文章编号:1672 - 058X(2012)07 - 0069 - 04

基于混沌理论的 HASH 函数研究现状综述*

杨 华 千
(重庆教育学院 网络管理中心,重庆 400067)

摘 要:首先综述了 HASH 函数的基本思想与应用情况,阐述了近年提出的一些传统 HASH 函数和基于混沌理论的 HASH 函数,并指出了存在的一些问题;最后分析了基于混沌理论的 Hash 函数进一步的研究方向。

关键词:混沌理论,Hash 函数,数字签名
中图分类号:O59 **文献标志码**:A

当前,以因特网为基础的电子商务应用正方兴未艾,如何保证网上信息的安全传输、存储等固然是研究的重点,但如何保证信息的公平和防抵赖也正成为一个亟待解决的问题。传统的手工签名是一种经典的防抵赖手段,然而在电子商务领域它正变得越来越不现实,因此,数字签名就应运而生。单向散列 HASH 函数是数字签名中的一个关键环节,可以大大缩短签名时间,在消息完整性检测,以及消息认证等方面有着广泛的应用。1968 年 Wilkes 首次提出了单向散列函数的概念,其最大特点就是正向计算简单、反向计算复杂,而且很难找到两个不同的输入对应于同一个输出。根据应用情况不同,HASH 函数一般分为无密钥单向散列函数和带密钥的单向散列函数两类。传统经典的单向 HASH 方法如 MD2、MD4、MD5 和 SHA 系列等多是采用有限域上的异或等逻辑运算或是用分组加密方法多次迭代得到 HASH 结果。但是,由于异或运算固有的缺陷,虽然每步运算简单,但计算轮数即使在被处理的文本很短情况下也很大。

1 HASH 函数的基本思想

单向散列函数(one-way hash function)的基本思想是单向散列函数 $h(x)$ 作用于任意长度的消息 x 上,它返回一个固定长度的散列值 y。即 $y = h(x)$。具有如下的几个性质:

(1) 压缩性质。对任意有限长度的输入消息位,都将输出固定长度的 HASH 值。

(2) 计算简单。在给定 HASH 函数 h 和输入 x,的情况下,$h(x)$ 的值很容易计算出来。

(3) 逆运算难。即给定 y,根据 $y = h(x)$ 反向计算 x 很难。

(4) 碰撞很难。给定 x,要找另一个消息 x' 并满足 $h(x) = h(x')$ 很难。

HASH 散列函数最根本的特点是变换的单向性,一旦消息序列被转换,就无法再以确定的方法从 HASH 序列获得其原始序列,从而无法控制变换得到的结果,达到防止信息被篡改的目的。正是由于 HASH 函数的这种单向不可逆性,使其非常适合被用来确定原文的完整性,从而被广泛用于数字签名。

收稿日期:2012 - 02 - 20;**修回日期**:2012 - 03 - 15.
* **基金项目**:重庆市科委自然科学基金(CSTC,2010BB2279)资助;重庆市教委科学技术研究项目(KJ091501,KJ091502,KJ101501)资助.
作者简介:杨华千(1972-),男,四川渠县人,副教授,博士,从事信息安全研究.

HASH 函数在数字签名中的应用步骤如下:

(1) 签名发送。发送方首先选择一个单向哈希函数将待发送的报文消息生成报文摘要,然后再用自己的私钥对这个摘要进行加密,加密后的摘要作为报文的数字签名和报文消息一起发送给接收方。

(2) 解密确认。接收方先用发送方的公钥来对报文附加的数字签名进行解密得到发送方传送过来的摘要,接着再用与发送方相同的哈希函数计算收到的原始报文的摘要。如果这两个摘要相同,接收方就能确认该数字签名是发送方的,并且原始报文传输过程中没有被篡改。

在实际应用中,针对数字签名一般会有两种攻击情况发生:一种就是针对一个现有报文伪造一个新的报文。根据上述的 HASH 函数性质4,这种攻击目前来说还相当困难;还有一种就是做两个不同的文件,它们有相同的签名,然后拿一个文件来骗一个签字,再拿另外一个文件作为证据,这种就是所谓的碰撞。

2 传统的 HASH 函数

MD5 与 SHA1 都属于 HASH 函数标准算法中两大重要算法,主要用于证明原文的完整性和准确性,是为电子文件进行数字签名的重要工具。MD5 和 SHA1 是目前应用最广泛的 HASH 算法,而它们都是以 MD4 为基础设计的。

(1) MD4。MD4(RFC 1320)是 MIT 的 Ronald L. Rivest 教授在 1990 年设计的一种信息摘要算法。它是一种用来测试信息完整性的密码散列函数,其摘要长度为 128 位。但由于 Hash 值较小,容易受到生日攻击。2004 年 Denboer 和 Bosselaers 以及其他学者很快发现了攻击 MD4 散列过程中第一步和第三步的漏洞,这直接导致了 MD4 退出应用市场。

(2) MD5。MD5(RFC 1321)是 Rivest 于 1991 年对 MD4 的改进版本。它在 MD4 的基础上增加了"安全-带子"(safety-belts)的概念。虽然 MD5 比 MD4 稍微慢一些,但却更为安全。但遗憾的是,该算法也于 2004 年 8 月被中国科学家王小云教授成功破解。因此,MD5 也正受到巨大的威胁。

(3) SHA1。SHA1 是由 NIST 设计为同 DSA 一起使用的,它产生长度为 160 bit 的散列值,因此抗穷举(brute-force)性更好。SHA-1 设计时基于 MD4 相同原理,并且模仿了该算法。该系列还有 SHA224、SHA256 和 SHA512 等。

3 基于混沌理论的 HASH 函数

离散混沌系统的迭代过程除了对初始条件敏感,产生复杂的轨迹外,迭代过程在一定意义上还具有单向性,离散混沌系统的这一属性在散列函数方面也得到了很好应用并已成为国内外的研究热点。目前,已有许多基于混沌理论的 Hash 函数构造方法被提出来[1-3]。

(1) 利用单个混沌映射构造 HASH 函数。D. Xiao 等人提出的基于参数可变的混沌系统的单向 HASH 函数构造方法[4]是这种方式的一个典型。该算法在混沌迭代的过程中,利用前次的迭代值和当前处理的明文消息不断调整分段线性混沌映射的控制参数,使得系统具有更好的混沌性。算法执行过程中,分别从前两轮的最后一次迭代值中抽取 40 bit,在最后一轮的迭代中抽取 48 bit 形成 128 bit 的 HASH 值。整个算法只需对明文消息进行三轮迭代。

(2) 利用时空混沌映射构造 HASH 函数。Y. Wang 等提出了基于二维耦合映像格子(2D CML)的单向 HASH 函数构造方法[6]。在算法中利用 Logistic 映射构造了一个 8×10 的耦合映像格子,为了破坏序列在迭代过程中的相关性,在每一步中都将 Logistic 映射迭代 45 次,然后再把输出与相邻的格子进行耦合。因而,算法的安全性和抵抗各种攻击的性能都比较高,但算法的效率有所降低。J. Zhang 等人进一步提出了一种基于"前馈-反馈"非线性数字滤波和时空混沌系统的单向 HASH 函数构造方法[7]。此外,郭现峰等提出了

一种基于混沌动态 S-BOX 的 HASH 函数构造方法[2]。

(3) 利用混沌映射网络构造 HASH 函数。在文献[8]中,提出了一种由多个简单混沌映射组成混沌映射网络的 HASH 函数。该算法的混沌网络由 16 个 Tent 映射构成,如图 1 所示。每次迭代输出 4 个值,当所有的消息序列 x 参与迭代完成后,从最后一列的输出值 $x_{i4}(i=0,1,2,3)$ 中分别抽取 40 bit,形成需要的散列值。由于消息序列 x 只参与一次迭代运算,其速度相当快。此外,从这种网状结构也可以发现,散列值对消息序列是相当敏感的。

虽然,基于混沌的 HASH 函数有较好的灵活性和运算速度,但也存在一些数字化实现过程中固有的缺陷。如计算机有效字长精度效应,使得混沌映射数字化后退化为周期较短的周期序列。

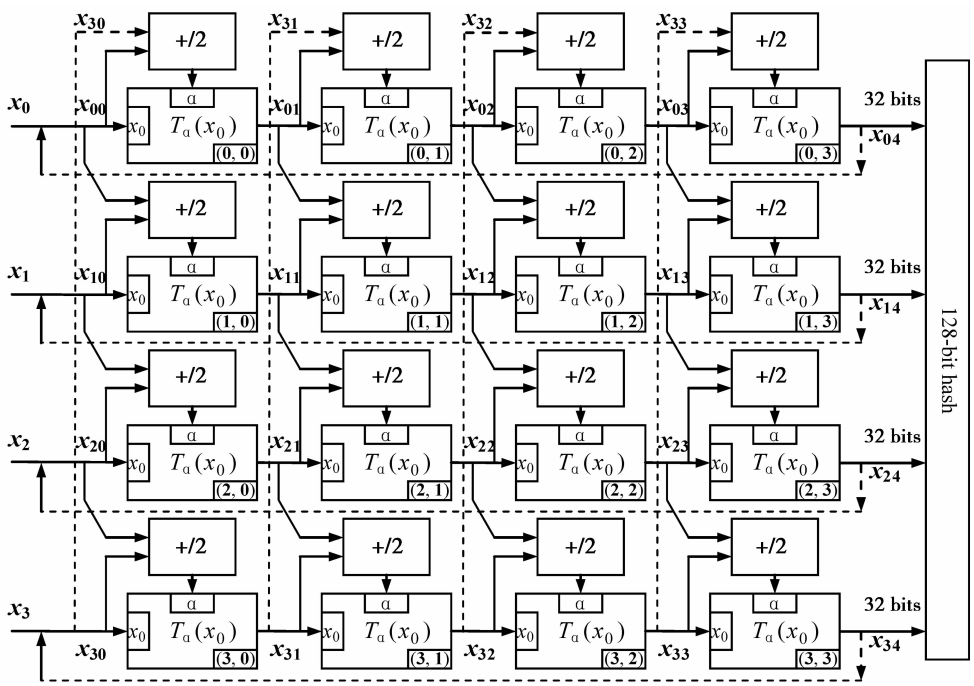

图 1　基于混沌映射网格的 PRNG

为了克服这些不足,如何利用离散混沌系统对初值和参数的极端敏感性、混沌序列的白噪声统计性和遍历性等优良特征并结合传统密码学中的经典理论来构造兼顾安全性和复杂度的混沌单向 HASH 算法仍将是今后研究的重点。

4　HASH 函数统计分析的几个指标

扩散与混淆是加密算法的两个基本条件,HASH 函数也不例外,也应该满足这两个基本条件。对于二进制格式的 HASH 值,每个比特可能的取值是 0 和 1,因此,对于一个理想的 HASH 函数来讲,初值的任何微小变化都将导致最后的 HASH 值的每个比特以 50% 的概率发生改变。通常,HASH 函数的安全性很难用理论来证明,一般采用概率统计的方法来分析 HASH 函数的安全性。常用 6 个判断指标[6]:最小改变比特数 $B_{\min}=\min(B_1,B_2,\cdots,B_i,\cdots,B_N)$;最大改变比特数 $B_{\max}=\max(B_1,B_2,\cdots,B_i,\cdots,B_N)$;平均变化的比特数 $\overline{B}=\frac{1}{N}\sum_{i=1}^{N}B_i$;平均变化概率:$P=(\frac{\overline{B}}{128})\times 100\%$;变化比特数的标准差:$\Delta B=\sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(B_i-\overline{B})^2}$;变化概率的标准差 $\Delta P=\sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(\frac{B_i}{128-P})^2}\times 100\%$,此处,$N$ 是测试次数,B_i 是第 i 次的比特变化数。

对于一个理想的 HASH 函数(128 bit 的 HASH 值),要求 $B_{\min}$、$B_{\max}$ 和 $\overline{B}$ 尽量接近 64,P 的理想值是 50%,

而 ΔP 和 ΔB 要尽可能小。

5 结 语

HASH 函数的构造及其分析是一对共生的矛盾体,相互之间不断推动对方的发展。基于混沌理论的 HASH 函数构造方法正成为目前的热点研究领域,为寻找理想的 HASH 函数构造方法提供了一个新的方向。在研究过程中,如何克服数值化实现过程中带来的系统性能退化问题将是一个重点研究问题。短期内,如何寻求一些新的混沌模式或结构来避免这个问题将有待进一步深入的研究。同时,如何从理论上的分析和评价基于混沌的 HASH 函数的性能也是未来的一个研究方向。

参考文献:

[1] 刘军宁,谢杰成,王普. 基于混沌映射的单向 Hash 函数构造[J]. 清华大学学报:自然科学版,2000,40(7):55-58

[2] 郭现峰,张家树. 基于混沌动态 S-BOX 的 Hash 函数[J]. 物理学报,2006,55(9):4442-4449

[3] 王继志,王英龙,王美琴. 一类基于混沌映射构造 Hash 函数方法的碰撞缺陷[J]. 物理学报,2006,55(10):5048-5054

[4] XIAO D,LIAO X,DENG S. One-way Hash function construction based on the chaotic map with changeable-parameter[J]. Chaos Solitons & Fractals,2005(24):65-71

[5] ZHANG H,WANG X,LI Z,et al. One way Hash function construction based on spatiotemporal chaos[J]. Acta Physica Sinica,2005,54(9):4006-4011

[6] WANG Y,LIAO X F,XIAO D,et al. One-way hash function construction based on 2D coupled map lattices[J]. Information Sciences,2008,178:1391-1406

[7] ZHANG J,WANG X,ZHANG W. Chaotic keyed hash function based on feedforward-feedback nonlinear digital filter[J]. Physics Letters A,2007,362:439-448

[8] YANG H Q,LIAO X F,WONG K W,et al. A fast image encryption and authentication scheme based on chaotic maps[C]. Communications in Nonlinear Science and Numerical Simulation,2010,15:3507-3517

Review of Researches on HASH Function Based on Chaotic Theory

YANG Hua-qian

(Centre of Network Management,Chongqing Education College,Chongqing 400067,China)

Abstract: In this paper, the application and main ideas of HASH function are reviewed, then some classic HASH functions and the HASH function based on chaotic theory in recent years are elaborated, some problems existed are pointed out, finally, the further research field of HASH function based on chaotic theory is forwarded.

Key words: chaotic theory; HASH function; digital signature

责任编辑:代小红