文章编号:1000-582X(2012)04-107-05

层叠式并行图像 Hash 函数的结构及其算法实现

周 庆,张燕贞
(重庆大学 计算机学院,重庆 400044)

摘 要:传统 Hash 函数采用链式结构,不能充分利用图形和图像的二维特征来提高处理速度,更难以支持并行计算。为克服这 2 个缺点,提出了一种 Hash 函数结构,其在并行计算平台上的时间复杂度仅为 $o(\log n)$。分析了该结构相关的基本问题,并设计了在该结构下基于细胞神经网络实现的 Hash 函数。实验结果表明该 Hash 函数具有优异的敏感性、随机性和抗碰撞能力。

关键词:图像认证;Hash 函数;并行计算;细胞神经网络
中图分类号:TP309.7 **文献标志码**:A

Design and implementation of cascaded structure for parallel image hash functions

ZHOU Qing, *ZHANG Yan-zhen*
(College of Computer Science, Chongqing University, Chongqing 400044, P. R. China)

Abstract: The traditional Hash functions use a chain-like structure, which can not make best use of the 2D property of graphics or images. The chain-like structure is low in efficiency when implemented on a parallel computing platform. A new structure of Hash function is proposed to overcome these shortcomings and the time complexity is as low as $o(\log n)$ on a parallel computing platform. Some fundamental problems regarding the structure are analyzed. With this structure, a Hash function based on cellular neural network is proposed, which shows satisfactory randomness sensitivity to input and resistance to collision with simulation experiments.

Key words: image authentication; Hash function; parallel computing; cellular neural network

多媒体对象,特别是图形、图像和视频被广泛用于当今生活的各个领域。当多媒体数据在公开网络上传输时,其数据容易被攻击者篡改。由于图像、音频或视频等多媒体对象的数据量庞大,研究针对多媒体对象的 hash 函数具有重要的意义。

一个 Hash 函数 H 对任意长度的消息 M 进行处理,得到一个固定的、长度较小输出 h,称为 Hash 值。安全的 Hash 函数应具备 2 个特征,一是敏感性,即消息微小的变化导致 Hash 值的显著变化,二是随机性,即 Hash 值应呈现出理想的均匀分布。

目前使用最多的 Hash 函数主要是 MD5 和 SHA 系列函数[1]。然而,2005 年前后,王小云等人发现了这类函数的安全缺陷[2]。为了进一步提高 Hash 函数的安全性,一些新的 Hash 函数被提出。如基于混沌的 Hash 函数[3-11],以及美国 NIST 机构 2007 年开始提议的、预计 2012 年发布的 SHA－3

收稿日期:2011-12-01
基金项目:国家自然科学基金资助项目(61003246,61003256);重庆市自然科学基金重点资助项目(CSTC,2009BA2024);重庆市自然科学基金资助项目(CSTC,2009BB2208,2010BB2242)
作者简介:周庆(1979-),男,重庆大学副教授,主要从事信息安全、多媒体技术方向研究,(Tel)13752812642;(E-mail)tzhou@cqu.edu.cn。

算法[12]等。

然而，无论是广泛应用的MD5，SHA系列算法还是新的基于混沌的Hash算法，均采用链式结构，因此当它们用于图像时存在2个缺陷：一是难以利用图形或图像的空间特性来提高处理速度；二是不能有效地支持并行计算。提出了一种新的Hash函数结构，并给出一个用细胞神经网络实现的Hash算法。实验结果表明，提出的算法具有良好的敏感性、随机性和抗碰撞能力。

1 层叠式并行图像Hash函数的结构及其分析

传统的Hash函数，如MD5，SHA系列等算法采用链式结构对数据进行处理。消息首先被分成若干个大小相等的分组(M_1，M_2，…，M_n)，依次处理每个分组得到一个输出值，然后将当前分组的输出值与下一个分组数据共同计算得到下一分组的输出。链式结构将最后一个分组的输出值作为整个消息的Hash值。然而，链式结构的Hash函数用于数字图像时尚存在2个不足，一是没有充分利用图像的空间特性或二维特性，二是难以支持并行计算。

为了弥补传统Hash函数的用于二维图形或图像的不足，提出了一个新的Hash函数的结构(见图1)。

Step 1：将图像进行分块，每块大小为$M\times N$；

Step 2：并行地对每个$M\times N$的数据块进行处理，得到$(M/a)\times(N/b)$的输出；

Step 3：若图像数据块个数大于1，将各块的输出组合成新的图像，转Step 1；

Step4：将Step2的输出转换成一维数据，作为整个图像的Hash值。

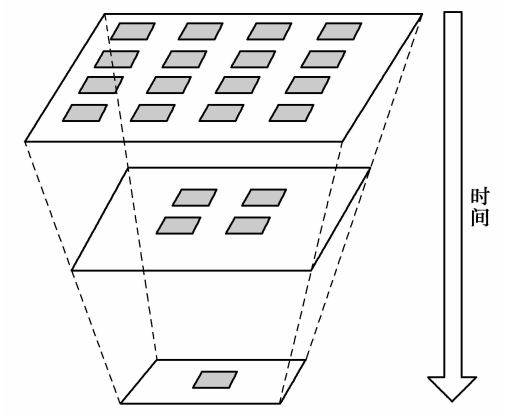

图1 层叠式并行Hash函数结构的处理方式

由图1和Step2可以知道，每次经Step 2处理后，图像大小减少ab倍，整个过程在外观上呈现一种层叠关系，又图像各块可并行处理，故称为“层叠式并行Hash函数”结构。

在传统Hash函数中，设每个分组(含m个像素)的处理时间为T_1，其时间复杂度等于

$$t_1 = T_1\frac{n}{m} = o(n)。 \tag{1}$$

层叠式并行Hash函数的时间复杂度主要由Step 2决定。假设图像含n个像素，由于每次经Step 2后图像的大小减少ab倍，Step 2运行的次数约为$\log_{ab}$n。假设$M\times N$数据块的处理时间为T_2，鉴于各块被并行运行，则Step 2的运行时间等于T_2，故该方法总的时间复杂度等于

$$t_2 = T_2\log_{ab}n = o(\log n)。 \tag{2}$$

因此，在时间复杂度上，层叠式并行Hash函数要远优于传统的Hash函数。

2 基于细胞神经网络的分块处理算法

分块处理算法是层叠式并行Hash函数设计的核心，要实现并行性、不可逆性、敏感性和随机性等多个目标。通过反复的比较和实验，发现细胞神经网络是实现以上目标的最佳选择之一。

细胞神经网络(cell neural network)是由Chua提出的一种局部互联的神经网络[13]，是一种能实时、高效地处理信号的大规模非线性模拟电路，具有易于VLSI(大规模集成电路)实现、能高速处理的信息的优点。图2显示了一个4×4的CNN网络结构。在图2中，每个格子即为一个“细胞”。单个细胞可由线性电容、线性电阻和非线性压控电流源等简单的元件组成，因此细胞神经网络很容易用电路实现。离散化后的CNN离散可用公式(3)表示，其中$E_{i,j}$表示电路中第i行第j列上“细胞”的状态。

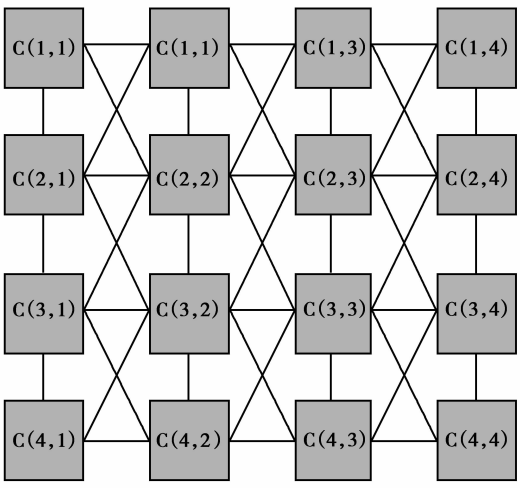

图2 一个4×4的CNN网络结构

$$E'_{i,j}=E_{i,j}+(-E_{i,j}/R_x+I+\sum_{(k,l)\in N_{i,j}} b_{k,l}I_{yx}(i,j;k,l))C'。\quad(3)$$

由于各个细胞可独立地处理信息,细胞神经网络是一个高速的并行计算系统。又因为细胞神经网络本身的非线性以及多个细胞耦合形成的复杂性,细胞神经网络是不可逆的。以往的研究和实验也证实了细胞神经网络优异的敏感性和随机性[14]。此外,细胞神经网络的拓扑结构非常适合处理图形和图像。这些特性使得细胞神经网络能很好地实现分块处理算法的要求。下面给出设计的一个基于细胞神经网络的分块处理算法。

算法:CNNblock。

输入:$M\times N$ 的图像矩阵 $\boldsymbol{W}$,

输出:$(M/a)\times(N/b)$的图像矩阵 $\boldsymbol{Z}$。

过程:

Step1:将 $\boldsymbol{W}$ 作为细胞神经网络的初始状态;

Step2:将细胞神经网络迭代 r 轮(r 取 M 与 N 中的较大值),得到全体细胞新的状态矩阵 $\boldsymbol{X}$;

Step3:将 $\boldsymbol{X}$ 按以下规则转换成 $M\times N$ 灰度矩阵 $\boldsymbol{Y}$

$$y_{i,j}=|x_{i,j}|\bmod 2^m.\quad(4)$$

其中 $x_{i,j}$和 $y_{i,j}$分别表示 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$ 第 i 行第 j 列的元素,$i=1,2,\cdots,M$,$j=1,2,\cdots,N$,　为下取整操作,m 为图像像素的精度;

Step4:将矩阵 $\boldsymbol{Y}$ 按以下规则转换成 8×8 的图像灰度矩阵 $\boldsymbol{Z}$ 并输出。

$$z_{i,j}=(\sum_{u=1}^{a}\sum_{v=1}^{b} y_{a(i-1)+u,b(j-1)+v})\bmod 2^m,\quad(5)$$

其中 $y_{i,j}$和 $z_{i,j}$分别表示 $\boldsymbol{Y}$ 和 $\boldsymbol{Z}$ 第 i 行第 j 列的元素,$i=1,2,\cdots,M/a$,$j=1,2,\cdots,N/b$.

3　层叠式并行 Hash 函数的缺陷与改进

直接采用层叠式并行 Hash 函数结构存在 2 个缺陷。

缺陷 1:设图像 I_0 为行数大于 M、列数均大于 N 的图像,又设 Hash 函数处理 I_0 时 Step 2 总共运行 n 次,且第 i 次进入 Step 2 时处理的图像为 I_i,($i=1,2,\cdots,n$),则 $I_1,I_2,\cdots,I_n$ 与 I_0 的 Hash 值全部相等。

缺陷 2:该 Hash 函数要求原始图像的大小必须等于 $(M(N)^{abk}$(k 为任意的正整数),这严重限制了算法的实用性。

通过对图像进行填充可弥补上述 2 个缺陷。与传统 Hash 函数不同,层叠式并行 Hash 函数的填充算法时在设计时须仔细考虑,以避免引入更多的缺陷。

算法:PAD

输入:待填充的图像、原始图像的行数和列数

输出:填充后的图像,其行数和列数分别为 M 和 N($N\geqslant 16$)的整数倍

过程:

Step1:获得待填充图像的行数 r 和列数 c;

Step2:在图像右部及下部填充 0 像素,使填充后的图像行数和列数分别为 M 和 N 的整数倍

a) 若如果 $r<M$,$c\leqslant N$ 或 $r=M$,$c\leqslant N/2$,使填充后的图像大小为 $M\times N$;

b) 否则使图像行数为 $r+M-r\%M$,列数为 $c+N-c\%N$ (%表示模运算);

Step3:填充原始图像的行数和列数:

a) 用图像最后一行倒数第 8,7,6,5 字节表示原始图像的行数;

b) 用图像最后一行倒数第 4,3,2,1 字节表示原始图像的列数。

4　实验结果

为了检验设计的 Hash 函数的安全性和有效性,做了详细的实验。在实验中,设 $M=16$,$N=16$,$a=2$,$b=2$,$m=8$,Hash 值的大小为 512 Bt;公式(4)中设 $R_x=2$,$C'=1$,$I=0$,规定细胞神经网络采用循环边界,每个细胞仅与其上、下、左、右 4 个方向的细胞相邻,且连接权值分别为 0.5,0.7,0.4 和 0.6。实验选取的处理对象为 120×120 的 Lena 图像。

4.1　敏感性测试

敏感性测试检查当消息发生微小变化时,Hash 值的改变情况。理想情况下,Hash 值应有接近一半比特发生变化,即所谓“雪崩效应”。随机改变原图像中的 1Bt 特,并检查 Hash 值的改变情况。表 1 列出了 1 000 次随机更改后,Hash 值比特变化率的平均值和标准差。在理想情况下,比特变化率的平均值等于 0.5,而标准差越小则说明算法的稳定性越高。表 2 说明 Hash 值的改变率非常接近理论值,且稳定性很高。

表 1　Hash 函数比特变化率的平均值和标准差　%

平均值	标准差
0.500 7	0.022 4

4.2 随机性测试

为了满足安全性要求，最终的 Hash 值应有很强的随机性。采用了美国 NIST 机构建议的随机数测试软件[15]对生成的 Hash 值进行严格的测试。NIST 随机数测试软件共包含 15 项测试。在 1 000 次实验中，若通过率大于 0.980 6 则认为通过该项测试。

实验采用前产生 1 000 个 Hash 值作为 1000 个序列，由于每个序列只包含 512 比特数据，NIST 测试软件中有 7 项测试可用于这些数据。表 2 给出了该算法在 7 种随机性测试中的通过率。从表 2 可以看出，所有测试的通过率均大于 0.980 6，可见提出的算法产生的 Hash 值满足随机性要求。

表 2 Hash 值的随机性检测结果

测试名称	通过率
frequency	0.995 0
block-frequency	0.988 0
cumulative-sums*	0.993 0
runs	0.993 0
longest-run	0.990 0
approximate entropy	0.985 0
serial*	0.983 0

4.3 碰撞分析

抗碰撞能力是 Hash 函数的一个重要指标。在碰撞分析实验中，随机改变图像中的 1 Bt，并检查 2 个 Hash 值中对应字节相等的个数。这里仍采用随机改变原图 1Bt 后产生的 1 000 个 Hash 值。表 3 列出了 1 000 个 Hash 值中，恰有 $i(i=0,1,2,3)$ 个字节相等的 Hash 值的个数。表 3 说明，设计的 Hash 函数非常接近理论值，具有良好的抗碰撞性能。

表 3 hash 值中字节相等个数的分布 ‰

Hash 值中字节相等的个数	0	1	2	3
理论值	778.4	195.4	24.1	1.95
实际值	778	200	19	2

5 结 论

传统的 Hash 函数在用于数字图像时未能利用其二维特性提高处理速度，更难以支持并行计算。提出的“层叠式并行 Hash 函数”结构和填充算法可很好地弥补这一缺陷，其时间复杂度在并行平台上远低于传统的 Hash 函数。选用细胞神经网络实现分块处理算法，以实现并行性、不可逆性、敏感性、随机性等要求。实验表明，基于提出的层叠式并行 Hash 函数结构和细胞神经网络的 Hash 函数可通过基本的安全性测试。

参考文献：

[1] SCHNEIER B. Applied cryptography[M]. New York: John Wiley & Sons, 1994.

[2] WANG X Y, YU H B. How to break MD5 and other hash functions[C]// Proceedings of the 24th Annual International Conference on Theory and Applications of Cryptographic Techniques, May 22-26, 2005, Aarhus, Denmark. Heidelberg: Springer-Verlag Berlin, 2005: 19-35.

[3] WANG X M, ZHANG J S, ZHANG W F. One way hash function construction based on the extended chaotic maps switch[J]. Acta Physica Sinica, 2003, 52(11):2737-2742.

[4] 韦鹏程，张伟，廖晓峰，等. 基于双混沌系统的带秘密密钥散列函数构造[J]. 通信学报，2006，27(9) 27-33.
WEI PENG-CHENG, ZHANG WEI, LIAO XIAO-FENG, et al. Design keyed hash function based on couple chaotic system[J]. Journal on Communications, 2006, 27(9) 27-33.

[5] 李红达，冯登国. 复合离散混沌动力系统与 Hash 函数[J]. 计算机学报，2003，26(4):460-467.
LI HONG-DA, FENG DENG-GUO. Composite nonlinare discrete chaotic dynamical systems and keyed hash functions[J]. Chinese Journal of Computers, 2003, 26(4):460-467.

[6] YI X. Hash function based on chaotic tent maps[J]. IEEE Transactions on Circuits and Systems II: Express Briefs, 2005, 52(6): 354-357.

[7] ZHANG J S, WANG X M, ZHANG W F. Chaotic keyed hash function based on feedforward feedback nonlinear digital filter[J]. Physics Letters A, 2007, 362(5-6): 439-448.

[8] XIAO D, SHIH F Y, LIAO X F. A chaos-based hash function with both modification detection and localization capabilities [J]. Communications in Nonlinear Science and Numerical Simulation, 2010, 15(9): 2254-2261.

[9] DENG S J, LI Y T, XIAO D. Analysis and improvement of a chaos-based hash function

construction[J]. Communications in Nonlinear Science and Numerical Simulation, 2010, 15(5): 1338-1347.

[10] ZHOU Q, LIAO X F, WONG K W, et al. True random number generator based on mouse movement and chaotic hash function[J]. Information Sciences, 2009, 179(19): 3442-3450.

[11] 王小敏, 张文芳, 张家树. 基于非线性数字滤波器的混沌 Hash 函数设计[J]. 计算机辅助设计与图形学学报, 2006, 18 (6): 870-875.
WANG XIAO-MIN, ZHANG WEN-FANG, ZHANG JIA-SHU. Design of chaotic hash function based on nonlinear digital filter[J]. Journal of Computer-aided Design & Computer Graphics, 2006, 18 (6): 870-875.

[12] US National Institute of Standards and Technology. Announcing request for candidate algorithm nominations for a new cryptographic hash algorithm (SHA-3) family[J]. Federal Register, 2007, 72: 62212-62220.

[13] CHUAN L O, YANG L. Cellular neural networks: theory [J]. IEEE Transactions on Circuit and Systems, 1988, 35(10):1257-1272.

[14] 周庆, 廖晓峰, 胡月. 采用细胞神经网络结构进行图像加密的框架及算法[J]. 计算机辅助设计与图形学学报, 2009, 21(11) : 1676-1681.
ZHOU QING, LIAO XIAO-FENG, HU YUE. An image encryption framework and an algorithm based on CNN [J]. Journal of Computer-aided Design & Computer Graphics, 2009, 21(11):1676-1681.

[15] RUKHIN A, SOTO J, NECHVATAL J, et al. NIST Special Publication 800-22 A statistical test suite for random and pseudorandom number generators for cryptographic applications[S]. Gaithersburg: National Institute of Standards and Technology, 2001.

(编辑　侯　湘)

(上接第 106 页)

[10] GENG Z Q, ZHU Q X. Multi-scale nonlinear principal component analysis (NLPCA) and its application for chemical process monitoring [J]. Industrial & Engineering Chemistry Research, 2005, 44 (10): 3585-3593.

[11] WANG D, ROMAGNOLI J A. Robust multi-scale principal components analysis with applications to process monitoring [J]. Journal of Process Control, 2005, 15(8): 869-882.

[12] LEE D S, PARK J M, VANROLLEGHEM P A. Adaptive multi-scale principal component analysis for on-line monitoring of a sequencing batch reactor[J]. Journal of Biotechnology, 2005, 116(2): 195-210.

[13] 周福娜, 文成林, 汤天浩, 等. 基于指定元分析的多故障诊断方法[J]. 自动化学报, 2009, 35(7): 971-982.
ZHOU FU-NA, WEN CHENG-LIN, TANG TIAN-HAO, et al. DCA based multiple faults diagnosis method[J]. Acta Automatica Sinica, 2009, 35(7): 971-982.

[14] 周东华, 胡艳艳. 动态系统的故障诊断技术[J]. 自动化学报, 2009, 35(6): 748-758.
ZHOU DONG-HUA, HU YAN-YAN. Fault diagnosis techniques for dynamic systems [J]. Acta Automatica Sinica, 2009, 35(6): 748-758.

[15] WANG X, KRUGER U, IRWIN G W. Process fault diagnosis using recursive multivariate statistical process control[C/OL] // The 16th IFAC World Congress in Prague, Prague Czech Republic, July 4-8, 2005: impact on http:// www. ifac-papersonline. net/Detailed/29131. html.

(编辑　侯　湘)