第 55 卷 第 11 期 2006 年 11 月
1000-3290/2006/55(11)/5688-06

物 理 学 报
ACTA PHYSICA SINICA

Vol.55,No.11,November,2006

基于混沌神经网络的单向 Hash 函数*

刘光杰† 单 梁 戴跃伟 孙金生 王执铨

(南京理工大学自动化学院,南京 210094)

(2006 年 1 月 10 日收到;2006 年 3 月 12 日收到修改稿)

提出了一种基于混沌神经网络的单向 Hash 函数,该方法通过使用以混沌分段线性函数作为输出函数的神经网络和基于时空混沌的密钥生成函数实现明文和密钥信息的混淆和扩散,并基于密码块连接模式实现对任意长度的明文序列产生 128 位的 Hash 值.理论分析和实验结果表明,提出的 Hash 函数可满足所要求的单向性,初值和密钥敏感性,抗碰撞性和实时性等要求.

关键词:混沌神经网络, Hash 函数, 分段线性混沌映射, 时空混沌

PACC:0545

1. 引 言

Hash 函数又称为单向散列函数,它在现代密码学中起着非常重要的作用[1].Hash 函数可以作为文件的唯一表示而用于内容标识和认证.随着互联网、电子商务以及数字化文档等应用的不断兴起和广泛应用,对 Hash 函数的要求越来越高,同时密码分析的手段也在不断的提高.传统的 Hash 函数如 MD2, MD4, MD5, SHA 等[1],需要进行大量复杂的异或和位操作运算,效率不高.最近,Wang 等人[2]成功破解了 MD5,SHA-1 等过去被认为足够安全的 Hash 函数,因此更为安全高效的 Hash 函数的研究是非常必要的.

混沌密码学是现代密码学中一个比较重要的分支,混沌由于其本身具有的丰富非线性复杂性,在混沌块密码[3]以及伪随机数发生器[4,5]的构造中都取得重要的应用.最近,基于混沌的 Hash 函数的研究也取得了一定的进展. Yi[6]提出了一种基于混沌帐篷映射的 Hash 函数构造方法;Xiao 等人给出了具有可变参数的分段线性混沌映的 Hash 函数构造方法[7],以及基于 Chebyshev 混沌映射的构造方法[8];王小敏等[9]也提出了一种基于广义混沌映射切换的单向 Hash 函数;李红达等人提出了基于符合混沌动力系统的构造方法[10];张瀚等人[11]提出一种基于时空混沌系统的 Hash 函数构造方法;彭飞等人[12]根据二维超混沌映射也设计了一种单向的 Hash 函数.

以上这些研究所采用的均为较简单的混沌动力系统,本文考虑混沌神经网络所具有的复杂非线性动力学特性,通过引入密码块连接模式,可产生任意长度明文序列的 128 位 Hash 值.神经网络本身所具有的混淆、扩散和压缩作用以及基于耦合映映像格子的密钥生成函数共同实现明文和密钥信息的混淆和扩散,理论和实验显示了该方法具有很好的密钥和明文敏感性,且具有一定的抗碰撞性.

2. 混沌神经网络用于 Hash 函数设计的可行性

单向 Hash 函数 $H(M)$作用于任意长度的明文消息 M,它返回一个具有固定长度的 Hash 值 h.即 Hash 函数应该有压缩特性以及单向特性,Hash 函数的单向性可描述为

1) 给定 M,很容易计算 h;

2) 给定 h,根据 $H(M) = h$ 计算 M 很难;

3) 给定 M,要找到另一消息 M'并满足 $H(M) = H(M')$很难.

此外,在一些其他的应用背景要求下,仅仅具有单向性是不够的,还需要抗碰撞性(collision-resistance),即要找出两个随机的消息 M 和 M'满足

* 国家自然科学基金(批准号:60374066);江苏省自然科学基金(批准号:BK2004132)资助的课题.

† E-mail:guangj-liu@yahoo.com.cn

$H(M) = H(M')$，在计算上很困难.

混沌神经网络较传统的混沌映射而言具有更为复杂的时空复杂度，其良好的混淆和扩展特性已被成功用于流密码[13]和块密码[14]的设计.且当神经网络的结构为多输入单输出时，它又具有很好的压缩特性.同时在混沌神经网络的内部参数确定的情况下，很容易根据输入计算输出，但由于混沌的初值和参数敏感性，根据输出计算输入却是非常困难的.混沌神经网络的不可逆性(单向性)、良好的混淆和扩散特性以及对密钥的敏感性使得其在理论上可以用于设计性能较好 Hash 函数.

3. 基于混沌神经网络的 Hash 函数

3.1. 混沌神经网络密码块结构

图 1 给出了用于实现单元 Hash 块的混沌神经网络结构，该网络具有输入和输出两层结构，输入层具有 8 个结点，输出层包含 4 个节点.输入层的每个节点都具有 4 个 8 比特数据的输入，通过每个节点上的固定权值 $W_1 = [1/2^8\ 1/2^{16}\ 1/2^{24}\ 1/2^{32}]$，输入层上节点可实现将 4 个字节的数据转化为 32 位的[0,1]之间的小数，并实现局部范围内的混淆和置乱.

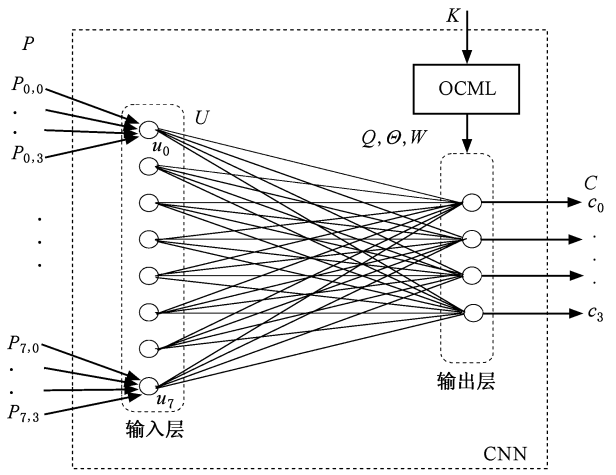

图 1　Hash 单元的混沌神经网络结构

输入层上神经元的传递函数

$$f(x,Q) = \begin{cases} x/Q, & 0 \leqslant x < Q, \\ (x-Q)/(0.5-Q), & Q \leqslant x < 0.5, \\ (1-Q-x)/(0.5-Q), & 0.5 \leqslant x < 1-Q, \\ (1-x)/Q, & 1-Q \leqslant x \leqslant 1 \end{cases} \tag{1}$$

为分段线性混沌映射(picewise linear chaotic map，PLCM)[15]，这里 Q 为满足 $Q \in (0,0.5)$ 的控制参数.

对输入 $P = [p_{0,0},\cdots,p_{0,3},p_{1,0},\cdots,p_{1,3},\cdots,p_{7,0},\cdots,p_{7,3}]^{\mathrm{T}}$，其中 $p_{i,j} \in \{0,1,\cdots,2^8-1\}$，可定义输出 $U = [u_0, u_1,\cdots, u_7]^{\mathrm{T}}$，$u_i \in [0,1]$.

$$u_i = f^{\tau}(W_1 \cdot [p_{i,0},p_{i,1},p_{i,2},p_{i,3}]^{\mathrm{T}},1/3),\quad i \in \{0,1,\cdots,7\}. \tag{2}$$

此处 PLCM 的控制参数取为 1/3，τ 表示映射 f 的迭代次数，为保证较好的置乱效果，本文取 $\tau = 40$.

输出层实现对输入数据的压缩，以及明文信息和密钥信息的在较大范围内的混淆和扩散.每个输出层神经元 i 连接着所有的输入层神经元.设从输入层神经元 j 到输出层神经元 i 之间连接的权值为 $W_2(i,j) \in (0,1)$，每个神经元 i 上的阈值为 $\Theta_i \in [0,1]$，其上 PLCM 传递函数的控制参数为 $Q_1 \in (0,0.5)$.参数 W_2,Θ,Q 均是由密钥 K 通过一个如(3)式的 4 维单向耦合映像格子(OCML)[16]产生.

$$\begin{aligned} x_1(i+1) &= (1-\varepsilon)g(x_1(i)) + \varepsilon g(x_4(i)), \\ x_2(i+1) &= (1-\varepsilon)g(x_2(i)) + \varepsilon g(x_1(i)), \\ x_3(i+1) &= (1-\varepsilon)g(x_3(i)) + \varepsilon g(x_2(i)), \\ x_4(i+1) &= (1-\varepsilon)g(x_4(i)) + \varepsilon g(x_3(i)), \end{aligned} \tag{3}$$

其中函数 g 为混沌 Logistic 映射：$x(i+1) = 4x(i)(1-x(i))$.本文取耦合系数 $\varepsilon = 1/3$.128 位密钥 K 首先按照先后顺序分解成 4 个 32 位的整数并通过除以 2^{32} 量化至[0,1]之间的小数：k_1，k_2，k_3，k_4.这四个小数作为系统初值进入如(3)式的 OCML 中进行迭代，连续取 10 个每隔 30 步迭代的系统状态值，其中前 8 个 4 维状态值作为权值 W_2，其后的 2 个状态值分别作为阈值矢量 Θ 和控制参数矢量 Q.

若设权值矩阵 W_2 的第 i 行为 W_2^i，对输入层的输出 U，输出 $C = [c_0, c_1, c_2, c_3]^{\mathrm{T}}$ 可记为

$$c_i = f^{\tau}(\mathrm{mod}(W_2^i \cdot U + \Theta_i,1),Q_i),\quad i \in \{0,1,\cdots,4\}. \tag{4}$$

3.2. 基于块连接的 Hash 函数构造

图 1 所示的 Hash 单元可将 256 位的明文数据映射为 128 位的 Hash 值，通过引入密码块链接模式(cipher block chaining，CBC)[1]，可对具任意长度的明文数据产生 128 位的 Hash 值.密码块链接模式由图 2 所示.

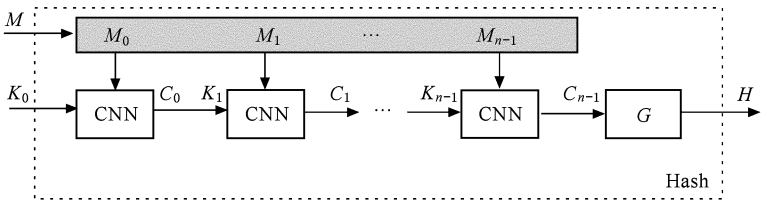

图 2 CBC Hash 函数模型

任意长的明文数据首先要进行位填充为 M，以保证 M 的长度为 256 的倍数．然后将 M 分成 n 个 256 位的子明文块，分别为 $M_0, M_1, \cdots, M_{n-1}$．其中第 i 个 CNN 的输出 C_i 可直接作为 $i+1$ 个 Hash 单元的密钥用来生成该 CNN 的权值、阈值和控制参数．整个基于 CBC 的 Hash 函数可描述为

$$C_i = \mathrm{CNN}(K_i, M_i),$$
$$H = G(C_{n-1}). \quad (5)$$

这里函数 G 实现将 C_{n-1} 的四个 32 位的小数通过位连接转换成 128 位的 Hash 值．

4. Hash 函数安全性与性能分析

4.1. 文本数据的 Hash 结果

根据本文提出的方法，选择密钥“1691AF0F13475A384CBCEA022ACF3F8A”，分别计算了下面五种情况下文本的 Hash 值：

1) Cryptographic hash functions play a fundamental role in modern cryptography. While related to conventional hash functions commonly used in non-cryptographic computer applications in both cases, larger domains are mapped to smaller ranges they differ in several important aspects. Our focus is restricted to cryptographic hash functions (hereafter, simply hash functions), and in particular to their use for data integrity and message authentication.

2) 将上述文字中的第一个“in”改为“on”.

3) 将“applications”改为“application”.

4) 在文本最后增加一个空格．

5) 将十六进制密钥“1691AF0F13475A384CBCEA022ACF3F8”，改为“1691AF0F13475A384CBCEA022ACF3F8”.

上面五种情况计算得到的 Hash 值如下：

1) 92169F4E53BE231608F5DF9DA8128BA4；

2) BB466E9D77B39E809E1CCAE138654862；

3) 18654E7143B361A955FAC9FB2C04B576；

4) C761966053EF1EA708C63F3CA83F587F；

5) 0BFF67CAEA315D599AD599B3378556C9.

从上面的仿真结果可以看出，本文的 Hash 函数具有很好的敏感，即使是很小的文本改变也会导致最终得到的 Hash 值发生很大的变化，此外 Hash 函数对密钥的变化也相当敏感．

4.2. 单向性分析

根据第 3 节的叙述可见，已知明文 M 和密钥 K，通过密码块连接方式计算 Hash 值是非常方便的．混沌神经网络中的传递函数 PLCM 和用于生成网络参数的 OCML 都是不可逆的混沌映射，因此根据最终的 Hash 值反过来计算明文 M 和密钥 K 则是非常困难的．从穷尽搜索攻击的角度来说，即使对仅有 10 比特的明文消息，由于使用 128 位的密钥，搜索也要在 2^{138} 的穷举空间中进行尝试，这在计算上是不可行的．

4.3. 明文和密钥敏感性分析

在本文算法中，神经网络的传递函数采用分段线性的混沌映射，网络参数则由一个 4 维的耦合映像格子通过迭代产生．通过将神经网络的复杂和压缩性与混沌的密钥和参数敏感性相结合，实现了密码编码所必须的混淆和扩散．这种良好的混淆和扩散作用，保证了 Hash 函数对统计攻击的安全性．理论上，系统的混淆和扩散特性越好，其相应的密钥和初值敏感性也越强．对一个二进制表示的 128 位 Hash 结果，其每个位置的值非 1 即 0，理想的敏感性应保证任何明文或者密钥的轻微改变将导致 Hash 比特发生 50% 的变化概率．

在明文敏感性实验中，对一个 1024 个比特的明文消息，每次改变其一位上的值，即将第 i 个比特的 0(1)改为 1(0)，计算改变后的明文消息的 Hash 值

h_i，然后将其和原始消息的 Hash 值 h_0 进行比较并计算 h_i 和 h_0 二进制表示的 Hamming 距离 $D(h_i, h_0)$，并最终得到 Hash 比特变化率

$$r(i) = \frac{D(h_0 h_i)}{128} \times 100\% . \quad (6)$$

图 3 给出了各个明文比特改变情况下，Hash 比特变化率的分布情况. 可以看到，比特变化率 r 非常接近理想的 50%（64 比特），这即是说，明文消息的任一微小变换都会导致 Hash 值发生较大的变化，进一步反映了所构造的 Hash 函数具有良好且稳定的明文敏感性. Hash 函数的这一性质保证了使得根据已知的明文-密文对很难伪造和推导出其他的明文-密文对非常困难，因此可以有效的抵抗所谓的选择明文攻击.

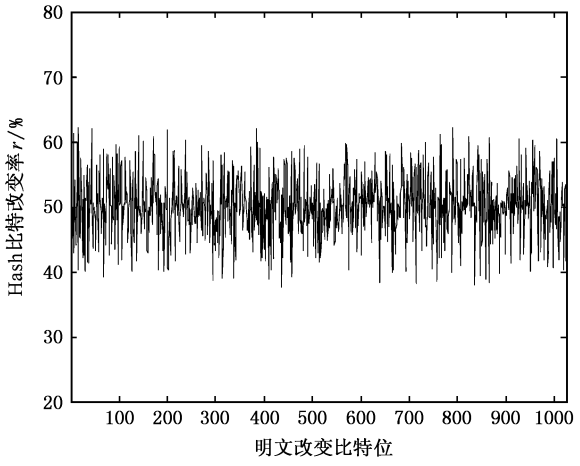

图 3　明文敏感性分析

混沌神经网络输出层的网络参数通过以 128 位密钥作为系统初值的时空混沌系统生成，因此密钥空间的大小为 2^{128}，这保证了密钥对任何的穷举攻击是安全的. 通过下面的实验可以看到 Hash 值对密钥同样具有很强的敏感性. 采用初始密钥“1691AF0F13475A384CBCEA022ACF3F8A”和如 4.1 节的 1）的文本明文，每次改变密钥中的一位，即将第 i 个比特的 0(1)改为 1 (0)，计算对应的 Hash 值，并计算如(6)式的 Hash 比特变化率. 图 4 给出了 Hash 比特变化率的分布情况.

从图 4 可以看到，Hash 值的比特变化率接近理想的 50%，因此具有较好的密钥敏感性保证了对统计分析的安全.

4.4. 抗生日攻击和碰撞攻击分析

对生日攻击而言，Hash 值的比特长度决定了密码系统的安全性. 对本文 Hash 函数而言，128 位 Hash 值长度意味着 2^{64} 的攻击难度. 这个数量级的攻击难度对一般应用来说是足够的. 由于本文神经网络结构的可扩展性，若增加输出层的节点的个数可得到更长的 Hash 值，能抵御更强的生日攻击.

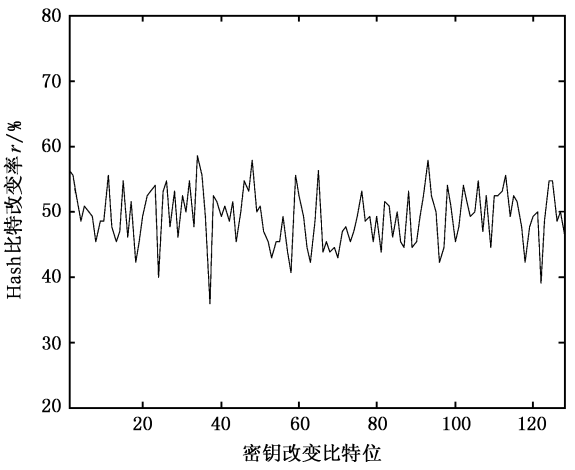

图 4　密钥敏感性分析

Hash 函数的抗碰撞性是指找到任意两个不同明文具有同样的 Hash 值在计算上是不可行的. 我们进行了如下的实验对本文提出的基于混沌神经网络的 Hash 函数（Chaotic Neural Network Hash，CNNH）的抗碰撞性进行了初步的测试. 选择 Hash 值的前 8 个比特作为 128 位 Hash 值的摘要比特 d，对应地取二进制明文消息的比特为 8，这样可使得 Hash 函数的原像空间等于像空间. 设像空间中具有 k 个原像点的像点数目为 $N(k)$. 从抗碰撞性的要求考虑，$N(1)$ 越大，发生碰撞的概率越小. 因此从 $N(k)$ 的分布情况，可观察到 Hash 函数的抗碰撞性能. 记 $n(k)$ 为

$$n(k) = \frac{N(k)}{\sum_{k=0}^{K} N(k)}, \quad (7)$$

其中，K 发生最大碰撞的数值，对一般 Hash 函数而言 K 不会超过 12，这里记 $K = 10$.

图 5(a)给出了 $d = 8$ 时 CNNH 的 $n(k)$ 分布的情况. 将同样的实验用于 MD5，SHA-1（基于 Java 的 Security.MessageDigest 函数），图 5(b)，(c)分别给出了 $d = 8$ 时 MD5 和 SHA-1 算法的 k-$n(k)$ 分布情况. 从图中可见，CNNH 的抗碰撞性好于传统的 MD5 和 SHA-1.

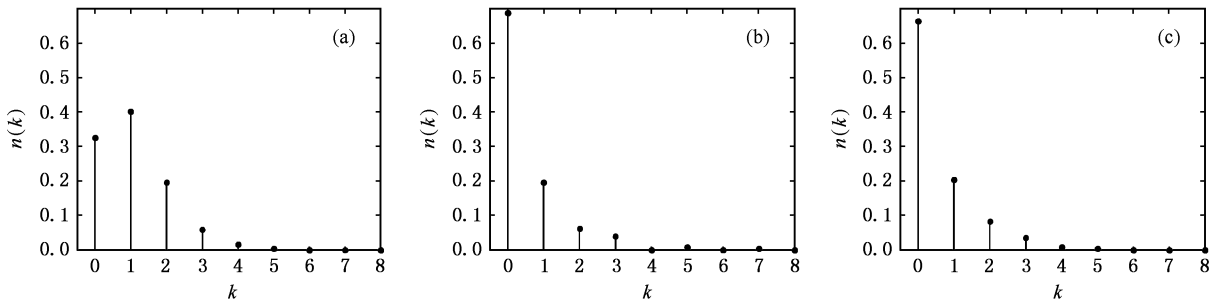

图 5 CNNH 与 MD5 和 SHA-1 的抗碰撞性对比 (a) CNNH;(b) MD5;(c) SHA-1

4.5. 时间性能分析

基于本文提出的算法,在 jdk1.4.2 上开发了本算法的 Java 类 cnnHash,在 Java 语言平台上同时实现文献[11]中基于时空混沌和文献[12]中基于二维超混沌的 Hash 算法的 Java 程序,将此三种算法同 jdk1.4.2 封装的 Security.MessageDigest 提供的 MD5,SHA-1 就算法的时间性能进行了比较.表 1 给出了在 P4 1.8GHz,256Mbit 内存环境下五种 Hash 算法计算长度为 100kbit 的明文序列所需时间.

表 1 几种 Hash 算法的时间性能比较

Hash 算法	本文算法	MD5	SHA-1	文献[11]算法	文献[12]算法
计算时间/ms	97	47	58	78	7.2×10^3
Hash 长度/比特	128	128	160	128	128

从表 1 可见,由于文献[11]中算法需要一个维数为 100k 的耦合映像格子的迭代因此计算速度较慢,本文算法和文献[12]中算法所使用时间大致相当,与传统 MD5,SHA-1 相比本文算法的时间性能相对较差.但神经网络计算的可并行化在多处理器环境下可进一步缩短计算的时间,因此本文算法仍具有时间效率上的优势.

5. 结 论

本文提出一种基于混沌神经网络的 Hash 函数构造方法.通过以混沌分段线性函数作为输出函数的神经网络和基于时空混沌的网络参数生成函数,并结合密码块连接方法,本文算法可实现对任意长的明文序列到 128 位 Hash 值的映射.由于混沌神经网络具有的复杂非线性行为,得到的 Hash 算法具有较好的明文和密钥敏感性,且实验表明本文提出的算法具有较好的抗碰撞性以及较好的时间性能.考虑到神经网络在输入层和输出层上的可扩展性,若增加网络输入和输出层的节点数,本文算法的 Hash 单元每次可处理更多的明文消息并提供更长的 Hash 长度.由于神经网络计算的可并行性,在多处理器的网络服务器端,本文提出的算法在时间性能上也具较大的提升空间.

[1] Vanstone S A, Menezes A J, Oorschot P C 1996 *Handbook of Applied Cryptography* (New York: CRC Press)

[2] Wang X Y, Yu H B 2005 *Lecture Notes in Computer Science* **3494** 19

[3] Pareek N K, Patidar V, Sud K K 2003 *Physics Letters* A **309** 1

[4] Sheng L Y, Cao L L, Sun K H, Wen J 2005 *Acta Phys. Sin.* **54** 4032[盛利元、曹莉凌、孙克辉、闻 姜 2005 物理学报 **54** 4032]

[5] Stojanovski T, Kocarev L 2001 *IEEE Trans. on Circuits and Systems* Ⅰ **48** 281

[6] Yi X 2005 *IEEE Transactions on Circuits and Systems* Ⅱ **52** 354

[7] Xiao D, Liao X F, Deng S J 2005 *Chaos Solitions & Fractals* **24** 65

[8] Xiao D, Liao X F, Tang G P, Li C D 2004 *Proceedings of International Symposium on Circuits and Systems* 11

[9] Wang X M, Zhang J S, 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]

[10] Li H D, Feng D G 2003 *Chinese Journal of Computers* **26** 460 (in Chinese)[李红达、冯登国 2003 计算机学报 **26** 460]

[11] Zhang H, Wang X F, Li C H, Liu D H 2005 *Acta Phys. Sin.* **54** 4006 (in Chinese)[张 瀚、王秀峰、李朝辉、刘大海 2005 物理学报 **54** 4006]

[12] Peng F, Qiu S S, Long M 2005 *Acta Phys. Sin.* **54** 4562 (in Chinese) [彭 飞、丘水生、龙 敏 2005 物理学报 **54** 4562]
[13] Yen J C, Guo J I 2002 *Pattern Recogition and Image Analysis* **12** 70
[14] Lian S G, Chen G R, Cheung A, Wang Z Q 2004 *Proceedings of International Symposium on Neural Network* 627
[15] Papadimitriou S, Bountis T, Mavroudi S, Bezerianos A 2001 International Journal on Bifurcation & Chaos **12** 3107
[16] Lv H P, Wang S H, Li X W, Tang G N, Kuang J Y, Ye W P, Hu G 2004 *Chinese Physics* **13** 625

One-way Hash function based on chaotic neural network*

Liu Guang-Jie Shan Liang Dai Yue-Wei Sun Jin-Sheng Wang Zhi-Quan
(*School of Automation, Nanjing University of Science & Technology, Nanjing* 210094, *China*)
(Received 10 January 2006; revised manuscript received 12 March 2006)

Abstract

In this paper, a new one-way Hash function is proposed based on chaotic neural network. With the neural network with piecewise linear chaotic map as the output function, the key generation function based on spatiotemporal chaotic system are used to realized the data confusion and diffusion. By the cipher block chaining mode, the proposed method can produce 128-bit Hash value for plaintext with arbitrary length. Theoretical analysis and experimental results indicate that the proposed Hash function satisfies the demands in performance, such as being one-way, having initial value and key sensitivity, collision resistance and real-time applicability.

Keywords: chaotic neural network, Hash function, piecewise linear chaotic map, spatiotemporal chaos
PACC: 0545

* Project supported by the National Natural Science Foundation of China (Grant No. 60374066), and the Natural Science Foundation of Jiangsu Province, China (Grant No. BK2004132).