第 54 卷 第 9 期 2005 年 9 月
1000-3290/2005/54(09)/4006-06

物 理 学 报
ACTA PHYSICA SINICA

Vol.54,No.9,September,2005

基于时空混沌系统的单向 Hash 函数构造*

张 瀚† 王秀峰 李朝晖 刘大海

(南开大学信息技术科学学院,天津 300071)

(2004 年 8 月 26 日收到;2004 年 10 月 20 日收到修改稿)

提出了一种基于时空混沌系统的 Hash 函数构造方法.以线性变换后的消息数作为一组初值来驱动单向耦合映像格子的时空混沌系统,产生时空混沌序列,取其空间最后一组混沌序列的适当项,线性映射为 Hash 值要求的 128bit 值.研究结果表明,这种基于时空混沌系统的 Hash 函数具有很好的单向性、弱碰撞性、初值敏感性,较基于低维混沌映射的 Hash 函数具有更强的保密性能,且实现简单.

关键词:时空混沌, Hash 函数, 单向耦合映像格子

PACC:0545

1. 引 言

单向 Hash 函数在公钥密码技术、数字签名、完整性验证、身份认证和动态口令鉴别等安全技术中的广泛应用,引起了人们极大的研究兴趣[1].基于复杂度假设的传统的单向 Hash 方法,如 MD2,MD4,MD5,SHA[2]等,需要进行大量复杂的异或等逻辑运算或是分组多次迭代得到 Hash 结果,即便在被处理的消息很短时运算量都很大.对此,人们提出了基于混沌映射模型的单向 Hash 算法[3].但是,这些方案存在以下缺陷:1)这些系统是基于某一种低维混沌系统来构造的,研究发现利用自适应同步预测相空间重构等各种混沌预测技术[4—7]可以成功分析预测,保密性能堪忧;2)由于实际实现中,计算机的有限精度效应,混沌映射所产生的混沌序列也会退化为大周期序列[8].因此,增加混沌信号的复杂度和减小有限精度效应的影响是提高混沌 Hash 单向性、置乱性和弱碰撞性的关键.由于高维混沌信号具有更高的随机性,其中时空混沌模型在时间方向和空间方向上都具有混沌行为.它不仅具有初始条件敏感性,而且具有边界条件敏感性.特别是时空混沌模型中的耦合映像格子,模型简单,运动状态极其复杂,适合于构造 Hash 算法.

针对以上问题,本文提出了一种基于时空混沌系统的 Hash 函数构造方法.以线性变换后的消息数作为初值来驱动单向耦合映像格子的时空混沌系统,产生时空混沌序列,取其空间最后一组混沌序列的适当项,线性映射为 128bit 的 Hash 值.初步分析了利用时空混沌映射实现单向 Hash 函数的不可逆性、防伪造性、初值敏感性等特点,理论分析与计算机仿真结果表明:这种基于时空混沌系统的 Hash 函数具有很好的单向性、弱碰撞性,比基于低维混沌映射的 Hash 算法具有更高的复杂度与安全性,且实现简单.

2. Hash 函数特性与混沌的应用

单向函数的定义:映射 $H:X\to Y$ 对所有的 $x\in X$,$H(x)$ 容易计算,但其逆过程,给定 $H(x)$ 要求出 x 在计算上是困难的,该函数称为单向函数.单向 Hash 函数是一种特殊的单向函数,它满足以下 4 个条件[9]:

1)能杂凑任意长度的 0,1 序列,但输出是固定长度的 0,1 序列;

2)不可逆性,已知 $c=\mathrm{Hash}(m)$,求 m 计算困难,除穷举外没有好办法;

3)防伪造性,已知 $c=\mathrm{Hash}(m)$,求 n 使 $\mathrm{Hash}(n)=c$ 计算困难;

4)初值敏感性,$c=\mathrm{Hash}(m)$ 中 c 的每一 bit 都

* 天津市自然科学基金(批准号:023601411)资助的课题.

† E-mail:zhanghan@mail.nankai.edu.cn

与 m 的每一 bit 相关,并有高度的敏感性,即每改变 m 的 1bit,都将对 c 产生明显影响.

由于消息空间的无限与 Hash 结果空间的有限,会有许多消息具有同样的 Hash 函数值,不可避免地发生所谓的碰撞. Hash 结果达到一定长度,例如为固定的 128bit 时,结果空间已有 2^{128} 个,若碰撞程度很低,则以现有的计算能力在这样大的空间穷举计算是困难的.可见碰撞程度低是 Hash 的重要特性.

混沌系统天然的密码学性质使得密码学界对其深感兴趣,混沌运动是非线性确定性系统内在随机性的表现,可以由十分简单的确定系统产生非常复杂的随机行为.由于混沌系统在迭代中的信息损失,使得混沌序列的信息量渐进趋于零,因此对混沌序列进行正确的长期预测是不太可能.以上性质使混沌序列拥有单向 Hash 函数所要求的较好的不可逆性、防伪造性、初值敏感性[10].但由于计算机的有限精度效应会使混沌映射所产生的混沌序列会退化为大周期序列,而且对有限长度的消息所进行的迭代在实际中不会操作无限次的,因此在混沌系统中混沌序列包含的信息量并不会趋于零,同时由于低维混沌吸引子的复杂程度较低,可以利用各种混沌预测技术对其分析.本文通过单向耦合映射格子模型产生时空混沌序列,极大增强了混沌序列的复杂度和减小计算机有限精度效应的影响,理论分析与仿真结果表明,此算法能有效地提高混沌型 Hash 函数的性能而不明显增加运算复杂度.

3. 基于时空混沌系统的单向 Hash 构造

3.1. 时空混沌的单向耦合映像格子模型及其统计特性

耦合映像格子(CML)[11]模型自提出后,由于其数字实验的高效率,在时空混沌研究工作中备受青睐.其中的单向耦合映像格子系统(one-way coupled map lattice,OCML)[11]是考虑一个有限维格子 $S = \{1,2,\cdots,N\}$,具有周期边界条件.任一格子 $i \in N$ 在时刻 n 的状态变量为 $x_n(i)$,它的局部更新规则由下面映射给出:

$$x_{n+1}(i) = F(x_n(i-1), x_n(i)),\quad i \in S, n \in Z, \tag{1}$$

其中 F 是局部演化的非线性映射连续函数.上述模型满足如下周期边界条件

$$x_n(N+i) = x_n(i), \forall N \in Z. \tag{2}$$

本文应用的是一种格子映射为 Logistic 映射的单向耦合映像格子模型

$$x_{n+1}(i) = (1-\varepsilon) f(x_n(i)) + \varepsilon \{f[x_n(i-1)]\}, \tag{3}$$

式中,n 为离散时间坐标;i 为离散空间坐标,$i = 1, 2, \cdots, N$(N 为 OCOML 的长度);ε 为耦合系数,且满足 $0 < \varepsilon < 1$.非线性函数 f 为 Logistic 映射,即 $f(x) = \mu x(1-x)$.初始条件为 $[0,1]$ 内的随机数.当参数 $\mu = 4.0$ 时,单个格子处于混沌状态.

当系统满足一定的初始条件与参数条件后会呈现如图 1 所示的混沌行为($\varepsilon = 0.8$).满足一定条件的系统产生的序列 $x_n(i)$ 在时间上自相关函数衰减很快,若 $|j-i| \geqslant 2$,序列 $x_n(i)$ 与 $x_n(j)$ 的互相关函数亦呈快速衰减[12—14].由于这种系统产生的时空混沌序列 $\{x_n(l_0+i), i = 1,2,\cdots,N\}$ 在计算机的精度内是混沌的,而且非周期,相关函数衰减快速,故由时空混沌序列线性映射产生的整数序列在一个很大的结果空间上近似为等概率分布.

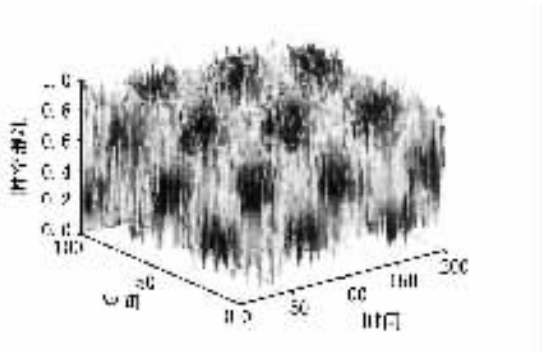

图 1　OCOML 的时空混沌图

3.2. 单向 Hash 构造

算法原理:将原始消息以字节为单位,经线性变换后作为输入的一组初值,利用时空混沌中的单向耦合映像格子映射迭代这组初值,取其空间最后一组混沌序列的三个适当项 $X_R(n)$,$X_{2R}(n)$,$X_{3R}(n)$,线性映射为 40bit,40bit,48bit 的三个大数,从而形成 128bit 的 Hash 值.

采用(3)式的单向耦合映像格子模型构造单向 Hash 的算法描述如下:

1)待处理消息按对应字节 $C_1 C_2 C_3 \cdots C_N$(C 为消息的的 ASCII 码)线性变换为 $[0,1]$ 范围内的数,整个消息变为一个数列,记为 $M_1 M_2 M_3 \cdots M_N$,其中数列个数即消息字节数为 N,计算公式如下:

$$M_1 = C_1/256, M_2 = C_2/256,$$
$$M_3 = C_3/256, \cdots, M_N = C_N/256.$$

2)令 $M_1 M_2 \cdots M_N$ 分别为 N 个格子的初值

$$X_0(1) = M_1, X_0(2) = M_2,$$
$$X_0(3) = M_3, \cdots, X_0(N) = M_N.$$

应用下述单向耦合映像格子模型迭代初值：

$$x_{n+1}(i) = (1-\varepsilon) f(x_n(i)) + \varepsilon \{f[x_n(i-1)]\},$$

其中 $f(x) = 4.0x(1-x)$，$\varepsilon = 0.8$，有 N 个格子，周期边界条件 $x_n(N+i) = x_n(i)$.

迭代生成时空混沌序列 N 组：

$$X_n(1), X_n(2), X_n(3), \cdots, X_n(N).$$

3)从迭代结果序列中取出最后一组序列的 $X_R(N), X_{2R}(N), X_{3R}(N)$，这里的 R 要远大于 N，将它们经线性变换和取整运算映射为两个 40bit，一个 48bit 的二进制数，合起来作为最后 128bit 的 Hash 结果.

3.3. 算法优点

由于时空混沌空间与时间的多自由度，只有当 N 个格子的输出在计算机有限精度的影响下同时退化为周期序列时，整个模型才会存在退化的周期序列，这样其输出时空混沌序列退化为周期序列的概率大大减小，因此该算法大幅度减弱了计算机有限精度的影响.高维数和高自由度的单向耦合映像格子模型映射使得在此应用混沌分析预测技术极为困难，预测低维混沌的成熟技术要想推广到时空混沌存在着很多难题.

时空耦合格子方程在格子间的映射是不可逆的，即已知第 $x_{n+1}(i)$ 项时，无法解析解出 $x_n(i)$，$x_n(i-1)$ 中的任何一项，这种性质保证了系统不可逆性和防伪造性.由图 1 可见，该时空混沌序列具有很好的噪声特性及其在时间与非相邻空间序列的相关函数快速衰减的性质[12] 适合用来构造单向 Hash 函数.其终值在结果空间呈不受迭代步数与初值影响的近似的等概率分布[12,14] .在已知 Hash 结果情况下，初值分布的概率比较均匀，只能以穷举方法搜索初值，因而保证了不可逆性和防伪造性.

同时算法具有混沌型 Hash 函数一般的优点，例如 Hash 结果的位数可以任意选取，而不是固定的 128bit，虽然时空混沌模型复杂于低维混沌，对于该算法的软件实现，因为有 N 组序列迭代，其运算量较之一维混沌型 Hash 函数至多增加 N 倍，并无显著提高.

4. 仿真结果

4.1. 文本 Hash 结果

初始文本 1 为“ In BECOMING AMERICA, Jon Butler synthesizes a generation of scholarship to produce a detailed exploration of the maturation of colonial North America after 1680. Despite its rural character and rudimentary technology, Butler asserts that eighteenth-century America was a modem place with a distinctive society ”，文本 2 将文本 1 中首字母的 I 改为小写，文本 3 将文本 1 中的 1680 改为 1681，文本 4 将 1 中的 Despite 写成 Despit，文本 5 将文本 1 的 character 写成 characters. Hash 结果用十六进制数表示，基于时空混沌映射算法的 Hash 结果分别为

文本 1 :BD4C C0A1AD0967A921D8523BEC34FE6；
文本 2 :A3F0B34F767B8C06B9A6E851D5E127ED；
文本 3 :FD741352B936837CE2C F4A 1074DB85AE；
文本 4 :16920C04E48B53C78F2 5A7EFD6A9B1D3；
文本 5 :1E6F13E484B5CA2BC56290A7D0839D7F7.

可见该算法的单向 Hash 性能是很好的，初值的微小扰动使得 Hash 结果以较大概率变化，具有高度的初值敏感性.

4.2. 混乱与扩散性质统计分析

在 Shannon 的信息论中提出了混乱与扩散的概念，加密体制中要求明文在密文空间中充分的扩散与混乱，Hash 函数也要求相应明文与所对应的 Hash 密文之间相关性很小. Hash 结果的二进制表示中每 bit 只取 0 或 1，因此理想的 Hash 函数应该是初值的扰动将导致 Hash 结果的每 bit 都以 50% 的概率变化.应用文献[15,16]中提到的如下定义来做混乱与扩散的统计分析.

考察算法在明文发生 1bit 变化的情况下，引起 Hash 密文结果的变化比特数定义平均变化比特数

$$\bar{B} = \frac{1}{N_t}\sum_{n=1}^{N_t} B_n, \tag{4}$$

记测试中 B_n 的最大值为 $B_{\max}$，B_n 的最小值为 $B_{\min}$.

定义平均变化概率

$$P = (\bar{B}/128) \times 100\%, \tag{5}$$

B 的均方差

$$\Delta B = \sqrt{\frac{1}{N_t - 1}\sum_{n-1}^{N_t}(B_n - \bar{B})^2}, \tag{6}$$

P 的均方差

$$\Delta P = \sqrt{\frac{1}{N_t - 1}\sum_{n=1}^{N_t}(B_n/128 - P)^2}, \quad (7)$$

其中 N_t 为统计次数，B_n 为第 n 次测试时结果的变化比特(bit)数，每次测试方法为 在明文空间中随机选取一段明文进行 Hash，然后改变明文 1bit 的值得到另一 Hash 结果，比较两个结果得到变化比特数 B_n.

在 $N = 1024$ 次测试所得置乱数分布情况如图 2 所示，明文 1bit 变化引起 128bit 的 Hash 值发生变化的 bit 数最小值为 47bit 和最大值为 79bit，平均 bit 变化数为 63.41 个，非常接近理想状况下的 64bit 变化数.其上下平均波动幅度很小，在 6bit 左右，最大波动幅度与最小波动幅度大致为 17bit，没有出现波动幅度十分剧烈的现象.

另 $N = 256, 512, 1024, 2048$ 次测试，得到多项指标 $\bar{B}$，P，ΔB，ΔP，B_{max}，B_{min} 的值，如表 1 所示.

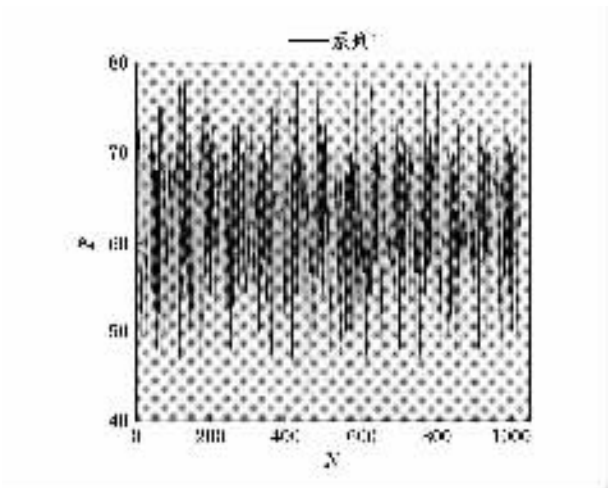

图 2　置乱数分布图

表 1　N 次测试的各项指标

$N=$	256	512	1024	2048	总平均
$\bar{B}$	63.35	64.62	63.41	64.43	63.95
ΔB	5.934	5.816	5.675	5.568	5.748
P/%	49.32	50.53	49.49	49.46	49.70
ΔP/%	5.013	4.927	4.684	4.506	4.783
B_{max}	69	73	78	85	76.25
B_{min}	51	48	47	44	47.5

由仿真结果可见，算法的平均变化比特数和每比特平均变化概率都趋近于理想状况下的 64bit 和 50%，可见时空混沌的单向耦合扩散机理使得算法充分均匀地混合了密文空间.攻击者在已知一些明文密文对时，若使用差分逼近方法对此算法亦不适用，因为明文的任何扰动，使得密文在统计上产生接近等概率的均匀分布.同时方差 ΔB 与 ΔP 都很小，这就保证了混乱与扩散程度聚集在一个平均稳定的水平上，这种分布的稳定性使得攻击者得不到有用的统计信息.

4.3. 算法的碰撞分析

碰撞指不同的初值 Hash 映射结果相同即多对一映射.应用文献[15,16]中提到的如下定义进行碰撞性分析.取初始文本为一字节，即 8bit，ASCII 码对应值为 0—255，Hash 结果取为 8bit，亦为 0—255 的整数，这样初值域与终值域相同.记终值域即像域中任一值对应初值域中原像的个数记为 k，记终值域中具有 k 个原像的点的个数记为 $n(k)$，$n(1)$越大，$n(0)$和其他各项越小，说明碰撞越少，混沌函数的散乱能力越强，用终值域与初值域的测度之比来定量衡量碰撞发生程度，令

$$P = [256 - n(0)]/256, \quad (8)$$

P 的值越接近 1，碰撞程度越低，等于 1 时，完全没有碰撞发生.由于在 128bit 等实际范围上进行碰撞分析，复杂度过大而不太可行，混沌映射构造 Hash 算法的特点可以将结果取为任意长度，从而在小范围 8bit 下进行算法碰撞程度的定量分析.

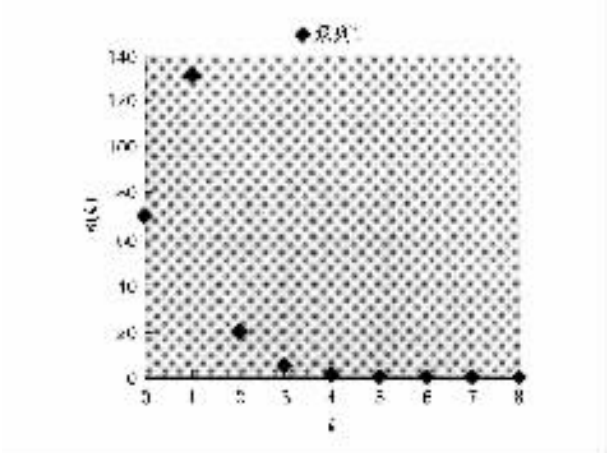

图 3　k-$n(k)$的分布图

经仿真，基于时空混沌系统的 Hash 算法碰撞值的分布为

$n(0) = 70$，$n(1) = 131$，$n(2) = 20$，$n(3) = 5$，$n(4) = 1$，$n(5) = 0$，$n(6) = 0$，$n(7) = 0$，$n(8) = 0$，$k > 8$，$n(k) = 0$，$P = 0.727$.图 3 为 k-$n(k)$的分布图.

由仿真结果知算法的碰撞程度较低,这是由时空混沌的类随机性与非相邻格子迭代序列的不相关性所产生的近似等概率分布结果决定的.

5. 结 论

本文提出了一种基于时空混沌系统的单向 Hash 函数构造方案 . 这种方案将待处理文本作为初值加载于高维数和高自由度的单向耦合映像格子模型的迭代中,仿真结果与理论分析表明:1)时空混沌丰富的态资源,使其避免了低维混沌构造 Hash 函数时单一混沌吸引子结构易于预测的缺点,使得该 Hash 算法具有很高的安全性.2)时空混沌系统确定性的模型使该算法实现简单,时空混沌系统对初值有高度的敏感性的特点,使得算法具有很好的单向 Hash 函数性能.3)由于时空耦合模型的关联结构及该算法的设计特点,从逻辑结构上看 N 个格子的关联迭代过程实为简单的算法同时处理所有消息,不同于传统 Hash 算法对消息分组后分别进行大量的多轮循环迭代与移位异或操作,该算法简洁,运算量少于如 MD5 与 SHA 等传统 Hash 算法.

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

[2] Knudsen L and Preneel B 2002 *IEEE Trans. Inform. Theor.* **48** 2524

[3] Heileman G L, Abdallah C, Hush D R *et al* 1993 *Proceedings of International Symposium on Nonlinear Theory and Its Applications* **1** 1183

[4] Short K M 1994 *Bifurc. Chaos* **4** 959

[5] Short K M 1994 *Bifurc. Chaos* **7** 1579

[6] Zhang J S and Xiao X C 2000 *China Phys.* **9** 408

[7] Zhang J S and Xiao X C 2000 *Chin. Phys. Lett.* **17** 88

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

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

[10] Frey D R 1993 *IEEE Trans. Circ. Syst.* Ⅱ **40** 660

[11] Wang S H, Kuang J Y, Li J H *et al* 2002 *Phys. Rev.* **66** 065202

[12] Xiao J H, Hu G and Qu Z 1996 *Phys. Rev. Lett.* **77** 4162

[13] Fang J and Jiang G P 2003 *J. Southeast University* (*Natural Science Edition*) **33** Supp. 78 (in Chinese) [房 建、蒋国平 2003 东南大学学报(自然科学版) **33** 增刊 78]

[14] Kuang J Y, Deng K and Huang R H 2001 *Acta Phys. Sin.* **50** 1856 (in Chinese) [匡锦瑜、邓 昆、黄荣怀 2001 物理学报 **50** 1856]

[15] Liu J N, Xie J C and Wang P 2000 *J. Tsinghua University* (*Natural Science Edition*) **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]

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

One way Hash function construction based on Spatiotemporal chaos*

Zhang Han　Wang Xiu-Feng　Li Zhao-Hui　Liu Da-Hai
(*College of Information Technical Science*, *Nankai University*, *Tianjin* 300071, *China*)
(Received 26 August 2004; revised manuscript received 20 October 2004)

Abstract

A Hash function construction method based on spatiotemporal chaos is proposed. We take the rumor after linearly transformed as a group of initial values and utilize the one-way coupled map lattice of spatiotemporal chaos to iterate the group of initial values concurrently. Then in spatiotemporal chaos convergence the proper items of the chaos convergence produced in the last space are linearly transformed into Hash value of 128 bits. The result obtained shows that the Hash function based on spatiotemporal chaos have advantages of irreversibility, weak collision and sensitivity to initial values. The method has a stronger secret performance than the Hash function based on low-dimensional chaos maps, and it is simple to be realized.

Keywords: spatiote mporal chaos, Hash function, one-way coupled map lattice
PACC: 0545

* Project supported by the Province Natural Science Foundation of Tianjin (Grant No. 023601411).