息分组;

m_i： 第 i 个32-bit消息子分组($0 \le i \le 15$); $H(H')$： 两块512-bit消息杂凑的结果;

a_i,b_i,c_i,d_i： 输入消息为 M_0 时， a,b,c,d 经第 i 次更新后的值($1 \le i \le 16$);

a_i',b_i',c_i',d_i'： 输入消息为 M_0' 时， a,b,c,d 经第 i 次更新后的值($1 \le i \le 16$);

$a_{i,[j_1,\cdots,j_k]}, d_{i,[j_1,\cdots,j_k]}, c_{i,[j_1,\cdots,j_k]}, b_{i,[j_1,\cdots,j_k]}$： a_i,b_i,c_i,d_i 的第 j_1 比特， …， 第 j_k 比特($1 \le i, j_1, \cdots j_k \le 32$);

ϕ_i： f 函数在第 i 步的输出结果($1 \le i \le 64$); $\phi_{i,j}$： ϕ_i 的第 j 比特($1 \le j \le 32$);

u_i： 循环左移前的结果($1 \le i \le 64$); v_i： 循环左移后的结果($1 \le i \le 64$);

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

$(u_i)_H, (u_i)_L$： $u_i = (u_i)_H 2^{n-k} + (u_i)_L$， $(u_i)_H$ 为 u_i 的高 k 比特， $(u_i)_L$ 为 u_i 的低 $(n-k)$ 比特;

λ_i： $(x+y) \bmod 2^n$ 中第 $(i-1)$ 比特向第 i 比特的进位，即 $\lambda_1 = 0$，对于 $i \ge 2$，有

$$\lambda_i = \begin{cases} 1, & x_{i-1} + y_{i-1} + \lambda_{i-1} \ge 2 \\ 0, & \text{其它} \end{cases}$$

3　碰撞攻击的充分条件集的冗余性分析

3.1 循环左移运算和模2^n减法可交换的充要条件

循环左移和模2^n减法运算是MD5碰撞攻击中用到的两个基本运算。在碰撞攻击中[3−6]，为使产生碰撞的差分路径一定满足，可以使循环左移运算和模2^n减法相互交换。下面将给出二者可交换的充要条件。

定理 1[7]　设 $x,\alpha \in Z/(2^n)$， $k \in Z$， 则有

$$[(x+\alpha) <<< k] - (x <<< k) =$$
$$\begin{cases} (\alpha <<< k), & \text{iff } x_L + \alpha_L < 2^{n-k} \text{和} x_H + \alpha_H < 2^k \\ (\alpha <<< k) + 1, & \text{iff } x_L + \alpha_L \ge 2^{n-k} \text{和} x_H + \alpha_H + 1 < 2^k \\ (\alpha <<< k) - 2^k, & \text{iff } x_L + \alpha_L < 2^{n-k} \text{和} x_H + \alpha_H \ge 2^k \\ (\alpha <<< k) - 2^k + 1, & \text{iff } x_L + \alpha_L \ge 2^{n-k} \text{和} x_H + \alpha_H + 1 \ge 2^k \end{cases}$$

推论 1　设 $\alpha,\beta \in Z/(2^n)$， $k \in Z$，则有

$$(\alpha <<< k) - (\beta <<< k)$$
$$= \begin{cases} (\alpha-\beta) <<< k, & \text{iff } \beta_L + (\alpha-\beta)_L < 2^{n-k} \\ & \quad \beta_H + (\alpha-\beta)_H < 2^k \\ [(\alpha-\beta) <<< k] + 1, & \text{iff } \beta_L + (\alpha-\beta)_L \ge 2^{n-k} \\ & \quad \beta_H + (\alpha-\beta)_H + 1 < 2^k \\ [(\alpha-\beta) <<< k] - 2^k, & \text{iff } \beta_L + (\alpha-\beta)_L < 2^{n-k} \\ & \quad \beta_H + (\alpha-\beta)_H \ge 2^k \\ [(\alpha-\beta) <<< k] - 2^k + 1, & \text{iff } \beta_L + (\alpha-\beta)_L \ge 2^{n-k} \\ & \quad \beta_H + (\alpha-\beta)_H + 1 \ge 2^k \end{cases}$$

由定理1和推论1可得循环左移运算和模2^n减法能够交换的充要条件。

定理 2　设 $x,\alpha,\beta \in Z/(2^n)$， $k \in Z$， 则 $[(x+\alpha-\beta) <<< k] - (x <<< k) = (\alpha <<< k) - (\beta <<< k)$ 成立的充要条件是

$$\begin{cases} x_L + (\alpha-\beta)_L < 2^{n-k} \text{且} x_H + (\alpha-\beta)_H < 2^k, \\ \quad \text{若 } \beta_L + (\alpha-\beta)_L < 2^{n-k} \text{ 且 } \beta_H + (\alpha-\beta)_H < 2^k \\ x_L + (\alpha-\beta)_L \ge 2^{n-k} \text{且} x_H + (\alpha-\beta)_H + 1 < 2^k, \\ \quad \text{若 } \beta_L + (\alpha-\beta)_L \ge 2^{n-k} \text{ 且 } \beta_H + (\alpha-\beta)_H + 1 < 2^k \\ x_L + (\alpha-\beta)_L < 2^{n-k} \text{且} x_H + (\alpha-\beta)_H \ge 2^k, \\ \quad \text{若 } \beta_L + (\alpha-\beta)_L < 2^{n-k} \text{ 且 } \beta_H + (\alpha-\beta)_H \ge 2^k \\ x_L + (\alpha-\beta)_L \ge 2^{n-k} \text{且} x_H + (\alpha-\beta)_H + 1 \ge 2^k, \\ \quad \text{若 } \beta_L + (\alpha-\beta)_L \ge 2^{n-k} \text{ 且 } \beta_H + (\alpha-\beta)_H + 1 \ge 2^k \end{cases}$$

3.2 文献[6]的充分条件集的冗余性分析

为保证给出的充分条件集一定能使产生碰撞的差分路径满足，文献[6]增加了若干条件，以便能使循环左移运算和模 2^{32} 减法运算可交换。我们发现其中的两个条件是冗余的。下面给出它们冗余的原因。

引理 1　设 $x,y,z \in Z/(2^n)$ 且 $(x+y) \bmod 2^n = z$，则当 $x_i \ne z_i$ 时，有 $\lambda_{i+1} = x_i$。

定理 3 设 $d_4, a_4, v_{14}, u_{14} \in Z/(2^{32})$， $a_{4,25}=1, d_{4,25}=0, a_{4,26}=0, d_{4,26}=0$， $\alpha=0$ 且 $\beta=2^{12}$，则有

$$[(u_{14}+\alpha-\beta)<<<12]-(u_{14}<<<12)=-(\beta<<<12)$$

证明 MD5 第 1 次迭代中第 14 步的运算为 $d_4=a_4+v_{14}$，其中 $v_{14}=(u_{14}<<<12)=(u_{14})_L 2^{12}+(u_{14})_H$。

因 $\beta=2^{12}, (-\beta)=2^{32}-2^{12}$，故 $\beta_L=2^{12}, (-\beta)_L=2^{12}+\cdots+2^{19}, \beta_H=0, (-\beta)_H=1+2+\cdots+2^{11}$，从而有 $\beta_L+(-\beta)_L=2^{12}+(2^{12}+\cdots+2^{19})=2^{20}$ 且 $\beta_H+(-\beta)_H+1=1+2+\cdots+2^{11}+1=2^{12}$。再由定理 2 知，$[(u_{14}+\alpha-\beta)<<<12]-(u_{14}<<<12)=-(\beta<<<12)$，当且仅当不等式 $(u_{14})_L+(-\beta)_L\geq 2^{20}$ 和 $(u_{14})_H+(-\beta)_H+1\geq 2^{12}$ 同时成立，即 $(u_{14})_L\geq 2^{12}$ 且 $(u_{14})_H\geq 0$ 成立。因为 $a_{4,25}=1$ 和 $d_{4,25}=0$，故由引理 1 知 a_4 与 v_{14} 相加的进位 $\lambda_{26}=1$。再由 $a_{4,26}=0, d_{4,26}=0$ 和 $d_{4,26}=a_{4,26}\oplus v_{14,26}\oplus\lambda_{26}$ 得到 $v_{14,26}=1$，故 $v_{14}\geq 2^{25}$。又因 $v_{14}=(u_{14})_L 2^{12}+(u_{14})_H$，所以 $(u_{14})_L\geq 2^{13}>2^{12}$。因此，$(u_{14})_H\geq 0$ 且 $(u_{14})_L\geq 2^{12}$。 证毕

类似于定理 3 的证明，利用定理 2 和引理 1 的结论可以证明下面的定理。

定理 4 设 $a_2, b_1, v_5, u_5 \in Z/(2^{32})$， $b_{1,31}=0, a_{2,31}=1, b_{1,32}=1$， $a_{2,32}=0, a_{2,[1,2,3,4,6]}=0$， $a_{2,[5,7]}=1$， $b_{1,[2,3,4,6,7]}=0$， $\beta=2^{18}+2$ 且 $\alpha=2^{30}+2^{26}+2^{25}+2^3$，则有

$$[(u_5+\alpha-\beta)<<<7]-(u_5<<<7)=(\alpha<<<7)-(\beta<<<7)$$

下面利用上述定理，证明文献[6]的充分条件集中 4 个条件是冗余的并给出冗余的原因，其中结论 3 和结论 4 是文献[4]的结论，但文献[4]没有给出这两个结论成立的原因。

结论 1 文献[6]中的条件 $d_{4,31}=0$ 是冗余的。

证明 文献[6]增加条件 $d_{4,31}=0$ 的目的是与充分条件集中的其它条件一起，使等式 $[(u_{14}-2^{12})<<<12]-(u_{14}<<<12)=-2^{24}$ 以概率 1 成立。但由定理 3 知，即使 $d_{4,31}=0$ 不成立，文献[6]中的条件 $a_{4,25}=1, d_{4,25}=0, a_{4,26}=0, d_{4,26}=0$ 也可保证等式以概率 1 成立，故 $d_{4,31}=0$ 是冗余的。

结论 2 条件 $b_{1,1}=c_{1,1}$ 可以替代文献[6]中的条件 $b_{1,1}=c_{1,1}=1$。

证明 文献[6]增加条件 $b_{1,1}=1$ 的目的是与定理 4 列出的条件一起，使得等式

$$[(u_5+2^{30}+2^{26}+2^{25}+2^3-2^{18}-2)<<<7]-(u_5<<<7)=[(2^{30}+2^{26}+2^{25}+2^3)<<<7]-[(2^{18}+2)<<<7]$$

以概率 1 成立，从而使第 2 次迭代中的差分路径满足。但由定理 4 知，即使 $b_{1,1}=1$ 不成立，文献[6]中的条件也可保证上述等式成立，故条件 $b_{1,1}=c_{1,1}$ 可以替代文献[6]中的条件 $b_{1,1}=c_{1,1}=1$。

为使 $\phi_{8,22}=\phi'_{8,22}$，文献[3]设置了 5 个条件 $c_{2,22}=1, a_{2,22}=0, a'_{2,22}=1, d_{2,22}=d'_{2,22}, c_{2,22}=c'_{2,22}$。为了减少条件，文献[4]在第 3 页证明了将 $a_{2,22}=0, a'_{2,22}=1$ 修改为 $a_{2,22}=a'_{2,22}$ 后仍可保证差分路径满足，同时指出可将上述 5 个条件修改为 3 个条件 $a_{2,22}=a'_{2,22}, d_{2,22}=d'_{2,22}, c_{2,22}=c'_{2,22}$，但没有说明上述 5 个条件能够被修改的原因。下面证明这种修改的正确性，从而说明条件 $c_{2,22}=1$ 是冗余的。

结论 3 条件 $c_{2,22}=1$ 是冗余的。

证明 因为 $d_{2,22}=d'_{2,22}, c_{2,22}=c'_{2,22}$，故若 $c_{2,22}=1$，则 $\Delta\phi_{8,22}=0$；若 $c_{2,22}=0$，则 $\Delta\phi_{8,22}=a'_{2,22}-a_{2,22}$。因此，条件 $d_{2,22}=d'_{2,22}, c_{2,22}=c'_{2,22}=1$ 可以保证 $\phi_{8,22}=\phi'_{8,22}$；条件 $a'_{2,22}=a_{2,22}$，$d_{2,22}=d'_{2,22}$，$c_{2,22}=c'_{2,22}=0$ 也可以保证 $\phi_{8,22}=\phi'_{8,22}$，由此可推出 $a'_{2,22}=a_{2,22}, d_{2,22}=d'_{2,22}, c_{2,22}=c'_{2,22}$ 可以保证 $\phi_{8,22}=\phi'_{8,22}$ 成立。这就证明了文献[4]所做出的上述修改的正确性，故条件 $c_{2,22}=1$ 是冗余的。

类似于上述的证明，我们可以证明下面的结论，该结论文献[4]已给出但未证明。

结论 4 条件 $c_{2,23}=1$ 是冗余的。

3.3 Yuto Nakano 等指出的两个冗余条件的不冗余性

Yuto Nakano 等在文献[4]中指出 Wang Xiaoyun 等给出的充分条件集[3]中的条件 $d_{16,26}=1$ 和 $c_{16,26}=1$ 是冗余的。文献[6]对 Wang Xiaoyun 等给出的充分条件集进行改进，给出了总能产生碰撞的充分条件集。表 1 给出了满足文献[6]中除 $d_{16,26}=1$ 和 $c_{16,26}=1$ 外的所有条件但不产生碰撞的反例。

下面从理论上证明这两个条件的不冗余性。

表 1 满足文献[6]除了 $d_{16,26}$=1 和 $c_{16,26}$=1 以外的所有条件但不产生碰撞的消息对

M_0	55441fb8	e0eca159	64566533	df6c1dde	a21406e9	f4e76e66	002638b8	434f6e33
	05332cd1	ffb43603	7d7aa8ab	9bea9564	024fb71e	842f4b13	3d34ddbd	b03f70d4
M_1	6f736703	9f91e7ce	fd3019d7	23544e60	261a374f	2dfe4cb7	56eb8936	70c158d2
	79b7278d	08c736bf	c03c831e	7db1c87d	352694ed	01a3c184	34215cf8	db53f600
H			f4fd959b	71b6bee6	213d4e21	671163fd		
M'_0	55441fb8	e0eca159	64566533	df6c1dde	221406e9	f4e76e66	002638b8	434f6e33
	05332cd1	ffb43603	7d7aa8ab	9beb1564	024fb71e	842f4b13	bd34ddbd	b03f70d4
M'_1	6f736703	9f91e7ce	fd3019d7	23544e60	a61a374f	2dfe4cb7	56eb8936	70c158d2
	79b7278d	08c736bf	c03c831e	7db1487d	352694ed	01a3c184	b4215cf8	db53f600
H'			f4fd959b	f1b83ce7	213d4c21	671163fd		

表 2 利用新充分条件集得到的 MD5 算法的 1024-比特碰撞对

M_0	52d308f9	0729df48	1e695aca	cf802cb1	3cd593f6	6f5a2c25	a399b835	be95adba
	06842dc5	a1d465f9	8010ec53	0fa9e341	88c31ef7	bedfee38	beacb577	c49ccc4d
M_0'	52d308f9	0729df48	1e695aca	cf802cb1	bcd593f6	6f5a2c25	a399b835	be95adba
	06842dc5	a1d465f9	8010ec53	0faa6341	88c31ef7	bedfee38	3eacb577	c49ccc4d
M_1	c8ecbb78	7dc5f3b2	54452ff7	c2593c0d	c62734c1	0cc41b47	5775351c	de00d4d1
	6c3ac81d	a863b862	aaf98a06	3c80f47f	513800fb	d09c1476	ea5aed9c	7090a773
M_1'	c8ecbb78	7dc5f3b2	54452ff7	c2593c0d	462734c1	0cc41b47	5775351c	de00d4d1
	6c3ac81d	a863b862	aaf98a06	3c80747f	513800fb	d09c1476	6a5aed9c	7090a773
H			6a0fcf70	bbca80ea	bd6ad3e2	635bca7b		

引 理 2　设 $X, X' \in Z/(2^{32})$ 且 $\Delta X = (X' - X) \cdot \bmod 2^{32}$，则 $\Delta X = 2^{31} - 2^{25}$ 的充要条件是存在 $q: 0 \le q \le 5$，使得 $\delta X = 2^{31} \oplus \oplus_{t=0}^{q} 2^{25+t}$。

定理 5　设 $b_{15}, a_{16}, d_{16}, c_{16} \in Z/(2^{32})$，$\Delta b_{15} = 2^{31}$，$\Delta a_{16} = 2^{31}, \Delta d_{16} = 2^{31} - 2^{25}, \Delta m_2 = 0, \Delta m_9 = 0$，则有

(1) 如果 $\delta d_{16} = 2^{31} \oplus \oplus_{t=0}^{i} 2^{25+t}$ 且 $0 \le i \le 5$，则 $\Delta c_{16} = 2^{31} - 2^{25}$ 的必要条件是对于满足 $0 \le t \le i$ 的 t，均有 $b_{15,(26+t)} = 0$。

(2) 如果 $\delta d_{16} = 2^{31} \oplus \oplus_{t=0}^{i} 2^{25+t}, \Delta c_{16} = 2^{31} - 2^{25}, \delta c_{16} = 2^{31} \oplus \oplus_{t=0}^{j} 2^{25+j}$ 且 $0 \le i, j \le 5$，则当 $i > j$ 时，有 $\Delta b_{16} \ne 2^{31} - 2^{25}$；当 $i \le j$ 时，$\Delta b_{16} = 2^{31} - 2^{25}$ 的必要条件是对于满足 $0 \le t \le i$ 和 $0 < s \le j - i$ 的 t, s，均有 $a_{16,(26+t)} = 1$ 和 $a_{16,(26+i+s)} = 0$。

结论 5　条件 $d_{16,26} = 1$ 和 $c_{16,26} = 1$ 在文献[3]给出的充分条件集中不是冗余的。

证明　设置条件 $d_{16,26} = 1$ 和 $c_{16,26} = 1$ 的目的是与条件 $b_{15,26} = 0$，$d_{16,32} = b_{15,32}$，$a_{16,26} = 1$，$c_{16,32} = a_{16,32}$，$\Delta a_{16} = 2^{31}, \Delta b_{15} = 2^{31}, \Delta d_{16} = 2^{31} - 2^{25}$ 一起，使 $\Delta c_{16} = 2^{31} - 2^{25}$ 和 $\Delta b_{16} = 2^{31} - 2^{25}$ 成立。

如果去掉条件 $d_{16,26} = 1$ 或 $c_{16,26} = 1$，则由定理 5 知必须增加关于 b_{15} 和 a_{16} 的新条件才可能保证 $\Delta c_{16} = 2^{31} - 2^{25}$ 和 $\Delta b_{16} = 2^{31} - 2^{25}$ 成立，故条件 $d_{16,26} = 1$ 和 $c_{16,26} = 1$ 不是冗余的。

4　文献[6]中充分条件之间的制约性分析

文献[6]给出的能产生 MD5 碰撞的充分条件集中有一些条件是相互制约的，在进行多重消息修改时，必须考虑这种制约性影响。

由于该充分条件集中的某些充分条件要求两个链接变量的一些对应比特位相等，所以当进行消息修改时不能仅对其中一个链接变量进行修改，否则将导致充分条件不再成立，从而使差分路径不再满足。例如，第二块消息应满足的充分条件集包括条件 $c_{1,4} = d_{1,4}$ 和 $c_{1,5} = d_{1,5}$。文献[6]提出的碰撞攻击算法在进行多重消息修改时，再次对 $d_{1,4}$ 和 $d_{1,5}$ 进行了修改，以便使第 2 轮的充分条件 $a_{5,32} = b_{4,32}$ 满足，但这可能导致第一轮已满足的条件 $c_{1,4} = d_{1,4}$ 和 $c_{1,5} = d_{1,5}$ 不再成立，从而使差分路径不一定成立。因此，在碰撞攻击算法进行多重消息修改时，不能再对 $d_{1,4}$ 和 $d_{1,5}$ 的值进行修改。

5　产生 MD5 碰撞的新充分条件集

文献[3]提出了能产生 MD5 碰撞的充分条件集，但该充分条件集实际上并不总能产生碰撞。文献[5]对该充分条件集进行了增改，但仍不能保证总产生碰撞。随后文献[6]增加了若干条件，从而得到了总能产生碰撞的充分条件集，该充分条件集中仍然保留了文献[4]指出的部分冗余条件。本文的研究和文献[4]中的结论说明文献[6]给出的充分条件集中含有冗余的条件。去掉这些冗余条件，同时将条件 $a_{2,21} = 0$ 改为 $a_{2,21} = 1$，将条件 $b_{1,1} = c_{1,1} = 1$ 放宽为 $b_{1,1} = c_{1,1}$ 之后，就得到新的充分条件集。这些冗余条件分别为

$$a_{2,22} = 0, a_{2,23} = 1, b_{1,22} = c_{1,22}, b_{1,23} = c_{1,23}, d_{2,22} = 1,$$
$$d_{2,23} = 1, c_{2,22} = 1, c_{2,23} = 1, d_{4,31} = 0, b_{16,26} = 1$$

本文在 Pentium4 2.5GHZ CPU 的 PC 机上进行了十余例实验，实验表明满足该充分条件集的所有条件且不满足上述任一冗余条件的消息对仍能产生碰撞，从而验证了新的充分条件集总能产生碰撞。表 2 给出了产生的一对碰撞消息，该消息满足新充分条件集的所有条件且不满足上述任一冗余条件。

6　结束语

本文通过理论分析和模拟实验对产生 MD5 碰撞的充分条件集进行了分析，证明了 Yuto Nakano 等给出的 16 个冗余条件中有两个并不冗余，且 Liang Jie 和 Lai Xuejia 增加的新条件中有两个是冗余的，指出 Liang Jie 和 Lai Xuejia 的碰撞攻击算法在多重消息修改时忽视了被修改条件之间的制约性，因而未必总能产生碰撞。本文对上述问题进行了修正，并给出了更加精简的充分条件集，最后通过实验验证了该充分条件集总能产生碰撞。

参 考 文 献

[1] Wang Xiaoyun, Feng Dengguo, and Lai Xuejia, *et al*.. Collisions for hash functions MD4, MD5, HAVAL-128 and RIPEMD[EB/OL]. Cryptology ePrint Archive, Report 2004/199, 2004.

[2] Hawkes, Paddon, and Rose G. Musings on the Wang *et al.* MD5 collision[EB/OL]. Cryptology ePrint Archive, Report 2004/264, 2004.

[3] Wang Xiaoyun and Yu Hongbo. How to break MD5 and other hash functions [C]. Eurocrypt' 05, Berlin, 2005, LNCS 3494: 19–35.

[4] Yuto Nakano, Hidenori Kuwakado, and Masakatu Morii. Redundancy of the Wang-Yu sufficient conditions [EB/OL]. Cryptology ePrint Archive, Report 2006/406, 2006.

[5] Jun Yajima and Takeshi Shimoyama. Wang's sufficient conditions of MD5 are not sufficient [EB/OL]. Cryptology ePrint Archive, Report 2005/263, 2005.

[6] Liang Jie and Lai Xuejia. Improved collision attack on hash function MD5 [EB/OL]. Cryptology ePrint Archive, Report 2005/425, 2005.

[7] Wang Zhangyi, Zhang Huanguo, and Qin Zhongping, *et al.*. A fast attack on the MD5 hash function [J]. *Journal of Shanghai Jiaotong University*, 2006, (2): 140–145.

[8] Yu Sasaki, Yusuke Naito, and Jun Yajima, *et al.*. How to construct sufficient condition in searching collisions of MD5 [EB/OL]. Cryptology ePrint Archive, Report 2006/074, 2006.

[9] Klima. Finding MD5 collisions on a notebook PC using multimessage modifications [C]. International Scientific Conference Security and Protection of Information, Brno, Czech Republic, May 2005: 53–62.

[10] Black J, Cochran M, and Highland T. A study of the MD5 attacks: Insights and improvements [C]. Fast Software Encryption 2006, Berlin, 2006, LNCS 4047: 262–277.

陈士伟： 女，1983 年生，硕士生，研究方向为密码学.

金晨辉： 男，1965 年生，教授，博士生导师，主要研究方向为密码学与信息安全.