立方攻击研究进展

王明兴[1,2]，朱玉倩[1]，苗三立[1]
（1.中国电子信息产业集团有限公司第六研究所，北京 102209；
2.密码科学技术国家重点实验室，北京 100878）

摘 要：立方攻击是一种新型的代数分析方法，刚提出时对密码算法的分析效果并不理想。但是在引入多重集合可分性、可分路径的概念之后，立方攻击的过程转化为求解混合整数线性规划问题，再使用数学软件进行计算，大大提高了其分析能力。梳理了立方攻击的技术脉络，论述了其最新进展，给出了立方攻击亟待解决的研究问题，这将有助于掌握立方攻击的最新技术，便于开展对分组密码、序列密码和哈希函数等密码算法的分析工作。

关键词：序列密码；立方攻击；可分性；混合整数线性规划
中图分类号：TP393　　文献标识码：A　　DOI：10.19358/j.issn.2096-5133.2020.10.006
引用格式：王明兴，朱玉倩，苗三立. 立方攻击研究进展[J].信息技术与网络安全，2020，39(10)：28-32.

Research advances on cube attack

Wang Mingxing[1,2], Zhu Yuqian[1], Miao Sanli[1]
(1.The 6th Research Institute of China Electronics Corporation, Beijing 102209, China;
2.State Key Laboratory of Cryptology, Beijing 100878, China)

Abstract: Cube attack is a new method of algebra analysis to cryptographic algorithm, which is weak crypt-analytic technique when it was first proposed. However, using the notions of division property of the multiset and division trail, the process of cube attack is transformed to solve the questions of mixed integer linear programming by mathematical software, which shows more powerful crypt-analytic of cube attack than before. In this paper, to keep abreast of advances in cube attack, the research summary via straightening up the context of its technology is proposed, and research questions to be solved in cube attack is given. The work will help to master the latest technique of cube attack and launch the analysis of cryptographic algorithms such as block cipher, stream cipher and hash function.

Key words: stream cipher; cube attack; division property; mixed integer linear programming

0 引言

DINUR I和SHAMIR A在2009年的欧密会上提出了一种新型的攻击手段[1]，称为立方攻击(Cube Attack)。立方攻击是一种选择明文攻击，其攻击思想是：密码算法被看成是一个未知的复杂的多元多项式，输入变量(明文或者初始向量)称为公开变量，密钥变量称为秘密变量；输出变量可以表示为公开变量和秘密变量的某种多项式的形式。只要至少有一位输出变量可以表示为秘密变量和公开变量的低次多项式，通过赋值一些公开变量，即可访问密码算法获取输出结果，得到秘密变量的简单的方程式，恢复密钥比特。

TODO Y等人[2]在2015年提出了多重集合的可分性(Division Property)的概念，它是分析分组密码积分特征的有力工具。在之后的一年，TODO Y等人[3]又提出了基于比特的多重集合的可分性。向泽军等人[4]在2016年的亚密会上提出了可分路径的概念，将可分路径的计算转化为求解混合整数线性规划(Mixed Integer Linear Programming，MILP)问题，提高了积分攻击的准确性和运算效率。在2017年的美密会上，TODO Y等人[5]提出了可分性的可分路径和立方攻击的超级多项式之间的联系，给出了求解超级多项式中的变量的算法。

王庆菊等人[6]进一步提升了MILP模型的计算

准确性，提出了超级多项式的最高次项的求解算法，降低了立方攻击的复杂度。王森鹏等人[7]采用了三子集可分性的概念，发展了求超级多项式的算法，降低了算法的时间复杂度。HAO Y L 等人[8]提出了基于完善的三子集的多重集合的可分性的定义，更进一步提升了 MILP 模型计算超级多项式的准确性。

本文的主要目的是全面回顾立方攻击的技术演进历程，介绍最新的立方攻击技术，展望未来立方攻击的发展方向和应用价值。

1 立方攻击的原理和过程

1.1 立方攻击的定义

立方攻击利用了多元布尔多项式的高阶差分的性质，即布尔多项式都可以表示成一种带余除法的形式，其中余式是不能包含除式的全部变量的多项式。

定义 1[1] 设布尔多项式 $f(x_1, x_2, \cdots, x_{n-1})$ 的代数正规型表示为：

$$f(x_1, x_2, \cdots, x_{n-1}) = \sum a_{j_1, j_2, \cdots, j_{n-1}} x_1^{j_1} x_2^{j_2} \cdots x_{n-1}^{j_{n-1}} \quad (1)$$

设 $I=\{i_1, i_2, \cdots, i_{|I|}\}$ 是集合 $\{1, 2, \cdots, n\}$ 的一个子集，I 中元素的个数记为 $|I|$，I 被称为是一个立方指数，$C_I=\{x_{i_1}, x_{i_2}, \cdots, x_{i_{|I|}}\}$ 称为一个立方(cube)，设 $t_I=\prod_{i \in I} x_i$，于是：

$$f(x_1, x_2, \cdots, x_{n-1}) = t_I \cdot P_{S(I)} \oplus r(x_1, x_2, \cdots, x_{n-1}) \quad (2)$$

其中 $P_{S(I)}$ 是与 t_I 没有相同变量的多项式，被称为立方 I 的超级多项式(Superpoly)；$r(x_1, x_2, \cdots, x_{n-1})$ 中的每一项至少缺失 t_I 中的一个变量。当 C_I 中的变量取遍所有的可能值，对式(2)的左端取连加和，得到 $P_{S(I)}$，即 $\sum_{C_I} f = P_{S(I)}$。

对攻击者而言，密码算法是一个关于秘密变量和公开变量的复杂的多元多项式，表示为 $f(v_1, v_2, \cdots, v_{m-1}, x_1, x_2, \cdots, x_{n-1})$，其中 x_i 是秘密变量，v_i 是公开变量。立方攻击是把多元布尔多项式表示成一种类似于式(2)的分解式，除式是某些公开变量的乘积，超级多项式是秘密变量的简单的多项式，然后通过下面的攻击过程恢复密钥。

1.2 立方攻击的攻击过程

攻击过程分为两个阶段：离线阶段和在线阶段。

离线阶段：攻击者可以选择秘密变量和公开变量，随机选择立方变量，执行密码算法，计算 $\sum_{C_I} f = P_{S(I)}$，使用一次或者二次超级多项式的判定准则(见第 2 节)，得到很多立方变量和相应的简单超级多项式，为在线阶段做准备。

在线阶段：秘密变量是固定的且是未知的，不允许选择，敌手利用离线阶段获得的立方变量，对其进行 0/1 赋值，去访问加密算法，得到输出比特；对所有输出比特进行连加求和，获得超级多项式的值，最后求解得到的代数方程组，可以恢复某些密钥比特；猜测剩余的密钥变量，完成密钥恢复攻击。

2 低次超级多项式的判别方法

文献[1, 9-10]中的超级多项式的次数是一次或二次的，判别方法也较为简单，下面分别叙述。

2.1 一次超级多项式判别方法

独立均匀随机地选择 $\boldsymbol{x}, \boldsymbol{y} \in \{0,1\}^n$，验证 $P_{S(I)}[\boldsymbol{0}] + P_{S(I)}[\boldsymbol{x}] + P_{S(I)}[\boldsymbol{y}] = P_{S(I)}[\boldsymbol{x}+\boldsymbol{y}]$ 是否成立。如果 $P_{S(I)}$ 是线性的，则测试总是成功的，如果 $P_{S(I)}$ 不是线性的，则测试以很高的概率失败，测试进行足够多次，敌手就可以确定 $P_{S(I)}$ 是非常接近线性的了。

2.2 二次超级多项式的判别方法

随机选择 $\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3 \in \{0,1\}^n$，验证 $P_{S(I)}[\boldsymbol{0}] + P_{S(I)}[\boldsymbol{x}_1] + P_{S(I)}[\boldsymbol{x}_2] + P_{S(I)}[\boldsymbol{x}_3] + P_{S(I)}[\boldsymbol{x}_1+\boldsymbol{x}_2] + P_{S(I)}[\boldsymbol{x}_1+\boldsymbol{x}_3] + P_{S(I)}[\boldsymbol{x}_2+\boldsymbol{x}_3] + P_{S(I)}[\boldsymbol{x}_1+\boldsymbol{x}_2+\boldsymbol{x}_3] = 0$ 是否成立。如果 $P_{S(I)}$ 是二次的，则测试总是成功的；如果 $P_{S(I)}$ 不是二次的(是更高次的)，则测试以很高的概率失败，测试多次就可以排除非二次的 $P_{S(I)}$。

2.3 非线性超级多项式的判别方法

在非线性超级多项式情形下，为确定超级多项式的项和系数，引入一种扩展的立方攻击理论(Extended Cube Attack)。

定义 2 (扩展立方体)[10] 在立方攻击中，对任意立方 C_I，如果存在另一个立方 C_H，且 $I \cap H = \phi$，则 C_I 与 C_H 可以得到一个扩展的立方 $C_{I \cup H}$。

下面的引理 1 和引理 2，判断哪些密钥变量存在于 $P_{S(I)}$ 中以及以怎样的单项式形式存在。

引理 1[10] 设 $f(x_1, x_2, \cdots, x_n)$ 是一个多项式，设 $I=\{i_0, \cdots, i_l\}$ 是集合 $\{1, 2, \cdots, n\}$ 的一个子集，$t_I=\prod_{i \in I} x_i$，令 $P_{S(I)}=\sum_{C_I} f$，则 t_I 以子项式或子项的一部分存在于 f 中，当且仅当至少存在一个向量 $\boldsymbol{x} \in \{0,1\}^{n-|I|}$，使得 $P_{S(I)}(\boldsymbol{x}) = \sum_{C_I} f(\boldsymbol{x}) = 1$。

引理 2[10] 单项式 $\prod_{i_j \in H} x_{i_j}$ 是 $P_{S(I)}$ 中的一个子项，当且仅当令所有 $x_i=0 (x_i \notin I \cup H)$，有 $P_{S(I \cup H)} = \sum_{C_{I \cup H}} f = 1$。

3 立方攻击的前沿研究及主要结论

上节介绍了求解低次超级多项式的判别方法，求解更高次的超级多项式的方法是本节要介绍的内容。

3.1 基于比特的多重集合的可分性

基于比特的可分性定义如下：

定义 3[2] 设 $X\subseteq F_2^n$ 是一个多重集合，$Y\subseteq F_2^n$ 是一个 n 维向量的集合，称多重集 X 有可分性 $D_Y^{1^n}$，如果满足下面的条件：

$$\sum_{x\in X}x^u=\begin{cases}\text{未定值，如果存在 } K\in Y,\text{ 满足 } u\succcurlyeq K\\0,\quad\text{其他}\end{cases}$$

这里 $u\succcurlyeq K$ 是指如果对于任意的 i，都有 $u_i\geqslant k_i$，其中，$x^u=x_0^{u_0}x_1^{u_1}\cdots x_{n-1}^{u_{n-1}}$。

定义 4 [4](可分路径) 假设可分性质的传播为 $\{K\}\triangleq Y_0\to Y_1\to\cdots\to Y_r$，进一步地，对任何向量 $K_{i+1}^*\in Y_{i+1}$ 必然存在一个向量 $K_i^*\in Y_i$，使得 K_i^* 可由可分性的传播规则传播到 K_{i+1}^*，更进一步地，$(K_0,K_1,\cdots,K_r)\in(Y_0\times Y_1\times\cdots\times Y_r)$，如果 K_i 可以传播到 K_{i+1}，$i\in\{0,1,\cdots,r-1\}$，就把 $(K_0,K_1,\cdots,K_r)$ 称为一条 r 轮的可分路径。

为了计算出所有的可分路径，需要先把密码算法中的复制运算(Copy)、异或运算(XOR)和且运算(AND)转化成 MILP 模型，再把整个密码算法建模成 MILP 模型，这样就可以使用最优化数学软件 Gurobi 来计算所有的可分路径。

3.2 基于可分路径寻找超级多项式

由第 2 节可知，立方攻击成功的关键在于找到简单的超级多项式，例如超级多项式中的变量的个数少，同时次数又是一次的或者二次的。下面的性质是利用可分路径求出超级多项式中所有的变量的依据。

性质 1[5] 设 $f(X,V)$ 是一个多元多项式，$X=(x_0,x_1,\cdots,x_{n-1})$ 是秘密变量，$V=(v_0,v_1,\cdots,v_{m-1})$ 是公开变量，设一个立方索引 $I=\{i_0,i_1,\cdots,i_{|I|-1}\}$，$K_I$ 是一个 m 维的向量，使得 $V^{K_I}=v_{i_0},v_{i_1},\cdots,v_{i_{|I|-1}}$，即如果 $i\in I$，则 $k_i=1$，否则 $k_i=0$；n 维向量 $e_j=(0,\cdots,0,1,0,\cdots,0)$，即在第 j 分量上取值为 1，而其他分量是 0；假设没有可分路径 $(e_j,K_I)\xrightarrow{f}1$，那么 x_j 不是超级多项式中的变量。

性质 1 说明，存在可分路径 $(e_j,K_I)\xrightarrow{f}1$，则 x_j 是超级多项式中的变量，那么可以得到超级多项式中的所有的变量，进而求出超级多项式的表达式。

离线阶段：设 J 是超级多项式中的变量下标的集合，集合 J 遍历所有可能的值组成的集合记为 C_J，集合 J 之外的变量赋值为 0，对于给定的值 $X_s=(x_s,0,0,\cdots,0)$，$x_s\in C_J$，敌手可以设定公开变量中除了立方变量之外的变量为某些常数，计算 $\sum_{CI}f(X_s,V)=P_{S(I)}$，得到超级多项式的真值表。值得注意的是，敌手可以多次选择除了立方变量以外的公开变量的值使得超级多项式是一次的。

在线阶段：访问加密算法，计算 $\sum_{C_I}f(X,V)=P_{S(I)}=\theta$，查询超级多项式的真值表，只保留那些 $x_j,j\in J$ 的值使得 $P_{S(I)}=\theta\in F_2$。

这里时间复杂度为 $2^{|I|+|J|}$，如果 $|I|+|J|$ 小于密钥长度，那么立方攻击就是有效的。

在文献[6]中，王庆菊等人注意到常数值 0 在运算中会消去多项式这一现象，提出了 Flag 技术，即定义了初始向量的常数值与变量的运算法则，修正了复制、异或、且运算的 MILP 模型，提高了可分路径的计算准确性，提出了超级多项式的最高次项的判断方法，使得立方攻击可以求出更高次数的超级多项式。

性质 2 设 $f(X,V)$ 是一个多元多项式，$X=(x_0,x_1,\cdots,x_{n-1})$ 是秘密变量，$V=(v_0,v_1,\cdots,v_{m-1})$ 是公开变量，设一个立方索引 $I=\{i_0,i_1,\cdots,i_{|I|-1}\}$，$K_I$ 是一个 m 维向量，使得 $V^{K_I}=v_{i_0}v_{i_1}\cdots v_{i_{|I|-1}}$，即如果 $i\in I$，则 $k_i=1$，否则 $k_i=0$；K_Λ 是一个 n 维向量，假设没有可分路径 $(K_\Lambda,K_I)\to1$，那么 X^{K_Λ} 不在超级多项式中。

性质 2 说明，如果存在可分路径 $(K_\Lambda,K_I)\to1$，那么 X^{K_Λ} 在超级多项式中，于是，在线性整数规划模型中可以求得超级多项式中的最高次项；再由性质 1，得到超级多项式中的所有的变量，就可以降低恢复超级多项式的复杂度。

3.3 基于三子集可分性

定义 5 [7](基于三子集可分性) 设 X 是一个多重集合，$X\subseteq F_2^n$，设 Y,L 是由 n 维比特向量组成的集合，当多重集 X 具有可分性 $D_{Y,L}^{1^n}$，是指下列的条件成立：

$$\sum_{x\in X} x^u=\begin{cases}\text{未定值,如果存在 } K\in Y,\text{满足 } u\geqslant K\\ 1, \quad \text{否则若存在 } l\in L,\text{满足 } u=l\\ 0, \quad \text{其他}\end{cases}$$

基于三子集可分性修改了复制运算、异或运算、与运算的 MILP 建模，王森鹏等人[7]提出了该可分性的修剪性质(Pruning Properties)，去掉无用的可分路径，并提出了可分路径的加速传播和停止准则，来提高可分路径的计算效率；最后提出了相似多项式(Similar Polynomial)的概念来计算超级多项式。但是这一概念与文献[10]中的引理 2 本质上是一致的。对密码算法 Trivium[11]的立方攻击的轮数是 839 轮，但时间复杂度为常数，攻击实际可行。

3.4 完善的三子集可分性

HAO Y L 等人[8]发现王森鹏等人的立方攻击的计算方法并非总是有效的，在求 841 轮的 Trivium 的超级多项式时，由于 L 的规模过大，在合理的时间内是无法求解的。于是，提出了完善三子集可分性的概念。

定义 6[8](完善三子集可分性) 设 X 是一个多重集合，其元素 $x\subseteq F_2^m$，设 L 也是一个多重集合，其元素 $u\in F_2^m$，X 有完善的三子集可分性 $T_{\tilde{L}}^{1^m}$，是指满足下面的条件：

$$\sum_{x\in X} x^u=\begin{cases}1,\text{存在奇数个 } u\in\tilde{L}\\ 0,\text{其他}\end{cases}$$

HAO Y L 等人进一步完善了复制运算、异或运算、与运算的 MILP 建模，提出了更有效的计算超级多项式的算法，对轮数是 841 轮的 Trivium，完全确定了其超级多项式的表达式，展示了立方攻击的强大的分析能力。

4 立方攻击技术比较

密码算法 Trivium 是基于非线性反馈移位寄存器的序列密码，迭代关系式是两次的，密钥长度是 80 bit，初始化向量是80 bit，初始化轮数是1 152 轮。表 1 以立方攻击分析 Trivium 算法为例，清晰地展示了立方攻击技术的分析能力。

通过表 1 可以发现，基于一次或者二次超级多项式判定方法的立方攻击，求得的超级多项式表达式简单，攻击轮数低，时间复杂度也低；而基于多重集合可分性的立方攻击，可以求得的超级多项式表达式复杂，攻击轮数高，但是时间复杂度不一定高。

究其原因，密码算法 Trivium 的迭代关系式是二次的，在初始化轮数较低时(小于 767 轮)，局部出现线性多项式的可能性较大，但是随着初始化轮数的增加(大于 832 轮)，输出比特的关系式的非线性程度将急剧增加，无论是代数次数还是项数都会增加很快，局部出现线性多项式的可能性较小，输出比特的关系式会非常复杂。这是两种立方攻击的分析效果截然不同的根本原因。这就是说，基于一次或者二次超级多项式的判定方法的立方攻击不可能分析到更高的轮数，即基于一次或者二次超级多项式的判定方法的立方攻击，不可能取得大的突破性进展，这是由密码算法自身的迭代关系式和轮数决定的。但是基于可分性，立方攻击可能分析到更高的轮数。这是因为后者有清晰的数学模型、扎实的数学理论支撑，而且计算过程使用了数学软件工具，能够求解的超级多项式从理论上讲是不限制次数和项数的，能否计算出超级多项式只取决于计算机的计算能力，所以后者方法可以求出更高轮数的超级多项式的表达式。

由此可知，未来的研究方向主要是要完善基于可分性的立方攻击的技术，提升其计算能力。亟待解决的研究问题有两个：一是进一步提高数学模型计算的准确性，因为有研究指出，针对 Trivium 的某些轮数，立方攻击不是密钥恢复攻击而是区分攻击；二是优化实现立方攻击的算法，降低其时间复杂度，提高其计算效率，因为目前的攻击算法在普通计算机上的运行时间需要几天或者时间复杂度

表 1 Trivium 算法的立方攻击结果比较

时间	基本原理	立方变量的数量	超级多项式次数(d)、项数(num)	攻击轮数	恢复超级多项式时间复杂度	参考文献
2009	线性关系式判定方法	30	$d=1$，num=4	767	c(常数)	[1]
2017	基于多重集合可分性	72	$d=1$，num=5	832	2^{77}	[5]
2018	基于多重集合可分性，修正数学模型	78	$d=1$，num=1	839	2^{79}	[6]
2019	基于三子集可分性	77	$d=3$，num=3	839	c(常数)	[7]
2020	基于完善三子集可分性	78	$d=4$，num=53	841	c(常数)	[8]

更大，还有进一步优化的必要。

5 结论

本文详细介绍了立方攻击的最新发展，展示了求超级多项式技术的进步，能够恢复超级多项式的次数从一次、二次到更高次。其中TODO Y等人的研究工作具有里程碑意义，打开了建立数学模型、利用数学软件计算超级多项式的大门，显著提高了立方攻击的分析能力。

从立方攻击的技术发展来看，立方攻击对反馈关系式为二次的序列密码的分析效果非常好。这说明在设计序列密码时，反馈关系式的代数次数应适当提高，使得输出比特的表达式随着轮数增加快速复杂化以抵抗立方攻击；未来密码算法能够抵抗立方攻击将成为必要的安全准则。

由于立方攻击技术的进步，将来可以分析分组密码和哈希函数抵抗立方攻击的能力，虽然这两类密码算法设计复杂，但也可能得到较理想的分析结果。

参考文献

[1] DINUR I, SHAMIR A. Cube attacks on tweakable black box polynomials[C]. EUROCRYPT 2009, Heidelberg, Springer, 2010, 5479: 278-299.

[2] TODO Y. Structural evaluation by generalized integral property[C]. EUROCRYPT 2015, Part I, Springer, 2015, 9056: 287-314.

[3] TODO Y, MORII M. Bit-based division property and application to Simon family[C]. FSE2016, Springer, 2016, 9783: 357-377.

[4] XIANG Z J, ZHANG W T, BAO Z Z, et al. Applying MILP method to searching integral distinguishers based on division property for 6 lightweight block ciphers[C]. ASIACRYPT 2016, Part I, Springer, 2016, 10031: 648-678.

[5] TODO Y, ISOBE T, HAO Y L, et al. Cube attack on nonblack box polynomials based on division property[C]. CRYPTO 2017, Springer, Heidelberg, 2017, 10403: 250-279.

[6] WANG Q J, HAO Y L, TODO Y, et al. Improved division property based cube attacks exploiting algebraic properties of superploy[C]. CRYPTO 2018, Springer, Cham, 2018, 8885: 143-160.

[7] WANG S P, HU B, GUAN J, et al. MILP-aided method of searching division property using three subsets and applications[C]. ASIACRYPT 2019, Part III. Springer, Cham, 2019, 11923: 398-427.

[8] HAO Y L, LEANDER G, MEIER W, et al. Modeling for three-subset division property without unknown subset improved cube attacks against Trivium and Grain-128AEAD[C]. UEROCRYPT 2020, Springer, 12105: 466-495.

[9] SONG H X, FAN X B, WU C K, et al. Cube attack on grain[J]. Journal of Software, 2012, 23(1): 171-176.

[10] ABDUL-LATIP S F, REYHANITABAR M R, SUSILO W, et al. Extended cubes: enhancing the cube attack by extracting low-degree non-linear equations[C]. Proceedings of the 6th ACM Symposium on Information, Computer and Communications Security, New York, ACM Press, 2011: 296-305.

[11] DE CANNIÈRE C. Trivium: a stream cipher construction inspired by block cipher design principles[C]. International Conference on Information Security, Heidelberg, Springer, 2006, 4176: 171-186.

（收稿日期：2020-07-02）

作者简介：

王明兴(1983-)，男，博士，工程师，主要研究方向：信息安全。

朱玉倩(1994-)，女，硕士研究生，主要研究方向：信息安全。

苗三立(1991-)，男，硕士，助理工程师，主要研究方向：信息安全。

版权声明