计算机研究与发展
Journal of Computer Research and Development
DOI:10.7544/issn1000-1239.2014.20130882
51(11):2513-2517, 2014

对 HTBC 杂凑函数的碰撞和第二原像攻击

马冰珂　　李　宝
(中国科学院信息工程研究所　北京　100195)
(bkma@is.ac.cn)

Collision and Second Preimage Attacks on the HTBC Hash Function

Ma Bingke and Li Bao
(*Institute of Information Engineering, Chinese Academy of Sciences, Beijing* 100195)

Abstract　A common way to build hash functions is combining a secure block cipher with an appropriate chaining structure. When the block cipher used has strong security properties such as AES, the security of this kind of hash functions mainly depends on the security of the certain chaining structure. HTBC is a hash function design which uses a special triple-block-chaining structure to be combined with a secure block cipher. The triple-block-chaining structure used by the HTBC hash function is not secure. Based on the serious flaws of the chaining structure, using particular properties of related operations, the collisions of the HTBC hash function can be directly constructed, and the time complexity to find a single collision is only 1. For the messages with certain length, a second preimage can be constructed as well. When AES-256 is used as the block cipher, the maximum time complexity to launch a second preimage attack is 2^{112}, which is lower than the brute force attack bound 2^{128}. For the particular weak messages, the overall time needed to find a second preimage is only 1. If the message is randomly chosen, the average time complexity to launch a successful second preimage attack is $2^{46.56}$.

Key words　hash function; triple-block-chaining; HTBC; collision; second preimage

摘　要　使用安全的分组密码作为基本组件,并搭配合适的链接结构是一种常用的杂凑函数设计方法.当使用安全性较强的分组密码,如高级加密标准(Advanced Encryption Standard, AES)等作为基本组件时,这类杂凑函数的安全性很大程度上取决于链接结构的安全性.基于三重分组链接的杂凑函数(triple-block-chaining-based hash function, HTBC),利用安全分组密码作为基本元件,采用一种特殊的三重链接结构,证明了 HTBC 杂凑算法的这种链接结构是不安全的.基于该链接结构的弱点,利用相关运算的特殊性质,直接构造出 HTBC 算法的碰撞,构造碰撞的时间复杂度为 1.对于长度满足特定要求的消息,可以构造该消息的第二原像.当使用 AES-256 作为底层的分组密码时,攻击的最大时间复杂度为 2^{112},低于穷举攻击所需要的时间复杂度 2^{128};对于某些满足特定性质的弱消息,攻击的时间复杂度仅为 1;如果消息的取值足够随机,攻击的平均时间复杂度为 $2^{46.56}$.

关键词　杂凑函数;三重链接;HTBC;碰撞;第二原像

中图法分类号　TP309

密码学杂凑(Hash)函数是现代密码学的重要组成部分,在消息认证、数字签名以及数据完整性校验等领域有着极为广泛的应用.杂凑函数可以用 $h=H(P)$ 表示,它将任意长度的明文信息 P 映射成

收稿日期:2013-06-26;**修回日期**:2013-12-04
基金项目:国家"九七三"重点基础研究发展计划基金项目(2013CB338002);国家"八六三"高技术研究发展计划基金项目(2013AA014002);中国科学院战略性先导专项(XDA06010702);中国科学院信息工程研究所密码研究专项基金(Y3Z0027103)

固定长度的杂凑值h.1 个安全的杂凑函数主要具有以下 3 点安全特性:

1) 抗原像攻击性. 给定输出值 h,计算消息 P 是困难的.

2) 抗第二原像攻击性. 给定消息 P,找到 1 个不同的消息 Q,使得 $H(P)=H(Q)$是困难的.

3) 抗碰撞性. 寻找任意 2 个不同的消息 P 和 Q,使得 $H(P)=H(Q)$是困难的.

杂凑函数的常用设计方法主要有两种:一种是定制杂凑函数的设计,这类杂凑函数的代表有 SHA-3[1],SHA-2,SHA-1[2],Whirlpool[3]以及 SHA-3 竞赛中涌现出的大量的杂凑函数算法;另一种经典的构造杂凑函数的方法是使用安全的分组密码作为基本组件,搭配特定的链接结构. Preneel 等人[4]对使用分组密码构造单块杂凑函数的所有可能链接结构进行了分析,证明其中 12 种结构的安全性. 由于单块杂凑函数的安全界较低,密码学者又相继提出了使用分组密码构造多块杂凑函数的链接结构,这类多块链接结构的代表有 MDC-2[5], Abreast-DM[6], Tandem-DM[6], Hirose's scheme[7]以及 MJH[8]等. 在底层分组密码安全性较强的情况下,对于这类杂凑函数的安全性分析主要侧重于链接结构的分析.

基于三重分组链接的杂凑函数(triple-block-chaining-based hash function, HTBC)[9]算法是由我国学者提出的一种基于三重分组链接的杂凑函数,它采用密钥长度是分组长度两倍的分组密码算法,如 AES-256[10]等. 本文证明了这种结构是不安全的,并且利用这种结构的安全性弱点,以时间复杂度 1 构造出 HTBC 算法的碰撞;对于满足特定消息长度的消息,可以构造其第二原像,以 AES-256 为例,第二原像攻击的最高时间复杂度为 2^{112},低于穷举攻击所需的时间复杂度 2^{128};对于某些特殊取值的弱消息,寻找其第二原像的时间复杂度仅为 1;如果消息的取值足够随机,则攻击的平均时间复杂度为 $2^{46.56}$.

1 HTBC 杂凑函数

符号约定:

① $C=E_K(P)$表示用密钥 K 将明文 P 加密成密文 C.

② T 表示每个消息大分组的长度.

③ L 表示消息总长度(单位 b),是 1 个长度为 128 b 的数.

④ $\|$ 表示串联操作.

⑤ $\oplus$表示异或操作.

⑥ $+$表示按字节模 256 加操作.

⑦ τ 表示最后 1 个消息分组的字节长度.

⑧ P_i 表示消息 P 的第 i 个分组.

⑨ $P[j]$表示消息 P 的第 j 个字节.

⑩ h 表示最终的杂凑值输出.

本文中使用的分组密码为 AES-256,其密钥长度为 256 b,恰为分组长度 128 b 的 2 倍. 由于本文是对于链接结构的安全性分析,因此对于其他满足密钥长度是分组长度两倍的分组密码,本文的攻击方法仍然有效. 以 AES-256 为例,HTBC 算法的消息大分组长度 T 为 384 b,杂凑值输出长度为 128 b,则 HTBC 算法 $h=HTBC(P)$的算法过程执行如下:

1) 补位. 若消息长度 L 正好是分组长度 T 的整数倍时,不需要补位;否则:

$$\text{For } i=1 \text{ to } T-\tau,$$
$$P[L+i]=P[i].$$

2) 分组. 将消息 P 按照 384 b 分成 n 个大组,每个大组按照 128 b 分成 3 个小组:P_{3i-2},P_{3i-1},P_{3i}.

3) 产生密钥 K 和初始值 Y_0.

$K=\sum_{i=1}^{1.5n}(P_{2i-1}\|P_{2i})$,加法为按字节模 256 加;

$Y=E_K(L)$,产生密钥时消息直接补 0.

4) 迭代计算:

$$Y_1=E_{(P_1\oplus Y_0\|P_2+Y_0)}(P_3\oplus Y_0).$$

if $n>1$,

$$Y_2=E_{(P_4+Y_0\|P_5\oplus Y_0)}(P_6\oplus Y_1);$$

if $n>2$, for $i=3$ to n

$$Y_i=E_{(P_{3i-2}\oplus Y_{i-3}\|P_{3i-1}\oplus Y_{i-2})}(P_{3i}\oplus Y_{i-1}).$$

5) 最终杂凑值输出:

$$h=E_K(Y_n)\oplus Y_{n-1}.$$

2 HTBC 杂凑函数的内部碰撞

图 1 为 HTBC 算法的 1 个 3 轮内部碰撞,灰色表示存在差分的位置,上述 3 轮内部碰撞的路径可用等式表示如下.

1) 消息 P

$$Y_{i-2}=E_{(P_{3i-8}\oplus Y_{i-5}\|P_{3i-7}\oplus Y_{i-4})}(P_{3i-6}\oplus Y_{i-3}); \quad (1)$$
$$Y_{i-1}=E_{(P_{3i-5}\oplus Y_{i-4}\|P_{3i-4}\oplus Y_{i-3})}(P_{3i-3}\oplus Y_{i-2}); \quad (2)$$
$$Y_i=E_{(P_{3i-2}\oplus Y_{i-3}\|P_{3i-1}\oplus Y_{i-2})}(P_{3i}\oplus Y_{i-1}). \quad (3)$$

2) 消息 P'

$$Y'_{i-2}=E_{(P'_{3i-8}\oplus Y_{i-5}\|P_{3i-7}\oplus Y_{i-4})}(P'_{3i-6}\oplus Y_{i-3}); \quad (4)$$

$$Y'_{i-1}=E_{(P_{3i-5}\oplus Y_{i-5}\,\|\,P_{3i-4}\oplus Y_{i-4})}(P'_{3i-3}\oplus Y'_{i-2});\quad(5)$$

$$Y_i=E_{(P_{3i-2}\oplus Y_{i-3}\,\|\,P'_{3i-1}\oplus Y'_{i-2})}(P_{3i}\oplus Y_{i-1}).\quad(6)$$

为了使上述 3 轮的碰撞路径成立,需要满足:

$$P_{3i-8}+P_{3i-6}=P'_{3i-8}+P'_{3i-6};\quad(7)$$

$$P_{3i-3}+P_{3i-1}=P'_{3i-3}+P'_{3i-1};\quad(8)$$

$$P_{3i-3}\oplus Y_{i-2}=P'_{3i-3}\oplus Y'_{i-2};\quad(9)$$

$$P_{3i-1}\oplus Y_{i-2}=P'_{3i-1}\oplus Y'_{i-2}.\quad(10)$$

其中式(7)、式(8)是保证初始密钥 K 保持不变,式(9)、式(10)是保证 Y_{i-1} 与 Y_i 的值保持不变.式(7)~(10)可以化简为

$$P_{3i-8}+P_{3i-6}=P'_{3i-8}+P'_{3i-6};\quad(11)$$

$$P_{3i-3}+P_{3i-1}=(P_{3i-3}\oplus Y_{i-2}\oplus Y'_{i-2})+(P_{3i-1}\oplus Y_{i-2}\oplus Y'_{i-2}).\quad(12)$$

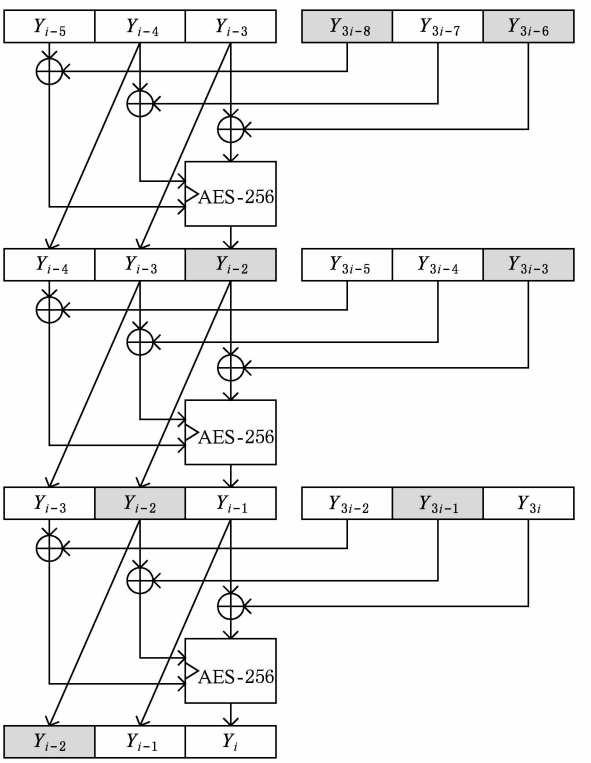

Fig. 1 Inner collision for 3 round HTBC.

图 1 HTBC 算法的 3 轮内部碰撞

由于 P_{3i-8},P_{3i-6},P'_{3i-8},P'_{3i-6},均可以自由选取,所以式(11)很容易满足.但 P_{3i-3},P_{3i-1} 的取值则需要满足式(12).由于 Y_{i-2} 与 Y'_{i-2} 的随机性,并且算法中采用的是按字节模 256 加法,所以求解式(12)即等价于求解:

$$\alpha+\beta=(\alpha\oplus\gamma)+(\beta\oplus\gamma),\quad(13)$$

其中,γ 是 1 个随机的长度为 8 b 的整数,α 和 β 是可以自由选取的长度为 8 b 的整数.经过计算可知,有 512 对(α,β)的取值,使得对于 γ 的所有 256 种取值式(13)都满足.将(P_{3i-3},P_{3i-1})中每个字节的取值固定为其中任意的 1 对(α,β)的取值,则式(12)恒成立.按照上述的方法选择$(P_{3i-8},P_{3i-6},P_{3i-3},P_{3i-1},P'_{3i-8},P'_{3i-6},P'_{3i-3},P'_{3i-1})$的取值,可以构造 HTBC 算法的 3 轮内部碰撞.更一般的情况,表 1 给出了式(13)的分布,表 1 中 $\#(\alpha,\beta)$表示(α,β)的取值对数,$\#\gamma$ 表示相应的满足式(13)的 γ 的取值的个数.$\#(\alpha,\beta)$所有元素相加之和正好对应了(α,β)的所有 65 536 种可能的取值.

Table 1 The Distribution of Equation (13)

表 1 式(13)的分布

$\#(\alpha,\beta)$	$\#\gamma$	$\#(\alpha,\beta)$	$\#\gamma$
512	256	17 920	16
3 584	128	10 752	8
10 752	64	3 584	4
17 920	32	512	2

3 对 HTBC 杂凑函数的碰撞攻击

利用 HTBC 算法的 3 轮内部碰撞可以构造出 HTBC 算法的碰撞,攻击的时间复杂度仅为 1,具体的攻击步骤如下:

1) 随机生成 1 个长度为 384 m(单位 b)的消息 P,其中 $m\geqslant 5$,将消息 P 分成 m 个长度为 384 b 的大分组,将每个大分组分成 3 个长度为 128 b 的小分组 P_{3i-2},P_{3i-1},P_{3i}.

2) 将 P_{3m-3} 与 P_{3m-1} 中的每个字节的取值,固定为使得式(12)恒成立的任意 1 对取值,计算消息 P 的杂凑值为

$$h=E_K(Y_m)\oplus Y_{m-1}.$$

3) 构造与消息 P 产生碰撞的消息 P'.将 P 的取值全部赋给 P',修改 P'_{3m-8} 与 P'_{3m-6} 的取值,但保证:

$$P_{3m-8}+P_{3m-6}=P'_{3m-8}+P'_{3m-6},$$

则 Y'_{m-2} 的值相应发生改变,对应修改 P'_{3m-3} 与 P'_{3m-1} 的值为

$$P'_{3m-3}=P_{3m-3}\oplus Y_{m-2}\oplus Y'_{m-2},$$

$$P'_{3m-1}=P_{3m-1}\oplus Y_{m-2}\oplus Y'_{m-2}.$$

4) 由于限制 P_{3m-3} 与 P_{3m-1} 的取值使得式(12)恒成立,故 P 的初始密钥 K 与 P' 的初始密钥 K' 完全相同.按照 HTBC 的 3 轮内部碰撞性质,计算 P' 的杂凑值为

$$h'=E_{K'}(Y'_m)\oplus Y'_{m-1}=E_K(Y_m)\oplus Y_{m-1}=h,$$

则(P,P')是 1 对碰撞消息.

因为碰撞消息是直接利用 HTBC 内部碰撞性质直接构造出的,所以上述攻击的时间复杂度为 1.

图 2 给出了 1 对 HTBC 算法的碰撞实例,下划线标出的部分代表有差分的位置.

Message	f8f6590a9c682f47e9b6e682564bc928	f8f6590a9c682f47e9b6e682564bc928
	b38c4154ba170037e84880bd8b02ac6d	b38c4154ba170037e84880bd8b02ac6d
	1bacdd92cd79079866ed35b84d9d32d8	1bacdd92cd79079866ed35b84d9d32d8
	e53689ac19c0b022392481fd04ece9bf	e53689ac19c0b022392481fd04ece9bf
	064be4cd245da9ce75af215453029ebe	064be4cd245da9ce75af215453029ebe
	b24cca5db100ded56f8f14c6202f5eac	b24cca5db100ded56f8f14c6202f5eac
	60db5804c69b7bb1bd05949d8f0475a3	60db5804c69b7bb1bd05949d8f0475a3
	c3d8ebada95ef01933922162065272fc	c3d8ebada95ef01933922162065272fc
	d165207fdebbe708e7f676dc2a2a204f	d165207fdebbe708e7f676dc2a2a204f
	bfe1d4e52a634fb62d349017e0dd8c76	22a9809dc99a80a41412966f137e9420
	03ef24869246549d9a8bad5a4cfc058a	03ef24869246549d9a8bad5a4cfc058a
	4f706d4c5b956374047d492ed35916e4	eca8c194bc5e32861d9f43d6a0b80e3a
	9b844c610ac32a367fcb225c1b038d1c	9b844c610ac32a367fcb225c1b038d1c
	1b8c9e2c657f941b607534ee094d770c	1b8c9e2c657f941b607534ee094d770c
	0e0e0e0e0e0e0e0e0e0e0e0e0e0e0e0e	0bde4cfc0efd79bdf38a9d09c5c9545c
	ca3d4fcb6fa64b72eb253aa1aa4117b6	ca3d4fcb6fa64b72eb253aa1aa4117b6
	71717171717171717171717171717171	74a13383718206c28cf5e276bab62b23
	058b6873a294ee424a94418dddbd5eac	058b6873a294ee424a94418dddbd5eac
Hash	2c5cc8421d923ca33945ef6dbe7620db	2c5cc8421d923ca33945ef6dbe7620db

Fig. 2 Collision example for the HTBC Hash function.

图 2 HTBC 杂凑函数的碰撞实例

4 对 HTBC 杂凑函数的第二原像攻击

利用 HTBC 的内部碰撞性质,可以对满足特定要求的消息进行第二原像攻击. 由于 HTBC 所采用的特殊填充算法只能对消息字节长度 L(单位为 b)满足

$$32+48m\leqslant L\leqslant 48+48m, m\geqslant 5$$

的消息进行第二原像攻击,具体的攻击步骤如下:

1) 将消息 P 分成 m 个长度为 384 b 的大分组,将每个大分组分成 3 个长度为 128 b 的小分组 $P_{3i-2}, P_{3i-1}, P_{3i}$.

2) 计算消息 P 的杂凑值为

$$h=E_K(Y_m)\oplus Y_{m-1}.$$

3) 构造消息 P 的第二原像消息 P',将 P 的取值全部赋给 P'. 随机修改 P'_{3m-8} 与 P'_{3m-6} 的取值,但保证:

$$P_{3m-8}+P_{3m-6}=P'_{3m-8}+P'_{3m-6},$$

则 Y'_{m-2} 的值相应发生改变,对应修改 P'_{3m-3} 与 P'_{3m-1} 的值为

$$P'_{3m-3}=P_{3m-3}\oplus Y_{m-2}\oplus Y'_{m-2},$$

$$P'_{3m-1}=P_{3m-1}\oplus Y_{m-2}\oplus Y'_{m-2}.$$

4) 对于得到的 (P'_{3m-3}, P'_{3m-1}),检查式(12)是否成立.

① 如果式(12)成立,则计算 P' 的杂凑值为

$$h'=E_{K'}(Y'_m)\oplus Y'_{m-1}=E_K(Y_m)\oplus Y_{m-1}=h,$$

则 P' 即为 P 的第二原像.

② 如果式(12)不成立,则重复 3).

当进行第二原像攻击时,由于消息 P 的取值已经完全确定,因此 P_{3m-3} 与 P_{3m-1} 的取值也已经固定,式(12)只能通过概率满足,对于特定的 P_{3m-3} 与 P_{3m-1} 的取值,可以通过计算式(13)的分布来估计第二原像攻击的复杂度. 最坏的情况下,对于 P_{3m-3}, P_{3m-1} 的每个字节,只有 2 个对应的 $Y_{m-2}\oplus Y'_{m-2}$ 的值满足式(12),式(12)成立的概率为 2^{-7},由于 P_{3m-3}, P_{3m-1} 的长度为 16 b,故式(12)成立的概率为 2^{-112},所以最坏情况下第二原像攻击的时间复杂度为 2^{112},但仍然优于穷举攻击的时间复杂度 2^{128}. 最优的情

况下,P_{3m-3},P_{3m-1}的取值使得式(12)恒成立,则第二原像攻击的时间复杂度仅为 1.当 P 的取值足够随机时,利用表 2 给出的分布,可以计算式(12)成立的平均概率为 $2^{-2.91\times16}=2^{-46.56}$.如果 P 的长度满足要求,且 P 的取值足够随机,则按照上述攻击步骤,找到 P 的第二原像 P'的时间复杂度为 $2^{46.56}$.

5 结束语

本文通过研究发现了 HTBC 算法的三重链接结构的漏洞,利用这种链接结构的安全性弱点,构造了 HTBC 算法的 3 轮内部碰撞,并利用 3 轮内部碰撞的特性构造出 HTBC 算法的碰撞,并给出了碰撞实例,攻击的时间复杂度仅为 1.对于特定长度的消息块,可以进行第二原像攻击,最坏情况下攻击的最大时间复杂度为 2^{112},优于穷举攻击的复杂度 2^{128};最优情况下构造第二原像的时间复杂度仅为 1;当消息随机取值时,攻击的平均时间复杂度为 $2^{46.56}$.

致谢 感谢 HTBC 算法的设计者黄玉划老师提供基于 AES-256 的算法源码,使得作者通过构造实际碰撞的方式检验本文攻击方法的正确性.

参 考 文 献

[1] Bertoni G, Daemen J, Peeters M, et al. The Keccak SHA-3 submission, submission to NIST [OL]. 2011[2013-06-26]. http://keccak.noekeon.org/Keccak-submission-3.pdf

[2] US Department of Commerce, National Institute of Standards and Technology. Secure Hash Standard (SHS): Federal Information Processing Standards Publication 180-3 [S]. Gaithersburg, USA: NIST, 2008

[3] International Organization for Standardization. ISO/IEC 10118-3: 2004, Information Technology-Security Techniques-Hash Functions—Part 3: Dedicated Hash Functions [S]. Geneva, Switzerland: ISO/IEC, 2004

[4] Preneel B, Govaerts R, Vandewalle J. Hash functions based on block ciphers: A synthetic approach [G] //LNCS 773: Proc of CRYPTO 1993. Berlin: Springer, 1994: 363-378

[5] International Organization for Standardization. ISO/IEC 10118-2: 1994, Information Technology-Security Techniques-Hash Functions-Part 2: Hash Functions Using an n-bit Block Cipher Algorithm [S]. Geneva, Switzerland: ISO/IEC, 1994

[6] Lai X, Massey J. Hash function based on block ciphers [G] //LNCS 658: Proc of EUROCRYPT 1992. Berlin: Springer, 1993: 55-70

[7] Hirose S. Some plausible constructions of double-block-length hash functions [G] //LNCS 4047: Proc of FSE 2006. Berlin: Springer, 2006: 231-246

[8] Lee J, Stam M. MJH: A faster alternative to MDC-2 [G] //LNCS 6558: Proc of CT-RSA 2011. Berlin: Springer, 2011: 213-236

[9] Huang Yuhua, Hu Aiqun, Wang Xingjian. Triple-block-chaining-based hash function and its performance analysis [J]. Journal of Computer Research and Development, 2006, 43(8): 1398-1404 (in Chinese)
(黄玉划,胡爱群,王兴建. 基于三重分组链接的散列函数及其性能分析[J]. 计算机研究与发展, 2006, 43(8): 1398-1404)

[10] US Department of Commerce, National Institute of Standards and Technology. Specification for the Advanced Encryption Standard (AES): Federal Information Processing Standards Publication 197 [S]. Gaithersburg, USA: NIST, 2001

Ma Bingke, born in 1989. PhD candidate. His main research interests include cryptanalysis of hash functions and block ciphers.

Li Bao, born in 1962. Professor and PhD supervisor in the Institute of Information Engineering, Chinese Academy of Sciences. His main research interests include security protocol and technology.