布尔函数的代数攻击

杨文峰，胡予濮，高军涛

(西安电子科技大学计算机网络与信息安全教育部重点实验室 西安 710071)

【摘要】基于代数攻击，提出了一种已知部分真值表还原整个布尔函数的方法。对于n元d次布尔函数，该方法的空间复杂度和数据复杂度均为$O(N)$，计算复杂度为$O(N^3)$，其中 $N=1+C_n^1+C_n^2+\cdots+C_n^d$ 。由复杂度可知，所求密码函数的代数次数越低，该方法的有效性越高。攻击方法表明密码设计中应该谨慎使用代数次数较低的布尔函数。

关 键 词 代数方法； 布尔函数； 密码分析； 密码学

中图分类号 TN918.1 **文献标识码** A doi:10.3969/j.issn.1001-0548.2010.06.006

Algebraic Attack on Boolean Functions

YANG Wen-feng, HU Yu-pu, and GAO Jun-tao

(Key Laboratory of Computer Networks and Information Security Ministry of Education, Xidian University Xi'an 710071)

Abstract Based on algebraic attack, a new reconstruction method of Boolean functions from the partial truth table is proposed. For the Boolean function with n variables and the degree of d, the proposed method requires $O(N)$ values in the truth table, and the computational complexity is $O(N^3)$, and the memory complexity is $O(N)$, where $N=1+C_n^1+C_n^2+\cdots+C_n^d$. From the above complexity, the lower the degree of the Boolean function is，the more efficient the method is. The proposed attack shows the designer of stream cipher should use Boolean functions with low degree carefully.

Key words algebraic approach; Boolean function; cryptanalysis; cryptology

作为许多流密码的核心部件，布尔函数的设计和分析一直是密码学中极为活跃的研究领域。在密码算法设计中，设计者通过构造满足各种密码学指标的布尔函数以增加密码的强度。而在密码分析中，攻击者往往通过寻找并利用布尔函数的设计漏洞和弱点实现对密码算法的攻击。密码设计者和攻击者这种坚持不懈的攻防争斗正是密码学发展的源动力所在[1-3]。各种密码攻击方法的不断出现迫使布尔函数的设计必须满足多方面的指标，而各种指标之间的相互制约关系往往又为攻击者寻找新的攻击方法提供了机会。本文提出的攻击方法就是一个基于该攻击思想的实例。

在流密码分析中，攻击者常常会遇到以下两种情形：一种情形是攻击者无法通过公开途径获得密码算法，尽管文献[4]指出流密码的安全性不应该依赖于密码算法的保密性，而应仅依赖于密钥的保密性，但实际中使用的流密码算法往往是不公开的[1]。另一种情形是密码算法使用的布尔函数是由密钥控制的[5]。对于这两种情形，攻击者的任务是不但要恢复密钥而且必须恢复布尔函数。密码分析过程中，布尔函数的恢复往往都是利用统计方法求取其真值表。统计方法的致命弱点是所需数据量巨大，使得攻击者往往因为数据量有限而只能求得布尔函数的部分真值表。因此，如何通过布尔函数的已知部分特性恢复出完整的布尔函数具有现实的应用背景。代数攻击思想[6-7]和布尔函数指标之间的内在制约关系为本文提供了一种基于部分真值表的布尔函数还原方法。只要布尔函数真值表的已知部分包含了确定该函数的所需的全部或者接近全部信息量，本文的攻击方法可以唯一确定该函数的代数表达式。布尔函数的代数表达式和真值表之间存在着一一对应关系，求出代数表达式就等于确定了所求的布尔函数。

1 密码学中的布尔函数

首先给出布尔函数的定义和两种表示方法，然后讨论布尔函数的相关密码学指标，最后阐述本文的攻击思路。

收稿日期：2009－04－10； 修回日期：2009－07－08

基金项目：国家自然科学基金(60833008，60803149)；国家973计划(2007CB311201)

作者简介：杨文峰(1971－)，男，博士生，副教授，主要从事密码学方面的研究.

定义 记二元域{0,1}为GF(2)，设n是任意一个正整数，称$GF(2)^n \to GF(2)$的任一映射为变元为n的($GF(2)^n$上的)布尔函数。即若记$x=(x_1,x_2,\cdots,x_n)\in GF(2)^n$，则有$f(x)=f(x_1,x_2,\cdots,x_n)\in GF(2)$。

布尔函数主要有真值表表示、小项表示、多项式表示和Walsh谱表示4种表示方法。本文只考虑真值表表示和多项式表示。

所谓的真值表表示也就是用穷举法将映射$f: GF(2)^n \to GF(2)$一一排列出来，即：

$$(f(0\cdots00),f(0\cdots01),\cdots,f(1\cdots11))$$

多项式表示法则是指任意n元布尔函数$f(x)$都可用其代数表达式表示为：

$$f(x)=a_0+a_1x_1+a_2x_2+\cdots+a_nx_n+a_{12}x_1x_2+\cdots+a_{n-1n}x_{n-1}x_n+\cdots+a_{12\cdots n}x_1x_2\cdots x_n$$

布尔函数的各种表示方法之间是相互等价的，只要知道了其中的一种表示方法就可以唯一确定其他的表示方法。

为了防止已知的各种攻击方法，密码学中使用的布尔函数必须满足一定的密码学指标，其中主要有次数(代数次数和非线性次数)、均衡性、相关免疫性、严格雪崩准则和扩散准则等。已有的研究结果表明，布尔函数各项指标之间是相互冲突、相互制约的，所以密码设计者总是无法设计出各项指标都能达到最优的布尔函数。由于攻击者是寻求利用布尔函数的最弱的一点采取攻击，这就使得设计者设计布尔函数时不得不采取各项密码学指标折衷的思想。充分挖掘利用密码学指标之间的制约关系正是解决本文中问题的基本思路。

n元均衡布尔函数$f(x)$的代数次数d和相关免疫阶m就存在着严格的制约关系，即如下定理。

定理 1[8] 若n元d次布尔函数$f(x)$，$x\in GF(2)^n$是m阶相关免疫的，则$d+m\leqslant n$；若更设$f(x)$是均衡的，且$1\leqslant m\leqslant n-2$，则$d+m\leqslant n-1$。

由上述定理可以看出，布尔函数的代数次数d必须满足$d\leqslant n-1-m$，即布尔函数的相关阶m越大，代数次数d就越小。当设计者为了防止相关攻击[9-10]而不得不提高布尔函数的相关免疫阶和牺牲布尔函数的次数时，就为利用布尔函数的部分真值表还原整个布尔函数提供了有利条件。只要已知的部分真值表中包含了$d(d\leqslant n-1)$次布尔函数的全部信息，应用代数攻击的思想首先求出该函数的代数表达式，然后把真值表中未知函数值的自变量代入代数正规型就可以补全其真值表。

2 攻击方法

代数攻击[6]是一种密码分析方法，常用于对流密码、分组密码及HFE公钥密码系统的安全性分析。代数攻击方法的步骤为：(1) 通过分析利用密码算法的代数结构，把密码算法的安全性(密钥恢复)规约为一个次数尽可能低的超定多元方程组的求解问题；(2) 求解步骤(1)中的多元高次方程组，代数攻击的复杂度主要取决于多元方程组的求解复杂度，现有的求解方法主要有Linearization、Relinearization、XL、Gröbner Base等；(3) 如果步骤(2)中顺利求出了多元方程组的解，结合步骤(1)中分析的密码算法的代数特征，密码算法的安全性就无法保证。

从上述代数攻击的基本思想可以看出，所谓的代数攻击实际上就是一个利用密码系统的代数结构建立方程组进而求解的过程。本文的攻击方法就是利用布尔函数的部分真值表和代数表达式建立一个(线性)方程组，然后求解方程组得到代数表达式的系数。主要分为以下3个步骤。

1) 根据真值表中已知函数值的个数$M(M<2^n)$，首先确定可求的n元布尔函数的最高次数d。求出满足不等式$1+C_n^1+C_n^2+\cdots+C_n^d \leqslant M$的最大的$d$，$d$即为所求函数的最大可能次数。布尔函数的代数表达式可表示为：

$$f(x)=a_0+a_1x_1+a_2x_2+\cdots+a_nx_n+a_{12}x_1x_2+\cdots+a_{n-1n}x_{n-1}x_n+\cdots+a_{n-d+1\cdots n}x_{n-d+1}x_{n-d+2}\cdots x_n$$

如果根据真值表的已知部分和上述代数表达式能够确定出多项式的系数a_0、a_1、…、$a_{n-d+1\cdots n}$，就求出了布尔函数。

2) 建立多元线性方程组，根据真值表中已知的函数值，把对应的自变量和函数值分别代入步骤1)的布尔函数的代数表达式，得到一个包含M个线性方程的多元线性方程组。

3) 求解线性方程组，利用Gauss消元法对步骤2)中得到的线性方程组进行化简求解。假设方程组中含有Q个线性无关的方程，求解的结果有4种情况。

(1) $Q<N$并且没有自由变量。说明该布尔函数的代数次数小于d，并且真值表包含的信息量足够唯一确定该布尔函数。这种情形下，根据化简的结果可直接写出所求布尔函数的代数表达式。

(2) $Q<N$并且含有自由变量。说明布尔函数的代数次数小于或者等于d，但是真值表中包含的信息量少于唯一确定该布尔函数所需的信息量。可以通过

穷举自由变量的办法获取多个布尔函数，其中有且只有一个为真。这种情形下，只有结合真值表以外的其他信息才能唯一确定布尔函数。

(3) $Q=N$。说明该布尔函数的代数次数正好为d并且真值表中包含所求布尔函数的信息量足够唯一确定布尔函数。

(4) $Q>N$。说明所求布尔函数的代数次数大于d，利用已在的真值表无法确定该布尔函数。

该攻击方法实现简单，计算复杂度低，有利于工程实现。攻击方法的主要计算就是用Gauss消元法求解一个线性方程组，其计算复杂度就是Gauss消元法的计算复杂度，即$O(N^3)$。对于n元d次布尔函数来说由于方程组中未知元的系数是由1、x_1、x_2、…、x_n、x_1x_2、…、$x_{n-1}x_n$、…、$x_{n-d+1}x_{n-d+2}\cdots x_n$组成的，输入变量的相互纠缠使得方程组中的线性无关的方程个数不能精确计算，也就无法给出唯一确定布尔函数的M的精确估计。但可以利用布尔函数的理论给出M的上下界。由攻击方法易知M的下界就是N。

定理 2[8]　任意两个不同的次数小于等于d的n元布尔函数的汉明距离至少为2^{n-d}。

由上述定理可知 M 的上界就是2^{n-d}。

3　攻击实例

举例说明攻击方法的有效性。假设已知8元布尔函数的部分真值表如下(自变量按照 $x_8,x_7,x_6,x_5,x_4,x_3,x_2,x_1$的字典序排列)：

...010.1.001..101.0..110....1001
.10.10..1.100..1....0..1.10.101.
011010.10.1..0.110..011010010110
..1..0.110.1..10.0010.100.10100.
.11..0.11001...0.11.10....01.11.
0011..001.000011.10000.10011.100
...001.0.00..00.10.1...1011.01.0
.1101.0.10.1011010.1011.0110.0..

首先根据攻击方法中的步骤1)，M=163，计算

$$1+C_8^1+C_8^2+\cdots+C_8^4=163\leqslant M$$

得到d=4。由此可以确定如果该布尔函数的代数次数$d\leqslant 4$，通过本文的攻击方法就可基本求出函数的代数表达式；如果$d>4$，本攻击方法无效。假设所求布尔函数的代数表达式为：

$$f(x)=a_0+a_1x_1+a_2x_2+\cdots+a_8x_8+a_{12}x_1x_2+\cdots+a_{78}x_7x_8+\cdots+a_{5678}x_5x_6x_7x_8$$

攻击的目的就是求出上述代数正规型中的所有系数a_0、a_1、a_2、…、a_8、a_{12}、…、a_{78}、…、a_{5678}。

根据攻击方法中的步骤(2)，把真值表中有函数值的x和$f(x)$代入上述代数表达式，得到一个包含163个线性方程的线性方程组。

接下来实施攻击方法中的步骤(3)，利用Gauss消元法化简上述方程组后，获得的方程组的秩为162，该方程组有一个自由变量a_{5678}，这属于攻击方法中的第二种情形。通过穷举自由变量a_{5678}得到两个布尔函数的代数表达式分别为：

$$f_1(x_1,x_2,x_3,x_4,x_5,x_6,x_7,x_8)=$$
$$x_1+x_2+x_3+x_4+x_5+x_2x_6+x_4x_7+x_5x_8+x_1x_6x_8+$$
$$x_2x_6x_7+x_2x_6x_8+x_3x_7x_8+x_4x_6x_7+x_4x_7x_8+x_5x_6x_8+$$
$$x_5x_7x_8+x_1x_6x_7x_8+x_1x_6x_7x_8+x_2x_6x_7x_8+x_3x_6x_7x_8+$$
$$x_4x_6x_7x_8+x_5x_6x_7x_8$$
$$f_2(x_1,x_2,x_3,x_4,x_5,x_6,x_7,x_8)=$$
$$x_1+x_2+x_3+x_4+x_5+x_2x_6+x_4x_7+x_5x_8+x_6x_7+$$
$$x_1x_6x_7+x_1x_6x_8+x_1x_7x_8+x_2x_3x_8+x_2x_5x_6+$$
$$x_2x_6x_8+x_3x_4x_8+x_3x_6x_7+x_3x_7x_8+x_4x_5x_6+$$
$$x_4x_7x_8+x_5x_6x_7+x_5x_6x_8+x_5x_7x_8+x_6x_7x_8+$$
$$x_1x_2x_3x_6+x_1x_2x_3x_8+x_1x_2x_5x_6+x_1x_2x_5x_8+$$
$$x_1x_2x_6x_7+x_1x_2x_7x_8+x_1x_3x_4x_6+x_1x_3x_4x_8+$$
$$x_1x_3x_6x_7+x_1x_3x_7x_8+x_1x_4x_5x_6+x_1x_4x_5x_8+$$
$$x_1x_4x_6x_7+x_1x_4x_7x_8+x_1x_5x_6x_7+x_1x_5x_7x_8+$$
$$x_1x_6x_7x_8+x_2x_3x_5x_6+x_2x_3x_5x_8+x_2x_3x_6x_7+$$
$$x_2x_3x_6x_8+x_2x_3x_7x_8+x_2x_5x_6x_8+x_3x_4x_5x_6+$$
$$x_3x_4x_5x_8+x_3x_4x_6x_7+x_3x_4x_6x_8+x_3x_4x_7x_8+$$
$$x_3x_5x_6x_7+x_3x_5x_7x_8+x_4x_5x_6x_8。$$

从已知的真值表所能提供的信息量看，应用本文的攻击方法无法确定所求布尔函数的代数表达式是$f_1(x)$还是$f_2(x)$。但是结合其他信息，例如流密码的简洁性原则或者密码学中布尔函数设计的经验，不难看出：

$$f_1(x_1,x_2,x_3,x_4,x_5,x_6,x_7,x_8)=(x_1,x_2,x_3,x_4,x_5)\cdot\varphi(x_6,x_7,x_8)$$

其中逻辑函数$\varphi(x_6,x_7,x_8)$的真值表如表1所示。

表1　逻辑函数$\varphi(x_6,x_7,x_8)$的真值表

$x_6x_7x_8$	000	001	010	011	100	101	110	111
$\varphi(x_6,x_7,x_8)$	11111	11110	11101	11011	10111	01111	11111	11111

这就是著名的Maiorana-McFarland函数[11-12]。由此可以得出，$f_1(x)$即为所求的布尔函数。当然，这只是对所攻击的布尔函数一种判断方法，最终正确与否还要结合密码攻击的实践进行判断。

4　结 束 语

流密码的安全往往取决于其密钥流生成器中布尔函数的设计质量。在未知密码算法或者由密钥控

制布尔函数的流密码分析过程中，布尔函数的恢复是首先要解决的问题。另外，密码分析中的许多问题都可以转化为布尔函数的分析。当攻击者占有的数据有限时，只能求出部分真值表。本文利用布尔函数的次数和相关免疫阶的制约关系提出了一种基于代数攻击思想的布尔函数还原方法。当已知的部分真值表包含所求布尔函数的全部信息时，该方法就可以唯一确定布尔函数。攻击方法的计算复杂度低，实现简单。但该攻击方法的复杂度与布尔函数阶的关系、小项数及其分布的关系等如何给定严格的界限，还有待进一步研究。

参 考 文 献

[1] KAHN D. The codebreakers: the story of secret writing[M]. New York: Macmillan, 1967.

[2] 汪小芬, 李胜强, 肖国镇. 认证群密钥协商协议的安全性分析与改进[J]. 电子科技大学学报, 2009, 38(1): 51-54.
WANG Xiao-fen, LI Sheng-qiang, XIAO Guo-zhen. Analysis and improvement of an authenticated group key agreement protocol[J]. Journal of University of Electronic Science and Technology of China, 2009, 38(1): 51-54.

[3] ZHANG J L, WANG Y M. Efficient membership revocation in ACJT group signature[J]. Journal of University of Electronic Science and Technology of China, 2008, 6(1): 39-42.

[4] SCHNEIER B. Applied cryptography second edition: protocols, algorithms, and source code in C[M]. [S.l.]: John Wiley & Sons, 1996.

[5] GOLIĆ J D, MORGARI G. On the resynchronization attack [C]//Proceedings of FSE'03, LNCS 2887. Berlin: Springer-Verlag, 2003: 100-110.

[6] COURTOIS N T, MEIER W. Algebraic attacks on stream ciphers with linear feedback[C]//Proceedings of Euro-crypt'03, LNCS 2656. Berlin: Springer-Verlag, 2003: 345-359.

[7] COURTOIS N T. Fast algebraic attacks on stream ciphers with linear feedback[C]//Proceedings of Crypt'03, LNCS 2729. Berlin: Springer-Verlag, 2003: 176-194.

[8] CLARLET C. Boolean functions for cryptography and error correcting codes[DB/OL]. [2009-9-10]. http://www-rocq.inria.fr /secret/Claude.Carlet/chap-fcts-Bool.pdf.

[9] SIEGENTHALER T. Decrypting a class of stream ciphers using ciphertext only[J]. IEEE Transactions on computers, 1985, 34(1): 81-85.

[10] MEIER W, STAFFELBACH O. Fast correlation attacks on stream ciphers[C]//Proceedings of Euro-crypt'88, LNCS 330. Berlin: Springer-Verlag, 1988: 301-314.

[11] GUPTA K C, SARKAR P. Efficient representation and software implementation of resilient Maiorana-McFarland S-boxes[C]//Proceedings of WISA'04, LNCS 3325. Berlin: Springer-Verlag, 2004: 317-331.

[12] CLARLET C. A larger class of cryptographic Boolean functions via a study of the Maiorana-McFarland construction[C]//Proceedings of Crypt'02, LNCS 2442. Berlin: Springer-Verlag, 2002: 549-564.

编 辑 税 红