对强化 MD 结构杂凑函数的一个新的“牧群”攻击

陈士伟 金晨辉
(信息工程大学电子技术学院 郑州 450004)

摘 要：该文构造了具有 2^k 个起始点的变长“钻石树”结构的多碰撞，并据此提出了对强化 MD 结构杂凑函数的一个新的选择目标强制前缀且原像长度为 $2k$+3 块的原像攻击(即“牧群”攻击)。由于增大了攻击过程中可利用的中间链接值的数量，故当 $k \ge n/4 - 1.05$ 时，新的牧群攻击可将该攻击的计算复杂性由现有结果 $O(2^{n-2(k+1)} + 2^{n/2+k+5/2})$ 降至 $O(2^{n-k}/3 + 2^{n/2+k+2})$。

关键词：密码学；杂凑函数；强化 MD 结构；原像攻击；牧群攻击；多碰撞
中图分类号：TN918.1 **文献标识码**： A **文章编号**：1009-5896(2010)08-1953-03
DOI: 10.3724/SP.J.1146.2009.01313

A New Herding Attack on Hash Functions with Strengthening Merkle-Damagard (MD) Construction

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

Abstract: This paper constructs a “diamond structure” multicollision with 2^k initial values and variant lengths, which is used to propose a new chosen target forced prefix preimage attack (herding attack) on hash functions with Strengthening Merkle-Damagard (SMD) construction to find a preimage with $2k$+3 blocks. Since the number of the chaining values available in herding attack is increased, the computational complexity of herding attack is reduced to $O(2^{n-k}/3 + 2^{n/2+k+2})$ from $O(2^{n-2(k+1)} + 2^{n/2+k+5/2})$ for $k \ge n/4 - 1.05$.

Key words: Cryptography; Hash functions; SMD construction; Preimage attack; Herding attack; Multicollision

1 引言

杂凑算法被广泛应用于数字签名、数据完整性等领域，许多杂凑算法如国际标准 MD5，SHA-0，SHA-1 等都具有相同的迭代结构，即 MD 结构。碰撞攻击是对杂凑函数的基本攻击方法。2004 年的欧洲密码年会上，王小云等人[1]提出了 MD5 等一系列 MD 结构杂凑算法的碰撞对。2005 年，王小云等[2]首次公布了针对 MD5 等杂凑函数的碰撞攻击算法。随后，许多密码研究者[3−6]针对王小云等的碰撞攻击算法又进行了详细系统的分析。与此同时，压缩函数随机条件下的 MD 结构杂凑函数的分析也成为杂凑函数的一个研究热点。2004 年，Joux[7]提出了寻找 MD 结构杂凑函数的 k-碰撞的方法，该方法可找到能产生相同杂凑值的 k 个长度相同的消息，其计算复杂性为 $2^{n/2}k$ (n 是杂凑值的比特数，下同)。2005 年，Kelsey 和 Schneier [8]提出了寻找(a,b)-可扩展消息的方法，它本质上类似于(b−a+1)-碰撞，只是找出消息的长度可以遍历 a 与 b 之间的所有值。2006 年，Kelsey 和 Kohno[9]指出杂凑函数应能抵抗选择目标强制前缀的原像攻击，即“牧群”攻击，它是指对于杂凑函数 H 和选择的杂凑值 h，对任给的消息前缀 M_{pre}，都能找到一个以 M_{pre} 为前缀且使 $H(M) = h$ 的消息 M。Kelsey 和 Kohno[9]指出 MD 结构杂凑函数不能抵抗“牧群”攻击，并给出了具体的攻击方法。其攻击思路是先构造一个具有 2^k 个起始点的深度为 k 的“钻石树”结构的 2^k-碰撞，且它们产生的杂凑值都为 H_0，并将 H_0 的值公开；然后对任给定的 M_{pre}，选取有一定意义的消息块 B，使得以 M_{pre} 为前缀的消息块 $M_{\text{pre}}||B$ 的杂凑值与已构造出的“钻石树”结构的 2^k-碰撞的 2^k 个起始值中的一个相等，最后再将该起始点对应的消息添加至 $M_{\text{pre}}||B$ 的后面，从而产生杂凑值 H_0 的以 M_{pre} 为前缀的原像。

针对 MD 结构杂凑函数的“牧群”攻击是一种选择目标强制前缀的原像攻击，但并不限定原像的长度。然而，在实际的应用中，有时还需限定原像的长度。本文将考察这类将原像长度限制为指定长度的“牧群”攻击问题，并考虑此时如何降低攻击

2009-10-09 收到，2010-02-09 改回
河南省杰出青年科学基金(0312001800)资助课题
通信作者：陈士伟 chenshiwei1012@sohu.com

算法的复杂性问题。本文将构造出具有 2^k 个起始点且长度可变的“钻石树”结构的多碰撞，并基于此结构提出对强化 MD 结构杂凑函数的一个新的“牧群”攻击，即找到给定杂凑值的长度为 $2k+3$ 块的一个原像。当 $k \ge n/4-1.05$ 时，对强化 MD 结构的新的“牧群”攻击的计算复杂性为 $O(2^{n-k}/3+2^{n/2+k+2})$，低于文献[9]的攻击算法在原像长度为 $2k$+3 块时的计算复杂性 $O(2^{n-2(k+1)}+2^{n/2+k+5/2})$。

2 MD 结构和强化 MD 结构杂凑函数的描述

设 $f:\{0,1\}^{n+m}\to\{0,1\}^n$ 为压缩函数，输入消息为 M。如果消息 M 的长度$|M|$不是 m 的整数倍，则在其后添加 1 个 1 及若干个 0，从而使填充后的消息 M^{pad} 的长度为 m 的整数倍。设 M^{pad} 可被划分为 l 个 m-bit 的消息块 $m_1,m_2,\cdots,m_l$，则 MD 结构杂凑函数的描述如下：

$$\text{MD}^f(IV,M^{\text{pad}})=f(f(\cdots f(f(IV,m_1),m_2),\cdots,m_{l-1}),m_l)$$

然而，MD 结构杂凑函数不能抵抗碰撞攻击，故需在消息 M 的后面添加 1 个 1 及若干个 0 使其长度为 m 的整数倍少于 L-bit，然后加上原消息长度的 L 位的二进制表示，即 MD 强化。例如对于 MD5 算法，L 为 64。设利用该方法填充过的消息 M^{pad_L} 可被划分为 l 个 m-bit 的消息块 $m_1,m_2,\cdots,m_l$，则强化 MD 结构杂凑函数的描述如下：

$$\begin{aligned}&\text{SMD}^f(IV,M^{\text{pad}\_L})\\&=f(f(\cdots f(f(IV,m_1),m_2),\cdots,m_{l-1}),m_l)\end{aligned}$$

下面将提出对强化 MD 结构杂凑函数的新的“牧群”攻击，即对于给定的杂凑算法 H，杂凑值 h 以及 消息前缀 M_{pre}，找到以 M_{pre} 为前缀的长度为 $2k+3$ 块的消息 M，使得消息 M 的杂凑值为 h。

3 强化 MD 结构杂凑函数的“牧群”攻击

Kelsey 和 Kohno[9]构造了具有 2^k 个起始点且深度为 k 的“钻石树”结构的多碰撞，并利用该多碰撞提出了对 MD 结构及强化 MD 结构杂凑函数的“牧群”攻击。由于强化 MD 结构杂凑函数的输入消息的尾部添加了原消息长度的二进制表示，因此“牧群”攻击找到的选择目标强制前缀的原像的长度应该与指定的长度相同。在利用文献[9]构造出的等长“钻石树”结构的多碰撞进行“牧群”攻击时，若要利用除起始点之外的中间链接值，则作者提出可在“钻石树”结构的尾部添加一个$(1,k)$-可扩展的消息，以保证构造出的原像具有指定的长度 k+2 块，且其算法的计算复杂性为 $O(^{n-k-1}+2^{n/2+k/2+2})$。下面将提出一种构造具有 2^k 个起始点且长度可变的“钻石树”结构的多碰撞的方法，使得在利用除起始点之外的中间链接值进行“牧群”攻击时，能保证找到的原像具有指定的长度 $2k$+3 块。

3.1 MD 结构的变长“钻石树”结构的多碰撞构造

以下约定$|B|$是消息 B 包含的 m-bit 块的数量。构造具有 2^k 个起始点的“钻石树”结构的多碰撞的具体算法 1 描述如下：

步骤 1　构造出 2^k 个不同的起始点：随机选择一个初始值 IV 和 2^k 个消息块，并分别计算出其杂凑值

$$H_i^{(0)}=f(IV,M_i),\ i=0,1,\cdots,2^k-1 \tag{1}$$

步骤 2　对 j 从 0 到 $k-1$，执行：

(a)对任意 $H_i^{(j)}(i=0,1,\cdots,2^{k-j}-1)$，利用碰撞攻击方法，找出两个不同的消息 $B_{1,i}^{(j)}$ 和 $B_{2,i}^{(j)}$，使得

$$\text{MD}^f(H_i^{(j)},B_{1,i}^{(j)})=\text{MD}^f(H_i^{(j)},B_{2,i}^{(j)})=HH_i^{(j)} \tag{2}$$

其中$|B_{1,i}^{(0)}|=1$，$|B_{2,i}^{(0)}|=2$，且对于 $t=1,\cdots,k-1$，有$|B_{1,i}^{(t)}|=1$和$|B_{2,i}^{(t)}|=2t+1$；

(b)对每个 $HH_i^{(j)}(i=0,1,\cdots,2^{k-j}-1)$，随机选择 $2^{n/2-(k-j)/2+1/2}$ 个单消息块，记其全体构成集合 Ω_i。利用碰撞攻击方法，找出集合 $\{0,1,\cdots,2^{k-j}-1\}$ 的不交二元子集 $\Lambda_i=\{k_i,l_i\}(i=0,1,\cdots,2^{k-j-1}-1)$，且对 $0\le i<2^{k-j-1}$，均找出 $M_{k_i}^{(j)}\in\Omega_{k_i}$，$M_{l_i}^{(j)}\in\Omega_{l_i}$，使得 $H_i^{(j+1)}=f(HH_{k_i}^{(j)},M_{k_i}^{(j)})=f(HH_{l_i}^{(j)},M_{l_i}^{(j)})$。不妨假设 $k_i=2i,l_i=2i+1$。

步骤 3　对 $0\le j\le k-1,0\le i<2^{k-j}$，选择 $B_{1,i}^{(j)}$ 或 $B_{2,i}^{(j)}$，从而产生一个具有 2^k 个起始点 $H_0^{(0)},\cdots,H_{2^k-1}^{(0)}$ 和 2^{k+1}–2 个可用中间链接值 $HH_i^{(j)}:0\le i<2^{k-j},0\le j\le k-1$的多碰撞“钻石树”结构。

由碰撞攻击理论易证，算法 1 的步骤 1 的计算量为 2^k 次 f 的计算；步骤(a)和步骤(b)的计算量分别近似为 $2^{n/2+k-j+1}$ 和 $2^{n/2+(k-j)/2+1/2}$ 次 f 的计算，故步骤 2 的计算量近似为 $\sum_{j=0}^{k-1}[2^{n/2+k-j+1}+2^{n/2+(k-j)/2+1/2}]$ 次 f 的计算，因而算法 1 的计算复杂性为 $O(2^{n/2+k+2}+2^{n/2+k/2+3/2})$ 次 f 的计算。

注　在步骤 2(a)中，利用碰撞攻击的方法可以找到两个不同的消息 $B_{1,i}^{(j)}$ 和 $B_{2,i}^{(j)}$ 产生相同的杂凑值，本文按生日攻击所需的计算量分析了该步的计算量。事实上，碰撞攻击的方法有许多种，例如王小云等[2]提出的碰撞攻击方法所需的计算量远远小于生日攻击的计算量。因而在实际的攻击中，构造出“钻石树”结构的多碰撞的计算量可以小于 $O(2^{n/2+k+2}+2^{n/2+k/2+3/2})$。

3.2 强化 MD 结构杂凑函数的新的“牧群”攻击

利用前面构造出的长度可变的“钻石树”结构的多碰撞，可以构造出对强化 MD 结构杂凑函数的一个新的“牧群”攻击，即对于给定的消息前缀 M_{pre}，杂凑值 $h=f(H_0^{(k)},w)$ (这里的 w 表示 1 个 1，m–L–1

个 0 和 $(2k+3)m$ 的 L-bit 的二进制联接而成的 m-bit 单消息块)及消息长度 $2k+3$ 块，可选择具有一定意义的消息 B，使得以 M_{pre} 为前缀的消息 $M_{\text{pre}}||B$ 的杂凑值 $\text{MD}^f(M_{\text{pre}}||B)$ 与“钻石树”结构的多碰撞中的某起始点 $H_i^{(0)}$ 或某中间链接值 $HH_i^{(j)}$ 相同，从而将 $M_{\text{pre}}||B$ 与“钻石树”中 $H_i^{(0)}$ 或 $HH_i^{(j)}$ 的一条后续路径相接。由于本文构造出的“钻石树”结构是变长的，故可依据相接点的位置选择适合长度的消息添加到 $M_{\text{pre}}||B$ 的尾部，以保证添加后的消息具有指定的长度，最后再在最后添加消息长度的二进制表示，从而产生以 M_{pre} 为前缀且杂凑值为 h 的一个消息。

具体的攻击算法 2 描述如下：

步骤 1　根据算法 1 构造具有 2^k 个起始点的变长“钻石树”结构的多碰撞；

步骤 2　选择有一定意义的单消息块 B 并计算出 $S=\text{MD}^f(M_{\text{pre}}\,||\,B)$。

若 $S\notin\{H_i^{(0)},HH_i^{(j)}:0\le i<2^{k-j},0\le j\le k-1\}$，则返回执行步骤 2。否则，定义 $i_0=i$，且对 $t=1,2,\cdots,k-1$，定义 $i_t=\lfloor i_{t-1}/2\rfloor$，并执行：

(1)若 $S=HH_i^{(j)}$，则在 $M_{\text{pre}}\,||\,S$ 的尾部依次添加 $M_{i_0}^{(j)}\,||\,B_{2,i_1}^{(j+1)}\,||\,M_{i_1}^{(j+1)}$ 及 $B_{1,i_2}^{(j+2)}\,||\,M_{i_2}^{(j+2)},\cdots,\ B_{1,i_{k-j-1}}^{(k-1)}\,||\,M_{i_{k-j-1}}^{(k-1)}$ 共 $[2(j+1)+1]+2[(k-1)-(j+1)+1]=2k+1$ 个消息块后，得到块数为$(2k+3)$的消息 M。

(2)若 $S=H_i^{(0)}$，则在 $M_{\text{pre}}\,||\,S$ 的尾部依次添加 $B_{2,i_0}^{(0)}\,||\,M_{i_0}^{(0)},B_{1,i_1}^{(1)}\,||\,M_{i_1}^{(1)},\cdots,B_{1,i_{k-1}}^{(k-1)}\,||\,M_{i_{k-1}}^{(k-1)}$ 共 $2k+1$ 个消息块后，得到块数为$(2k+3)$的消息 M。

步骤 3　计算知 $H_0^{(k)}=\text{MD}^f(M)$，故在消息尾部添加 1 个 1，$m-L-1$ 个 0 以及 $(2k+3)m$ 的 L-bit 的二进制表示，则可产生给定的杂凑值 h。

由 3.1 节知算法 2 的步骤 1 的计算复杂性为 $O(2^{n/2+k+2}+2^{n/2+k/2+3/2})=O(2^{n/2+k+2})$。由于步骤 2 利用的起点 $H_i^{(0)}$ 和中间链接值 $HH_i^{(j)}$ 的总个数为 $2^{k+1}+2^k-2\approx 3\times 2^k$，故由碰撞攻击理论知，步骤 2 平均需要选择 $2^{n-k}/3$ 个单消息块。因此，针对长度为 $2k+3$ 块的原像，新的“牧群”攻击的计算量约为 $O(2^{n-k}/3+2^{n/2+k+2})$。而利用文献[9]的算法找到长度为 $2k+3$ 块的原像的计算复杂性约为 $O(2^{n-2(k+1)}+2^{n/2+k+5/2})$。又因当 $k\ge n/4-1.05$ 时，$2^{n-k}/3+2^{n/2+k+2}\le 2^{n-2(k+1)}+2^{n/2+k+5/2}$，故此时本文提出的新的“牧群”攻击低于文献[9]提出的“牧群”攻击的计算复杂性。

4　结束语

本文构造了具有 2^k 个起始点且长度可变的“钻石树”结构的多碰撞，并据此提出了对强化 MD 结构杂凑函数的新的“牧群”攻击，该攻击方法找到了给定杂凑值的长度为 $2k+3(k\ge n/4-1.05)$ 块且前缀固定的原像，并将现有“牧群”攻击的计算复杂性由 $O(2^{n-2(k+1)}+2^{n/2+k+5/2})$ 降至 $O(2^{n-k}/3+2^{n/2+k+2})$。基于变长“钻石树”结构设计“牧群”攻击的思想能否用于对杂凑函数的其它结构的分析，是值得进一步研究的问题。

参 考 文 献

[1] Wang X Y, Feng D G, and Lai X J, *et al*.. Collisions for hash functions MD4, MD5, HAVAL-128 and RIPEMD[EB/OL]. Cryptology ePrint Archive, Report 2004/199, 2004.

[2] Wang X Y and Yu H B. How to break MD5 and other hash functions [C]. Eurocrypt' 05, Berlin, 2005, LNCS 3494: 19–35.

[3] Yu S, Yusuke N, and Jun Y, *et al*.. How to construct sufficient condition in searching collisions of MD5 [EB/OL]. Cryptology ePrint Archive, Report 2006/074, 2006.

[4] Xie T, Feng D G, and Liu F B. A new collision differential for MD5 with its full differential path[EB/OL]. Cryptology ePrint Archive, Report 2008/230, 2008.

[5] Chen S W and Jin C H. An improved collision attack on MD5 algorithm[C]. Third SKLOIS Conference, Inscrypt 2007, Xining, China, August 31- September 5, 2007, Lecture Notes in Computer Science, 2007, Vol. 4990: 343–357.

[6] 陈士伟, 金晨辉. MD5 碰撞攻击的多重消息修改技术研究. 通信学报, 2009, 30(8): 89–95.
Chen S W and Jin C H. Research on the multi-message modification techniques on MD5[J]. *Journal on Communications*, 2009, 30(8): 89–95.

[7] Joux A. Multicollisions in Iterated hash functions[C]. CRYPTO 2004, Berlin: Springer- Verlag, 2004. LNCS: 3152, 306–316.

[8] Kelsey J and Schneier B. Second preimages on n-bit hash functions for much less than 2^n work[C]. Eurocrypt 2005, Berlin: Springer-Verlag, 2005, LNCS 3494: 19–35.

[9] Kelsey J and Kohno T. Herding hash functions and the Nostradamus attacl[C]. Eurocrypt 2006, Berlin: Springer-Verlag, 2006, LNCS 4004: 183–200.

陈士伟：　女，1983 年生，博士生，研究方向为密码学.

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