基于消息网络的Hash函数构造

王继敏[1]，宋玉蓉[2]，蒋国平[2]
（1.南京邮电大学 计算机学院，江苏 南京 210003；
2.南京邮电大学 自动化学院，江苏 南京 210003）

摘 要：为了提高Hash函数的敏感性和运算速度，利用明文通过某种规则构造权重网络，并建立权重网络与混沌复杂动态网络的映射关系。将权重网络的邻接矩阵用到混沌系统中，经过特定量的迭代运算，将本次的输出对称交换后作为下个消息块运算的输入，类似的处理所有的消息块，最后的输出经过线性变换和进制转换，得到一定长度的Hash值，其长度与网络的大小有关。理论分析和数值仿真表明，提出的算法具有良好的初值敏感性、单向性、置乱性和强的抗碰撞性。

关键词：消息网络；Hash；混沌映射
中图分类号：TP393 **文献标识码**：A **文章编号**：1673-629X(2011)09-0024-04

Hash Function Construction Based on Message Network

WANG Ji-min[1], SONG Yu-rong[2], JIANG Guo-ping[2]
(1. College of Computer, Nanjing University of Posts and Telecommunications, Nanjing 210003, China;
2. College of Automation, Nanjing University of Posts and Telecommunications, Nanjing 210003, China)

Abstract: In order to improve Hash algorithm's sensitivity and speed, a new Hash construction algorithm is proposed, where a weighting network is constructed based on the message with some specific rules, and a mapping is established from the weighting network to the chaotic complex dynamic network. By a bit iterative operation, the chaotic system with the adjacent matrix of the message weighting network can get an output, which be symmetric exchanged and considered as the input of the next message block's operation. All the message blocks can be handled similarly. By linear transformation and disables conversion, the output of the last message block can produce a certain length Hash value, the length of which relates to the size of the network. Simulations show that the algorithm is extremely sensitive to the initial values and has excellent performance in one-way, confusion, diffusion and collision resistance.

Key words: message network; Hash; chaotic map

0 引 言

Hash函数，又称为杂凑函数，是将任意长的输入压缩成固定长度的输出，它在现代密码学中起着非常重要的作用[1]。混沌系统是一种确定的非线性系统，但是具有初值敏感性，能够产生貌似随机的运动轨迹，这些特性正好满足Hash的特性要求[2]。近年来，已经有许多的基于混沌耦合映像格子理论的Hash函数构造方法被提出来[3-8]，并取得了很好的效果。然而，耦合映像格子又称为最近邻耦合网络，在恶意软件病毒传播的研究中[9]，发现病毒的传播速度和网络的拓扑结构之间有着一定的联系，且病毒在随机网络或小世界网络中的传播速度远远大于在最近邻耦合网络中的传播速度。可见，网络中各节点间的关系越复杂，联系越紧密，其传播速度也就越大。由耦合映像格子的模型不难发现，耦合映像格子是一种规则网络，只是本节点与它相邻的两个节点之间有联系。若将一个节点的变化扩散到整个消息网络中，需要经过多次扩散，速度比较慢。文献[10~12]介绍了金融、数学等领域的复杂网络构建方法，将需要处理的数据信息构造成具有某种拓扑结构的网络，实现了从网络的角度分析和处理问题，效果很好。在此，本算法将耦合映像格子模型进化为具有某种拓扑结构的权重网络，实现单向Hash函数的构建。

文中提出了一种新的基于消息网络的单向Hash函数构造方法，利用明文通过某种规则构造权重网络，建立了权重网络与混沌复杂动态网络的映射关系。

收稿日期：2011-02-25；**修回日期**：2011-05-03

作者简介：王继敏(1986-)，女，硕士研究生，研究方向为信息安全、网络安全；宋玉蓉，副教授，研究方向为信息安全、复杂网络、病毒传播；蒋国平，教授，研究方向为复杂动态网络。

1 混沌耦合映射网络模型

1.1 混沌耦合映射网络的时空混沌行为

格子映射为单峰映射的耦合映射网络模型[13]的

形式为:$x_{i+1}^i = f(x_i^i) + \sum_{j=1}^{N} e_{ij}x_i^j$　　(1)

其中x_i^i为输入;e_{ij}是描述网络系统中元素之间的权重;N为网络的大小;$f(x)=\mu x(1-x)$,其中$0<x<1$,当参数$3.569945<\mu\leqslant 4$时,系统处于混沌状态。为了满足x的取值范围,每次迭代都要做去整运算,即$x=x-\mathrm{int}(x)$。式(2)可以更直观地描述:

$$x_{i+1}^i=\mu x_i^i(1-x_i^i)+\sum_{j=1}^{N}e_{ij}x_i^j-\mathrm{int}\left(\mu x_i^i(1-x_i^i)+\sum_{j=1}^{N}e_{ij}x_i^j\right) \quad (2)$$

当系统满足一定的初始值和参数条件时,呈现如图1所示的混沌行为。

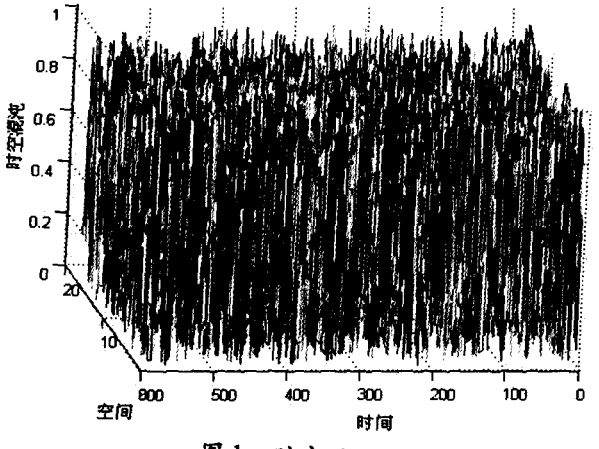

图1　时空混沌

系统迭代1000次时,变量x的分布情况如图2所示,可以发现,变量x在系统内部的分布比较均匀。

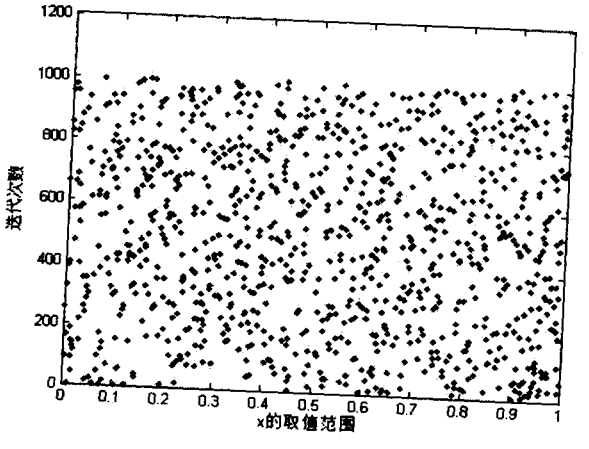

图2　$\mu=3.99$变量x的分布图

1.2　网络的构造

首先介绍网络,网络是由顶点集合及顶点间的关系集合(边)组成的,即$G=(V,E)$。在文中,消息网络中的顶点v_{ij}是消息m_{ij}线性变换的$\overline{m_{ij}}$(表示消息子块M'_i的第$i\times N+j$个字节的ASCII码线性变换后的小数);消息网络中的边e_{ij}表示节点v_{ij}所在行与所在列之间的乘积,也称为节点间的权值,其邻接矩阵为$E=(e_{ij})_{N\times N}$。e_{ij}的取值范围应该为[0,1]。为了保证e_{ij}的值有效,需要去除整数,只保留小数,如式(3)表示:

$$\begin{cases} temp=(\overline{m_{i1}}\ \ \overline{m_{i2}}\ \ \cdots\ \ \overline{m_{iN}})(\overline{m_{1j}}\ \ \overline{m_{2j}}\ \ \cdots\ \ \overline{m_{Nj}})^{\mathrm{T}} \\ e_{ij}=temp-\mathrm{int}(temp) \end{cases} \quad (3)$$

$\overline{m_{ij}}$的微小变化能影响其整个邻接矩阵,实现了将微小的变化迅速扩散到网络中。

1.3　基于消息网络的单向Hash函数构造

按照图3描述的过程来处理明文。基于消息网络构造单向Hash函数的一般过程如下:

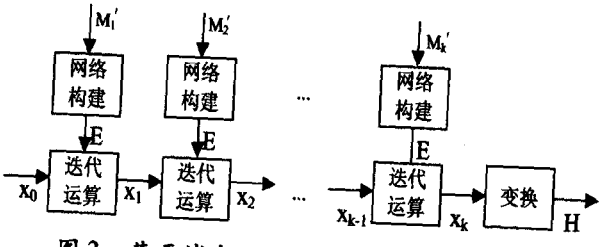

图3　基于消息网络的单向Hash函数模型

1)首先将明文消息预处理:$M'=M+r_{(\mathrm{reversal})}(M)+l_{\mathrm{length}}(M)$,式中+表示字符的连接运算,函数$r_{(\mathrm{reversal})}(\cdot)$表示将消息串逆排。将预处理明文消息$M'$分割成$k$个子块$M'=M'_1,M'_2,\cdots,M'_k$,消息子块的长度为$N\times N$。若消息$M'$的长度不是$N\times N$的整数倍,则用消息$M$来填充。

2)将待处理消息块M'_i按对应字节m_{ij}的ASCII码线性变换到[0,1]区间,记为$\overline{m_{ij}}$,对应的消息块记为$\overline{M'_i}$,计算公式:$\overline{m_{ij}}=m_{ij}/256$。

3)读入消息子块$\overline{M'_i}$,按照上文描述的方法,通过消息子块$\overline{M'_i}$得到相应的权重网络E。

4)读入初始值x_0,将网络的邻接矩阵代入式(2)中,迭代若干次,迭代次数为$\mathrm{int}(e_{NN}\times 100)$,可得到一个长度为$N$的数列。

5)将迭代运算后得到的输出进行对称交换,再作为下一个消息块的初始值。

6)按步骤2)~5)迭代处理其他明文块,将最终得到数列中的小数线性变换到[0,256]的整数,再将这列整数转换成十六进制的字符串,就得到一个长度为$N\times 8$的Hash值。

2　仿真实验结果与分析

参数分别取值为$x_0=[0.1\ \ 0.1\ \ \cdots\ \ 0.1]^{\mathrm{T}}$,$x_0$的长度$N=16$,$\mu=3.99$。

2.1　数据敏感性分析

Hash函数应该具有良好的初值敏感性。对如下

消息文本做散列试验，Hash 值长度为 128 位。

初始文本记为文本 1："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, Butler asserts that eighteenth-century American was a modem place with a distinctive society."

测试方法为：变换一，将文本 1 中大写的 I 改为小写，记为文本 2；变换二，将文本 1 中的 1680 改为 1681，记为文本 3；变换三，将文本 1 中的 Despite 写成 Despit，减去一个字符，记为文本 4；变换四，将文本 1 的 character 写成 characters，增加一个字符，记为文本 5。最后得到的 Hash 值都用十六进制数表示，结果分别为：

文本 1：8F28B4ED5C8B87CD180FDA0806E844D2

文本 2：9F6B8153469B28F6832A64A482BBB8E3

文本 3：F2D1F485FC6CC5C9539973909F84BD95

文本 4：DFB04BB9C4B6D414D87621519708233E

文本 5：AB27A17C5F2A352913745623B5FC88A8

由此可见，该算法还是比较敏感的，明文消息的微小改变都可以引起 Hash 结果发生较大的变化。所以，这种算法对初值很敏感。

2.2 混沌与扩散性能统计分析

Hash 值的二进制表示中每 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。混乱与扩散的统计分析用到的评价指标[3~8]：平均变化位数为 $\bar{B}=\frac{1}{|N|}\sum_{i=1}^{|N|}B_i$，平均变化率为 $P=\frac{\bar{B}}{128}\times 100\%$，平均变化位数的均方差为 $\Delta B=\sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(B_i-\bar{B})^2}$，平均变化率的均方差为 $\Delta P=\sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(B_i-P)^2}\times 100\%$，其中 N 为统计次数，B_i 为第 i 次测试结果变化的比特数。测试方法：在明文空间中随机选取一段明文进行 Hash，然后改变明文 1 bit 的值得到另一 Hash 结果，比较两个结果得到变化比特数 B_i。进行 $N=256,512,1024,2048$ 次测试，得到明文 1 bit 变化时的各项统计值如表 1 所示。

表 1 N 次测试的各项指标

统计值	$N=256$	$N=512$	$N=1024$	$N=2048$	平均
平均变化位数	63.87	64.09	64.24	63.94	64.04
平均变化位数的均方差	5.32	5.78	5.64	5.58	5.58
平均变化率	49.90	50.07	50.19	49.96	50.03
平均变化率的均方差	4.16	4.52	4.41	4.36	4.36

在 $N=1024$ 时，置乱数分布图并且如图 4 所示，B 为 64.24，P 为 50.19%，极为接近理想值 64 和 50%，ΔB 值和 ΔP 值也都比较小，其值越小说明算法的混淆与扩散特性越稳定。

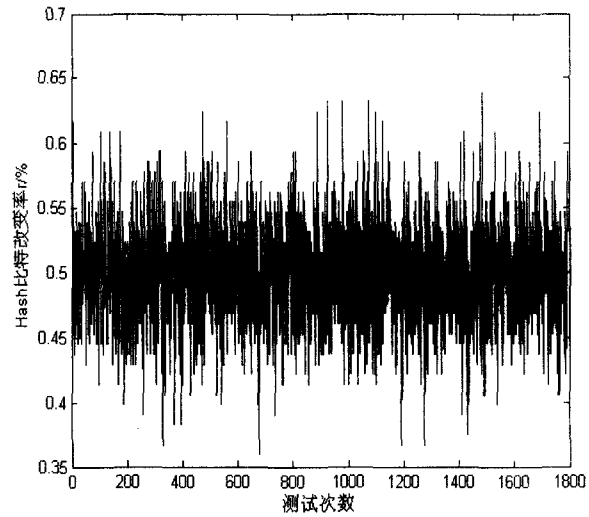

图 4 变化比特率的分布图

2.3 算法的抗碰撞性分析

碰撞是指不同的初始值，其 Hash 映射结果相同，即发生了多对一映射。通过以下的实验来定量测试本算法的抗碰撞能力：在明文消息空间中随机地选取一段明文，求出其 Hash 值，并用 ASCII 码的形式进行存储，然后随机地选择明文消息中的 1 bit，并改变这 1 bit 的值，可以得到另外一个新的 Hash 值，同样用 ASCII 码的形式进行存储。比较改变 1 bit 前后的两个 Hash 值，若改变 1 bit 前后的两个 Hash 值在相同位置上有相同的 ASCII 码字符，则称为被击中一次，然后统计被击中的次数。用公式 $d=\sum_{i=1}^{N}|t(e_i)-t(e_i')|$ 计算改变 1 bit 前后两个 Hash 值的绝对差异度，其中 e_i 和 e_i' 分别是改变 1 bit 前的 Hash 值和改变 1 bit 后的 Hash 值的第 i 个 ASCII 码字符，N 为 Hash 值对应 ASCII 字符的个数，在这里 $N=16$，而函数 $t(\cdot)$ 将 e_i 和 e_i' 的 ASCII 码字符转化为它们相应的十进制数值。对这个算法做了 2048 次这样的实验，我们发现，其中 1951 次测试没有发生碰撞，95 次测试中发生一次碰撞，有 2 次测试中发生了两次碰撞，可见发生碰撞的概率是很低的。在相同位置上有相同值的 ASCII 码字符的分布图如图 5 所示，d 的最大值、最小值、平均值在表 2 中。

表 2 两个 Hash 值的绝对差异度

最大值	最小值	平均值
2141	631	1312

2.4 算法执行效率分析

执行效率是决定 Hash 算法是否具有实际应用价

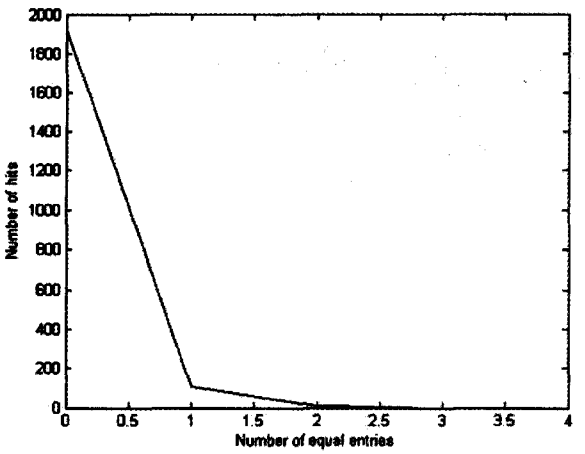

图 5　Hash 值在相同位置上有相同值的 ASCII 码字符的分布图

值的重要标志之一。文中的算法是利用明文消息来构建权重网络,将消息中微小的变化直接扩散到整个网络中,与基于混沌耦合映像格子的 Hash 算法相比,扩散速度更快,从而可以减少混沌运算迭代的次数。然而,本算法是基于混沌系统的 Hash,主要采用浮点数运算,所需算术运算较多,对算法执行效率有所影响。

3　结束语

提出了消息网络的思想,在此基础上构造了单向 Hash 函数。该算法是利用消息明文通过特定的规则构造成具有某种结构的权重网络,并建立消息权重网络与混沌复杂动态网络的映射关系。将消息权重网络的邻接矩阵用到混沌系统中,经过迭代运算,将本次的输出对称交换后作为下个消息块运算的输入,类似的处理所有的消息块,最后一个消息块运算得到的数,经过线性变换和进制转换,得到一定长度的 Hash 值,其中 Hash 值的长度与网络的大小有关。理论分析和数值仿真表明,所提出的新算法具有良好的单向性、置乱性和强的抗碰撞性,增强了消息之间的耦合度,使消息中任何微小的变化都能直接扩散到整个网络中,提高了算法对初始明文的敏感性,降低了算法的复杂度。在网络信息安全领域,该算法可以很容易地完成数字签名、身份认证等功能。

参考文献:

[1] Stallings W. 密码编码学与网络安全——原理与实践[M]. 北京:电子工业出版社,2006.

[2] 郭　雷,许晓鸣. 复杂网络[M]. 上海:上海科技教育出版社,2006.

[3] 刘建东,付秀丽. 基于耦合帐篷映射的时空混沌单向 Hash 函数构造[J]. 通信学报,2007,28(6):30-38.

[4] 张　瀚,王秀峰,李朝晖,等. 基于时空混沌系统的单向 Hash 函数构造[J]. 物理学报,2005,54(9):4006-4011.

[5] 刘光杰,单　梁,孙金生,等. 基于时空混沌系统构造 Hash 函数[J]. 控制与决策,2006,21(11):1244-1248.

[6] 赵　耿,袁　阳,王　冰. 基于交叉耦合映像格子的单向 Hash 函数构造[J]. 东南大学学报(自然科学版),2009,39(4):728-732.

[7] 程艳云,宋玉蓉. 基于耦合映像格子混沌系统的 Hash 函数构造[J]. 应用科学学报,2010,28(1):44-48.

[8] Song Y R, Jiang G P. Constructing hash function based on coupled network generated by logarithmic map[J]. Journal of Nanjing University of Posts and Telecommunications (Natural Science), 2010, 30(1):6-10.

[9] 宋玉蓉,蒋国平. 基于一维元胞自动机的复杂网络恶意软件传播研究[J]. 物理学报,2009,58(9):5911-5918.

[10] Liu J, Chi K T, He K Q. Fierce stock market fluctuation disrupts scale free distribution[J]. Quantitative Finance, 2009, 10(80):7688-7696.

[11] Chen C, Lu J A, Wu X Q. Complex networks constructed from irrational number sequences[J]. Physica A, 2010, 389(13):2654-2662.

[12] Zhang J, Sun J F, Luo X D, et al. Characterizing pseudoperiodic time series through the complex network approach[J]. Physica D, 2008, 237(22):2856-2865.

[13] García P, Parravano A, Cosenza M G, et al. Coupled map networks as communication schemes[J]. Physical Review E, 2002, 65(4):5201-5204.

(上接第 23 页)

由算法[J]. 电子与信息学报,2007,29(2):340-344.

[9] Bell J E, Mcmullen P R. Ant colony optimization techniques for the vehicle routing problem[J]. Advanced Engineering Informatics, 2004, 18(1):41-48.

[10] 潘达儒,袁艳波. 一种基于 AntNet 改进的路由算法[J]. 小型微型计算机系统,2006,27(7):1169-1174.

[11] 张　千,梁　鸿,李　振. 基于改进蚂蚁算法的网格资源管理的研究[J]. 微电子学与计算机,2009,26(9): 71-74.

[12] Yang H C, Dasdan A, Hsiao R L. Map-reduce-merge: Simplified relational data processing on large clusters[C]// Proc of the 2007 ACM SIGMOD International conference on management of data. Beijing, China:[s. n.], 2007:1029-1040.

[13] Calheriros R N, Ranjan R, Buyya R. CloudSim: A Novel Framework for Modeling and Simulation of Cloud Computing Infrastructures and Services[R]. Melbourne: Grid Computing and Distributed Systems Laboratory, The University of Melbourne, 2009.