第 52 卷 第 11 期 2003 年 11 月
1000-3290/2003/52(11)/2737-06

物 理 学 报
ACTA PHYSICA SINICA

Vol.52,No.11,November,2003
©2003 Chin. Phys. Soc.

# 基于广义混沌映射切换的单向 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)

(2003 年 2 月 10 日收到;2003 年 3 月 13 日收到修改稿)

如何设计快速高效的单向 Hash 函数一直是现代密码学研究中的一个热点.提出了一种基于广义混沌映射切换的 Hash 函数构造方法.这种方法首先构建产生多种混沌序列的广义混沌映射模型,然后在明文信息的不同位置根据切换策略产生不同的混沌序列,并用线性变换后的信号信息对混沌参数进行调制来构造单向 Hash 函数.初步分析了利用混沌映射实现单向 Hash 函数的不可逆性、防伪造性、初值敏感性等特点.研究结果表明:这种基于广义混沌映射切换的 Hash 函数具有很好的单向性、弱碰撞性,较基于单一混沌映射的 Hash 函数具有更强的保密性能,且实现简单.

**关键词**:Hash 函数,混沌,混沌映射切换

**PACC**:0545

## 1. 引 言

随着电子商务的迅速发展,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为代表的安全技术中得到了广泛应用,并成为研究热点[1,2].传统的单向 Hash 方法有 MD2,MD4,MD5,SHA 等标准[2,3],但它们大多是基于复杂度假设的,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多次迭代[1]得到 Hash 结果,后一种方法运算量很大,难以找到快速同时可靠的加密方法,而前一种方法中由于异或运算中固有的缺陷,虽然每步运算简单,但计算轮数即使在被处理的文本很短的情况下也很大.

对此,有人提出了基于混沌映射模型的单向 Hash 算法[4,5].但是,这些方案均存在以下不足:(1)这些系统能够实现摘要提取,是基于混沌同步的鲁棒性,而这种鲁棒性也可为攻击者所利用,从而降低了系统的保密性[6];(2)这些系统均是基于某一种特定的混沌系统来实现的,可以利用各种混沌预测技术成功地破译、提取信息信号[7—13],保密性能并非所说的那样好;(3)由于实际实现中的有效字长精度效应,混沌映射本身所产生的混沌序列也会退化为周期序列[6].因此,如何增加混沌信号的复杂度和减小有效字长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主要问题.

针对上述问题,本文提出了一种基于广义混沌映射切换的混沌 Hash 方法,这种方法首先构建产生多种混沌序列的广义混沌映射模型,然后在原始文本的不同位置根据切换策略来产生不同混沌序列作为混沌载波,并用线性变换后的信号信息对混沌参数进行调制来构造单向 Hash 函数.理论分析与计算机仿真结果表明:这种基于广义混沌映射切换的单向 Hash 算法比基于单一混沌映射的 Hash 算法具有更好的置乱性,且易于实现.

## 2. 混沌映射构造 Hash 函数的可行性

单向函数的定义:如映射 $H:X\to Y$ 对所有的 $x\in X$,$H(x)$ 都容易计算,但反过来,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[1]:

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

* 国家自然科学基金(批准号:60272096)、四川省青年基金(批准号:03ZQ026-033)及西南交通大学基础学科研究基金(批准号:2001B08)资助的课题.

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.

我们知道,报文空间可以是无限的,而 Hash 结果总是一段定长字节的数字,会有无数的报文具有同样的 Hash 函数值,但在 Hash 结果达到一定长度,比如结果为固定的 128bit 长时,结果空间已有 $2^{128} \approx 3.4028 \times 10^{28}$个,以现有的计算能力在这样大的空间穷举计算是困难的.

混沌[14] 是指确定性非线性动力系统的长期行为对系统初始状态或系统参数异常敏感,却又不发散,而且无法精确重复的现象,它是非线性系统普遍具有的一种复杂动力学行为.混沌序列具有对初始条件和混沌参数的极端敏感性、白噪声的统计特性和混沌序列的遍历特性.由于混沌系统在迭代中的信息损失,使得混沌序列包含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很困难.

以上特点使混沌序列天然地拥有单向 Hash 函数所要求的较好的散布性和不可逆性、防伪造性、初值敏感性[15].但由于(1)实际实现中的有效字长精度效应,混沌映射本身所产生的混沌序列会退化为周期序列[6];(2)为了提高运算速度,对有限长度的报文,不可能进行大量的迭代.因此在单一混沌系统中,其混沌系列包含的信息量并不会趋于零,可以利用各种混沌预测技术对其破译[7—13].本文通过构建产生多种混沌序列的广义混沌映射模型,大大增强了混沌信号的复杂度和减小有效字长精度效应的影响,实验结果表明,这种方式能有效地提高混沌 Hash 函数的性能而不增加运算次数.

## 3. 基于广义混沌映射切换的单向 Hash 构造

### 3.1. 广义混沌映射模型的定义

要在统一的结构下实现多种混沌映射,一个主要的问题就是应保证各个混沌映射的输入不能超出它的值域范围[6],因此,应优先考虑值域范围一致的混沌映射来构造这个广义混沌映射.好在有一些离散混沌映射的值域范围能够满足这一要求,例如虫口映射、立方映射、锯齿映射和 tent 映射等的值域范围都在 $-1$—$1$ 之间.为此,定义一个如下的广义混沌映射模型:

$$x_{n+1} = g(x_n) = a_0 + a_1 x_n + a_2 x_n^2 + a_3 x_n^3 + a_4 |x_n| + a_5 \mathrm{sign}(x_n - b_0) + a_6((1 + b_1(x_{n-1} - b_2) + b_3 x_n^2) \bmod 2 - 1). \quad (1)$$

当 $a_i (i = 0,1,\cdots,6)$和 $b_j (j = 0,1,2,3)$取不同值时,(1)式就分别对应于不同的混沌映射,例如,当 $a_0 = 1, a_2 = -2$,其余的 $a_i, b_j = 0$ 时,$x_{n+1} = g(x_n)$即为虫口映射;而当 $a_1 = 3, a_3 = -4$,其余的 $a_i, b_j = 0$ 时,$x_{n+1} = g(x_n)$即为立方映射.显然,(1)式中 $a_i (i = 0,1,\cdots,6)$和 $b_j (j = 0,1,2,3)$的不同组合能够满足通过混沌映射切换方式的这种参数调制方案的需求,因而,(1)式可以用于这种广义混沌 Hash 算法.

### 3.2. 单向 Hash 构造

算法原理 将原始输入报文以字节为单位,经线性变换后对当前混沌映射的参数进行调制,同时对迭代次数进行计数,计数到一定次数时切换混沌映射,并用上次混沌映射的最后迭代结果作为本次映射的初始值继续迭代,当迭代到一定次数时再次切换混沌映射,直到报文结束.最后将 $x_R, x_{2R}, x_{3R}$ 三个迭代结果分别映射成 40bit,40bit,48bit 的三个大数,从而形成 128bit 的 Hash 值.

从上面原理可知,虽然整个迭代过程中使用了多个混沌映射,但在任一时刻只有一个混沌映射进行工作,因此对整个广义模型分析时可从分析单个混沌映射模型入手.为方便起见,不妨令(1)式中的 $a_6 = 1, b_1 = 0.3, b_2 = 1.08, b_3 = 2917$,其余的 $a_i, b_0 = 0$,则(1)式变为

$$x_{n+1} = (1 + 0.3(x_{n-1} - 1.08) + 2917x_n^2) \bmod 2 - 1. \quad (2)$$

注意到这个方程在已知第 $n$ 项时是无法解析解出 $n-1, n-2$ 中的任何一项的,这是系统不可逆性和防伪造性的保证之一.图 1 为系统在任一初值下经 4000 次迭代的结果.由图 1 可见,该混沌序列具有很好的类噪声特性.其分布特性很适合用来构造单向 Hash 函数,不仅终值的分布平稳,与迭代步数无关,而且与初值无关,分布也较为均匀,这样在已知终值情况下,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.

采用(1)式的广义混沌映射模型构造单向 Hash 算法的系统框图如图 2 所示,算法描述如下:

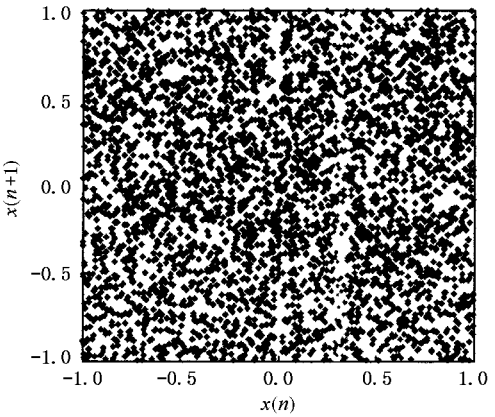


图 1　(2)式映射混沌吸引子的分布

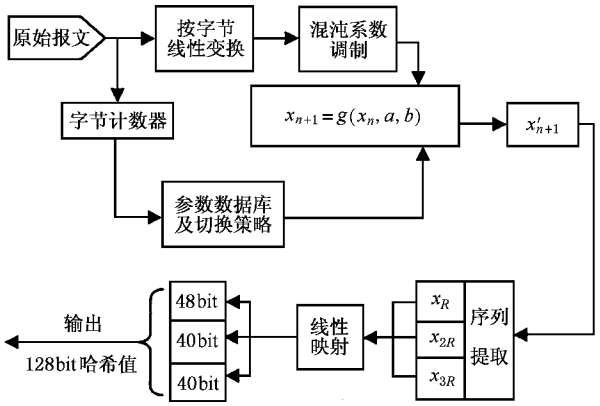


图 2　基于广义混沌映射切换的单向 Hash 构造系统

1) 将待处理报文按对应字节的 ASCII 码转换为数字,线性变换为 −1—1 范围内的数,这样整个报文得到一个大数组,记为 $S$,记数组长度即报文字节数为 $N$;

2) 令 $x_1 = 2 \times S_1/(k-1) - 0.8$, $x_2 = 2 \times S_2/(k-1) - 0.8$,其中 $k = 256$;

3) 迭代　令迭代轮数为 $r = R \times ([N/R] + 1)$,其中[　]为取整运算,$R = 32$;

i) $j$ 从 3 到 $r$ 计算

当 $j \leqslant r/5$ 时,取 $a_6 = 1$, $b_1 = 0.3$, $b_3 = 2917$,其余的 $a_i$, $b_0 = 0$,则广义混沌模型按

$$x_j = (1 + 0.3(x_{j-2} - c) + 2917x_{j-1}^2) \bmod 2 - 1 \tag{3}$$

进行迭代;

同理,当 $r/5 < j \leqslant 2r/5$, $2r/5 < j \leqslant 3r/5$, …, $4r/5 < j \leqslant r$ 时,分别改变 $a_i$, $b_i$ 的值,使系统在不同时段按不同的混沌映射进行迭代;

若当前迭代模型为(3)式,在 $j \leqslant N$ 时,令 $c = 2S_j/(k-1) - 0.8$,在 $j > N$ 时,令 $c = x_N$.

同理,对于当前迭代模型为其他混沌映射时,也可按上面的方法用原始文本信息对迭代参数进行调制;

ii)进一步迭代,$j$ 从 3 到 $3R$ 计算

若当前迭代模型为(3)式时,令 $c = x_{r-2}$, $x_1 = x_{r-1}$, $x_2 = x_r$,计算

$$x_j = (1 + 0.3(x_{j-2} - c) + 2917x_{j-1}^2) \bmod 2 - 1.$$

若当前迭代模型为其他模型时可依此类推;

4)从迭代结果序列中取出 $x_R$, $x_{2R}$, $x_{3R}$,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.同时还可看出,若将 $x_R$, $x_{2R}$, $x_{3R}$,映射为两个 56bit,一个 48bit 的二进制数,在不增加任何附加运算量情况下,即可得到 160bit 的 Hash 结果,因此该算法较以往的 Hash 算法更为灵活.

## 4. 仿真研究

### 4.1. 文本 Hash 结果

对以上单向 Hash 函数构造算法进行计算机仿真,取初始文本 1 为‘ AppWizard has created this improved application for you. This file contains a summary of what you will find in each of the files that make up your improved application. The improved dsp file (the project file) contains information at the project level and is used to build a single project or subproject. ”

文本 2 将文本 1 中首字母 $A$ 改为 $B$,文本 3 将文本 1 中的 information 改为 informatioN,文本 4 将文本 1 最后的句号改为逗号,文本 5 在文本 1 的最后再加一个空格.基于广义混沌映射算法的 Hash 结果用十六进制数表示如下,用 0,1 序列的图形化表示,如图 3 所示.

初始文本 1： 3E7D077EFFB5FB2066BBEB482DAAB475

文本 2： F8C1CD9990B761445F78EBAA21117CCE

文本 3： 6A97B992595447998A0D773730182EDE

文本 4： E44E7C526EA97C52CF884E62C5FC9E82

文本 5： 948086E2FFFF367B50820958FD818F56

从以上数据可看出,初值的每 bit 变动,结果都以很大的概率发生较大变化,可见该算法的单向 Hash性能很好,具有十分高的初值敏感性.

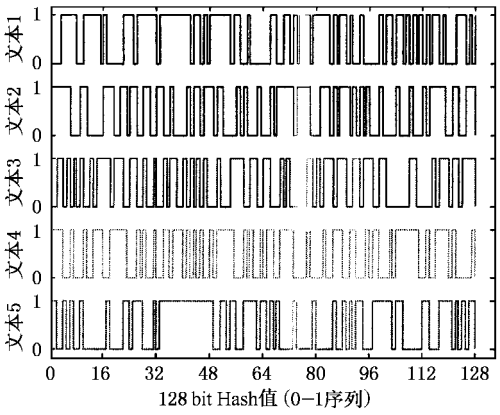


图 3 文本 1—5 的 Hash 值比较图

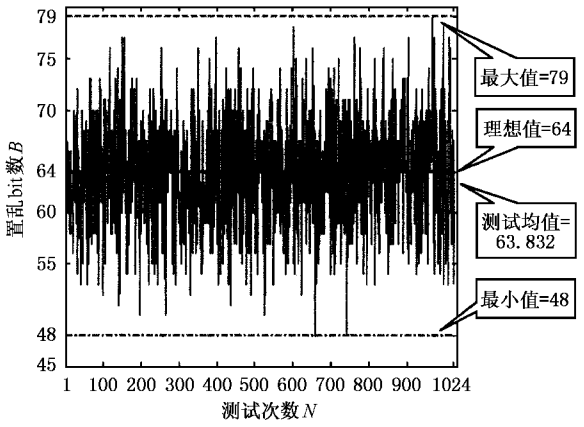


图 4 置乱数分布图

### 4.2. 混乱与散布性质统计分析

为了隐藏明文消息的冗余度,Shannon 提出了混乱与散布的概念,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. 要尽量做到相应明文对应的 Hash 密文不相关,而对于结果的二进制表示,每 bit 只有 0 或 1 两种可能,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 bit 都以 50% 的概率变化. 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 bit 数. 定义:

平均变化 bit 数

$$\overline{B} = \frac{1}{N}\sum_{i=1}^{N} B_i , \tag{4}$$

平均变化概率

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

$B$ 的均方差

$$\Delta B = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(B_i - \overline{B})^2} , \tag{6}$$

$P$ 的均方差

$$\Delta P = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(B_i/128 - P)^2} \times 100\% , \tag{7}$$

其中 $N$ 为统计总次数,$B_i$ 为第 $i$ 次测试时结果的变化 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 测试,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 bit 数 $B_i$. 在 $N = 1024$ 次测试下得到的置乱数分布情况如图 4 所示.

由图 4 的测试结果可知,在 $N = 1024$ 次测试中,明文 1bit 变化引起 128bit 的 Hash 值发生变化的 bit 数位于 48 和 79bit 之间,平均 bit 变化数为 63.832 个,非常接近理想状况下的 64bit 变化数. 从图 4 还可看出,1024 次测试的 bit 变化数集中在 60—70bit 之间,即紧紧云集在理想值 64bit 附近,从而表明该算法对明文的置乱能力强而稳定.

另外,经 $N = 256, 512, 1024, 2048$ 次测试,得到基于该算法明文 1bit 变化下的平均密文变化 bit 数 $B, P, \Delta B, \Delta P$ 的值,如表 1 所示.

表 1 平均密文变化 bit 数

| | $N = 256$ | $N = 512$ | $N = 1024$ | $N = 2048$ | 总平均 |
|---|---|---|---|---|---|
| $\overline{B}$ | 64.346 | 63.514 | 63.832 | 63.767 | 63.865 |
| $\Delta B$ | 5.913 | 5.831 | 5.705 | 5.642 | 5.773 |
| $P$/% | 50.27 | 49.62 | 49.87 | 49.82 | 49.90 |
| $\Delta P$/% | 4.772 | 4.555 | 4.461 | 4.396 | 4.546 |

由表 1 中数据可见,该算法的平均变化 bit 数和每 bit 平均变化概率都已非常接近理想状况下的 64bit 和 50% 的变化概率,相当充分和均匀地利用了密文空间,从统计效果来看,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,密文从统计上来看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 $\Delta B, \Delta P$ 标志着 Hash 混乱与散布性质的稳定性,越接近 0 就越稳定,文中算法的 $\Delta$ 都已很小,从而也可看出基于广义混沌映射切换的 Hash 构造算法对明文的混乱与散布能力强而稳定.

## 5. 算法的快速性和碰撞分析

以上算法只是迭代所选用的混沌方程不同及因混沌方程性能差异而选取的参数值不同，而算法设计的基本思想是一致的.

首先，算法与原始文本长度基本成正比，无需对原始文本进行补齐操作，迭代步数满足不等式：原始文本长度 $+3R\leqslant$迭代步数$\leqslant$原始文本长度 $+4R-1$（其中 $R=32$，若选取的混沌方程初值敏感性高时，可取 $R=16$，且可将每 3 个字符映射为一个 $-1$—1 之间的数，其余都与前面的算法相同，这样迭代的步数将进一步减少很多），以往的算法往往要将原始文本补齐一固定长度的序列，在原始文本很短的情况下也要进行相当多的计算，而本方法在原始文本很短时只需很小的迭代步数. 其次，本算法具有一定并行的特点，直接迭代出 3 个大数，每个对应最终结果的 40bit 左右，这就约相当于 5 个字节并行计算. 更重要的是，算法可以容易地改造为真正的并行，只要将原始文本分割为适当长度的子段，分别迭代，再将结果组成新的序列进行迭代，这在原始文本十分长的情况下，可以成倍地缩短时间.

所谓碰撞，是指不同的初值 Hash 映射结果相同，即发生了多对一映射. 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，即也为 0—255 的数，这样初值空间与终值空间相同. 记终值空间即像空间中任一值对应初值空间中原像的个数记为 $k$，记终值空间中具有 $k$ 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，用值域空间与定义域空间的测度之比来定量衡量碰撞发生程度，令

$$P=\frac{256-n(0)}{256}, \tag{8}$$

$P$ 的值越接近 1，碰撞程度越低，等于 1 时，完全没有碰撞发生.

基于广义混沌映射算法的碰撞分布图如图 5 所示. 图 5 中 $n(0)$ 至 $n(8)$ 依次为 81，105，61，7，2，0，0，0，0，$k>8$ 亦均为 0，$P=0.68359375$. 可见，该算法的碰撞程度较低. 参数 $P$ 的计算目前较难与其他算法比较，因为其他算法的设计多与结果长度相关，改变结果长度 Hash 性能变化难以预测，相当于重新设计算法，而在 128bit 等应用尺度上的碰撞分析，计算量过于庞大而不现实. 而本算法可以将结果取为任意长度，从而可以方便地在小尺度下进行算法碰撞程度的定量分析，这正是一个独特的优点.

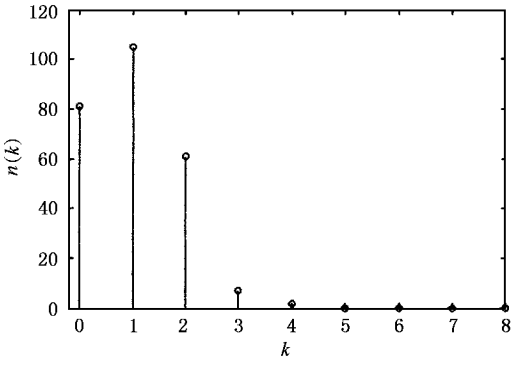


图 5　$k$-$n(k)$ 分布图

## 6. 结束语

本文研究了一种基于广义混沌映射切换的单向 Hash 函数构造方案. 这种方案首先建立能够产生多种混沌序列的广义混沌映射模型，然后在不同时段根据权值切换策略来更换产生混沌序列的混沌映射，并将待处理文本作为初值和调制信号加载于混沌映射的迭代中. 研究结果表明：(1) 混沌系统固有的特点使该算法实现简单，对初值有高度的敏感性，具有很好的单向 Hash 函数性能；(2) 利用原始文本对混沌参数的调制和混沌映射本身的变换相结合，使得混沌载波的吸引子结构无明显的单一混沌吸引子结构，从而使该 Hash 算法具有很高的安全性. 此外，该方法迭代步数与初始文本长度基本成正比，并且易于并行实现，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 and Sadeghiyan B 1993 *Design of Hashing Algorithm* (Berlin: Springer)

[3] Knudsen L and Preneel B 2002 *IEEE Trans. Inform. Theor.* **48** 2524
[4] Hayes S, Grebogi C and Ott S 1993 *Phys. Rev. Lett.* **70** 3031
[5] Heileman G L *et al* 1993 *Proceedings of International Symposium on Nonlinear Theory and Its Applications* **1** 1183
[6] Zhang J S and Xian X C 2001 *Acta Phys. Sin.* **50** 2121 (in Chinese) [张家树、肖先赐 2001 物理学报 **50** 2121]
[7] Short K M *et al* 1994 *Bifurc. Chaos* **4** 959
[8] Short K M *et al* 1997 *Bifurc. Chaos* **7** 1579
[9] Zhang J S and Xian X C 2000 *Acta Phys. Sin.* **49** 403 (in Chinese) [张家树、肖先赐 2000 物理学报 **49** 403]
[10] Zhang J S and Xian X C 2000 *Acta Phys. Sin.* **49** 1221 (in Chinese) [张家树、肖先赐 2000 物理学报 **49** 1221]
[11] Zhang J S and Xian X C 2001 *Acta Phys. Sin.* **50** 1248 (in Chinese) [张家树、肖先赐 2001 物理学报 **50** 1248]
[12] Zhang J S and Xiao X C 2000 *Chin. Phys.* **9** 408
[13] Zhang J S and Xiao X C 2000 *Chin. Phys. Lett.* **17** 88
[14] Yang T and Shao H H 2002 *Acta Phys. Sin.* **51** 742 (in Chinese) [杨 涛、邵惠鹤 2002 物理学报 **51** 742]
[15] Frey D R 1993 *IEEE Trans. Circ. Syst.*-II **40** 660

# One way Hash function construction based on the extended chaotic maps switch *

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

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

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

(Received 10 February 2003; revised manuscript received 13 March 2003)

Abstract

How to design an efficient one-way Hash function is always the hot point in modern cryptography researches. In this paper, a Hash function construction method based on extended chaotic maps switch is proposed. The extended chaotic model is first built to generate various kinds of chaotic signals at different parts of the original signals according to the switching schemes, and then chaotic parameters of one-way Hash function is modulated by the linear-transformed signals. The advantages of irreversibility, resistance to imitations and sensitivity to initial values, etc., are also discussed. Simulation results show that this chaotic Hash function based on extended chaotic maps switch has good one-way, weak collision property, better security than the chaotic Hash function based on single chaotic map, and it can be realized easily.

**Keywords**: Hash function, chaos, chaotic maps switch

**PACC**: 0545

* Project supported by the National Natural Science Foundation of China (Grant No. 60272096), the Foundation for Young Scientists of Sichuan Province, China (Grant No. 03ZQ026-033), and the Basic Research Foundation of Southwest Jiaotong University, China (Grant No. 2001B08).