文章编号：2096-1472(2017)-11-27-03

基于局部碰撞算法的SHA-1改进算法设计与研究

刘 坤，杨正校
(苏州健雄职业技术学院软件与服务外包学院，江苏 太仓 215411)

摘 要：SHA-1是一种哈希函数，它被广泛使用在电子商务这样的现代安全领域，特别是应用于数据加密通信、数字签名。很多的密码协议、标准中都包括了SHA-1算法，如著名的SSL、IPsec和PKCS。本文通过深入分析SHA-1算法及碰撞算法原理，找出SHA-1算法内部碰撞的原因，对算法中逻辑函数和压缩函数进行改进设计，得到基于局部碰撞算法的SHA-1改进算法。

关键词：哈希函数；SHA-1算法；局部碰撞算法；压缩函数
中图分类号：TP309 **文献标识码**：A

Design and Research of the Improved SHA-1 Algorithm Based on the Local Collision Algorithm

LIU Kun,YANG Zhengxiao
(*Institute of Software and Service Outsourcing,Suzhou Chine-Shiung Institute of Technology*,*Taicang* 215411,*China*)

Abstract: As a hash function,SHA-1 is widely used in modern security fields such as electronic commerce,especially for data encrypted communication and digital signature.The SHA-1 algorithm is applied in many cryptographic protocols and standards,such as the famous SSL,IPsec and PKCS.Through the in-depth analysis of the SHA-1 algorithm and the collision algorithm principle,the paper identifies the causes of internal collision in the SHA-1 algorithm,improves the logical function and compression function in the algorithm,and achieves the improved SHA-1 algorithm based on the local collision algorithm.

Keywords:hash function;SHA-1 algorithm;Local collision algorithm;compression function

1 引言(Introduction)

Hash函数主要有两个系列，分别是MDx系列和SHA系列，其中MDx系列包括MD4、MD5等，SHA系列主要包括SHA-1、SHA-2等。MD5、SHA-1和SHA-2算法在数据加密、数字签名方面被广泛应用。1990年MD4算法被提出，但是很快发现MD4算法存在严重的安全问题，在1992年MD4算法被MD5算法取代。MD5算法在之后的十几年内被软件行业广泛使用，直到2004年我国密码学家王小云在国际密码讨论年会(CRYPTO)上展示了MD5算法的碰撞，并给出了第一个实例[1]。该攻击复杂度很低，在普通计算机上只需要几秒钟的时间。在2005年王小云教授与其同事又提出了对SHA-1算法的碰撞算法[2]，不过计算复杂度为2的69次方，在实际情况下难以实现。2008年的Chaos Communication Congress大会上，研究人员展示了利用MD5碰撞来伪造合法CA证书，从而攻破了HTTPS的安全体系。2012年在中东大范围爆发的火焰(Flame)病毒，包含了一个伪造的数字签名，就是利用MD5碰撞伪造了合法的微软签名来逃避杀毒软件的查杀。

2017年2月23日，荷兰阿姆斯特丹(CWI)研究所和Google公司的研究人员在谷歌安全博客上发布了世界上第一例公开的SHA-1哈希碰撞实例，在经过两年的联合研究和花费了巨大的计算机时间之后，研究人员在他们的研究网站SHAttered上给出了两个内容不同，但是具有相同SHA-1消息摘要的PDF文件，这就意味着在理论研究长期以来警示SHA-1算法存在风险之后，SHA-1算法的实际攻击案例也浮出水面，同时也标志着SHA-1算法终于走向了生命的末期。从这些事件上可以看出，MD4、MD5和SHA-1已经不安全。本文主要根据近几年国内外对SHA-1算法的碰撞算法进行分析研究，给出算法中压缩函数和逻辑函数的改进描述，以提高SHA-1抗碰撞性。

2 SHA-1算法内部碰撞原理(Internal collision principle of SHA-1 algorithm)

SHA-1算法通过一系列的迭代计算把任意长度的比特串

基金项目：本文是2017年太仓科技局科技计划项目基础研究计划《基于攻击原型建模的工业控制网络安全技术研究》的研究成果.

压缩成长度160位的位串，而且一般认为它的计算过程在密码学意义上是单向的，也就是很难找到两个不同的位串可以压缩成相同的160位串[3]。正因为SHA−1算法具有良好的特性，它被广泛使用在电子商务这样的现代安全领域，特别是应用于公钥密码系统的数字签名中，很多的密码协议、标准中都包括了SHA−1算法，如著名的SSL、IPsec和PKCS。当今社会移动终端技术快速发展，推动了电子商务的发展，因此SHA−1算法的安全性直接影响了使用它作为协议的密码系统安全性，也将影响到电子商务活动中数字证书的安全性。

针对哈希函数的攻击方式很多，具体分类如图1所示，其中最常用的是碰撞攻击。所谓碰撞攻击也就是假设哈希函数为H，攻击者尝试找到两个信息M和M'，假设$M\neq M'$，但$H(M)=H(M')$[4]。根据Hash函数的值域与定义域相比规模要小得多，是“多对一”映射，找出两个不同的消息，使其产生相同的Hash结果，这称为碰撞攻击。一个具有n比特输出长度的Hash函数共有2^n个可能的输出值，用穷举法只要计算$2^{n/2}$个消息，就能期望找到一对碰撞。因此，值$2^{n/2}$决定了Hash函数抗强行攻击的强度[5]。如果一个输出长度为n比特的Hash函数可以用小于$2^{n/2}$的计算找到一对碰撞，则该Hash函数理论上被认为是可破解的。对于SHA−1来说，利用穷举法寻找它的碰撞至少需要进行2^{80}次运算，而最新的研究已经将碰撞次数减低到了$2^{57.5}$，也就是大大提高了SHA−1碰撞可能性，并且已经找到具体碰撞实例[6]，这说明SHA−1碰撞处理方面有严重的安全缺陷。

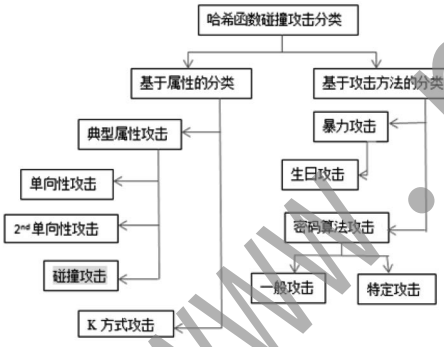

图1 哈希函数碰撞攻击分类

Fig.1 Hash function collision attack classification

在明文空间中随机选取一段明文求出其Hash值，并以单字节字符的方式来表示，然后随机地选择并改变明文中任意1比特的值得到另一新的Hash结果。定义两个Hash值之间的距离为：

$$d=\sum_{i=0}^{S}|t(ei)-t(ei')|$$

其中，ei和ei'分别是最初和新的Hash值的第i个字符，S为Hash值对应字符的个数，函数t()将ei和ei'转换成对应的十进制数。若两个Hash分别由两个独立的均匀分布的随机序列所组成，则理论上Hash值的单位字符的平均距离为85.33。取输入长度n=512比特，随机选择输入样本，测试其输出的单位字符的平均距离。有关实验数据表明SHA−1算法在迭代在30步之后，其输出的单位字符的平均距离才趋于稳定[6]。

SHA−1算法的关键是压缩函数。在SHA−1的内部结构中，链接变量A、B、C、D、E经过6个操作步的传递、混合后，又回到A、B、C、D、E，过程如图2所示。这样就容易出现局部碰撞。例如若在SHA−1算法第i步存在1比特消息差分，这1比特差分将在随后的5操作步中依次影响5个链接变量。由于存在这样的规律性，若能阻止差分传播，则可构造一个局部碰撞差分链，进而产生6操作步的局部碰撞。

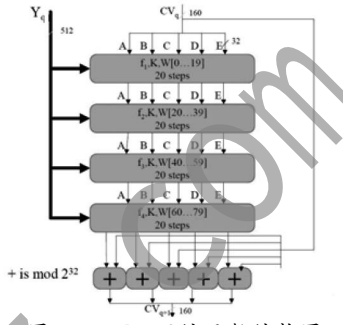

图2 SHA−1压缩函数结构图

Fig2. Structure diagram of SHA−1 compression function

3 SHA-1算法压缩函数和逻辑函数分析(Analysis of SHA-1 algorithm's compression function and\ logical function)

3.1 SHA-1算法逻辑函数分析

在SHA−1算法中逻辑函数ft具有良好的混淆性，但是自身的抗碰撞性较弱。SHA−1压缩函数逻辑结构有四轮循环模块，每一轮使用一个函数如f_1、f_2、f_3、f_4来表示四轮循环对应的逻辑函数，函数如下式所示[7]：

$$f_1=(x\wedge y)\oplus(\bar{x}\wedge z)$$
$$f_2=x\oplus y\oplus z$$
$$f_3=(x\wedge y)\vee(z\wedge(x\vee y))$$
$$f_4=x\oplus y\oplus x$$

从中可以知道这四个函数中f_2和f_4是相同的，这样会造成防御风险。

3.2 SHA-1算法压缩函数分析

SHA−1算法能够将任意长的输入压缩成160bit的输出，但SHA−1算法中的基本迭代只能处理512bit的数据块。因此首先需要将输入的消息每512bit分成一块，并将最后一块不足的消息按一定规则补齐。SHA−1算法的核心部分是压缩函数。SHA−1算法压缩函数的功能结构，是将字符串以512位分组为单位进行处理，主循环由四轮循环模块构成，每轮20个步骤的运算，输入当前分组q的16个字，和由前一个分组输出的160位缓存值[8]。

SHA−1压缩函数的工作过程是首先5个中间变量a，b，c，d，e中置入特定初值，压缩函数的输入值为IHVin=(a,b,c,d,e)。然后512位信息块B被分割成16个连续的32位字，如W_0，W_1，…，W_{15}，利用扩展函数公式(1)，将字

从16个扩展到80个。

$$W_t = RL(W_{t-3} \oplus W_{t-8} \oplus W_{t-14} \oplus W_{t-16}, 1), 16 \le t \le 79 \quad (1)$$

其中，RL表示向左循环移动1位。对于每一步t=0，1，…，79，用5个32位字Q_t，Q_{t-1}，Q_{t-2}，Q_{t-3}，Q_{t-4}来计算一个新的中间状态字Q_{t+1}。先对Q_t进行初始化，公式(2)如下所示：

$$(Q_0, Q_{-1}, Q_{-2}, Q_{-3}, Q_{-4}) = (a, b, (c, 30), RR(d, 30), RR(e, 30)) \quad (2)$$

其中，RR(x,30)表示向右循环移动30位。对于t=0,1,…,79,Q_{t+1}是这样被计算出来的：

$$F_t = f_t(Q_{t-1}, RL(Q_{t-2}, 30), RL(Q_{t-3}, 30)) \quad (3)$$

$$Q_{t+1} = F_t + AC_t + W_t + RL(Q_t, 5) + RL(Q_{t-4}, 30) \quad (4)$$

其中，

$$AC_t = \begin{cases} 0x05a82799 & 0 \le t \le 19 \\ 0x6ed9eba1 & 20 \le t \le 39 \\ 0x8f1bbada & 40 \le t \le 59 \\ 0xca62c1d6 & 60 \le t \le 79 \end{cases}$$

$$f_t(x, y, z) = \begin{cases} (x \wedge y) \oplus (\bar{x} \wedge z)(xy) & 0 \le t \le 19 \\ x \oplus y \oplus z & 20 \le t \le 39 \\ (x \wedge y) \vee (z \wedge (x \vee y)) & 40 \le t \le 59 \\ x \oplus y \oplus z & 60 \le t \le 79 \end{cases}$$

最终循环结束输出的值为：

$$\delta IHV_{in} = (a + Q_{80}, b + Q_{79}, c + RL(Q_{78}, 30), d + RL(Q_{77}, 30), e + RL(Q_{76}, 30)) \quad (5)$$

4 基于局部碰撞的SHA-1改进算法研究(Research on improved SHA-1 algorithm based on local collision)

通过对SHA−1算法和碰撞攻击原理分析，找到SHA−1算法产生内部碰撞的原因，这里主要从两个方面即SHA−1算法的逻辑函数和压缩函数进行算法改进，以提高SHA−1抗碰撞攻击能力。

4.1 SHA-1逻辑函数改进研究

SHA−1每轮循环实际上只使用了三个函数。这样随着现在攻击者技术和计算能力的不断提高，SHA−1的安全性受到了很大的影响。本文采用变换其中一个函数，也就是使f2和f4不一样，加强SHA−1算法的完备性，以及增加整体算法的抗差分分析能力。这里针对函数f4的表达式进行修改，见表1。这样改进后，对其运算速度几乎没有影响，同时大大提高了SHA−1的安全性和抵抗SHA−1算法局部碰撞攻击的能力。

表1 改进后的逻辑函数表达式

Tab.1 Improved logic function expression

轮数	逻辑函数	函数表达式
第一轮(0≤t≤19)	$f_1(x, y, z)$	$(x \wedge y) \oplus (\bar{x} \wedge z)$
第二轮(20≤t≤39)	$f_2(x, y, z)$	$x \oplus y \oplus z$
第三轮(40≤t≤59)	$f_3(x, y, z)$	$(x \wedge y) \vee (z \wedge (x \vee y))$
第四轮(60≤t≤79)	$f_4(x, y, z)$	$(x \wedge y) \oplus z$

4.2 SHA-1压缩函数逻辑结构改进研究

从上面的SHA−1压缩过程中得知，压缩函数逻辑结构对其进行四轮、每轮20个步骤的运算存在一个规律性的安全隐患，这里通过改变压缩函数逻辑运算来抵抗强碰撞攻击，同时对计算量几乎没有影响。针对上面所述SHA−1碰撞出现的原因，这里将SHA−1压缩函数进行改进的算法是在第一轮的压缩函数中，用简单的逻辑判断、逻辑取反、移位操作引入到混合函数，目的在于加速首轮的差分扩散，并且打破原来固定的链接变量传递方式带来的规律性，使传递过程具有很强的随机性，从而消除局部碰撞的依从条件。

5 结论(Conclusion)

针对SHA−1碰撞算法研究必将导致不久的将来SHA−1算法被破解，攻击者可以通过伪造数字证书等手段破坏密码系统、数字证书系统的电子商务领域的安全性、可靠性，严重可导致经济损失或有政治攻击目的。因此对SHA−1算法的研究和改进具有重大意义。SHA−1算法的改进可以有效提高利用SHA−1算法进行加密、数字签名的信息安全性、可靠性。

本文根据SHA−1算法内部缺陷找到碰撞原因，对SHA−1算法中逻辑函数和压缩函数进行改进描述，以提高SHA−1算法抗碰撞性。下一步可以将改进算法在硬件电路上进行设计实现，测试改进算法的抗碰撞性和运算速度。

参考文献(References)

[1] Xiaoyun Wang,Hongbo Yu.How to Break MD5 and Other Hash Functions,EUROCRYPT(Ronald Cramer,ed.)[M]. Lecture Notes in Computer Science,2005,3494:19−35.

[2] Xiaoyun Wang,Yiqun Lisa Yin,Hongbo Yu.Finding Collisions in the Full SHA−1,CRYPTO(Victor Shoup,ed.)[??].Lecture Notes in Computer Science,2005,3621:17−36.

[3] 黄谆,白国强,陈弘毅.快速实现SHA−1算法的硬件结构[J].清华大学学报(自然科学版),2005,45(1):123−125.

[4] 张斌,徐名扬.SHA−1算法及其在FPGA加密认证系统中的应用[J].中国集成电路,2011,145(6):57−61.

[5] 林雅榕,侯整风.对哈希算法SHA−1的分析和改进[J].计算机技术与发展,2006,16(3):124−126.

[6] 刘建东,余有明,江慧娜.单向Hash函数SHA−1的统计分析与算法改进[J].计算机科学,2009,10(15):141−145.

[7] Christophe De Cannière,Florian Mendel,Christian Rechberger. Collisions for 70−Step SHA−1:On the Full Cost of Collision Search,SelectedAreas in Cryptography(Carlisle M.Adams, Ali Miri,and Michael J.Wiener,eds.),Lecture Notes in Computer Science,Springer,2007,4876:56−73.

[8] Stevens M.New collision attacks on SHA−1 based on optimaljoint local−collision analysis.In EUROCRYPT,2013.

作者简介：

刘　坤(1979−)，女，硕士，讲师.研究领域：计算机网络技术，网络安全技术.

杨正校(1963−)，男，硕士，教授.研究领域：软件技术.