doi:10.13582/j.cnki.1674-5876.2016.01.015

代数方法在分组密码分析中的应用

杨璇,刘金旺

(湖南科技大学 数学与计算科学学院,湖南 湘潭 411201)

摘 要:主要对分组密码的不可能差分分析中所使用到的一些代数方法进行了研究,包括线性方程组的求解、布尔代数、有限域上的多项式理论以及Groebner基理论等;对原有的代数自动化搜索方法进行了改进,改进后的算法能更好的评价分组密码抵抗不可能差分分析的能力,为矿山的系统安全提供技术保障.

关键词:分组密码;不可能差分分析;代数方法

中图分类号:O15 **文献标志码**:A **文章编号**:1672-9102(2016)01-0072-04

Application of algebraic methods into cryptanalysis of block cipher

YANG Xuan, LIU Jinwang

(School of Mathematics and Computational Science, Hunan University of Science and Technology, Xiangtan 411201, China)

Abstract: This paper mainly analyzes algebraic methods that are applied in impossible difference cryptanalysis of block cipher, including linear equations, Boolean algebra, polynomial theory in finite fields and groebner basis theory. Improving the original method of algebra automated search, the improved algorithm can better evaluate the ability of block cipher to resistance impossible differencial cryptanalysis and provide technical support for system safety of the mines.

Key words: block cipher; impossible difference cryptanalysis; algebraic methods

密码学是信息安全的核心技术,在计算机科学、电子通信以及矿业工程的系统安全分析和安全评价中都有广泛应用.分组密码是密码学的重要组成部分,分组密码的研究主要包括分组密码的设计与分析2个方面.在分组密码的分析研究中,不可能差分分析[1-3]作为差分[4]分析的变体,是分组密码分析的重要的分析方法.本文就在寻找分组密码不可能差分区分器[5,6]的过程中所使用到的一些代数方法进行了研究,包括线性方程组的求解、布尔代数[7]、有限域上的多项式理论[8]以及Groebner基理论等,这些代数方法的综合运用使得密码分析和实现成为可能.

1 求解不可能差分区分器的数学模型

在基于S盒的迭代型分组密码中,其基本部件通常包括2大部分:线性变换层P和非线性代换层S盒层.基于这2个基本部件可以建立差分传播系统Φ,差分传播系统包括线性部分Φ_L和非线性部分Φ_P,这个系统描述了全轮分组密码的差分传播行为.

以SPN算法为例,对于1个分组长度为n的R轮SPN结构的分组密码ε,假设第i轮S盒的输入和输出差分分别为$X_i=(x_{i,1},x_{i,2},\cdots,x_{i,n})$,$Y_i=(y_{i,1},y_{i,2},\cdots,y_{i,n})$,差分传播系统建立如下:

收稿日期:2015-12-08

基金项目:湖南省研究生科研创新资助项目(CX2015B463)

通信作者:刘金旺(1964-),男,湖南常德人,博士,教授,研究方向:代数与符号计算. E-mail: jwliu64@aliyun.com

$$\begin{cases} \Phi_{\mathrm{P}_{ij}} & \text{其中 } l \leqslant i \leqslant R, 1 \leqslant j \leqslant m; \\ X_{i+1}^{T} \oplus P \cdot Y_i^{T} = 0 & \text{其中 } l \leqslant i \leqslant R-1. \end{cases} \tag{1}$$

这里 $\Phi_{\mathrm{P}_{ij}}$ 是第 i 轮第 j 个 S 盒对应的非线性表达式,$\boldsymbol{P}(n \times n)$ 是置换层的系数矩阵,在这个系统中,X_1 是输入差分,Y_R 是输出差分.

不可能差分和差分传播系统的关系由以下定理给出:

设 R 轮分组密码算法 ε 的差分传播系统为 Φ,ΔP 和 ΔC 分别是 ε 的输入差分和输出差分,$\Phi(\Delta P, \Delta C)$ 表示用 ΔP 和 ΔC 初始化的差分传播系统.

定理 1. $\Phi(\Delta P, \Delta C)$ 无解当且仅当 $\Delta P \rightarrow \Delta C$ 是 1 条 R - 轮不可能差分,记作 $\Delta P \nrightarrow_R \Delta C$.

证明:充分性.(反证法)假设 $\Delta P \rightarrow \Delta C$ 是一条 R - 轮可能差分,则存在某个密钥 K 使得差分方程 $\varepsilon_K(x) \oplus \varepsilon_K(x \oplus \Delta P) = \Delta C$ 有解,不妨设 $(p, p \oplus \Delta P = p')$ 是该方程的一对解,将它们在 ε_K 加密下的每一轮中间状态作差分,将全部 R 轮中间状态的差分提取出来即为 $\Phi(\Delta P, \Delta C)$ 的解,矛盾.

必要性.(反证法,这里以 SPN 算法为例)假设 $\Phi(\Delta P, \Delta C)$ 有解,不妨设 $(\Delta P = X_0, X_1, Y_1, X_2, Y_2, \cdots, X_R, Y_R = \Delta C)$ 是 $\Phi(\Delta P, \Delta C)$ 的 1 个解,将中间状态所有输入差分提取出来得到 $(\Delta P = X_0, X_1, X_2, \cdots, X_R, \Delta C)$,这是一条以 ΔP 为起点以 ΔC 为终点且概率不为 0 的差分特征,而差分 $(\Delta P, \Delta C)$ 的概率等于所有以 ΔP 为起点以 ΔC 为终点的差分特征的概率总和,因此 $\Delta P \rightarrow \Delta C$ 是一条可能差分,矛盾.

因此,要说明 $\Delta P \rightarrow \Delta C$ 是不可能差分,只需要证明 $\Phi(\Delta P, \Delta C)$ 无解即可,这样就完全把不可能差分的判别转化到差分传播系统的求解上.

2　模型求解中的代数方法

2.1　线性子系统的求解

在差分传播系统式(1)的求解中,最基本的便是线性方程子系统的求解. 对于一个初始化的分组长度为 n 的 R 轮 SPN 结构的分组密码 ε,其差分传播系统的线性子系统由 $(2R-2)n$ 个变量和 $(R-1)n$ 个线性方程组组成,这是一个庞大的线性方程组,因为,即使对于分组长度为 64 比特的轻量级分组密码而言,当 $R=8$ 时,该线性方程组也涉及到 896 个变量和 448 个线性方程,而对于这样庞大的线性方程组,手工计算几乎无法办到. 另外,由于线性方程组中所有的未知数均为比特值,因此线性方程组的求解还必须在有限域 GF(2)上进行.

高斯消元法能有效的解决线性子系统的求解问题,这是因为高斯消元法不仅可以在任何有限域中进行,也可以通过计算机编程实现. 对于一个具有 m 个未知数和 s 和方程的线性方程组

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1m}x_m = b_1; \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2m}x_m = b_2; \\ \cdots\cdots\cdots\cdots \\ a_{s1}x_1 + a_{s2}x_2 + \cdots + a_{sm}x_m = b_m. \end{cases} \tag{2}$$

其中 $a_{ij}, b_j \in GF(2), 1 \leqslant i \leqslant s, 1 \leqslant j \leqslant m$.

通过提取式(2)的系数构造增广矩阵式(3),然后对增广矩阵进行初等变换将其化为最简阶梯形式(4),这样就求得了线性方程组的解.

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2m} & b_2 \\ \vdots & \vdots & & \vdots & \vdots \\ a_{s1} & a_{s2} & \cdots & a_{sm} & b_s \end{pmatrix} \quad (3) \quad \Rightarrow \quad \begin{pmatrix} c_{11} & 0 & \cdots & c_{1r} & \cdots & c_{1m} & d_1 \\ 0 & c_{22} & \cdots & c_{2r} & \cdots & c_{2m} & d_2 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & c_{rr} & \cdots & c_{rm} & d_r \\ 0 & \cdots & 0 & d_{r+1} & & & \\ 0 & & & \cdots & & & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & & & \cdots & & & 0 \end{pmatrix}. \tag{4}$$

2.2 非线性子系统的求解

差分传播系统的线性子系统可以通过高斯消元法完全求解,但是由于线性子系统中的变量个数远大于线性方程组的个数,在实际求解过程中还需要从非线性系统中获取更多的信息来判断矛盾的发生.

2.2.1 非线性子系统的代数表达

非线性子系统的信息恢复一直是不可能差分分析的重点和难点,此前的 U - 方法[9]和 UID - 方法[10]都是基于"中间相错"思想来构造不可能差分,但是这种方法容易错失非线性系统更多的信息.

为了避免非线性系统信息的丢失,通常的做法是将 S 盒的差分分布表转化为真值表来构造对应的布尔函数多项式. 对于一个 $r \times r$ 的 S 盒(输入和输出差分分别设为 $X = (x_1, x_2, \cdots, x_r)$, $Y = (y_1, y_2, \cdots, y_r)$),其布尔多项式构造的具体步骤如下:

步骤 1. 求出 S 盒的差分分布表 $\boldsymbol{T}$.

步骤 2. 将差分分布表中的所有非零元替换为 1,所有取值为零的元素保持不动,得到布尔函数真值表 $\boldsymbol{T}'$.

步骤 3. 从真值表 $\boldsymbol{T}'$ 中恢复布尔函数 f,f 可以通过布尔函数的小项表示[7]得到.

注:由于不可能差分分析只利用了差分分布表中的所有取值为零的元素,所以步骤 2 中将差分分布表中所有非零元替换为 1 并不会影响不可能差分分析的结果.

此时得到的布尔多项式 f 是一个代数次数为 $2r$,且包含 2^{2r} 项的高次多元多项式,它满足 S 盒的所有输入比特与输出比特之间的关系,即: $X \to Y$ 是不可能差分当且仅当 $f(x_1, \cdots, x_r, y_1, \cdots, y_r) \oplus 1 \neq 0$.

另外,由于 f 中的变量都是单个比特,只能在 0 和 1 之间取值,即满足 $x_i^2 + x_i = 0, y_i^2 + y_i = 0, (1 \leqslant i \leqslant r)$. 由此,我们得到了一个由 $2r$ 个变量和 $2r+1$ 个多项式组成的多元多项式集合 $\boldsymbol{I}$:

$$\boldsymbol{I} = \{f + 1, x_i^2 + x_i, y_i^2 + y_i\} \quad i = 1, \cdots, r. \tag{5}$$

在这个集合 $\boldsymbol{I}$ 中,所有多项式的系数只能在有限域 GF(2) 中取值,因此,集合 $\boldsymbol{I}$ 又构成了有限域 GF(2) 上多元多项式环的一个理想. 至此,差分传播方程的非线性子系统最终可以描述为有限域GF(2)上多元多项式环的理想 $\boldsymbol{I}$:

$$\boldsymbol{I} = < f + 1, x_i^2 + x_i, y_i^2 + y_i > \quad i = 1, \cdots, r. \tag{6}$$

通过将非线性系统多项式化,将复杂的分组密码划归为一个代数问题,并且利用目前已有的相关代数理论研究成果加以研究,是解决分组密码不可能差分分析的有效途径.

2.2.2 非线性子系统的线性化与 Groebner 基理论

求解非线性子系统的最终目的是为了获取更多比特的信息,从而更好的用于不可能差分分析的矛盾判定. 差分传播系统的非线性子系统虽然可以描述为有限域 GF(2) 上多元多项式环的一个理想,但是这个理想中的多项式都是复杂的高次多元多项式,要直接求解非常困难. 由于计算机能解决的更多的是线性代数的问题,这里考虑通过将高次多元多项式降阶的方式,来获取对应的线性表达式,进而得到更多的有用信息.

计算代数中的 Groebner 基理论可以有效的将多元多项式环理想中的多项式进行降阶处理,在上述理想 $\boldsymbol{I}$ 中,以分次字典序为多项式序,在有限域 GF(2) 上求取理想 $\boldsymbol{I}$ 的约化 Groebner 基 G. 此时,G 中的多项式就是对理想 $\boldsymbol{I}$ 中多项式进行充分降次后得到的多项式,并且 G 生成的理想就是 $\boldsymbol{I}$.

在理想 $\boldsymbol{I}$ 的约化 Groebner 基 G 中出现线性的表达式就是通过非线性系统分析得到的"更多信息",把所有这些线性表达式添加到差分传播系统的线性子系统中可以更好的对线性子系统进行求解. 最后,将线性系统求得的新的解代入到非线性系统中再次求取对应理想的约化 Groebner 基,反复进行这个操作,直到不能获取更多信息为止.

通上述方法可以充分利用 Groebner 基理论求解差分传播系统,而且不会遗漏掉任何有用信息. 但是,由于高次多元多项式方程组求解的复杂性高,目前通过 Groebner 基理论只能实现差分传播方程非线性子系统的部分求解,非线性子系统的更多信息恢复甚至完全求解还有待更进一步的研究.

3　结论

1)差分传播系统无解当且仅当 $\Delta P \rightarrow \Delta C$ 是 1 条不可能差分.

2）利用布尔函数的小项表示将差分传播系统的非线性系统转化为有限域上的多元多项式环中的理想,有利于进一步的研究.

3）通过 Groebner 基理论对非线性系统进行了部分求解,这是目前通过代数方法求解非线性方程结果最好的方法.

4）利用代数方法对分组密码进行不可能差分分析比传统的 U－方法和 UID－方法得到的结果更好.

参考文献:

[1] Nyberg K, Knudsen L R. Provable security against differential cryptanalysis[J]. Journal of Cryptology, 1995,8(1):27－37.

[2] Biham E, Biryukov A, Shamir A. Cryptanalysis of Skipjack reduced to 31 rounds using impossible differentials[J]. Journal of Cryptology, 2005,18(4):291－311.

[3] Biham E, Biryukov A, Shamir A. Miss in the middle attacks on IDEA and khufu[C]//FES, 1999:124－138.

[4] Biham E, Shamir A. Differntial cryptanalysis of the data encryption standard[M]. Heidelberg: Springer, 1993.

[5] 吴生宝.差分和线性分析的代数自动化方法[D].北京:中国科学院软件研究所,2014:27－58.

[6] Wu S B, Wang M S. Automatic search of truncated impossible differentials for word－oriented block ciphers[C]//Indocrypt, 2012:283－302.

[7] 温巧燕,钮心忻,杨义先.现代密码学中的布尔函数[M].北京:科学出版社,2000:2－16.

[8] Wan Z X. Lectures on finite fields and galois rings[M]. London: World Scientific Publishing Co Pte Ltd, 2006:161－189.

[9] Kim J, Hong S, Sung J, Lee C, Lee S. Impossible differential cryptanalysis for block cipher structures[C]//Indocrypt, 2003: 82－96.

[10] Luo Y, Wu Z, Lai X, Gong G. A unified method for finding impossible differentials for block cipher structures[J]. Information Sciences, 2014,271:211－220.