第 54 卷 第 12 期 2005 年 12 月
1000-3290/2005/54(12)/5566-08

物 理 学 报
ACTA PHYSICA SINICA

Vol.54,No.12,December,2005

基于复合非线性数字滤波器的 Hash 函数构造*

王小敏[1)] 张家树[1)] 张文芳[2)]

1)(西南交通大学信号与信息处理四川省重点实验室,成都 610031)

2)(西南交通大学计算机安全与通信保密研究所,成都 610031)

(2005 年 1 月 6 日收到;2005 年 7 月 4 日收到修改稿)

在对多个满足 Kelber 条件的滤波器组成的复合系统进行初步分析的基础上,提出了一个基于复合非线性数字滤波器的带密钥的 Hash 算法.算法首先构建能产生高维混沌序列的复合滤波器系统,然后在明文作用的复合序列控制下随机选择滤波器子系统,并以复合系统的初态作为密钥,以粗粒化的量化迭代轨迹作为明文的 Hash 值.讨论了复合系统实现 Hash 函数的不可逆性、防伪造性、初值敏感性等特点.研究结果表明:基于复合非线性数字滤波器的 Hash 算法简单快速,比基于单一混沌映射的 Hash 算法有着更高的安全性,同时滤波器结构中没有复杂的浮点运算,比一般复合混沌系统更易于软硬件实现.

关键词:Hash 函数,混沌,非线性自回归数字滤波器

PACC:0545

1. 引 言

随着网络应用的发展,Hash 算法得到了广泛应用并成为研究的热点[1,2].但经典 Hash 算法如 MD5,SHA 等,大多是基于复杂度假设,需要进行大量复杂的异或等逻辑运算或是用分组加密方法进行多次迭代[1],运算量很大.由于混沌具有对初始条件敏感、伪随机和类噪声等优良密码特性,被广泛应用在加密和随机数生成算法中.近几年来,混沌也被应用到 Hash 算法的研究中,并取得了某些研究成果[3—5].如文献[3]提出了基于混沌映射模型的 Hash 算法,但该算法基于某一特定的混沌系统,易被混沌预测技术[4—9]破译;同时有效字长精度效应将导致混沌序列的短周期行为,使得算法的性能蜕化[10].针对文献[3]中的问题,文献[4,5]分别提出采用广义混沌映射切换和复合混沌的方法来克服上述问题,取得了较好的效果.但以上方法中的混沌映射都涉及到复杂的浮点运算,影响了运算速度,也不利于硬件实现.

对此,本文提出一种基于复合非线性数字滤波器的混沌 Hash 新方法.首先构建能产生高维混沌序列的非线性数字滤波器组形成复合混沌系统,然后在消息 M 产生的复合序列控制下,随机选择迭代过程中的滤波器子系统,将消息 M 巧妙地调制在其迭代轨迹中,并以迭代轨迹的粗粒化量化作为 M 的 Hash 值.同时算法将复合滤波器系统的迭代初始点作为 Hash 函数的私有密钥,能够保证带密钥的 Hash 函数(K-HF)的安全性完全由密钥的安全性决定,满足 K-HF 的安全性要求.理论分析与仿真结果表明:这种复合滤波器系统比单一的混沌系统具有更好的密码特性,不仅能保证迭代轨迹与初始条件之间的复杂敏感的非线性关系,而且利用复合序列增加了迭代参数选择的随机性,比基于单一混沌映射的 Hash 算法具有更好的置乱性、更强的抗破译能力,且算法简单、易于软硬件实现.

2. 复合非线性数字滤波器的设计

2.1. 非线性数字滤波器结构

虽然传统的单峰混沌映射便于理论分析,但研究表明在有限精度条件下,它所产生的混沌序列安全性很脆弱[6—10].1993 年,Frey[8],Lin 和 Chau[11] 等提出了 n 维非线性数字滤波器结构(图 1).在某种条件下,这种结构能产生高维混沌信号[11—13],较单

* 国家自然科学基金(批准号:60272096)和四川省青年基金(批准号:03ZQ026-33,51430804Q72201)资助的课题.

峰混沌映射复杂,安全性更好.

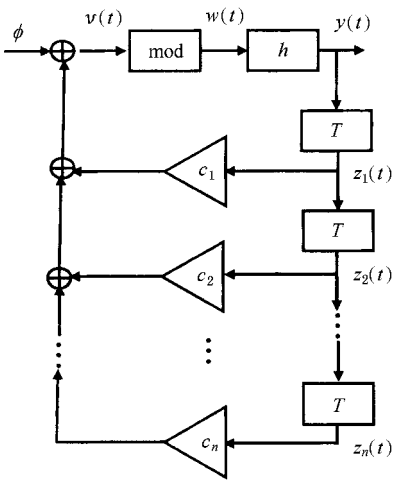

图 1　n 维非线性滤波器结构

图 1 的状态方程记为

$$z_1(t+1) = h\circ\mathrm{mod}\Big(\sum_{i=1}^{n} c_i z_i + \phi\Big),$$
$$z_i \in I,\quad \phi \in \Phi = \Re, \tag{1}$$
$$z_k(t+1) = z_{k-1}(t),\quad k = 2,3,\cdots,n,$$

式中,分段线性映射

$$h: I \to I,$$
$$h(w) = m_k \cdot w + r_k,$$
$$w \in W_k \subseteq I,\quad k \in \{1,\cdots,M\}.$$

硬件溢出函数

$$\mathrm{mod}(v) = v - 2\cdot\left[\frac{v+1}{2}\right] = v - 2\cdot l$$
$$v \in [-1+2\cdot l, 1+2\cdot l),$$
$$l \in G. \tag{2}$$

将(1)式写成向量形式

$$\varphi(z) = A_{lk}\cdot z + b_{lk}, \tag{3}$$

式中,

$$A_{lk} = \begin{bmatrix} m_k c_1 & m_k c_2 & \cdots & m_k c_{n-1} & m_k c_n \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ & & \ddots & & \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}_{n\times n},$$
$$b_{lk} = [r_k - 2\cdot l\cdot m_k + \phi \quad 0 \quad 0 \quad \cdots \quad 0]^{\mathrm{T}},$$
$$z = (z_1, z_2, \cdots, z_n)^{\mathrm{T}} \in Z = I^n,$$
$$\phi \in \Phi = \Re.$$

不失一般性,令

$$h: [-1,1) \to [-1,1),$$
$$\Phi = [-1,1).$$

设 $\lambda_1,\lambda_2,\cdots,\lambda_n$ 为 A_{lk} 的 n 个特征值,则有

$$\lambda_1\cdot\lambda_2\cdots\lambda_n = |A_{lk}| = m_k c_n. \tag{4}$$

Kelber[13] 证明了当系数 $c_n \in Z, |c_n| > 1$,而其他系数不为零时,非线性滤波器可以产生混沌序列.若 $|\lambda_i| \neq 1$,且 $h(\cdot)$ 具备均匀分布特性,则滤波器是遍历的而且保持 n 维均匀分布.

满足 Kelber 条件的 n 维滤波器产生的序列,具有很好的密码学特性,但仍存在以下不足:随着级数的增大,求解系统的特征值满足 $|\lambda_i| \neq 1$ 变得困难甚至是不可行,因而系统蜕化成简单系统级联的可能性大大增加;基于固定参数 c_i 的混沌序列,可以通过预测方法[6—9]对其破译;周期不够长,作为密码用的序列,周期越长越好,但考虑到硬件资源的限制,滤波器阶数不可能太高.

为了在不增加滤波器阶数的情况下,保证系统迭代轨迹与初始值之间复杂且敏感的非线性关系,本文提出采用变系数的方式形成复合滤波器系统,通过复合序列的控制产生更为复杂的混沌序列.

2.2. 复合非线性数字滤波器的设计与分析

针对固定系数 c_i 存在的问题,本文先提出一种变系数的复合非线性数字滤波器构造方法,并将其用于后面的 Hash 函数构造中.该复合滤波器的基本结构如图 2 所示.

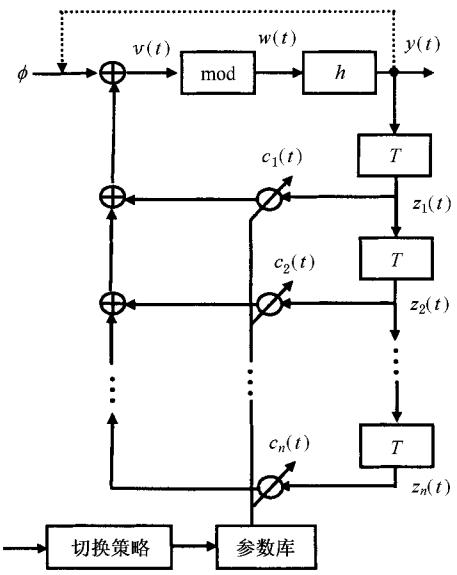

图 2　n 维复合非线性滤波器结构

首先建立满足 Kelber 条件的 k 个 n 维系数组

$\{c_i=[c_{i1},c_{i2},\cdots,c_{in}],i\in[1,2,\cdots,k]\}$，并将其放入参数库.滤波器每迭代一次时，就在切换策略的作用下选择一个系数组 c_i，而本次的输出作为下次的输入，在新的系数组的作用下继续迭代，循环往复.这种将固定系数 c_i 变成时变的系数 $c_i(t)$，相当于将 1 个滤波器变成 k 个滤波器切换的复合系统.下面将证明这种复合系统也具有 n 维均匀分布和遍历的特性.

定义 1 设 $y_i=\varphi_q(\boldsymbol{y},\phi,c_q)$，$\boldsymbol{y}=(y_{i-1},y_{i-2},\cdots,y_{i-n})$，$c_q$ 为第 q 组系数，$q=1,2,\cdots,k$ 是 k 个滤波器系统，对任意序列 $R=(r_1,r_2,\cdots)\in(1,2,\cdots,k)^{\infty}$，称 $y_i=\varphi_{r_i}(\boldsymbol{y},\phi,c_{r_i})$，$i=1,2,\cdots$ 为滤波器组在序列 R 下的复合系统，记为 $(\varphi_1,\varphi_2,\cdots,\varphi_k,R)$，其中 R 称为复合序列，$y_i=\varphi_q(\boldsymbol{y},\phi,c_q)$，$q=1,2,\cdots,k$ 为子系统.

复合系统的动力行为与复合序列 R 有关，若从某个 i 开始，r_i 为常数，则复合系统蜕化为单一滤波器系统.一般地，复合系统保持了所有子系统的特性，比单个子系统的行为要复杂得多.

引理 1 $\forall z,\phi$ 为统计独立，且 $z\in[-1,1)$ 为均匀分布，则对所有的 $c\in Z\setminus\{0\}$，$w=\mathrm{mod}(c\cdot z+\phi)\in[-1,1)$ 为均匀分布且与 ϕ 的分布无关.

证明 z 为均匀分布，不妨设 z 的分布密度为

$$f_z(z)=\begin{cases}2^{-1}, & z\in I,\\ 0, & z\notin I,\end{cases}$$

因为

$$w=\mathrm{mod}(c\cdot z+\phi)=\mathrm{mod}(\mathrm{mod}(c\cdot z)+\phi),$$

由(2)式的性质易知，

$$\forall c\in Z\setminus\{0\},$$

$$f_{\mathrm{mod}(c\cdot z)}(z)=f_z(z)=\begin{cases}2^{-1}, & z\in I,\\ 0, & z\notin I,\end{cases}$$

因此有

$$\begin{aligned}f_w(w)&=\mathrm{mod}(f_z(z)*f_\phi(\phi))\\&=\sum_{j=-\infty}^{\infty}\int_{-\infty}^{\infty}f_z(z)\cdot f_\phi(w+2j-z)\mathrm{d}z\\&=\begin{cases}2^{-1}, & z\in I,\\ 0, & z\notin I.\end{cases}\end{aligned}\tag{5}$$

证毕.

引理 2 记 c_{qn} 为第 q 个系数组的 c_n 系数，若 $c_{qn}\in Z\setminus\{0\}$ 且 $h(\cdot)$ 服从均匀分布，则第 q 个子系统 $y_i=\varphi_q(\boldsymbol{y},\phi,c_q)$ 服从 n 维均匀分布且与 ϕ 无关.

文献[13]对此进行了证明，此处略.利用文献[13]的结果，q 子系统的分布密度为

$$\begin{aligned}&f_q(\boldsymbol{y},\phi,c_q,t)\\&=f_q(\boldsymbol{y},c_q)\\&=\begin{cases}2^{-n}, & \boldsymbol{y}\in I^n,\\ 0, & \boldsymbol{y}\notin I^n.\end{cases}\end{aligned}\tag{6}$$

推论 1 若 $y_i=\varphi_q(\boldsymbol{y},\phi,c_q)$ 子系统服从 n 维均匀分布，则 $y_i=\varphi_{r_i}(\boldsymbol{y},\phi,c_{r_i})$ 复合系统亦服从 n 维均匀分布.

证明 图 2 中，滤波器每迭代一次，就切换一组系数，同时本次的输出作为下次的输入，但 $\mathrm{mod}(\cdot)$，$h(\cdot)$ 不变.不失一般性，记第 i 次迭代时的滤波器状态为 $\boldsymbol{y}^{(i)}=\{y_i,y_{i-1},y_{i-2},\cdots,y_{i-n}\}$，使用的系数组为 c_{r_i}.

(1)由 Kelber 条件知，第 i 次的输出服从 n 维均匀分布.由引理 1 知，c_{r_i} 的改变不影响 $\mathrm{mod}(\cdot)$，$h(\cdot)$ 的分布特性，再结合引理 2 易知，在 $c_{r_{i+1}}$ 作用下第 $i+1$ 次的状态 $\boldsymbol{y}^{(i+1)}=\{y_{i+1},y_i,y_{i-1},\cdots,y_{i-n+1}\}$ 也服从 n 维均匀分布.

(2)假设在 $c_{r_{i+k}}$ 作用下，第 $i+k$ 次的输出服从 n 维均匀分布，则由文中的复合迭代规则和以上所述可知，在 $c_{r_{i+k+1}}$ 作用下，第 $i+k+1$ 次的输出亦服从 n 维均匀分布.证毕.

引理 3 对于服从 n 维均匀分布的 q 子系统 $y_i=\varphi_q(\boldsymbol{y},\phi,c_q)$，若 $|c_{qn}|>1$ 且子系统的特征值 $|\lambda_i|\neq1$，则 q 子系统是遍历的.

证明 文献[13]从拓扑条件和特征值条件出发，得出了状态子空间 I_i 上的分布密度通过迭代后将弥散到整个系统的状态空间 I^n 上，

$$\begin{aligned}&f_q(\boldsymbol{y},t+1)\\&=\frac{1}{c_{qn}}\sum_{lk\in S(\boldsymbol{y},\prod\varphi)}\frac{1}{|m_k|}\cdot f_q(A_{lk}^{-1}(\boldsymbol{y}-b_{lk}),t).\end{aligned}\tag{7}$$

设子系统的初始分布为任意分布

$$f(\boldsymbol{y},0)=\begin{cases}f_0(\boldsymbol{y})>0, & \boldsymbol{y}\in I_0\subseteq I^n,\\ 0, & \text{其他},\end{cases}$$

则在(7)式中，随着 t 的增加，分布密度将趋于均匀不变分布[13]，

$$\begin{aligned}&\lim_{t\to\infty}f_q(\boldsymbol{y},t)\\&=\lim_{t\to\infty}P^tf(\boldsymbol{y},0)\\&=\begin{cases}2^{-n}, & \boldsymbol{y}\in I^n,\\ 0, & \boldsymbol{y}\notin I^n,\end{cases}\end{aligned}\tag{8}$$

式中 P 为 Frobenius-Perron-Operator(FPO)积分算子[14].证毕.

推论 2　若 $y_i=\varphi_q(\boldsymbol{y},\phi,c_q)$ 子系统是遍历的,则 $y_i=\varphi_{r_i}(\boldsymbol{y},\phi,c_{r_i})$ 复合系统也是遍历的.

证明　设 $N(q)$ 表示复合序列 $R=\{r_i\}$ 前 N 个元素中 q 的个数,则复合系统迭代过程中 q 子系统的使用次数为 $N(q)$,并记 $\alpha(q)=\lim\limits_{N\to\infty}\dfrac{N(q)}{N}\cong P(r_i=q)$.

为了方便,复合系统迭代过程统一表示为 $\boldsymbol{y}_i=F(\boldsymbol{y}_{i-1},t)$,那么对任意的 $\boldsymbol{y}\subseteq I_{lk}$,有

$$\begin{aligned}&P(F(\boldsymbol{y},t)\subseteq I_{lk})\\&=\sum_{r=0}^{k}P(\varphi_q(\boldsymbol{y},t)\subseteq I_{lk},q=r)\\&=\sum_{r=0}^{k}P(\varphi_r(\boldsymbol{y},t)\subseteq I_{lk})\cdot\alpha(r)\\&=\sum_{r=0}^{k}\alpha(r)\int\limits_{\varphi_r^{-1}(\boldsymbol{y},t)\subseteq I_{lk}}f_r(\boldsymbol{y},t)\,\mathrm{d}t,\end{aligned}$$

则复合系统分布密度为

$$\begin{aligned}f(\boldsymbol{y},t)&=\frac{\mathrm{d}}{\mathrm{d}\boldsymbol{y}}P(F(\boldsymbol{y},t)\subseteq I_{lk})\\&=\sum_{r=0}^{k}\alpha(r)\frac{\mathrm{d}}{\mathrm{d}\boldsymbol{y}}\int\limits_{\varphi_r^{-1}(\boldsymbol{y},t)\subseteq I_{lk}}f_r(\boldsymbol{y},t)\,\mathrm{d}t.\end{aligned}$$

由文献[15]有

$$\begin{aligned}f(\boldsymbol{y},t)=&\alpha(0)f_0(\boldsymbol{y},t)+\alpha(1)f_1(\boldsymbol{y},t)+\cdots\\&+\alpha(k)f_k(\boldsymbol{y},t).\end{aligned}\tag{9}$$

由 $\alpha(q)$ 的定义知 $\sum\limits_{q=0}^{k}\alpha(q)=1$,由引理 3 知,各子系统为遍历的且均服从不变分布密度(8)式,则(9)式复合系统的不变分布密度为

$$\begin{aligned}&\lim_{t\to\infty}f(\boldsymbol{y},t)\\&=[\alpha(0)+\alpha(1)+\cdots+\alpha(k)]\cdot\lim_{t\to\infty}f_r(\boldsymbol{y},t)\\&=\lim_{t\to\infty}P^tf_r(\boldsymbol{y},0)\\&=\begin{cases}2^{-n}, & \boldsymbol{y}\in I^n,\\0, & \boldsymbol{y}\notin I^n.\end{cases}\end{aligned}$$

证毕.

从推论 1 和推论 2 可知,满足 Kelber 条件的 k 个数字滤波器构成的复合滤波器也具有 n 维均匀分布和遍历的特性.因此采用参数切换形成的复合系统能够很好地解决固定系数结构中的三个主要缺陷,其保密性能明显优于单峰映射混沌系统和固定参数的高维混沌系统.在已知终值的情况下,初值分布的概率比较均匀,只能以穷举方法搜索初值,保证了不可逆性和防伪造性,可以用来构造性能优良的 Hash 函数.

3. 基于复合非线性数字滤波器的 K-HF 构造

Berson 等[16]首先提出了 K-HF 安全性的一般要求,并给出了一个 K-HF 强定义.Bakhtiari 等[17]认为该定义太强,并给出了一个可满足实际需要的安全性要求.

(1)给定一组 M 及对应的 $h_k(M)$,求出其他消息 M' 的 $h_k(M')$ 或其他摘要 $h_k(M')$ 的消息 M' 是困难的,则称 $h_k(\cdot)$ 为安全的带密钥的单向 Hash 函数.

(2)秘密密钥的长度应不小于 128 b,以防止密钥穷尽搜索攻击.

(3)消息的 Hash 值长度也应不小于 128 b,以防止生日攻击.

(4)Hash 值均匀分布,以抵御统计分析.

K-HF 一般由已有的加密算法或 Hash 函数构造,本文则采用图 2 中的复合系统 $(\varphi_1,\varphi_2,\cdots,\varphi_k,R)$ 来构造 K-HF,使其满足安全性要求(1)—(4).

不失一般性,记 Hash 值的长度为 L,为抵御生日攻击,要求 $L\geqslant 128$ b.假设初始消息为 M',用零填充后的消息为 M,并使 M 的长度满足 $|M|=(||M'|/L|+1)L\cong sL$.将 M 按长度 L 分组,记为 $M=(M_1,M_2,\cdots,M_s)$,其中 $M_i=m_i^1m_i^2m_i^3\cdots m_i^L$.为方便起见,定义 $\boldsymbol{y}^{(i)}=(y_i,y_{i-1},\cdots,y_{i-n})$ 为第 i 次迭代后滤波器状态,其中 $y_0^{(i)}=y_i$ 表示第 i 次迭代的滤波器输出.

算法的基本思想是:按照 Kelber 条件,预先在参数库中放置 k 个滤波器系数组 $C=(c_{i1},c_{i2},\cdots,c_{in})$,$i=1,2,\cdots,k$.取长为 L 的零序列 H_0 为初始向量,给定密钥 $\mathrm{SK}=\{\phi_0,\boldsymbol{y}^{(0)}\}$,$\phi_0$ 为滤波器的初始输入值,$\boldsymbol{y}^{(0)}$ 为滤波器的初始状态.以 $H_0\oplus M_1=\{r_1,r_2,\cdots,r_L\}$ 作为复合序列 R,通过 R 选择系数组 C,迭代得到复合系统的输出轨迹 $\{y_0^{(i)}\}=\{y_i\}_1^L$,量化为二进制序列作为 M_1 的 Hash 值 H_1.然后再以 $\phi_1=y_L$ 作为迭代的初始值,以 $\boldsymbol{y}^{(L)}$ 作为初始状态,以 $H_1\oplus M_2$ 作为复合序列,得到 H_2.重复上述过程至消息结束,得到 M 的 Hash 值 H_s.Hash 算法的一般过程可描述为(图 3)

$$(H_i,\phi_i)=F(\phi_{i-1},H_{i-1}\oplus M_i),\quad i=1,2,\cdots,s,$$
$$H(M)=H_s. \tag{10}$$

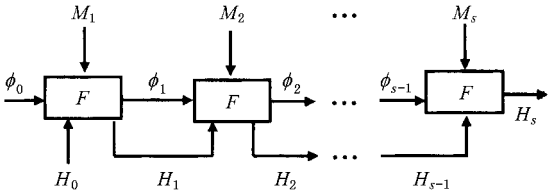

图 3 Hash 算法结构图

在这个过程中,需要一个将滤波器输出轨迹量化为 0,1 序列的变换,定义这个变换为

$$T_n(x)=\begin{cases}1, & x\in\bigcup_{d=0}^{2^{n-1}-1}I_{2d},\\ 0, & x\in\bigcup_{d=0}^{2^{n-1}-1}I_{2d+1},\end{cases}$$

其中 $n>0$ 为任意正整数,$I_0,I_1,\cdots$ 为 $[-1,1]$ 上的 2^n 个等分区间.(10)式和图 3 中的 $F(\cdot,\cdot)$ 表示求某一消息块的 Hash 过程,也就是复合数字滤波器的迭代过程.

1)从 $i=1$ 到 $i=s$,执行以下步骤:

(ⅰ) $\phi=\phi_{i-1}$.

(ⅱ) $q=H_{i-1}^j\oplus m_i^j$,
$y^{(j)}=\varphi_q(y^{(j-1)},\phi,c_q)$,
$\phi=y_0^{(j)}$,
$H_i^j=T_n(y_0^{(j)})$,
$j=1,2,\cdots,L$.

(ⅲ) $H_i=H_i^1H_i^2\cdots H_i^L$,
$\phi_i=y_0^{(L)}$.

2)输出 $H(M)=H_s=H_s^1H_s^2\cdots H_s^L$.

由于滤波器能产生高维混沌序列,因此复合系统对初始状态的敏感性和迭代过程的随机性,使得 Hash 结果与消息有着复杂而敏感的非线性关系,而且最后的 L 次迭代,使得最终 Hash 值的每比特都与消息 M 的所有比特有关,M 的任何微小变化都将引起 Hash 值的极大变化.若密钥 $\mathrm{SK}=\{\phi_0,y^{(0)}\}$ 在精度允许范围内发生微小改变,复合系统的迭代过程将使差异不断放大,经过第一轮的迭代就可使差异大到足以影响 Hash 结果,最终得到完全不同的 Hash 值.从上述算法的描述可知,基于复合滤波器的 K-HF 的安全性完全依赖于密钥 SK,即迭代初始值,这也符合算法公开、密钥保密的设计思想.

4. 实验仿真

加密体制中要求充分且均匀地利用密文空间,Hash 函数同样如此.因此理想 Hash 的散布效果应该是初值的细微变化将导致结果的每比特都以 50% 的概率变化[4].考察算法在明文发生 1 b 变化的情况下,引起 Hash 结果的变化比特数 B,定义最小变化比特数

$$B_{\min}=\min(B_i),$$

最大变化比特数

$$B_{\max}=\max(B_i),$$

平均变化比特数

$$\bar{B}=\frac{1}{N}\sum_{i=1}^{N}B_i,$$

平均变化概率

$$P=(\bar{B}/128)\times 100\%,$$

B 的均方差

$$\Delta B=\sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(B_i-\bar{B})^2},$$

P 的均方差

$$\Delta P=\sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(B_i/128-P)^2}\times 100\%,$$

其中 N 为统计总次数,B_i 为第 i 次测试时 Hash 结果的变化比特数.

仿真时采用一个二阶滤波器,参数库预存两组系数 $\{c_1=[3.57,4],c_2=[5.7,7]\}$,分段线性映射为

$$h(w)=\begin{cases}-1+2(w-d_i)/(d_{i+1}-d_i), & w\in(d_i,d_{i+1}),\\ 1, & w=1,\\ h(-w), & w\in[-1,0),\end{cases}$$

$L=128$,滤波器的初始值即密钥

$$\mathrm{SK}=\{\phi_0,y^{(0)}\}=\{\phi,y_1,y_2\}$$
$$=\{\phi=0.5648,\ y_1=-0.564,\ y_2=0.679\}.$$

测试方法为 在明文空间中随机选取一段明文进行 Hash,然后任意改变明文 1 b 后得到另一 Hash 结果,比较两个结果得到变化比特数 B.在 $N=1024$ 次测试下得到的变化比特数分布情况如图 4 所示.

图 4 表明,1024 次测试下的 $\bar{B}=63.6748$,非常接近理想状况下的 64.另外,B_i 的最小值为 47,最大值为 82,且集中在理想值 64 附近,表明算法对明文的敏感性强而稳定.

经 $N=128,256,512,1024,2048$ 次测试,得到 $B_{\min},B_{\max},\bar{B},\Delta B,P,\Delta P$ 的统计值如表 1 所示.

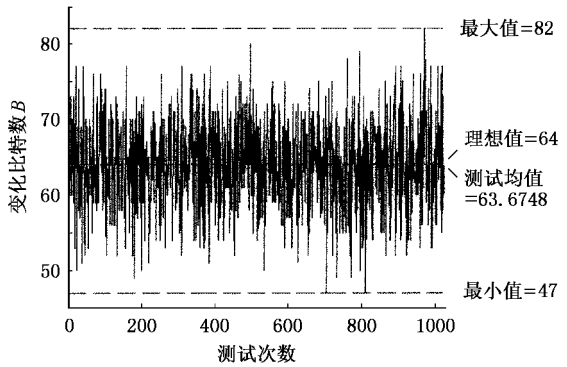

图 4　1024 次测试的变化比特数分布

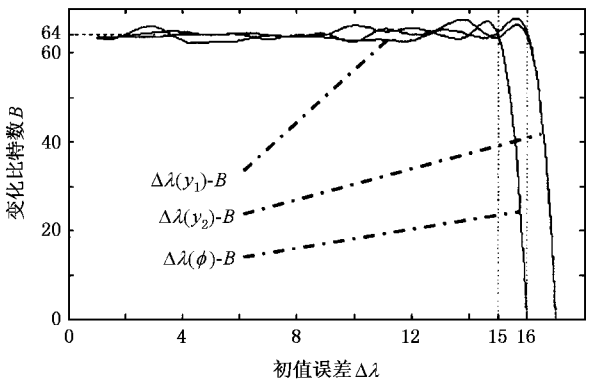

图 5　$\Delta\lambda$-B 曲线

表 1　Hash 性能统计

测试次数 N	B_{min}	B_{max}	$\bar{B}$	ΔB	P/%	ΔP/%
128	50	75	63.0469	5.6894	49.26	4.98
256	47	85	63.8672	5.9611	49.90	4.66
512	47	83	63.4355	5.6120	49.56	4.38
1024	47	82	63.6748	5.6257	49.75	4.40
2048	47	82	63.8170	5.7021	49.86	4.45

由表 1 数据可知,该算法的 $\bar{B}$ 和 P 都非常接近理想状况下的 64 和 50% 的变化概率,这相当充分和均匀地利用了密文空间.从统计效果看,攻击者在已知一些明文密文对,对其伪造或反推其他明文密文对没有任何帮助,因为明文的任何细微变化,从统计上看密文在密文空间中都是接近等密度的均匀分布,从而得不到任何密文分布的有用信息.而 ΔB, ΔP 标志着 Hash 混乱与散布性质的稳定性,越接近零就越稳定,文中算法的 ΔB,ΔP 都已很小,说明算法对明文的混乱与散布能力强而稳定.

另外,为了考察密钥 $SK=\{\phi_0,\boldsymbol{y}^{(0)}\}=\{\phi,y_1,y_2\}$ 对 Hash 结果的影响,定义 $\Delta\lambda$ 为 ϕ,y_1 或 y_2 的微小变化量,B 为对应的 Hash 变化比特数.测试时 ϕ,y_1 或 y_2 各自按 10^{-1} 的速率递减,考察相应变化量下 B 的大小.在赛扬 4,Windows2000 环境下,测试的 $\Delta\lambda$-B 关系如图 5 所示.

图 5 中,横坐标为 ϕ,y_1 或 y_2 变化量的负对数表示,纵坐标为相应的 Hash 值比特变化数 B.测试结果表明,当 $\Delta\lambda(\phi)=10^{-15}$ 时,$B\approx64$;当 $\Delta\lambda(\phi)=10^{-16}$ 时,$B\approx0$.因此算法对输入初始值 ϕ 的敏感度为 10^{-15} 数量级.同理,算法对初始状态 y_1,y_2 的敏感度为 10^{-16} 数量级.这说明算法对密钥高度敏感,在[-1,1]的实数范围内,密钥空间是很大的.

与文献[5]算法相比,本文算法具有如下特点:(1)文献[5]中要求混沌映射是互补的,而满足此关系的混沌源相对较少,对于性能优良的混沌源,一般情况下其混沌方程较复杂,难以找到互补的对等方程 ;若使用滤波器,只需选取满足 Kelber 条件的系数即可,而这种系数的选取容易且数量多,若要提高序列的复杂度,只需增加滤波器的阶数即可.(2)文献[5]中只能对明文逐比特运算,没有扩展能力.基于滤波器结构下,若提供 2^n 组系数,可由复合序列 R 中的 n b 子序列完成系数组 C 的选择,从而对明文按 n b 运算,成倍提高运算速度.(3)即使两种算法都是逐比特运算,文献[5]的迭代次数为 $2\cdot L\cdot(s-1)$,而本文算法为 $L\cdot s$ 次,当明文较大时(即 s 较大),其迭代次数只有文献[5]中的一半,考虑运算的复杂度不同,本文算法具有更快的运算速度.(4)基于滤波器结构的算法实现简单快捷,没有复杂的浮点运算,比基于其他混沌模型的算法更易于扩充和硬件实现.

5. 结　语

本文提出了一种基于复合非线性数字滤波器的单向 Hash 函数算法.研究结果表明 :(1)算法充分利用了混沌系统对初始条件敏感和迭代过程的单向性,以及由明文产生的复合序列对迭代子系统选择的随机性,使得 Hash 结果的每比特都与明文及密钥有着敏感、复杂的非线性强耦合关系,可以有效抵抗线性分析.(2)密钥 SK 在精度允许范围内(10^{-15}—10^{-16})发生微小改变,将导致 Hash 结果有近一半的比特位发生变化,对同一明文用不同的密钥,将得到完全不同的 Hash 值.由于具有很大的密钥空间,可以抵抗密钥的强力攻击.(3)复合滤波器

产生的混沌序列周期长且满足 n 维均匀分布，通过明文的调制和轨迹的粗粒化量化，使得 Hash 结果在散列空间中均匀分布，可以抵抗统计攻击.此外，基于滤波器的算法实现简单，比基于其他混沌模型的算法更易于扩充和软硬件实现，具有成为一种快速实用的单向 Hash 算法的潜力.

[1] Kou W D 1997 *Network Security and Standards* (Boston: Kluwer Academic)

[2] Pieprzyh J, Sadeghiyan B 1993 *Design of Hashing Algorithm* (Berlin: Springer)

[3] Liu J N, Xie J C, Wang P 2000 *J. Tsinghua Univ.* (Sci. Tech.) **40** 55 (in Chinese) [刘军宁、谢杰成、王 普 2000 清华大学学报(自然科学版) **40** 55]

[4] Wang X M, Zhang J S 2003 *Acta Phys. Sin.* **52** 2737 (in Chinese) [王小敏、张家树 2003 物理学报 **52** 2737]

[5] Li H D, Feng D G 2003 *Chin. J. Comput.* **26** 460 (in Chinese) [李红达、冯登国 2003 计算机学报 **26** 460]

[6] Zhang J S, Xiao X C 2000 *Acta Phys. Sin.* **49** 1221 (in Chinese) [张家树、肖先赐 2000 物理学报 **49** 1221]

[7] Zhang J S, Xiao X C 2001 *Acta Phys. Sin.* **50** 2121 (in Chinese) [张家树、肖先赐 2001 物理学报 **50** 2121]

[8] Frey D R 1993 *IEEE Trans. Circuits Syst.* Ⅱ **40** 660

[9] Zhang J S, Xiao X C 2001 *Chin. Phys. Lett.* **18** 337

[10] Zhou H, Ling X T 1997 *Acta Elec. Sin.* **25** 95 (in Chinese) [周 红、凌燮亭 1997 电子学报 **25** 95]

[11] Lin T, Chua L O 1993 *Int. J. Circ. Theory* **21** 473

[12] Gotz M, Kelber K, Schwarz W 1997 *IEEE Trans. Circuits Syst.* Ⅰ **44** 963

[13] Kelber K 2000 *IEEE Trans. Circuits Syst.* Ⅰ **47** 1413

[14] Lasota A, Mackey M C 1994 *Chaos, Fractals and Noise-Stochastic Aspects of Dynamics* (New York: Springer)

[15] Baranovsky A, Daems D 1995 *Int. J. Bifur. Chaos* **5** 1585

[16] Berson T A, Gong L 1993 *Secure, Keyed and Collisionful Hash Functions* (California: SRI International Lab.)

[17] Bakhtiari S, Safavi N R, Pieprzy K J 1996 *Lecture Notes in Computer Science* (Berlin: Springer) p201

Keyed Hash function based on composite nonlinear autoregressive filter*

Wang Xiao-Min[1)] Zhang Jia-Shu[1)] Zhang Wen-Fang[2)]

1)(*Key Laboratory of Signal and Information Processing of Sichuan Province, Southwest Jiaotong University, Chengdu* 610031, *China*)

2)(*Computer Security and Communication Secrecy Institute, Southwest Jiaotong University, Chengdu* 610031, *China*)

(Received 6 January 2005; revised manuscript received 4 July 2005)

Abstract

In recent years, chaotic dynamic systems are widely applied in information security because of its characteristic that the trajectory is sensitive to initial conditions and seems to be random though it is really a determinate process. The nonlinear autoregressive digital filters satisfying the Kelber conditions can also produce chaotic sequences just like chaotic maps do. The composite filter system consists of several above-mentioned filters, and the filter used for iteration is completely decided by a predetermined sequence called composite sequence. Consequently, the trajectory of the composite filter system is not only sensitive to its initial conditions, but also related with the composite sequence, which determines the choice of iterated sub-filter system in the iterating process. So the trajectory is more complex than that of general chaotic systems or single filter system. After analyzing some properties of composite filter systems such as N-dimensional uniform distribution and invariant distribution density function, a new keyed Hash algorithm based on composite system is presented. The approach selects the sub-filter system with the composite sequence obtained from message to be hashed, uses the initial iteration value of composite system as the secret key, and the coarse-grained trajectory as Hash value. Because of the sensitivity to initial value and randomicity of the iteration process, there is a very complex nonlinear relation between Hash value and the corresponding message and secret key, and then every bit of the Hash value derived from the message M is related with every bit of M. Furthermore, the filter-based algorithm is simple enough without complex operations, so it can be realized easily.

Keywords: Hash function, chaos, nonlinear autoregressive digital filter

PACC: 0545

* Project supported by the National Natural Science Foundation of China (Grant No. 60272096) and the Outstanding Young Researchers Foundation of Sichuan Province, China (Grant Nos. 03ZQ026-33, 51430804Q72201).