一类基于拟群的 Hash 函数

池相会，徐允庆*
（宁波大学 理学院，浙江 宁波 315211）

摘要：基于 Hash 函数是用于信息安全领域中的加密算法, 因此利用剩余类环和有限域理论给出一种基于拟群运算的具有良好抗碰撞性的 Hash 函数, 并对其安全性作出分析.

关键词：拟群; Hash 函数; 置换; 抗碰撞性
中图分类号：O157.2; TP309.7　　**文献标识码：**A　　**文章编号：**1001-5132（2012）02-0072-04

密码学上的 Hash 函数通常用来构造数据的短"指纹", 一旦数据改变, 指纹就不再正确; 但即使将数据放在不安全的地方, 通过重新计算数据的指纹并验证指纹是否改变, 就能检测数据的完整性. 它的应用范围很广, 包括误差检测、消息认证、数据完整性检验和公钥密码等. 近几十年来, 由于公钥密码系统和数字签名方案的快速发展, 用于密码学的 Hash 函数变得越来越重要.

Hash 函数的单向性概念最早是由 Diffie 和 Hellman 提出[1], 单向性、强抗碰撞性和弱抗碰撞性都是 Hash 函数的重要安全性质. 一直以来, 寻找具有这些安全性质的 Hash 函数吸引着众多研究人员. 拟群可以用来加密、产生伪随机数等良好性质, 使得其能够用于 Hash 函数的构造. 1992 年, Dénes 和 Keedwell 首次运用拟群理论提出了一类新的 Hash 函数[2]. 随后, Dawson 和 Donovan 对这种 Hash 函数的安全性作了分析并得出结论: 将消息划分成"块"的方法对 Hash 函数的安全性有着重要影响[3]. 2006 年, Meyey 构造了基于非结合拟群的消息认证码[4] (带密钥的 Hash 函数). 笔者给出了一类新的基于拟群的 Hash 函数, 并证明了它具有单向性、强抗碰撞等性质.

1　相关的定义及基础知识

定义 1　Hash 函数 h 是集合 X 到集合 Y 的映射, 并满足以下性质:

(1) 压缩性: $|X|>|Y|$, 并且 $h(x)$ 的长度为固定值 n;

(2) 易于计算性: 给定 $x\in X$, $h(x)$ 容易计算;

(3) 单向性: 给定 $y\in Y$, 求 $x\in X$ 使得 $h(x)=y$ 是计算上不可行的(指计算量非常大, 以致于无法完成);

(4) 弱抗碰撞性: 给定 $x\in X$, 求 $x'\neq x$ 使得 $h(x)=h(x')$ 是计算上不可行的;

(5) 强抗碰撞性: 求 x,x' 使得 $h(x)=h(x')$ 是计算上不可行的.

定义 2　1 个拟群是 1 个有序对 $(Q,*)$, 其中 Q 是 1 个非空集合, $*$ 是 Q 上的二元运算使得对每对 $a,b\in Q$, 方程 $a*x=b$ 和 $y*a=b$ 都有唯一解.

我们称基数 $|Q|$ 为拟群 $(Q,*)$ 的阶. 关于更多有关拟群和 Hash 函数的知识, 可参阅文献[5-7].

设 n 是 1 个整数, $(\mathbb{Z}_n,+,\cdot)$ 是模 n 剩余类环, 其中 $\mathbb{Z}_n=\{0,\ 1,\ 2,\ \cdots,\ n-1\}$. 若 $Q=\mathbb{Z}_n$, $b\in Q$ 且 $b\neq 0$, 定义 Q 上的二元运算 $*$ 如下: $x*y=bx+y$ ($\forall x,y\in Q$). 这里, 我们将 $b\cdot x$ 简记为 bx.

引理 1　设 $b\in Q=\mathbb{Z}_n$ 且 $(b,n)=1$, 则 $(Q,*)$ 是 1 个拟群.

证明　若 $(b,n)=1$, 则 b 是剩余类环 $(\mathbb{Z}_n,+,\cdot)$ 中的乘法可逆元. 记其逆元为 b^{-1}, 方程 $a*x=c$ 和 $y*a=c$ 分别有唯一解 $x=c-ba$ 和 $y=b^{-1}(c-a)$. 由定义 2 可知 $(Q,*)$ 是拟群, 证毕.

例 1　令 $Q=\mathbb{Z}_6$, $b=5$ 及 $x*y=5x+y$. 则图

收稿日期：2011−10−13.　　宁波大学学报（理工版）网址: http://3xb.nbu.edu.cn
第一作者：池相会（1979－）, 男, 浙江临海人, 在读硕士研究生, 主要研究方向: 组合设计与密码学 . E-mail: 283066183@qq.com
*通讯作者：徐允庆（1959－）, 男, 河南南阳人, 博士/教授, 主要研究方向: 组合设计与密码学. E-mail: xuyunqing@nbu.edu.cn

1 中的乘法表 $(Q,*)$ 是 1 个 6 阶拟群.

$*$	0	1	2	3	4	5
0	0	1	2	3	4	5
1	5	0	1	2	3	4
2	4	5	0	1	2	3
3	3	4	5	0	1	2
4	2	3	4	5	0	1
5	1	2	3	4	5	0

图1　6 阶拟群

推论 1　设 n 是 1 个素数幂, $(\mathbb{F}_n,+,\cdot)$ 是 1 个 n 阶有限域, 其中 $\mathbb{F}_n=\{0, 1, a, \cdots, a^{n-2}\}$, a 是 $\mathbb{F}_n$ 的 1 个生成元. 若 $Q=\mathbb{F}_n$, $b\in Q$ 且 $b\neq 0$, 定义 Q 上的二元运算 $*$ 如下: $x*y=bx+y$, $\forall x,y\in Q$, 由此 $(Q,*)$ 是 1 个拟群.

设 $(Q,*)$ 是 1 个拟群, σ 和 τ 都是 Q 上的置换. Q 上的二元运算 $\otimes$ 定义如下:

$x\otimes y=a\sigma(x)+\tau(y)$ 当且仅当 $x*y=ax+y$.

需特别指出, 当 σ 和 τ 都是恒等置换时, 有 $\otimes=*$. 容易证明: $(Q,\otimes)$ 是 1 个拟群的充要条件是 $(Q,*)$ 是 1 个拟群.

引理 2　二元运算 $*_1$ 与 $*_2$ 分别定义为: $x*_1 y=a_1x+y$, $x*_2 y=a_2x+y$. 若 $(Q,*_1)$ 与 $(Q,*_2)$ 都是拟群, 则 $(Q,*_1)$ 可由 $(Q,*_2)$ 经过一个行置换得到.

证明　事实上, 只须证明: $\forall x_1\in Q$, $\exists x_2\in Q$ 使得 $\forall y\in Q$, $x_1*_1 y=x_2*_2 y$ 都成立. 由引理 1 知 a_2 可逆. 令 $x_2=a_2^{-1}(a_1x_1)$ 即可, 证毕.

在构造 Hash 函数之前, 我们先给出下面 2 个拟群向量运算的定义.

定义 3　拟群向量坐标乘法: 若 $A=(x_0,x_1,\cdots,x_{n-1}),B=(y_0,y_1,\cdots,y_{n-1})$, 定义:

$A\otimes B=(x_0\otimes y_0, x_1\otimes y_1, \cdots, x_{n-1}\otimes y_{n-1})$.

定义 4　拟群循环向量函数 $C:Q^n\to Q^n$: 若 $A=(x_0,x_1,\cdots,x_{n-1}),B=(y_0,y_1,\cdots,y_{n-1})$, 则 $B=C(A)\Leftrightarrow y_i=(\cdots((x_i\otimes x_{(i+1)\bmod n})\otimes x_{(i+2)\bmod n})\otimes\cdots\otimes x_{(i+n-1)\bmod n})$, $0\leqslant i\leqslant n-1$.

以下我们证明拟群循环向量函数 C 是 1 个单向函数. 即给定向量 $A\in Q^n$, 计算 $C(A)$ 是容易的; 但是反过来, 若只给定向量 $B\in Q^n$, 求 A 使得 $B=C(A)$ 却是计算上不可行的, 即随着 n 的增长, 计算量至少呈指数增长.

定理 1　设 $(Q,\otimes)$ 是 1 个拟群, 则函数 $C:Q^n\to Q^n$ 是单向函数.

证明　令 $A=(x_0,x_1,\cdots,x_{n-1}),B=(y_0,y_1,\cdots,y_{n-1})$. 由于 b, σ 和 τ 都是给定的, $(Q,\otimes)$ 是唯一确定的. 明显地, 对于给定的向量 A 和向量 $B=C(A)$ 也是唯一确定的, 并且只需要 $O(n^2)$ 次运算便可得到向量 B.

反过来, 给定 B, 求 A 使得 $B=C(A)$. 这相当于解下列方程组:

$$\begin{cases} y_0=(\cdots((x_0\otimes x_1)\otimes x_2)\otimes\cdots\otimes x_{n-1}), \\ y_0=(\cdots((x_1\otimes x_2)\otimes x_3)\otimes\cdots\otimes x_0), \\ \quad\vdots \\ y_0=(\cdots((x_{n-1}\otimes x_0)\otimes x_1)\otimes\cdots\otimes x_{n-2}), \end{cases}$$

由于置换是非线性的, 且 Q 的元素之间没有什么特殊的关系(除了满足拟群的要求外), 这就意味着要解出上述方程组, 就必须对 b, σ 和 τ 的所有可能值逐一假设, 然后验证才能得到方程的解. 因为 σ 和 τ 都是 n 阶置换, 所以要作的假设次数至少是 $(n!-1)[(n-1)!-1]$. 由 Stirling 公式可知其复杂度为 $O((n/e)^{2n+1})$. 证毕.

2　Hash 函数的构造

本节将给出基于拟群循环向量函数 C 的 Hash 函数 CH 的定义. 它将长度为 l 字节(l 是任意长的)的消息 M 映射成长度为 n 字节的 Hash 值($l>n$).

填充: 对消息 M 进行填充得到 M', 使 M' 长度为 n 的倍数. 具体来说, 给定消息 $M=b_1b_2\cdots b_l$ (l 是个长度 8 字节的数), 对其进行填充得到新消息 $M'=b_1b_2\cdots b_ll_1l_2\cdots l_8b_1b_2\cdots b_j$, 其中, $l=l_1l_2\cdots l_8$, 而 j 是使 $l+8+j\equiv 0 \bmod n$ 成立的最小非负整数. 令 $k=(l+8+j)/n$, 则可将 M' 表示为连续长度为 n 字节的块: $M'=M_1M_2\cdots M_k$, 其中, 每块 M_i 的长度都是 n 字节. 也可以用其他方法对 M' 进行划分, 而划分的方法对 Hash 函数的安全性质会产生一定的影响[3].

使用下述算法反复计算 Hash 函数:

从 $i=1$ 到 k, 计算 $H_i=C(M_i\otimes H_{i-1})$,

其中, 初值 $H_0=(255-n+1,255-n+2,\cdots,255)$, 它的长度为 n 字节.

最后令 $CH(M)=H_k$. 最终得到的 Hash 值 $CH(M)=H_k$ 即为所谓"指纹".

3　Hash 函数的安全性分析

这里讨论 Hash 函数的安全性主要从抗碰撞性

的强弱方面来考虑. 由于 Hash 函数的压缩性, 不可避免地会产生碰撞. 如何使碰撞概率达到最小, 是 Hash 函数设计重要问题. Hash 函数 CH 可以看作是集合 Q^+ 到 Q^n 的 1 个映射, 其中 Q^+ 是 Q 上字符串的集合，记 $|\{CH(M):M\in Q^+\}|=m$, 即 Hash 函数值域的基数为 m.

定理 2 [8] 对任意的 $M\in Q^+$, $|M|=l$, 找到 Hash 函数的碰撞的成功概率为:

$$\varepsilon=1-\left(\frac{m-1}{m}\right)\left(\frac{m-2}{m}\right)\cdots\left(\frac{m-l+1}{m}\right).$$

设 x 是 1 个小实数, 则 $1-x\approx e^{-x}$. 若 l/m 是个很的小实数, 则有:

$$\prod_{i=1}^{l-1}\left(1-\frac{i}{m}\right)\approx\prod_{i=1}^{l-1}e^{\frac{i}{m}}=e^{-\sum_{i=1}^{l-1}\frac{i}{m}}=e^{\frac{l(l-1)}{2m}},$$

将它代入定理 2 中公式, 化简可得:

$$l^2-l=2m\ln\left(\frac{1}{1-\varepsilon}\right),$$

当 l 很大时, 上式中的 $-l$ 可以忽略不计. 于是可以估计出:

$$l\approx\sqrt{2m\ln\left(\frac{1}{1-\varepsilon}\right)}.$$

由定理 2 可知, Hash 函数的值域基数越大(即 m 越大), 碰撞发生的概率就会越小. 特别是当定义 4 中的函数 C 是双射时, 其值域基数达到最大值 $m=n^n$. 另一方面, 若函数 C 是双射, 则 Hash 函数 CH 的值域呈均匀分布. 这种特性使得 Hash 函数 CH 能够抵抗以统计学原理为基础的攻击. 以下我们分析函数 C 为双射的条件.

设 $(Q,*)$ 是拟群, $A=(x_0,x_1,\cdots,x_{n-1})\in Q^n, B=(y_0,y_1,\cdots,y_{n-1})\in Q^n$. 若 $B=C(A)$, 则:

$$\begin{aligned}y_0&=(\cdots((x_0*x_1)*x_2)*\cdots*x_{n-1})=\\&(\cdots((bx_0+x_1)*x_2)*\cdots*x_{n-1})=\\&(\cdots(b^2x_0+bx_1+x_2)*\cdots*x_{n-1})=\cdots=\\&b^{n-1}x_0+b^{n-2}x_1+\cdots+x_{n-1}.\end{aligned}$$

同样地, 可以得到下列方程组:

$$\begin{cases}y_0=b^{n-1}x_0+b^{n-2}x_1+\cdots+x_{n-1},\\ y_1=b^{n-1}x_1+b^{n-2}x_2+\cdots+x_0,\\ \quad\vdots\\ y_{n-1}=b^{n-1}x_{n-1}+b^{n-2}x_0+\cdots+x_{n-2}.\end{cases}$$

计算上述方程组的系数行列式，得:

$$D=\begin{vmatrix}b^{n-1}&b^{n-2}&\cdots&b&1\\1&b^{n-1}&\cdots&b^2&b\\\vdots&\vdots&&\vdots&\vdots\\b^{n-2}&b^{n-3}&\cdots&1&b^{n-1}\end{vmatrix}=$$

$$\begin{vmatrix}b^{n-1}&b^{n-2}&\cdots&b&1\\1-b^n&0&\cdots&0&0\\\vdots&\vdots&&\vdots&\vdots\\0&0&\cdots&1-b^n&0\end{vmatrix}=(b^n-1)^{n-1},$$

当 $D\neq0$ 时, 函数 $C:Q^n\to Q^n$ 是双射. 从而我们有下面 2 个引理.

引理 3 正整数 $n=p_1^{l_1}p_2^{l_2}\cdots p_r^{l_r}$ $(r\geqslant2)$, 其中, p_i 是互不相同的素数且整数 $l_i\geqslant1, 1\leqslant i\leqslant r$. 设 $Q=\mathbb{Z}_n=\{0,1,2,\cdots,n-1\}$ 是模 n 剩余类环, $b\in Q$. 若 $(b,n)=1$, 且 $b-1$ 不是 $p_1p_2\cdots p_r$ 的倍数, 则基于拟群 $(Q,*)$ 的函数 $C:Q^n\to Q^n$ 是双射.

证明 由引理 1 知: 当 $(b,n)=1$ 时, $(Q,*)$ 是拟群. 令 $b=1+c$, 于是 $b^n=(1+c)^n=1+c^n$, 从而 $D=(b^n-1)^{n-1}=(1+c^n-1)^{n-1}=c^{n(n-1)}$. 当 $c=b-1$ 不是 $p_1p_2\cdots p_r$ 的倍数时, 易知 $D\neq0$, 证毕.

例 2 设 $n=60=2^2\cdot3\cdot5$, 则满足引理 3 中条件的 b 的集合为{7, 11, 13, 17, 19, 23, 29, 37, 41, 43, 47, 53, 49, 59}.

引理 4 n 是素数幂, $Q=\mathbb{F}_n=\{0, 1, a, \ldots, a^{n-2}\}$ 是 1 个有限域, 其中 a 是 Q 的 1 个生成元. 若 $b\in Q$ 且 $b\neq0,1$, 则基于拟群 $(Q,*)$ 的函数 $C:Q^n\to Q^n$ 是双射.

证明 由前面分析可知: 当 $b\neq0$ 时, $(Q,*)$ 是拟群. 因为 Q 是个有限域, 则 $\forall b\in Q$, $b\neq0$ 有 $b^{n-1}=1$. 当 $b\neq1$ 时, $b-1\neq0$, 从而 $D=(b^n-1)^{n-1}=(b-1)^{n-1}=1\neq0$, 证毕.

二元运算 $\otimes$ 定义如下: $x\otimes y=a\sigma(x)+\tau(y)$. 选取适当的置换 σ 和 τ, 我们可以得到一系列基于拟群 $(Q,\otimes)$ 的双射 C. 例如, 设 $n=5$, $x*y=2x+y$, τ 是恒等. 若 $\sigma\in\{I$, (1243), (1342), (1423), (01)(24), (01234), (0132), (0143), (0231), (02)(34), (02413), (0214), (0341), (0324), (03)(12), (03142), (04321), (0423), (0412), (04)(13)}, 其中, I 是恒等置换, 则基于拟群 $(Q,\otimes)$ 的函数 C 是双射. 当 $\sigma=(03)(12)$ 时, 拟群 $(Q,\otimes)$ 满足: $x\otimes y=a\sigma(x)+y=3x+y$.

若定义 4 中函数 C 是双射, $|Q|=n$, 则有 $m=n^n$ 且

$$l \approx \sqrt{2m\ln(1/(1-\varepsilon))} = \sqrt{2n^n\ln(1/(1-\varepsilon))} = \sqrt{2\ln(1/(1-\varepsilon))}n^{n/2}.$$

由于 ε 是找到碰撞的成功概率，不妨限定 $\varepsilon \in [1/1000, 1/2]$，则有 $l \geqslant 0.045 \times n^{n/2}$．即以概率 $\varepsilon = 1/1000$ 成功找到 1 个碰撞所需输入的消息数量 $l \geqslant 0.045 \times n^{n/2}$．$l$ 的增长比指数增长要快，从而基于双射 C 的 Hash 函数 CH 具有较强的抗碰撞能力.

当 C 是双射时，对于取值不同的 ε，l 和 m 之间的关系见表 1．容易看到：l 的值受 ε 值的影响有限(数量级在 2 个单位以内，差别不大)．下面来考虑 n 值对 l 值的影响．表 2 列出了 ε=0.5 时，n 取不同值时 l 的取值．容易看到：n 值对 l 值的影响很大(比指数增长还要快得多)．即使取 $n = 40$，以概率 $\varepsilon = 1/1000$ 找到一个碰撞所需输入的消息数量将达到 1.29×10^{32}．因此，找到碰撞是目前计算上不可行的(目前世界上计算机的最快运行速度为 8.162×10^{15} 次·s^{-1}).

表 1　ε, l 和 m 之间的关系

ε	0.01	0.1	0.125	0.25	0.5	0.75
$l/\sqrt{m}$	0.14	0.46	0.52	0.78	1.17	1.67

表 2　n 与 l 间的关系

n	20	30	40	100	256
l	1.20×10^{13}	1.68×10^{22}	1.29×10^{32}	1.17×10^{100}	2.1×10^{308}

4　总结

Hash 函数 CH 有很多优点：它只有 3 个参数，所需内存很小；计算速度快(计算次数为 $O(n^2)$)；可以非常容易得到一系列新的 Hash 函数(只须改变这 3 个参数即可)；另外，CH 的一个突出优点是输出长度是任意的并且长度是可以改变的，这样我们就可以按实际需要随时改变输出的长度.

Hash 函数 CH 具有很好的强抗碰撞性质，能够抵抗生日攻击，可以选择输出长度的良好性质使得它可以广泛地应用到实际中．拟群可以用来保证安全通信，产生伪随机码，这种基于拟群的 Hash 函数 CH 也用到了相似的原理，所以它也可以应用于这些领域．此外，函数 CH 还可以用于消息认证码、数字签名.

参考文献:

[1] Diffe W, Hellman M. New directions in cryptography [J]. IEEE Transactions Information Theory, 1976, 26(2): 644-654.

[2] Dénes J, Keedwell A D. A new authentication scheme based on Latin squares[J]. Discrete Math, 1992, 106/107: 157-161.

[3] Dawson E, Donovan D, Offer A. Quasigroups, isotopisms, and authentication schemes[J]. Australasian Journal of Combinatorics, 1996, 13:75-88.

[4] Meyer K A. A new message authentication code based on the non-associativity of quasigroup[D]. Iowa: Iowa State University, 2006.

[5] Bakhtiari S, Safavi N R, Pieprzyk J. Cryptographic Hash functions: A survey[C]//Department of Computer Science, Universityof Wollongong, 1995.

[6] Markovski S, Gligoroski D, Bakeva V. Quasigroup string processing: Part 1. contributions[J]. Sec Math Tech Sci, 1999, 20 (1/2):13-28.

[7] Markovski S, Kusakatov V. Quasigroup string processing: Part 2. contributions[J]. Sec Math Tech Sci, 2000, 21 (1/2):15-32.

[8] Stinson D R. Cryptography theory and practice[M]. 3rd ed. New York: CRC Press, 2006:123-126.

A Class of Hash Functions Based on Quawigroups

CHI Xiang-hui, XU Yun-qing*

(Faculty of Science, Ningbo University, Ningbo 315211, China)

Abstract: Hash functions are encryption algorithms used in information security. In this paper, using the theory of finite field and residue class ring, a class of hash functions based on quasigroups is given, the analysis of the security is also presented.

Key words: quasigroup; hash function; permutation; collision resistance

（责任编辑　章践立）