浅析分组密码分析方法的关联性

王美琴, 孙 玲, 陈怀风, 刘 瑜

山东大学密码技术与信息安全教育部重点实验室 济南 中国 250100

摘要 自从差分分析和线性分析相继被提出以后，许多基于其之上的分析方法陆续出现，各式各样的分析方法通常利用了自算法中所提取的数据的不均匀性来获取密钥的信息。人们在利用这些分析方法对分组密码的安全性进行评估时，经常会发现一些相似的现象，如：某两种区分器的轮数总是相同、两种统计分析方法中所利用的统计量存在数学关系等。所以，在建立新的分析方法的同时，人们渐渐将关注点转移到研究各种已有的分析方法的关联性上。尽管在处理和分析方式有着形式上的不同，但经过仔细分析之后发现许多看似不同的分析方法之间有着一些关联性，研究这种关联性不管是从理论上还是从分析分组密码安全性的角度都是非常有必要的。近几年，各种分析方法之间的关联性逐渐被建立起来。这些关联性的建立一方面有助于我们对已知的分组密码分析方法进行分类，另一方面这些关联性可能会给出分组密码安全性的补充信息。本文中简要介绍了一些已有的分析方法，并总结了已有的分析方法之间的关联性。

关键词 分组密码；分析方法；关联性
中图法分类号 TP309.7

A Brief Analysis of Links between Different Types of Cryptanalytic Methods for Block Cipher

WANG Meiqin, SUN Ling, CHEN Huaifeng, LIU Yu

Key Lab of Cryptologic Technology and Information Security Ministry of Education, Shandong University, Jinan 250100, China

Abstract Many cryptanalytic methods have gradually appeared since the successive foundation of linear cryptanalysis and differential cryptanalysis. Various cryptanalytic methods usually use the nonuniformity of the data extracted from the ciphers to get the information of the secret key. Some similar phenomena can be found when using these cryptanalytic methods to analyze the security of block ciphers. For example, the number of rounds covered by certain two distinguishers is always the same, and there exist some mathematical links between certain two statistical cryptanalysis methods. So, people are gradually shifting their concern to discover the links between various existing cryptanalytic methods while proposing new cryptanalytic methods. Despite the formal differences lie in the management and the cryptanalysis, there exist some links between many cryptanalytic methods which may look different after carefully research. Discovering this kind of links is necessary not only from the point of theoretical but also from the perspective of estimating the security of block ciphers. Many links between cryptanalytic methods have gradually been built in recent years. The establishing of these links, on the one hand, can help us classify existing cryptanalysis methods of block ciphers. On the other hand, these may give some supplement information of the security of block ciphers. In this paper, we briefly introduce some existing cryptanalytic methods and summarize the links between existing cryptanalytic methods.

Key words block ciphers; cryptanalytic methods; links

1 引言

分组密码算法现已作为许多密码元件(如: 伪随机数生成器、哈希函数等)的构造单元, 对于这些密码元件安全性的分析通常归结于对分组密码算法的攻击。在各种各样的分析方法中, 统计的分析方法通常利用了从算法中提取的数据的不均匀性来获取密钥的信息。差分分析和线性分析是统计分析方法中最基本、最重要的两种方法。

差分分析[1]最初由 Biham 和 Shamir 发表在 1990 年的美洲密码年会上, 此后基于此理论的研究层出不穷。直至今日, 差分分析仍是分组密码最有效的分析方法之一, 也是衡量一个分组密码安全性的重要指标。基本的差分分析利用一条高概率的差分路线

通讯作者：王美琴, 博士, 教授, Email: mqwang@sdu.edu.cn。
本课题得到 973 计划(No. 2013CB834205); 国家自然科学基金(No. 61133013 和 No. 61572293); 新世纪优秀人才支持计划(No. NCET-13-0350)资助。收稿日期：2015-12-01；修改日期：2015-12-22；定稿日期：2016-01-08

$(\Delta A \to \Delta B)$来恢复密钥。随着密码学的发展，许多基于其之上的分析方法陆续被提出，如：截断差分分析[2]、不可能差分分析[3,4]、飞去来器攻击[5]和矩形攻击[6]等。

有些时候对某些算法不存在给定轮数的高概率差分特征，但我们可以以较大优势决定一个分组密码算法经过一定轮数之后的一部分输出差分，截断差分[2]正是利用这种差分特征来恢复密钥信息。不可能差分分析由 Knudsen[3]和 Biham[4]等人独立的提出来，与传统差分分析基于高概率的差分特征不同，不可能差分利用不可能出现的(概率为 0 的)差分特征来进行区分攻击或获取密钥信息。在不可能差分中最常用的实际上是截断不可能差分，这种不可能差分区分器与算法中S盒的选择无关。其基本的构造方法是利用中间相错的方法，即分别以从前往后和从后往前的方式以概率为 1 的方式延拓出两条截断差分特征然后使这两段差分特征中间连接的部分出现矛盾。许多文献给出了这种截断不可能差分的自动化搜索方法，如$\mathcal{U}$方法[7]、UID方法[8]。

在 1993 年的欧洲密码年会上，日本学者 Matsui 提出了对 DES 算法的新的攻击方法——线性分析[9]。线性分析是一种已知明文攻击，他通过研究明文和密文之间的线性关系来恢复密钥。经过二十几年的发展和完善，线性分析同差分分析一样，已经成为现代分组密码设计时必须考虑的重要设计准则。同样的，许多建立在线性分析理论之上的分析方法也层出不穷，如：多重线性分析[10,11]、多维线性分析[12]、零相关线性分析[13,14,15]等。

要改进基本的线性分析，一个自然的想法是利用多条线性路线。Matsui 最先提出了这种改进，在文献[10]中，他同时利用了两条线性逼近改进了 DES 的攻击。同年，Burton 等人[11]正式的建立了基于多条线性逼近的线性分析模型，该模型基于一个较强的假设，即所使用的线性逼近统计独立。但 Murphy[16]指出，这一假设一般情况下不成立。2009 年，Hermelin[12]等人给出了多维线性分析的模型，该模型避免了多重线性的模型中的独立性假设。

零相关线性分析由 Bogdanov 和 Rijmen[13]于 2012 年首次提出，与传统的线性分析不同，零相关线性分析利用相关度为零的线性逼近来进行区分攻击或恢复密钥。随着零相关线性分析的提出、发展以及应用，零相关线性分析已经成为密码分析中的一种重要工具。虽然零相关线性分析与不可能差分分析在理论和技术两方面有很大的不同，但零相关线性分析可以看做不可能差分分析在线性分析领域的一种对偶方法。

然而基本的零相关线性分析所需要的数据量为整个明文空间或2^{n-1}个选择明文，其中n为分组密码算法的分组长度。一方面，极高的数据复杂度需求极大地限制了零相关理论的应用，另一方面由于分组密码中一般存在大量的零相关线性逼近，对于零相关线性逼近的不充分应用也限制了零相关这一理论的发展。以此为出发点，Bogdanov 和王美琴[14]在 2012 年提出了利用多条零相关线性逼近来进行区分攻击或密钥恢复攻击的多重零相关分析模型。假设存在ℓ条零相关线性逼近，则多重零相关线性分析所需的数据复杂度约为$O\left(\frac{2^n}{\sqrt{\ell}}\right)$。与基本的零相关线性分析相比，多重零相关线性分析的数据复杂度显著降低。但与多重线性分析模型类似的是，多重零相关的分析模型也基于较强的独立性假设。Bogdanov[15]等人又提出了积分零相关区分器和多维零相关线性分析新模型。多维零相关线性分析模型与多重零相关线性分析模型相比，在维持数据复杂度基本不变的基础上，不再依赖ℓ条零相关线性逼近相互独立这一强假设条件，极大地完善了零相关线性分析理论。

除了差分和线性这两条基本的主线之外，许多形式上与上述统计分析方法看似不同的统计分析方法也在密码学的发展过程中被提出，如：积分分析[17,18]、统计饱和度分析[19]。积分[17,18]是继差分分析和线性分析之后，密码学界公认的最有效的密码分析方法之一。这种攻击方法更多的与算法的结构有关，而与算法部件的具体取值关系不大，作为主要针对面向字节运算算法安全性的密码分析方法，积分攻击从其出现就受到密码学界的广泛关注。统计饱和度攻击[19]的主要想法是利用某些特殊的分组密码算法中较弱的扩散性，通过固定明文比特的某些比特输入，考虑密文部分比特的分布。

人们在利用这些分析方法对分组密码的安全性进行评估时，经常会发现一些相似的现象，如：某两种区分器的轮数总是相同、两种统计分析方法中所利用的统计量存在数学关系等。所以，在建立新的统计分析方法的同时，人们渐渐将关注点转移到研究各种已有的统计分析方法的关联性上。尽管在分析和统计方式有着形式上的不同，但经过仔细分析之后发现许多看似不同的统计分析方法之间有着一些关联性，研究这种关联性不管是从理论上还是从分析分组密码安全性的角度都是非常有必要的。近几年，各种统计分析方法之间的关联性逐渐被建立起来。这些关联性的建立一方面有助于我们对已知的分组密码分析方法进行分类，另一方面这些关联性

可能会给出分组密码安全性的补充信息。

早在 1994 年, Chabaud 和 Vaudenay[20]就提出了线性分析和差分分析之间的数学关系, 但这一数学关系后来并没有实际应用于差分概率的计算, 因为直接的应用需要涉及2^{n+m}个线性逼近相关度的计算, 其中n和m分别为明文空间和密文空间的维数, 当n和m比较大时, 这种计算方法并不可行。

直到 2013 年的欧密会上, Blondeau 和 Nyberg[21]将这一数学关系进行推广, 应用到划分后的明文空间和密文空间, 得到了特定形式的截断差分的概率和多维线性中相关度之间的关系。并给出了特殊情况下, 截断不可能差分和零相关线性分析的一种等价性, 这种等价性并不依赖于分组密码的结构。同样的, 由于这种等价性的成立要求所使用的截断不可能差分的维数和零相关线性逼近的维数均为n, 所以零相关与不可能差分之间的等价性(或不等价性)并没有得到实际的解决。Blondeau 和 Nyberg 在文献[22]中又利用推广后的数学关系给出了截断差分的概率和多维线性中容度的关系, 并将文献[23]中对于差分分析和线性分析数据复杂度的确切评估应用到截断差分和多维线性的分析中来, 得出了截断差分分析和多维线性分析各种攻击复杂度之间的关系。

2014 年, Blondeau[24]等人对两类特殊的结构, Feistel 结构和 Skipjack 结构, 给出了不可能差分区分器和零相关区分器的等价性。而后, 孙兵等人[25]将上述结论进行推广, 通过引入结构和对偶结构的概念, 对 Feistel 结构和代换置换结构(Substitution Permutation Network, SPN)证明了不可能差分和零相关的等价关系。

另一方面虽然统计饱和度攻击对于轻量级算法 PRESENT[26]给出了最好攻击, 但在文献[19]中并没有给出攻击复杂度的一种确切评估。直到 2011 年的欧密会上, Leander[27]证明统计饱和度分析和多维线性分析在本质上是相同的, 这种关系允许我们正确的评估统计饱和度攻击中使用的偏差, 从而确切的评估攻击的复杂度。从这一方面来讲, 研究统计分析方法之间的关系, 也有助于我们用较为完善的理论体系丰富和发展密码体系中较为薄弱的环节。后来 Blondeau 和 Nyberg 在文献[22]通过对截断差分分析和统计饱和度分析中所使用的统计量的观察, 发现这两种统计量只相差一个常数, 从而证明截断差分分析和统计饱和度分析在本质上是相同的。

零相关分析和积分分析之间的关系最早由 Bogdanov 等人[15]提出, 作者给出了特定形式的零相关区分器与积分区分器之间的等价关系。文献[15]指出: 积分区分器可以无条件的转换为一个零相关区分器, 而当零相关区分器转换为积分区分器时要求零相关区分器的输入掩码和输出掩码相互独立。后来, 孙兵等人[25]证明一个零相关区分器可以无条件的转换为一个积分区分器, 从而去掉了[15]中的独立性要求。除此之外, 作者还证明[15]中给出的“积分区分器可以无条件的转换为零相关区分器”这一结论仅适用于平衡向量布尔函数的积分性质, 但这并不是一般使用的积分性质。对于积分区分器和不可能差分区分器之间的等价关系, 文献[25]中也给出了相关结果。

作为差分分析的两种推广——不可能差分分析和截断差分分析之间的关系的研究, 李雷波等人[28]也做出了一些结果。他们通过将截断不可能差分中间相错的关键步骤以一定的概率转换为中间相遇, 从而将不可能差分转变为截断差分, 并给出了这种截断差分的概率计算公式。得益于截断不可能差分的自动化搜索方法, 这一关系的发现将可能有助于促进截断差分搜索的自动化。

我们在图 1 中给出上述的所有关系。

本文剩下的部分将详细介绍分组密码不同分析方法之间的关联性。第二节将简要介绍分组密码的不同分析方法并给出一些符号定义。第三节关注基本的线性分析与基本的差分分析之间的关系。第四节详细阐述了多维线性、截断差分和统计饱和度攻击之间的关联性。第五节给出零相关、不可能差分和积分攻击之间的关系。第六节介绍了由不可能差分向截断差分的转换。第七节进行总结。

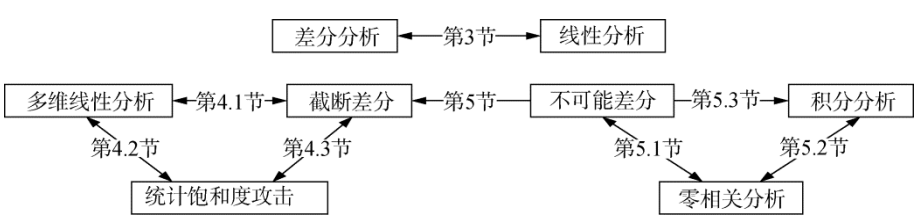

图 1　分组密码统计分析方法的关联性

2 分组密码不同分析方法简介

差分分析和线性分析是分析分组密码安全性的两种重要手段。基于这两种基本的统计分析方法, 许多新的统计分析方法又陆续被提出。本节我们将简要的介绍一些后面分析会用到的统计分析方法。

2.1 分组密码的差分分析方法综述

2.1.1 基本的差分分析

差分分析[1]最初由 Biham 和 Shamir 发表在 1990 年的美密会上, 此后基于此理论的研究层出不穷。直至今日, 差分分析仍是分组密码最有效的分析方法之一, 也是衡量一个分组密码安全性的重要指标。

设X和X'为两个长度为n的比特串, X和X'的差分定义为

$$\Delta X = X \otimes (X')^{-1},$$

其中$\otimes$为比特串的运算并且该运算取决于密钥介入的方式, X^{-1}为X关于运算$\otimes$的逆。基本的差分分析基于在密钥介入前后, 中间状态的差分相同, 然而在一个加密算法中, 总存在相对于$\otimes$运算是非线性的运算。差分分析利用了如下观测: 对于一个确定的输入差分, 输出差分的分布是不均匀的。

定义 1(差分路线)**.**

一个 s 轮的差分路线为一个 $s+1$ 元数组 $\{\beta_0, \beta_1, \cdots, \beta_s\}$, 其中$\Delta P = \beta_0, \Delta C_i = \beta_i (1 \leqslant i \leqslant s)$。

设p_i是输入差分为β_{i-1}输出差分为β_i的概率。文献[29]中引入了 Markov 密码的概念, 在 Markov 密码中这一概率与轮函数的具体输入无关并且遍历所有轮密钥的可能取值得到。并且他们证明在 Markov 密码中若轮密钥K_i是相互独立的, 则p_i也相互独立并且有

$$\mathbf{Pr}\{\Delta C_s = \beta_s | \Delta P = \beta_0\} = \prod_{i=1}^{s} \mathbf{Pr}\{\Delta C_i = \beta_i | \Delta C_{i-1} = \beta_{i-1}\}. \quad (1)$$

大量的实验证明等式⑴确实是差分特征概率计算的一个好的近似。我们称s轮的差分路线为一个区分器, 在具体的攻击时, 通常在区分器前后分别加几轮, 然后利用一定数量的选择明密文对恢复这几轮中涉及的部分密钥。

在各种各样的统计分析方法中, 通常会涉及两类错误概率:

1. 正确密钥被当作错误密钥的概率(α_0),
2. 错误密钥被当作正确密钥的概率(α_1)。

而攻击的成功率$P_S = 1 - \alpha_0$。统计攻击中常常会涉及信噪比、优势的概念。

定义 2(差分分析中的信噪比)**.**

差分分析的密钥恢复阶段, 我们枚举选择明密文对对候选密钥进行推荐, 信噪比定义为: 正确密钥被推荐的次数与随机选择的错误密钥被推荐的次数之比。

定义 3(优势)**.**

对于针对k比特密钥的密钥恢复攻击, 优势为a指的是正确密钥位于前2^{k-a}个候选密钥当中。

对于优势a, 我们有$\alpha_1 = 2^{-a}$。

文献[23]中给出了攻击的成功率P_S, 数据量N和优势a之间的数量关系:

定理 1([23, 定理 3])**.**

对于一个关于针对k比特密钥的差分攻击, 所利用的差分特征的概率为p, 设P_S为攻击的成功率, S_N为信噪比, N为攻击中使用的明密文对的数量, a为攻击的优势。假设密钥的计数器是独立的并且它们对于所有的错误密钥同分布, 则对有充分大的k和N有,

$$N = \frac{\left(\sqrt{S_N+1}\Phi^{-1}(P_S) + \Phi^{-1}(1-2^{-a})\right)^2}{S_N} p^{-1}. \quad (2)$$

由等式⑵我们可以看出, 差分攻击中所需的选择明密文对的个数正比于$\frac{1}{p}$。

注意到虽然这一差分路线是固定的, 但在区分攻击或密钥恢复攻击中我们只利用了这一路线输入差分和输出差分的信息, 而且当输入差分和输出差分固定时, 这样的差分路线通常不止一条, 这就引出了差分特征的概念。

定义 4(差分特征)**.**

我们称差分对$\beta_0 \to \beta_s$为一个s轮差分特征, 其中$\Delta P = \beta_0, \Delta C_s = \beta_s$。这一差分特征的概率定义为

$$\mathbf{Pr}\{\Delta C_s = \beta_s | \Delta P = \beta_0\} = \sum_{\beta_1} \cdots \sum_{\beta_{s-1}} \prod_{i=1}^{s} \mathbf{Pr}\{\Delta C_i = \beta_i | \Delta C_{i-1} = \beta_{i-1}\}. \quad (3)$$

由差分路线概率的计算公式⑴和差分特征概率的计算公式⑶的比较我们可以看出: 差分特征的概率一般大于等于相应的一条差分路线的概率。

2.1.2 截断差分

与基本的差分分析不同, 有些时候确定所有比特上的高概率差分特征是困难的, 但我们可以以较大优势决定一个分组密码算法经过一定轮数之后的一部分输出差分, 截断差分[2]正是利用这种差分特征来恢复密钥信息。文献[2]中给出了截断差分的确切定义。

定义 5(截断差分)**.**

一个部分已知的n比特差分称为截断差分。形式的有, 设(α, β)为一个r轮差分, 若Γ_α为α构成的集合,

Γ_β为β构成的集合, 则称$\Gamma_\alpha \to \Gamma_\beta$为一个$r$轮的截断差分。

在进行统计分析方法关联性分析的过程中我们常常需要对明密文空间进行划分。一个n比特的加密算法, 通常划分为下面的形式(如图 2 所示)

$$F: \mathbb{F}_2^s \times \mathbb{F}_2^t \to \mathbb{F}_2^q \times \mathbb{F}_2^r: (x_s, x_t) \mapsto (y_q, y_r) = F(x_s, x_t), s + t = q + r = n.$$

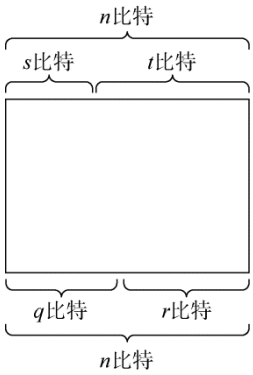

图 2　分组密码明密文空间的划分

在具体的应用中, 我们所使用的截断差分的输入差分和输出差分将满足一定的结构特性, 如构成线性空间。由于本文中只讨论这种情况, 所以我们在下面给出一些符号说明。

设截断差分的输入有2^t种可能且形式$(0, \delta_t) \in \{0\} \times \mathbb{F}_2^t$, 有$2^r$种可能的输出差分且其形式为$(0, \Delta_r) \in \{0\} \times \mathbb{F}_2^r$, 即共考虑$2^{t+r}$个不同的差分特征。我们知道输入差分为$(0, \delta_t)$、输出差分为$(0, \Delta_r)$的确定的差分特征, 其概率为

$$\mathbf{Pr}\{(0, \delta_t) \to (0, \Delta_r)\} = 2^{-n} \#\left\{x \in \mathbb{F}_2^n \middle| F(x) \oplus F\left(x \oplus (0, \delta_t)\right) = (0, \Delta_r)\right\}.$$

定义上述截断差分$TD[(0, \delta_t), (0, \Delta_r)]_{\delta_t \in \mathbb{F}_2^t, \Delta_r \in \mathbb{F}_2^r}$的概率$p$定义为

$$p = 2^{-t} \sum_{\delta_t \in \mathbb{F}_2^t, \Delta_r \in \mathbb{F}_2^r} \mathbf{Pr}\{(0, \delta_t) \to (0, \Delta_r)\}.$$

2.1.3　不可能差分

不可能差分分析由 Knudsen[3]和 Biham[4]等人独立的提出来, 与传统差分分析基于高概率的差分特征不同, 不可能差分分析利用不可能出现的差分特征来进行区分攻击或恢复密钥信息。在不可能差分中最常用的实际上是截断不可能差分, 这种不可能差分区分器与算法中S盒的选择无关。其基本的构造方法是利用中间相错的方法, 即分别以从前往后和从后往前的方式以概率为 1 的方式延拓出两条差分特征然后使这两段差分特征中间连接的部分出现矛盾。如图 3 所示, 我们将加密算法E分解为$E = E_1 \circ E_0$, 对E_0存在概率为 1 的截断差分$\Gamma_0 \to \Gamma_1$, 对E^{-1}存在概率为 1 的截断差分$\Gamma_2 \to \Gamma_1^*$, 如果Γ_1与Γ_1^*不重合, 则$\Gamma_0 \to \Gamma_2$构成了E的一个不可能差分区分器。许多文献给出了这种截断不可能差分的自动化搜索方法, 如$\mathcal{U}$方法[7]、UID方法[8]。在进行密钥恢复攻击时, 若部分加解密得到的中间状态差分满足了该不可能差分, 则可以将这一候选密钥排除。

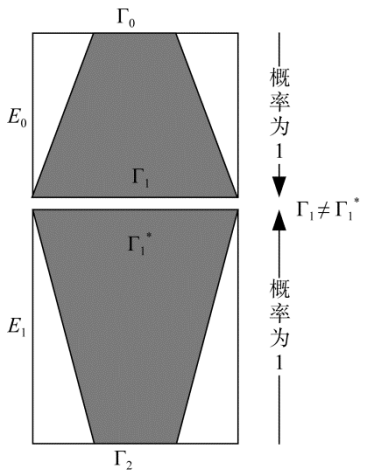

图 3　分组密码不可能差分区分器的构造

2.2　分组密码的线性分析方法综述

在 1993 年的欧密会上, 日本学者 Matsui 提出了对 DES 算法的一种新的攻击方法—“线性密码分析”[9]。线性分析是一种已知明文攻击方法, 它通过研究明文和密文之间的线性关系来恢复密钥。经过十几年的发展和完善, 线性分析已经成为现代分组密码设计时的重要设计准则之一。

2.2.1　基本的线性分析

线性攻击的基本思想是攻击者希望找到一个形如⑷的线性表达式,并且该表达式成立的概率与$\frac{1}{2}$之间的差距足够大,

$$P \cdot \alpha \oplus C \cdot \beta = K \cdot \gamma, \tag{4}$$

其中“·”表示标准的内积运算, α称为输入掩码, β称为输出掩码, 这样的一个表达式称为一条线性路线。简单来说, 我们期待在明文比特和密文比特之间找到一个概率线性关系, 即存在一个比特子集使得其中元素的异或和表现出非随机的分布。

一条线性路线的强弱可以用相关度的概念来衡量。

定义 6(布尔函数的相关度)**.**

一个布尔函数$f: \mathbb{F}_2^n \to \mathbb{F}_2$的相关度定义为

$cor(f) = 2^{-n}(\#\{x \in \mathbb{F}_2^n | f(x) = 0\} - \#\{x \in \mathbb{F}_2^n | f(x) = 1\})$.

定义 7(线性逼近的相关度)**.**

给定函数$F: \mathbb{F}_2^n \to \mathbb{F}_2^m$, 输入掩码为$a$, 输出掩码

为b的线性逼近$b\cdot F(x)\oplus a\cdot x$的相关度$cor(a,b)$定义为

$$cor(a,b)=\mathbf{Pr}\{b\cdot F(x)\oplus a\cdot x=0\}-\mathbf{Pr}\{b\cdot F(x)\oplus a\cdot x=1\}. \quad (5)$$

与相关度相关的一个概念是函数的傅里叶变换

$$\hat{F}(a,b)=\sum_{x}(-1)^{b\cdot F(x)\oplus a\cdot x}.$$

由上面的定义, 我们有

$$c=\frac{\hat{F}(a,b)}{2^n}.$$

线性路线的偏差ε与相关度c满足$c=2\varepsilon$。

给定一条形如⑷的线性路线(不失一般性, 假设偏差为正)和 N 个明文, 可以用下面的过程恢复部分密钥信息。

1. 遍历所有的明文, 记录使得等式⑷左侧等于 0 的明文个数T。

2. 若$T>N/2$, 则猜测$K\cdot\gamma=0$; 否则, 猜测$K\cdot\gamma=1$。

通过上述攻击我们可以得到关于密钥的一比特信息$K\cdot\gamma$。

在实际的攻击中, 我们通常将高偏差的线性逼近放在被攻击算法的中部, 猜测部分密钥后, 通过从前往后(从后往前)部分加(解)密得到中间状态的值, 并由此判断线性表达式成立与否。这种攻击过程与上述攻击过程基本类似, 但由于部分加解密的过程需要牵扯到对部分子密钥的猜测, 我们对每个猜测密钥建立计数器并记录线性表达式成立的次数, 按照相关度从大到小的顺序对候选密钥进行排序。文献[23]中给出了成功率P_S, 数据量N和优势a之间的数量关系:

定理 2([23, 定理 2]).

设P_S为攻击的成功率, 攻击中所使用的线性路线的偏差为ε, 对于一个针对k比特密钥的线性攻击, 所利用的线性路线的偏差为ε, 设P_S为攻击的成功率, N为攻击中使用的明文的数量, a为攻击的优势, 假设线性路线是否成立关于密钥独立, 则有

$$N=\left(\frac{\Phi^{-1}(P_S)+\Phi^{-1}(1-2^{-a-1})}{2\varepsilon}\right)^2. \quad (6)$$

由等式⑹我们可以看出, 线性攻击中所需的明文个数正比于$\frac{1}{\varepsilon^2}$。

注意到等式⑷的右侧确定了一条线性路线内部状态的掩码, 但在进行攻击时, 我们并没有对线性路线的内部做要求, 所以与差分分析中差分特征的概念类似, 线性分析中也存在线性壳的概念。文献[30]中作者基于下面的定理给出了线性壳的概念, 令$X\in\mathbb{F}_2^m, K\in\mathbb{F}_2^l, Y=Y(X,K), Y\in\mathbb{F}_2^n$为关于 X 和 K 的随机变量。

定理 3 ([30, 定理 1]).

若X和K相互独立, 且K均匀分布, 则对于所有的$\alpha\in\mathbb{F}_2^m,\beta\in\mathbb{F}_2^n,\gamma\in\mathbb{F}_2^l$有

$$2^{-l}\sum_{k\in\mathbb{F}_2^l}\left|\mathbf{Pr}_X\{X\cdot\alpha\oplus Y(X,K)\cdot\beta=0\}-\frac{1}{2}\right|^2=2^{-l}\sum_{k\in\mathbb{F}_2^l}\left|\mathbf{Pr}_X\{X\cdot\alpha\oplus Y(X,K)\cdot\beta\oplus k\cdot\gamma=0\}-\frac{1}{2}\right|^2=\sum_{c\in\mathbb{F}_2^l}\left|\mathbf{Pr}_{X,K}\{X\cdot\alpha\oplus Y(X,K)\cdot\beta\oplus K\cdot c=0\}-\frac{1}{2}\right|^2.$$

称线性路线族

$$P\cdot\alpha\oplus C\cdot\beta\oplus K\cdot\gamma,\gamma\in\mathbb{F}_2^l$$

为分组密码算法的线性壳。从上述定理中我们可以得到线性壳的偏差与相对应的线性路线的偏差间的关系, 但由于这种线性壳偏差的计算需要遍历整个密钥空间, 而算法分析的情境通常是在一个固定密钥下的, 所以该公式并不能实际的用来计算线性壳的偏差。当前计算线性壳偏差的方式是将所有输入掩码、输出掩码相同的线性路线的偏差(带符号)相加, 即便如此, 有时因为所涉及的线性路线条数较多, 而大多数的线性路线的偏差绝对值较小, 我们通常只取其中的一条或几条偏差绝对值较大的线性路线作为代表近似地计算线性壳的偏差。

2.2.2 多重线性分析

要改进基本的线性分析, 一个自然的想法是利用多条线性路线。Matsui 最先提出了这种改进, 在文献[10]中, 他同时利用了两条线性逼近改进了 DES 的攻击。同年, Burton 等人[11]建立了基于多条线性逼近的线性分析模型, 该模型基于一个较强的假设, 即所使用的线性逼近统计独立。

假设我们可以得到m条相互独立的、偏差有优势的线性逼近, 与基本的线性分析类似, 我们对每一条线性逼近统计使得该线性逼近成立的明文个数$T_i, i=1,2,\cdots,m$, 然后利用T_i计算下面的统计量C

$$C=\sum_{i=1}^{m}a_i\frac{T_i}{N},$$

其中a_i为对应线性逼近的权重, 文献[11]中证明, 当a_i与其对应的线性路线的偏差成比例时, 统计量C的区分优势最大。并在统计量之下给出了多重线性分析所需明文量N的计算公式。

定理 4 ([11, 定理 3]).

在独立性的假设之下, 多重线性分析的成功率为

$$P_S=\Phi\left(2\sqrt{N}\sqrt{\frac{\sum_{i=1}^{m}\varepsilon_i^2}{1-4\sum_{i=1}^{m}\varepsilon_i^2}}\right).$$

到目前为止，多重线性分析所需的数据量N并没有更好的计算公式。注意到与之前差分分析和线性分析数据量N的计算公式不同，这里的公式并没有涉及优势a，这是由于作者默认统计量C最大的密钥为正确密钥。

2.2.3 多维线性分析

注意到在多重线性分析模型建立在线性逼近统计独立的假设之上，但 Murphy[16]指出这一假设一般情况下不成立。2009 年的 FSE 会议上，Hermelin 等人[12]给出了多维线性分析的模型，该模型避免了多重线性模型中的独立性假设。

同截断差分一样，在多维线性分析的具体的应用中，我们所使用的线性逼近输入掩码和输出掩码满足一定的结构特性，本文中只讨论其构成线性空间的情况。这里我们仍利用图 2 中给出的明密文空间的划分，设多维线性区分器的输入掩码有2^s种可能的情况且形式为$(a_s,0)\in\mathbb{F}_2^s\times\{0\}$，输出掩码有$2^q$可能的情况且形式为$(b_q,0)\in\mathbb{F}_2^q\times\{0\}$，即共考虑$2^{s+q}$个不同的线性逼近。

多维线性区分器的强度可以用容度的概念来刻画，上述多维线性区分器$ML[(a_s,0),(b_q,0)]_{a_s\in\mathbb{F}_2^s,b_q\in\mathbb{F}_2^q}$的容度定义为

$$C=\sum_{(a_s,b_q)\neq(0,0)} cor^2\left(a_s\cdot x_s\oplus b_q\cdot y_q\right).$$

容度也可以用(x_s,y_q)的概率分布与$\mathbb{F}_2^s\times\mathbb{F}_2^q$上的均匀分布之间的$L^2$距离来计算[31]，严格的说，假设$(x_s,y_q)$的分布为

$$p_\eta=\mathbf{Pr}\{(x_s,y_q)=\eta\},\eta=0,1,\cdots,2^{s+q}-1,$$

则容度为

$$C=2^{s+q}\sum_{\eta=0}^{2^{s+q}-1}\left(p_\eta-\frac{1}{2^{s+q}}\right)^2.$$

根据分析中所使用的统计量的不同，多维线性分析又可以分为卡方方法和 LLR 方法。由于本文的讨论并不涉及多维线性分析的具体分析过程，这里不做过多介绍。

2.2.4 基本的零相关线性分析

零相关线性分析由 Bogdanov 和 Rijmen[13]于 2012 年首次提出，与传统的线性分析不同，零相关线性分析利用相关度为零的线性逼近来进行区分攻击或恢复密钥。虽然零相关线性分析与不可能差分分析在理论和技术两方面有很大的不同，但零相关线性分析可以看做不可能差分分析在线性分析领域的一种对偶方法。

利用零相关线性逼近进行分组密码算法与随机置换的区分时，需要计算给定线性逼近的相关度。通常的计算方法是遍历所有的明文输入，计算相应的输出，从而求得给定线性逼近的相关度。对于分组密码的零相关线性逼近求相关度，求得的相关度一定为 0，但对于随机置换，有下面的命题:

命题 1 ([13，命题 3]).

对n比特置换函数上非平凡的线性逼近相关度为 0 的概率可以如下近似计算

$$\frac{1}{\sqrt{2\pi}}2^{\frac{4-n}{2}},n\geqslant 5.$$

根据这一命题，我们可以确定最终密钥候选的数量，最后用一定数量的明密文对对候选密钥进行筛选，唯一确定正确密钥。

2.2.5 多重零相关线性分析

基本的零相关线性分析所需要的数据量为整个明文空间或2^{n-1}个选择明文，其中n为分组密码算法的分组长度。一方面，极高的数据复杂度需求极大地限制了零相关理论的应用，另一方面由于分组密码中一般存在大量的零相关线性逼近，对于零相关线性逼近的不充分应用也限制了零相关的发展。以此为出发点，为降低零相关线性分析模型中的数据复杂度, Bogdanov 和王美琴[14]在 2012 年的 FSE 会议上提出了利用多条零相关线性逼近来进行区分攻击或密钥恢复攻击，以此来降低数据复杂度。

多重零相关线性分析的出发点是利用较少的数据量来估计大量线性逼近的相关度，建立统计量来区分正确密钥与错误密钥。模型构建中的一个基本问题是对两个正态分布的区分。假设已知样本s取自正态分布$\mathcal{N}(\mu_0,\sigma_0)$或$\mathcal{N}(\mu_1,\sigma_1)$，其中$\mu_0$和$\mu_1$为正态分布的均值，$\sigma_0$和$\sigma_1$为标准差，不失一般性，假设$\mu_0<\mu_1$，需要判断样本来自哪个正态分布。判定的方法是选定一个阈值τ，若$s\leqslant\tau$，则判定$s\in\mathcal{N}(\mu_0,\sigma_0)$；否则判定$s\in\mathcal{N}(\mu_1,\sigma_1)$。这样存在两类错误概率

$$\alpha_0=\mathbf{Pr}\{s\in\mathcal{N}(\mu_1,\sigma_1)|s\in\mathcal{N}(\mu_0,\sigma_0)\},$$
$$\alpha_1=\mathbf{Pr}\{s\in\mathcal{N}(\mu_0,\sigma_0)|s\in\mathcal{N}(\mu_1,\sigma_1)\}.$$

假设攻击者拥有N个已知明文，ℓ条零相关线性逼近，假设$T_i,1\leqslant i\leqslant\ell$，是$N$个明文中满足第$i$条线性逼近的个数，计算下面的统计量

$$C=\sum_{i=1}^{\ell}\left(2\frac{T_i}{N}-1\right)^2.$$

多重零相关线性分析模型的建立基于下面的命题

命题 2 ([14，命题 2、3]).

设有分组密码算法上的ℓ个非平凡零相关线性逼近和N个已知明文，设$T_i,1\leqslant i\leqslant\ell$是$N$个明文中满

足第i条线性逼近的个数。假设T_i之间相互独立, 则对足够大的ℓ, N和n, 在正确密钥下

$$C_{cipher}\sim\mathcal{N}\left(\frac{\ell}{N},\frac{\sqrt{2\ell}}{N}\right),$$

在错误密钥下

$$C_{random}\sim\mathcal{N}\left(\frac{\ell}{N}+\frac{\ell}{2^n},\frac{\sqrt{2\ell}}{N}+\frac{\sqrt{2\ell}}{2^n}\right).$$

命题 3 ([14, 定理 1]).

用ℓ个非平凡的零相关线性逼近, 以错误概率α_0(正确密钥当成错误密钥的概率)和α_1(错误密钥当成正确密钥)来区分正确密钥和错误密钥, 需要的已知明文的个数需满足如下关系:

$$\frac{2^{n+0.5}}{N\sqrt{\ell}}\left(q_{1-\alpha_0}+q_{1-\alpha_1}\right)+\frac{q_{1-\alpha_1}\sqrt{2}}{\sqrt{\ell}}=1,$$

其中$q_{1-\alpha_0}$和$q_{1-\alpha_1}$为标准正态分布的分位数, 判定的阈值为: $\tau=\mu_0+\sigma_0 q_{1-\alpha_0}=\mu_1-\sigma_1 q_{1-\alpha_1}$。

由命题 3 可知, 假设存在ℓ条零相关线性逼近, 则多重零相关线性分析所需的数据复杂度约为$O\left(\frac{2^n}{\sqrt{\ell}}\right)$。与基本的零相关线性分析相比, 多重零相关线性分析的数据复杂度显著降低。

2.2.6 多维零相关线性分析

虽然多重零相关线性分析的数据复杂度较之前基本的零相关线性分析显著降低, 然而, 多重零相关的分析模型与多重线性分析模型类似, 基于一个强烈的假设, 即要求所使用的ℓ条零相关线性逼近相互独立, 这一条件通常难以满足。在 2012 年的亚洲密码年会上, Bogdanov[15]等人提出积分零相关区分器和多维零相关线性分析新模型。多维零相关线性分析模型与多重零相关线性分析模型相比, 在维持数据复杂度基本不变的基础之上, 不再依赖ℓ条零相关线性逼近相互独立这一强假设条件, 极大地完善了零相关线性分析理论。

设$(\alpha_i,\beta_i), 1\leqslant i\leqslant m$为$\mathbb{F}_2^n\to\mathbb{F}_2^t$上的 m 个相互独立的线性逼近, 记$x\in\mathbb{F}_2^n$为明文, $y\in\mathbb{F}_2^t$为加密过程中的部分中间状态值, 用一对(x,y)对 m 个(α_i,β_i)求值, 得到 m 比特的值, 记为$z_i, 1\leqslant i\leqslant m$:

$z=(z_1,z_2,\cdots,z_m)$, $z_i=\alpha_i\cdot x\oplus\beta_i\cdot y$,

多维零相关线性分析考虑z的概率分布。设在 N 个明密文对下, 利用计数器$V[z]$记录z出现的次数, 经过分析有: 在正确密钥下, $V[z]$服从多变量超几何分布; 在错误密钥下, $V[z]$服从多项分布。

利用$V[z]$构造统计量T

$$T=\sum_{z=0}^{2^m-1}\frac{\left(V[z]-\frac{N}{\ell}\right)^2}{N\cdot\frac{1}{\ell}},$$

通过概率统计中的相关知识, 我们有下面的命题:

命题 4 ([15, 命题 2]).

设可以得到N个明密文对以及ℓ条零相关线性逼近, 则统计量T服从$\mathcal{X}^2$分布。在正确密钥下, 其期望和方差为

$$\mu_0=(\ell-1)\frac{2^n-N}{2^n-1},$$

$$\sigma_0=2(\ell-1)\left(\frac{2^n-N}{2^n-1}\right)^2,$$

在错误密钥下, 其期望和方差为

$$\mu_1=(\ell-1),\sigma_1=2(\ell-1).$$

推论 1([15, 推论 2]).

n比特的分组密码算法的ℓ条零相关线性逼近构成了一个$\log_2\ell$维的线性空间, 以错误概率α_0(正确密钥当成错误密钥的概率)和α_1(错误密钥当成正确密钥)来区分正确密钥和错误密钥, 需要的已知明文个数需满足如下关系:

$$N=\frac{(2^n-1)\left(q_{1-\alpha_0}+q_{1-\alpha_1}\right)}{\sqrt{\frac{\ell-1}{2}}+q_{1-\alpha_0}}+1,$$

其中$q_{1-\alpha_0}$和$q_{1-\alpha_1}$为标准正态分布的分位数, 判定的阈值为: $\tau=\mu_0+\sigma_0 q_{1-\alpha_0}=\mu_1-\sigma_1 q_{1-\alpha_1}$。

2.3 分组密码的其他分析方法

除了差分和线性这两条基本的主线之外, 许多形式上与上述统计分析方法看似不同的统计分析方法也在密码学的发展过程中被提出来, 这里我们只简要介绍两种: 积分攻击和统计饱和度攻击。

2.3.1 积分分析

积分区分器首先由 Kundsen[17]在攻击 Square 算法时提出的, 由于这个原因, 积分区分器[18]也称为 Square 区分器。积分是继差分分析和线性分析之后, 密码学界公认的最有效的密码分析方法之一。最初的攻击方法更多的与算法的结构有关, 而与算法部件的具体取值关系不大, 作为主要针对面向字节运算算法安全性的密码分析方法, 积分攻击从其出现就受到密码学界的广泛关注。

积分分析中经常会涉及到平衡性的概念。

定义 8 (平衡性).

称函数$F:\mathbb{F}_2^n\to\mathbb{F}_2^k$是平衡的若所有原像集的势都相同, 即集合

$$F^{-1}(y)=\{x\in\mathbb{F}_2^n|F(x)=y\}$$

的大小与y无关。

积分攻击考虑一系列状态求和, 考虑到二元域上, 差分的定义对应为两个元素求和, 因此积分攻击可以看做差分攻击的一种推广。最初的积分区分器基于如下观测: 在固定明文的一些比特, 遍历明

文的另外一部分比特时, 得到的密文的一些比特是平衡的。传统的积分攻击方法对基于比特运算的算法是无效的, 有鉴于此, Z'aba 等人在文献[32]中提出了基于比特的积分攻击方法, 其实质为一种计数方法, 通过分析特定比特位置上元素出现次数的奇偶性来确定这一比特的积分特性, 此时的积分特性通常表现为这些比特位置的异或和为常数(通常为 0)。

2.3.2 统计饱和度攻击

统计饱和度攻击[19]主要想法是利用某些特殊的密码算法中较弱的扩散性, 通过固定明文的某些比特, 考虑密文部分比特的分布。

给定一个加密函数

$$F: \mathbb{F}_2^n \to \mathbb{F}_2^n,$$

统计饱和度攻击考虑当F的输入部分固定时, 其输出的分布。这里我们同样考虑对明文空间和密文空间的分割,

$$F: \mathbb{F}_2^s \times \mathbb{F}_2^t \to \mathbb{F}_2^q \times \mathbb{F}_2^r$$

$$F(x,y) = \left(F^{(1)}(x,y), F^{(2)}(x,y)\right)$$

为了方便, 另h_y表示将F输入的后t比特固定为y, 且在输出的前q维空间上的限制, 即:

$$h_y = \mathbb{F}_2^s \to \mathbb{F}_2^q$$

$$h_y(x) = F^{(1)}(x,y)$$

在统计饱和度攻击中我们需要考虑h_y的容度C_h, 而攻击所需的数据复杂度为$O\left(\frac{1}{C_h}\right)$。

虽然统计饱和度攻击对于轻量级算法 PRESENT[26]给出了最好攻击, 但文献[19]中并没有给出容度的C_h的计算公式, 从而攻击的复杂度并不能确切的评估。直到 2011 年的欧密会上, Leander[27]证明统计饱和度攻击与多维线性攻击在本质上是相同的, 这种关系允许我们正确的评估统计饱和度攻击中使用的容度, 从而更准确的衡量攻击的复杂度。作者利用统计饱和度攻击和多维线性攻击之间的关系, 给出了C_h的理论计算公式。

定理 5 ([27, 定理 5]).

统计饱和度攻击中的平均容度与线性逼近的相关度有如下关系:

$$\overline{C_h} = \sum_{a \in \{0\} \times \mathbb{F}_2^t, b \in \mathbb{F}_2^q \times \{0\}} cor^2(a,b).$$

人们在利用各种统计分析方法对算法的安全性进行评估时, 经常会发现一些相似的现象, 如: 某两种区分器的轮数总是相同的、两种统计分析方法中所利用的统计量存在数学关系等。所以, 在建立新的统计分析方法的同时, 人们渐渐将关注点转移到研究各种已有的统计分析方法的关联性上。尽管在分析和统计方式有着形式上的不同, 但经过仔细分析之后发现许多看似不同的统计分析方法之间有着一些关联性, 研究这种关联性不管是从理论上还是从分析分组密码安全性的角度都是非常有必要的。近几年, 各种统计分析方法之间的关联性逐渐被建立起来。这些关联性的建立一方面有助于我们对已知的分组密码分析方法进行分类, 另一方面这些关联性可能会给出分组密码安全性的补充信息。后面我们将分四节介绍已有的统计分析方法之间的关联性。

3 差分与线性分析

早在 1994 年, Chabaud 和 Vaudenay[20]就提出了线性分析和差分分析之间的数学关系, 他们严格的证明了差分分析中差分特征的概率与线性分析中线性逼近的相关度之间的数学关系。

定理 6 ([20, 定理 1]).

设$F: \mathbb{F}_2^n \to \mathbb{F}_2^m$为一个向量布尔函数, 则有 $\mathbf{Pr}\{\delta \to \Delta\} =$

$$2^{-m} \sum_{a \in \mathbb{F}_2^n} \sum_{b \in \mathbb{F}_2^m} (-1)^{a \cdot \delta \oplus b \cdot \Delta} cor^2(a,b).$$

但这一数学关系后来并没有实际应用于差分概率的计算, 因为直接的应用需要涉及2^{n+m}个相关度的计算, 当n和m都比较大时, 这种计算方法是不可行的。直到 2013 年的欧密会上, Blondean 和 Nyberg[21]将这一数学关系进行推广, 应用的划分后的明密文空间, 得到了特定形式的截断差分的概率与多维线性中容度之间的关系。

定理 7 ([21, 定理 3]).

对任意的$\delta_s \in \mathbb{F}_2^s, \Delta_q \in \mathbb{F}_2^q$下面的式子成立

$$2^{-t} \sum_{\delta_t \in \mathbb{F}_2^t, \Delta_r \in \mathbb{F}_2^r} \mathbf{Pr}\{(\delta_s, \delta_t) \to (\Delta_q, \Delta_r)\} =$$

$$2^{-q} \sum_{a_s \in \mathbb{F}_2^s, b_q \in \mathbb{F}_2^q} (-1)^{a_s \cdot \delta_s \oplus b_q \cdot \Delta_q} cor^2\left((a_s, 0), (b_q, 0)\right).$$

利用这一关系, 作者令$q = t$, 使所有非平凡的线性逼近的相关度和差分的概率为零, 导出了特殊情况下, 截断不可能差分和零相关线性分析的一种等价性, 注意到这种等价性并不依赖于分组密码的结构。

推论 2([21, 推论 2]).

设$F: \mathbb{F}_2^s \times \mathbb{F}_2^t \to \mathbb{F}_2^q \times \mathbb{F}_2^r, t = q$为向量布尔函数, 则所有的非平凡线性逼近$(a_s, 0) \cdot x \oplus (b_q, 0) \cdot F(x), a_s \in \mathbb{F}_2^s, b_q \in \mathbb{F}_2^q\{0\}$相关度为 0 当且仅当所有的非平方凡差分$(0, \delta_t) \to (0, \Delta_r), \delta_t \in \mathbb{F}_2^t\{0\}, \Delta_r \in \mathbb{F}_2^r$为不可能差分。

但正如文献[24]中指出的那样, 虽然这一等价性与底层算法的结构无关, 但对于大多数的不可能差分区分器和零相关区分器, 推论 2 并不能直接应用, 因为这一等价性的成立需要满足区分器的维数为n, 即区分器包含2^n个不可能差分特征或零相关路线。对于大多数算法而言, 只有极少数的不可能差分和零相关路线, 所以零相关与不可能差分之间的等价性并没有得到实际的解决。

由于线性分析与差分分析的攻击条件不同(线性分析为已知明文攻击, 差分分析为选择明文攻击), 研究二者之间的关联性有助于我们选择更适合攻击的区分器进行算法攻击。但因为现存的数学关系不能有效的应用, 所以二者之间更实用的关联性有待进一步发掘。

4 多维线性、截断差分与统计饱和度攻击

4.1 多维线性与截断差分

Blondeau 和 Nyberg[22]进一步应用定理 7 的结论, 考虑令$\delta_s = 0, \Delta_q = 0$, 而不对线性逼近的相关度和差分概率做假设, 得到了特定形式(见图 4)的截断差分概率与多维线性容度之间的关系。

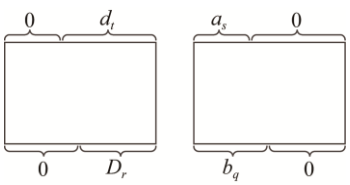

图 4 多维线性区分器与其对应的截断差分

推论 3([22, 推论 1]).

设截断差分的概率为$p = 2^{-t}\sum_{\delta_t \in \mathbb{F}_2^t,\, \Delta_r \in \mathbb{F}_2^r} \mathbf{Pr}\{(0, \delta_t) \to (0, \Delta_r)\}$, 多维线性分析中所使用的区分器的容度为$C = \sum_{(a_s, b_q) \neq (0,0)} cor^2\left(a_s \cdot x_s \oplus b_q \cdot y_q\right)$。则有

$$p = 2^{-q}(C + 1).$$

为导出多维线性分析和截断差分分析在攻击时数据复杂度的关联性, 作者分别利用定理 1 和定理 2 给出了利用上述截断差分和多维线性区分器进行攻击时的数据复杂度。

命题 5([22, 命题 2]).

假设上述截断差分的概率为p, 攻击的成功率为 50%, 优势为a比特, 则攻击所需的数据量为

$$N^{TD} = \frac{2^{-q+1}}{S \cdot (p - 2^{-q})^2} \cdot \Phi^{-1}(1 - 2^{-a}),$$

其中S为每个结构体(在差分分析中, 为了以较大的效率利用选择的明密文对, 通常会构造结构体(structure))中选择明文的数量。

命题 6([22, 命题 1]).

假设上述多维线性逼近的容度为C, 则利用其进行攻击时所需的数据量为

$$N^{ML} = \frac{2^{\frac{s+q+1}{2}}}{C} \cdot \Phi^{-1}(1 - 2^{-a}).$$

命题 5、6 结合上推论 3, 作者得到了上述多维线性和截断差分分析数据量之间的关系。

推论 4([22, 推论 2]).

上述截断差分与多维线性分析中所使用的数据量满足下面的关系:

$$N^{TD} = \frac{(N^{ML})^2}{S \cdot 2^s} = \frac{2^{q+1}}{S \cdot C^2} \cdot \Phi^{-1}(1 - 2^{-a}).$$

特别的, 当$S = 2^t$时, 我们有

$$N^{TD} = 2^{-n}(N^{ML})^2.$$

现存的截断差分分析和多维线性分析攻击时所使用的数据量之间的关系可以帮助我们判断哪一种攻击方法更适用于目标算法。另一方面, 二者在攻击方式上的不同使得我们可以通过将截断差分区分器转换为相应的多维线性区分器, 从而将选择明文的要求降低为已知明文。不过由于现存的等价性不依赖与算法的内部结构, 所以这种等价性虽具有一般性, 但欠缺实用性。这是因为这种等价性要求截断差分非零的部分与多维线性非零的部分互补, 这就使得若其中一种区分器的维数比较小, 另一种区分器的维数会非常大, 这种区分器在密钥恢复阶段并不具备优势。

4.2 多维线性与统计饱和度

在 2.3.2 节关于统计饱和度攻击的介绍中, 定理 5 给出的平均容度与相关度的数学关系说明统计饱和度攻击与多维线性攻击在本质上是相同的, 这种关系允许我们对统计饱和度攻击复杂度做一个更为准确的评估。从这一方面来讲, 研究统计分析方法之间的关系, 也有助于我们用较为完善的理论体系丰富和发展密码体系中较为薄弱的环节。

4.3 截断差分与统计饱和度

Blondeau 和 Nyberg[22]通过对截断差分分析和统计饱和度分析中所使用的统计量的观察, 发现这两种统计量只相差一个常数, 从而证明截断差分分析和统计饱和度分析在本质上也是相同的。

5 零相关、不可能差分与积分攻击

5.1 零相关与不可能差分分析

对于以字为单位的分组密码, 零相关和不可能差分分析是两种非常有效的攻击方法。尽管在密钥

恢复阶段这两种分析方法存在一些差异, 但这两种区分器通常会覆盖相同的轮数。然而在一些特殊的情况下, 这两种区分器的轮数并不相同。文献[20]中给出了特定形式的零相关区分器和不可能差分区分器之间的数学关系(上述推论 2), 这一数学关系并不依赖与分组密码的结构, 正如文献[24]中所指出的那样, 虽然这一数学关系与底层的算法结构无关, 但对于大多数的不可能差分区分器和零相关区分器, 推论 2 并不能直接应用, 因为这一数学关系需要满足区分器的维数为n, 即区分器包含2^n个不可能差分特征或零相关区分器。对于大多数算法而言, 只有极少数的不可能差分路线或零相关路线, 所以零相关与不可能差分之间的等价性或不等价性并没有得到实际的解决。

Blondeau 等人在文献[24]中解决了对于 Feistel 和 Skipjack 类型的分组密码算法这两种分析方法的条件等价性, 由于这种等价性考虑了分组密码的结构, 所以这一结论虽不具有一般性, 但更具实用性。首先对广义 Feistel 结构, 作者利用矩阵方法给出了零相关与不可能差分的等价条件。在描述这些关联性之前, 我们首先给出一些基本的符号表示。

令n为面向字结构的分组密码算法的分组长度, 算法的每个状态有b个字, 每个字的长度为s, 即有$n = b \cdot s$。n比特的状态X可以表示为$X = (X_1, X_2, \cdots, X_b)$, 每个$X_i$的长度为$s$。对于 Feistel 结构的算法, 其轮函数满足: 每一个输入分支或者可以是一个非线性函数的输入或者与一个这样的非线性函数的输出异或, 通常可以表示为一个非线性层(F层)和一个线性层(P层)。

定义 9([24, 定义 1])**.**

不考虑密钥和常数加的过程, 具有b个分支的 Feistel 结构算法的轮函数可以表示为两个$b \times b$的矩阵$\mathcal{F}$、$\mathcal{P}$的组合, 其中$\mathcal{F}$和$\mathcal{P}$矩阵中元素由$\{0,1,F_i\}$构成, 其中$\{F_i\}_{i \leqslant I}$表示轮函数中的非线性函数。

- 非线性层表示($\mathcal{F}$层): 矩阵$\mathcal{F}$对角线上的元素为 1; 如果函数F_i的输入为第k分支, 其输出与第j分支进行异或, 将$\mathcal{F}$矩阵的第j行第k列设置为F_i; 其他元素为 0。这同时意味着在每一行每一列中至多含有一个F_i。
- 置换层表示($\mathcal{P}$层): 矩阵$\mathcal{P}$是一个置换矩阵, 在每一行每一列中只有一个非零元素。

则一个 Feistel 结构的轮函数可以表示为上面的两个矩阵的组合, 即$\mathcal{R} = \mathcal{P} \cdot \mathcal{F}$。

用矩阵定义的Feistel结构来表示r轮的不可能差分$\Gamma_0 \to \Gamma_1$, 则存在l和$m(l + m = r)$使得输入差分Γ_0正向传播l轮得到的与输出差分Γ_1逆向传播m轮得到的存在着不一致的部分, 也就是:

$$\mathcal{R}^l \cdot \Gamma_0 \nleftrightarrow \mathcal{R}^{-m} \cdot \Gamma_1.$$

定义 10([24, 定义 2])**.**

对于给定轮函数$\mathcal{R} = \mathcal{P} \cdot \mathcal{F}$的Feistel结构的算法, 我们定义$\mathcal{R}$的镜像函数为$\mathcal{M} = \mathcal{P} \cdot \mathcal{F}^T$, 其中$\mathcal{F}^T$为函数$\mathcal{F}$的转置。

Blondeau 等人在下面的定理中给出了 Feistel 结构算法的零相关路线和不可能差分路线的等价条件。

定理 8([24, 定理])**.**

令$\mathcal{R}$为 Feistel 结构算法的轮函数的矩阵表示, $\mathcal{M}$为$\mathcal{R}$的镜像矩阵, 如果存在一个$b \times b$的置换矩阵$\mathcal{Q}$满足:

$$\mathcal{R} = \mathcal{Q} \cdot \mathcal{M} \cdot \mathcal{Q}^{-1} \; or \; \mathcal{R} = \mathcal{Q} \cdot \mathcal{M}^{-1} \cdot \mathcal{Q}^{-1}.$$

我们可以得出: 分组算法存在m条r轮的不可能差分路线构成的不可能差分区分器当且仅当该分组算法也存在m条r轮零相关路线构成的零相关区分器。

Skipjack 类型的算法轮函数中的非线性函数的输入输出在同一个分支上, 非线性函数部分为F层; 在F层后将不同分支的状态进行线性操作的部分为X层; 最后一个分支位置的置换层P层, 其矩阵定义如下。

定义 11([24, 定义 3])**.**

对于具有b个分支的 Skipjack 类型的算法, 其轮函数可以表示为 3 个$b \times b$矩阵$\mathcal{G}$、$\mathcal{X}$、$\mathcal{P}$的组合, 这些矩阵中的元素在$\{0,1,F\}$中取值, 其中F表示非线性操作。

- F层表示: 函数$\mathcal{G}$的对角线上元素为 1 或F, 如果非线性操作F作用到第j分支上, 则矩阵$\mathcal{G}$的第j个对角元素为F。
- X层表示: 矩阵$\mathcal{X}$在对角线上的元素为 1, 而且每行每列至多两个 1。
- P层表示: 矩阵$\mathcal{P}$是一个置换矩阵, 每行每列只有一个非零元素。

则一个 Skipjack 类型的算法轮函数可以表示为一个$b \times b$的矩阵$\mathcal{R} = \mathcal{P} \cdot \mathcal{X} \cdot \mathcal{G}$。

定义 12([24, 定义 4])**.**

给定轮函数为$\mathcal{R} = \mathcal{P} \cdot \mathcal{X} \cdot \mathcal{G}$的 Skipjack 类型算法, 该轮函数的镜像函数为$\mathcal{M} = \mathcal{P} \cdot \mathcal{X}^T \cdot \mathcal{G}$。

对于 Skipjack 类型的分组算法, 下面的定理给出了零相关和不可能差分等价的条件。

定理 9([24, 定理 2])**.**

$\mathcal{R}$为Skipjack类型算法的轮函数矩阵表示, $\mathcal{M}$为$\mathcal{R}$的镜像函数矩阵表示, 如果存在一个$b \times b$的置换矩阵$\mathcal{Q}$满足:

$$\mathcal{R} = \mathcal{Q} \cdot \mathcal{M} \cdot \mathcal{Q}^{-1} \ or$$
$$\mathcal{G} \cdot \mathcal{P} \cdot \mathcal{X} = \mathcal{Q} \cdot \mathcal{M}^{-1} \cdot \mathcal{Q}^{-1},$$

我们可以得出: 分组算法存在m条r轮的不可能差分路线构成的不可能差分区分器当且仅当该分组算法也存在m条r轮零相关路线构成的零相关区分器。

对于其他的一些具有广义 Feistel 结构的算法, 也存在类似定理 8 和定理 9 的结论。

定理 10([24, 定理 3]).

$\mathcal{R} = \mathcal{P} \cdot \mathcal{X} \cdot \mathcal{F}$为广义 Feistel 结构算法的轮函数矩阵表示, $\mathcal{M}$为$\mathcal{R}$的镜像函数矩阵表示$\mathcal{M} = \mathcal{P} \cdot \mathcal{X}^T \cdot \mathcal{F}^T$, 如果存在一个$b \times b$的置换矩阵$\mathcal{Q}$满足:

$$\mathcal{R} = \mathcal{Q} \cdot \mathcal{M} \cdot \mathcal{Q}^{-1} \ or$$
$$\mathcal{R} = \mathcal{Q} \cdot \mathcal{M}^{-1} \cdot \mathcal{Q}^{-1} \ or$$
$$\mathcal{F} \cdot \mathcal{P} \cdot \mathcal{X} = \mathcal{Q} \cdot \mathcal{M}^{-1} \cdot \mathcal{Q}^{-1}.$$

我们可以得出: 分组算法存在m条r轮的不可能差分路线构成的不可能差分区分器当且仅当该分组算法也存在m条r轮零相关路线构成的零相关区分器。

在 2015 年的美洲密码年会上, 孙兵等人[25]对具有 SP 结构的 Feistel 算法以及 SPN 结构的算法的零相关分析和不可能差分分析之间的联系进行了研究。首先给出一类算法的定义。

定义 13([25, 定义 2]).

令$E: \mathbb{F}_2^n \to \mathbb{F}_2^n$为一个分组算法, 其中非线性部分由双射$S$盒构成。

1. $\mathbb{F}_2^n$上的一类结构$\mathcal{E}^E$定义为一组分组密码算法E'的集合, 这些算法除了S盒取遍相应域上所有可能的双射变换之外, 其他部分完全相同。

2. 令$a, b \in \mathbb{F}_2^n$, 对于任意$E' \in \mathcal{E}^E$, $a \to b$是E'的一个不可能差分路线(零相关路线), 则$a \to b$被称为结构$\mathcal{E}^E$的一个不可能差分路线(零相关路线)。

定义 14([25, 定义 3]).

令$\mathcal{F}_{SP}$为分支数为 2 的、具有 SP 结构轮函数的 Feistel 结构, P是轮函数中线性变换矩阵。σ进行交换 Feistel 状态左半边与右半边的操作, 则$\mathcal{F}_{SP}$的对偶结构$\mathcal{F}_{SP}^{\perp}$定义为$\sigma\mathcal{F}_{P^TS}\sigma$。

令$\mathcal{E}_{SP}$为 SPN 结构, 线性变换可逆矩阵为P, 则$\mathcal{E}_{SP}$的对偶结构$\mathcal{E}_{SP}^{\perp}$定义为$\mathcal{E}_{S(P^{-1})^T}$。

对于$\mathcal{F}_{SP}$结构, 零相关和不可能差分的等价性见下述定理。

定理 11([25, 定理 1]).

$a \to b$是$\mathcal{F}_{SP}$的一条r轮不可能差分路线当且仅当它是$\mathcal{F}_{SP}^{\perp}$的一条r轮零相关路线。

对$\mathcal{E}_{SP}$结构具有类似结论。

定理 12([25, 定理 2]).

$a \to b$是$\mathcal{E}_{SP}$的一条r轮不可能差分路线当且仅当它是$\mathcal{E}_{SP}^{\perp}$的一条r轮零相关路线。

当中的矩阵P是一个可逆矩阵时, 由于存在等式

$$\mathcal{F}_{P^TS} = ((P^T)^{-1}, (P^T)^{-1})\mathcal{F}_{SP^T}(P^T, PT),$$

可以得到如下推论。

推论 5([25, 推论 2]).

令$\mathcal{F}_{SP}$为具有SP结构轮函数的 Feistel 结构, 如果轮函数中的线性矩阵P是可逆的, 则寻找$\mathcal{F}_{SP}$的零相关路线等价于寻找$\mathcal{F}_{SP}^{\perp}$的不可能差分路线。

进一步, 如果$\mathcal{F}_{SP} = \mathcal{F}_{SP^T}$或$\mathcal{E}_{SP} = \mathcal{E}_{S(P^{-1})^T}$则有如下推论。

推论 6([25, 推论 3]).

令$\mathcal{F}_{SP}$为具有SP结构轮函数的 Feistel 结构, 如果轮函数中的线性矩阵P是可逆的, 且有$P = P^T$, 则在不可能差分和零相关之间存在一一对应关系。对于 SPN 结构$\mathcal{E}_{SP}$, 如果$P^TP = E, a \to b$是一条不可能差分路线当且仅当它是一条零相关线性路线。

通过研究零相关与不可能差分区分器之间的关联, 在已知其中一类区分器的情况下, 我们可以利用该关联性判断是否存在与之类似的另一类区分器, 并且在存在的情况下直接将等价区分器推导得出。在研究零相关与不可能差分关系的文献中, 许多新的区分器被提出来。另一方面这两种区分器在攻击方式上也不同, 我们可以从等价的区分器中选择更适合攻击的区分器进行算法攻击。

不过这些等价性也存在一些不足: 这些关联覆盖的结构还不够完整, 而且等价关系存在的条件比较苛刻。例如: 文献[24]中, 等价性建立在$\mathcal{Q}$是置换的条件之下, 所以其无法用 SMS4 或 MARS 算法自己的结构解释零相关与不可能差分区分器的等价性, 这就限制了等价关系的应用; 文献[25], 主要针对SP结构的 Feistel 算法和 SPN 算法。

5.2 零相关与积分攻击

积分区分器包括两类:

- 在选择明文(或密文)满足一定条件下, 输出的某些比特异或和为 0(称其为零和积分区分器)。
- 在选择明文(或密文)满足一定条件下, 输出的某些比特构成的数出现次数完全相等(称其为平衡积分区分器)。

零相关线性区分器主要与平衡积分区分器存在关联。

2012 年的亚洲密码年会上, Bogdanov[13]等人给出了特定形式的零相关区分器与积分区分器之间的关系。与一般的积分区分器考虑输出某部分异或和的性质不同, 文献[13]中所考虑的积分区分器要求输出在某一空间内满足平衡特性。在给出具体关联之前, 我们根据文献[13]先给出一些定义。

如前面做的一样, 我们将函数$H\colon \mathbb{F}_2^n \to \mathbb{F}_2^n$的输入输出分别做划分, 表示如下

$$H\colon \mathbb{F}_2^r \times \mathbb{F}_2^s \to \mathbb{F}_2^t \times \mathbb{F}_2^u$$
$$H(x,y) = \big(H_1(x,y), H_2(x,y)\big).$$

定义函数T_λ如下:

$$T_\lambda\colon \mathbb{F}_2^s \to \mathbb{F}_2^t$$
$$T_\lambda = H_1(\lambda, y).$$

根据上述定义, 我们可以表示出平衡积分区分器和零相关线性区分器的数学表示形式。

- 平衡积分区分器: 当x固定为常数, y遍历所有可能的取值, 输出的前t比特是平衡的(也就是每一个数出现的次数相等)表示为:

$$\forall \lambda, b_1, \widehat{T_\lambda}(0, b_1) = 0.$$

- 零相关区分器: 当函数H的输入掩码为$a = (a_1, 0)$, 输出掩码为$b = (b_1, 0)$时, $a \to b$是零相关的表示为:

$$\forall a_1, b_1, \widehat{H}(a, b) = 0.$$

实际上, 上述零相关区分器是多维的, 表示只要a_1、b_1不同时为零的情况下, $a \to b$是零相关的。

命题 7([13, 命题 1])**.**

若输入掩码a与输出掩码b相互独立, 则下面两条结论等价:

1. 对任意的$a = (a_1, 0)$和$b = (b_1, 0) \neq 0$, 线性逼近$a \cdot x + b \cdot H(x)$是零相关线性逼近;

2. 对任意的λ, T_λ是平衡的。

文献[13]指出: 积分区分器可以无条件的转换为一个零相关区分器, 而当零相关区分器转化为积分区分器是要求零相关区分器的输入掩码与输出掩码相互独立。这一等价性非常有趣, 因为零相关分析和积分分析不管是在区分器的构造上还是密钥恢复攻击上看起来都十分不同。首先, 零相关分析是已知明文攻击, 而积分攻击是选择明文攻击。其次, 在零相关区分器前面加轮数并不会使攻击的数据复杂度增加, 但在积分区分器前面加轮数一般会导致所需的选择明文的个数增加, 这是因为区分器的头部要求某些比特为固定值, 为满足这一条件就需要以牺牲数据复杂度为代价。除此之外, 满足平衡特性的积分区分器一般可以通过将平衡特性放宽为零和特性(或使得最后输出某部分的函数不具有最高的代数次数)而使区分器加长一轮, 但对于零相关区分器, 还没有发现这种可以加长一轮的性质。

上面的零相关区分器与积分区分器相互推导的条件限制着命题 7 的一般性应用, 观察命题 7 中的条件, 要求输入输出掩码是相互独立的, 而且掩码是由非零任意值与 0 级联而成, 这就出现了两种无法直接应用命题 7 的情况:

1. 输入掩码中要求某些比特掩码值必须为 1 时的零相关路线(在 ARX 结构的算法或是比特级操作的算法会出现此种路线)。

2. 输入输出掩码不独立, 例如要求$a_1 = b_1$或$a_1 \neq b_1$条件下的零相关路线。

对于第 1 种情况, 温隆和王美琴[33]对其作了进一步研究, 给出的结果主要是从零相关路线推导积分路线, 由于在输入掩码中含有必须为 1 的部分, 其将算法H'的输入分割成了三部分, 表示如下:

$$H'\colon \mathbb{F}_2^r \times \mathbb{F}_2 \times \mathbb{F}_2^s \to \mathbb{F}_2^t \times \mathbb{F}_2^u$$
$$H'(x,y,z) = \big(H_1'(x,y,z), H_2'(x,y,z)\big).$$

相应的函数$T_{\lambda\|\lambda'}\colon \mathbb{F}_2^s \to \mathbb{F}_2^t$定义为$T_{\lambda\|\lambda'}(z) = H_1'(\lambda, \lambda', z)$。

命题 8([33, 命题 2])**.**

当线性逼近的掩码a和b相互独立, 且对任意$a = (a_1 \| 1, 0), a_1 \in \mathbb{F}_2^r$、$b = (b_1, 0) \neq 0, b_1 \in \mathbb{F}_2^t$, 线性逼近$a \to b$是零相关的(零相关区分器),我们可以得到一条零和积分区分器: 对任意的λ, 函数$H_1'(x,y,z)$的输出异或和为 0, 也就是$\bigoplus_{y\|z} H_1'(\lambda, y, z)$。

对于第 2 种情况, 输入输出掩码不独立的情况下, 从零相关路线推导积分区分器的问题由孙兵等人在文献[25]中得到解决。

推论 7([25, 推论 4])**.**

令$F\colon \mathbb{F}_2^n \to \mathbb{F}_2^n$是$\mathbb{F}_2^n$上的函数, A是$\mathbb{F}_2^n$上的子空间, $b \in \mathbb{F}_2^n \backslash \{0\}$。假设$A \longrightarrow b$是函数$F$的零相关线性壳, 则对任意的$\lambda \in \mathbb{F}_2^n, b \cdot F(x \oplus \lambda)$在$A^\perp$是平衡的。

从推论 7 可以看出, 只要零相关线性路线的输入掩码构成一个子空间, 就可以从中导出一条积分路线, 进一步, 输入掩码构成子空间这一条件也可以去掉, 得出如下定理。

定理 13([25, 定理 3])**.**

分组密码算法的任一非平凡的零相关线性壳总是可以推导出一条积分路线。

定理 13 成立主要是因为: 对任一条零相关线性路线$a, \{0, a\}$就可以构成一个子空间, 再根据推论 7 就得到了定理 13 的结论。

零相关分析与积分攻击之间的关联性有助于我们利用这种关联性构造新的区分器, 并且根据攻击方式的不同, 我们可以从等价区分器中选择更适合攻击的区分器。现有关联性之间的不足在于: 目前的结果虽然可以无条件的将零相关区分器转化为积分区分器, 但是是在减少使用的零相关路线条数的情况下实现的, 这可能会导致找到的积分区分器需要较大的数据量, 影响了区分器攻击的有效性。例如对于 ARX 结构的算法, 零相关路线的掩码中会存在必须为 1 的部分, 无法构成大的子空间, 如何将更多的

零相关路线应用到应用的积分区分器的构造中值得进一步思考。另一方面, 从积分区分器到零相关的转化目前为止局限于平衡积分区分器方面, 是否可以从零和区分器构造零相关路线还未被解决。

5.3 不可能差分与积分攻击

孙兵等人在文献[25]中讨论了不可能差分区分器和积分区分器之间的关联, 这些结果主要是从上面零相关路线与不可能差分路线之间的关联以及零相关路线与积分路线之间的关联推导而出。

由于$\mathcal{E}^{\perp}$的一条零相关线性壳总可以导出$\mathcal{E}^{\perp}$的一条积分路线, 而$\mathcal{E}^{\perp}$的零相关线性壳可以从$\mathcal{E}$的不可能差分路线导出, 从而得出下述定理。

定理 14([25, 定理 4]).

令$\mathcal{E} \in \{\mathcal{F}_{SP}, \mathcal{E}_{SP}\}$, 则$\mathcal{E}$的一条不可能差分路线总是可导出$\mathcal{E}^{\perp}$的一条积分路线。

当A_1、A_2为线性变换, $\mathcal{E}^{\perp} = A_2\mathcal{E}A_1$时, 可以得到不可能差分路线与积分路线的直接关联:

推论 8([25, 推论 5]).

令$\mathcal{F}_{SP}$为轮函数为SP结构的 Feistel 结构算法, 且轮函数中线性变换为P。如果P是可逆的, 且对任意的$(x_0, x_1, \dots x_{t-1}) \in \mathbb{F}_2^{s\times t}$, 存在对这$t$个元素的置换$\pi$, 使得:

$$P(x_0, x_1, \dots x_{t-1}) = \pi^{-1}P^T\pi(x_0, x_1, \dots x_{t-1}),$$

则对于$\mathcal{F}_{SP}$, 一条不可能差分路线可以导出一条积分路线。

对于 SPN 结构的算法, 也存在着不可能差分与积分路线之间的直接关联。

推论 9([25, 推论 6]).

令$\mathcal{E}_{SP}$为 SPN 结构的算法, 线性层表示为P, 如果$P^TP = diag(Q_1, Q_2, \dots Q_t)$, 其中$Q_i \in \mathbb{F}_2^{s\times s}$, 则对于$\mathcal{E}_{SP}$, 一条不可能差分路线可以导出一条积分路线。

如果线性层P是一个比特置换矩阵, 则一定满足$P^TP = E$是对角矩阵, 则得出如下结论。

推论 10([25, 推论 7]).

对任意的以比特置换为线性混乱层的 SPN 结构的算法, 一条r轮的不可能差分路线可以导出一条r轮积分路线。

目前为止, 不可能差分与积分攻击之间的关联主要是从不可能差分与零相关以及积分与零相关之间的关联性推导而出, 这两者之间是否存在其他的、两者独有的关联值得进一步思考。

6 不可能差分与截断差分分析

李雷波等人[28]受到不可能差分的启发, 通过将截断不可能差分中间相错的关键步骤以一定的概率转为中间相遇(图 4), 同时保证差分$\Gamma_0 \to \Gamma_2$成立的概率足够大(可以与随机置换区分), 这样就完成了不可能差分到截断差分的转换。得益于截断不可能差分的自动化搜索方法, 这一关系的发现将可能有助于促进截断差分搜索的自动化。作者也将该方法应用于 CLEFIA[34]和 Camellia[35]的攻击中, 并得到了较好的结果。

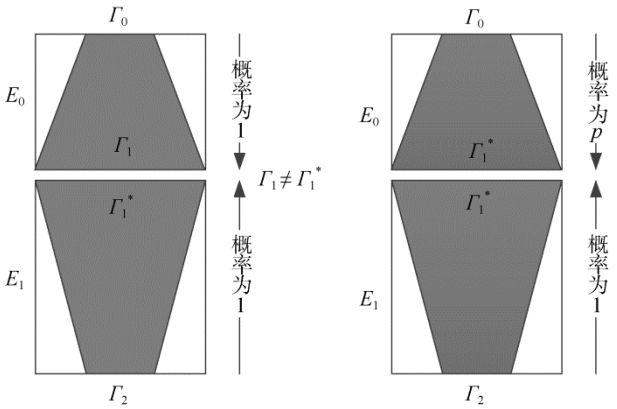

图 5 不可能差分转为截断差分

对于随机置换而言, 截断差分$\Gamma_0 \to \Gamma_2$成立的概率为$\mathbf{Pr}\{\Gamma_0 \to \Gamma_2\} = \frac{|\Gamma_2|}{2^{n-1}}$, 其中$n$为分组长度。当同样的截断差分对于算法的概率$\mathbf{Pr}\{\Gamma_0 \to \Gamma_2\} = \frac{|\Gamma_2|}{2^{n-1}} + \varepsilon(\varepsilon > 0)$时, 则这一截断差分可以当做区分器来恢复密钥。文献[28]假设E为 Markov 算法[29], 即差分特征的概率用每一轮的概率相乘得到。

想要计算$\mathbf{Pr}\{\Gamma_0 \to \Gamma_2\}$, 由

$$\mathbf{Pr}\{\Gamma_0 \to \Gamma_2\} = \mathbf{Pr}\{\Gamma_0 \to \Gamma_1\} \times \mathbf{Pr}\{\Gamma_1 \to \Gamma_2\}, \quad (7)$$

已知$\mathbf{Pr}\{\Gamma_2 \to \Gamma_1\} = 1$, 为计算$\mathbf{Pr}\{\Gamma_1 \to \Gamma_2\}$, 作者引入了下面的假设。

假设 1([28, 假设 1]).

对于截断差分$\Gamma_{in} \to \Gamma_{out}$, 若$E$为 Markov 算法, $a \in \Gamma_{in}, b \in \Gamma_{out}$时, 我们有

$$\mathbf{Pr}\{a \to b\} = \mathbf{Pr}\{b \to a\}.$$

利用假设 1, 我们可以计算$\mathbf{Pr}\{\Gamma_1 \to \Gamma_2\}$。

命题 9([28, 命题 1]).

假设截断差分$\Gamma_2 \to \Gamma_1$成立的概率为 1, 则在假设 1 之下, 截断差分$\Gamma_1 \to \Gamma_2$成立的概率为

$$\mathbf{Pr}\{\Gamma_1 \to \Gamma_2\} = \frac{|\Gamma_2|}{|\Gamma_1|},$$

其中$|\Gamma_2| \leqslant |\Gamma_1|$。

根据命题 9, 当我们将不可能差分转为截断差分时, 截断差分的概率可以用下面命题计算。

命题 10([28, 命题 3]).

对于分组密码$E = E_1 \circ E_0$, 假设有两段高概率的截断差分, 即$\mathbf{Pr}\{\Gamma_0 \to \Gamma_1\} = p, \mathbf{Pr}\{\Gamma_2 \to \Gamma_1\} = q$。则

截断差分$\Gamma_0 \to \Gamma_2$的概率大于$pq \times \frac{|\Gamma_2|}{|\Gamma_1|}$。

利用中间相遇的方法找到的截断差分也可以用上面的命题计算概率，但由于不可能差分的寻找可以得益于自动化搜索的方法，所以这一命题将可能有助于截断差分搜索的自动化。这样的特性在许多分组密码中存在，特别是 Fesitel 结构的密码，如 CLEFIA、Camellia 等。

7 小结

本文中，我们简要介绍了一些已有的统计分析方法，并总结了已有的统计分析方法之间的关联性。这些关联性的建立一方面有助于我们对已知的分组密码分析方法进行分类，另一方面这些关联性可能会给出分组密码安全性的补充信息。除此之外，研究统计分析方法之间的关系，也有助于我们用较为完善的理论体系丰富和发展密码体系中较为薄弱的环节。

纵观所有已经提出的关联性，研究这些关联性的价值有以下几方面:

1. 通过研究两种统计分析方法的关联性，在已知其中一类区分器D_1的情况下，我们可以利用该关联性直接导出与之相等价的另一类区分器D_2。但直接推导区分器D_2可能较为困难，利用等价性可以便于我们发现更长轮数的区分器。

2. 两种具有等价关系的统计分析方法在攻击方式上可能存在不同(一种是已知明文攻击，另一种是选择明文攻击)，我们可以从等价的区分器中选择更适合攻击的区分器进行算法攻击。

3. 研究统计分析方法之间的关系，也有助于我们用较为完善的理论体系丰富和发展密码体系中较为薄弱的环节。例如: 通过多维线性和统计饱和度攻击在本质上的统一性，我们可以确切的衡量统计饱和度攻击中使用的数据量。

当然，现存的各种统计分析方法之间的关联性也存在一些不足:

1. 一些现有的关联性完全由数学公式推导而得，并不依赖于底层算法的结构，这种等价性虽具有一般性，但缺少实用性。

2. 现有的关联性覆盖的结构还不够完整，而且等价关系存在的条件也比较苛刻，这些缺陷也限制了这些等价性的应用。

有些关联性(如: 不可能差分与积分攻击之间的关联性)是从一些现有的关联性推导而来，而所涉及的统计分析方法之间是否存在其他的、二者独有的关联值得进一步思考。

参考文献

[1] E. Biham, and A Shamir, "Differential Cryptanalysis of DES-like Cryptosystems," *Journal of Cryptology*, vol. 4, no. 1, pp. 3-72, 1991.

[2] L. R. Knudsen, "Truncated and High Order Differentials," in *Fast Software Encryption*, vol. 1008,pp.196-211, 1995.

[3] L. R. Knudsen, "DEAL - A 128-bit Block Cipher". *Technical report, DEPARTMENT OF INFORMATION, University of Bergen*, 1998.

[4] E. Biham, A. Biryukov, and A. Shamir, "Cryptanalysis of Skipjack Reduced to 31 Rounds Using Impossible Differential," in *Advances in Cryptology – EUROCRYPT 1999 (EUROCRYPT'99)*, pp. 12-23, 1999.

[5] D. Wagner, "The Boomerang Attack" in *Fast Software Encryption - FSE 1999*, pp. 156-170, 1999.

[6] E. Biham, O. Dunkelman, and N. Keller, "The Rectangle Attack - Rectangling the Serpent," in *Advances in Cryptology – EUROCRYPT 2001 (EUROCRYPT'01)*, pp. 340-357, 2001.

[7] J. Kim, S. Hong, J. Sung, S. Lee, and J. Jim, "Impossible Differential Cryptanalysis for Block Cipher Structures," in *Indocrypt 2003*, pp. 82-96, 2003.

[8] Y. Luo, X. Lai, Z. Wu, and G. Gong, "A Unified Method for Finding Impossible Differentials of Block Cipher Structures," in *Information Sciences*, vol. 263, no. 1, pp. 211-220, 2014.

[9] M. Matsui, "Linear Cryptanalysis Method for DES Cipher," in *Advances in Cryptology – EUROCRYPT 1993 (EUROCRYPT'93)*, pp. 386-397, 1994.

[10] M. Matsui, "The First Experimental Cryptanalysis of the Data Encryption Standard," in *Advances in Cryptology - CRYPTO 1994 (CRYPTO'94)*, pp. 1-11, 1994.

[11] J. Burton, S. Kaliski, and M. J. B. Robshaw, "Linear Cryptanalysis Using Multiple Approximations," in *Advances in Cryptology - CRYPTO 1994 (CRYPTO'94)*, pp. 26-39, 1994.

[12] M. Hermelin, J. Y. Cho, and K. Nyberg, "Multidimensional Extension of Matsui's Algorithm 2," in *FSE 2009*, vol. 5665, pp. 209-227, 2009.

[13] A. Bogdanov, and V. Rijmen, "Linear Hulls with Correlation Zero and Linear Cryptanalysis of Block Ciphers," in *Designs, Codes and Cryptography*, vol. 70, no. 3, pp. 369-383, 2014.

[14] A. Bogdanov, and M. Wang, "Zero Correlation Linear Cryptanalysis with Reduced Data Complexity," in *Fast Software Encryption - FSE 2012*, pp. 29-48, 2012.

[15] A. Bogdanov, G. Lender, K. Nyberg, and M. Wang, "Integral and Multidimensional Linear Distinguishers with Correlation Zero," in *Advances in Cryptology - ASIACRYPT 2012*, pp. 244-261, 2012.

[16] S. Murphy, "The Independence of Linear Approximation in Symmetric Cryptology," *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5510-5518, 2006.

[17] J. Daemen, L. Knudsen, and V. Rijmen, "The Block Cipher SQUARE," in *Fast Software Encryption - FSE 1997*, pp. 149-165, 1997.

[18] L. R. Knudsen, and D. Wagner, "Integral Cryptanalysis," in *Fast Software Encryption - FSE 2002*, pp. 112-127, 2002.

[19] B. Collard, and F. X. Standaert, "A Statistical Saturation Attack against the Block Cipher PRESENT," in *CT-RSA 2009*, vol. 5473, pp. 195-210, 2009.

[20] F. Chabaud, and S. Vandenay, "Links between Differential and Linear Cryptanalysis," in *EUROCRYPT 1994 (EUROCRYPT'94)*, pp. 356-365, 1994.

[21] C. Blondeau, K. Nyberg, "New Links between Differential and Linear Cryptanalysis," in *EUROCRYPT 2013 (EUROCRYPT'13)*, pp. 388-404, 2013.

[22] C. Blondeau, K. Nyberg, "Links between Truncated Differential and Multidimensional Linear Properties of Block Ciphers and Underlying Attack Complexities," in *EUROCRYPT 2014 (EUROCRYPT'14)*, pp. 165-182, 2014.

[23] A. A. Selçuk, "On Probability of Success in Linear and Differential Cryptanalysis," *Journal of Cryptology*, vol. 21, no. 1, pp. 131-147, 2008.

[24] C. Blondeau, A. Bogdanov, M. Wang, "On the (In) Equivalence of Impossible Differential and Zero Correlation Distinguishers for Feistel- and Skipjack-type Ciphers," in *ACNS 2014*, pp. 271-288, 2014.

[25] B. Sun, Z. Liu, V. Rijmen, R. Li, L. Cheng, Q. Wang, …, and C. Li, "Links among Impossible Differential, Integral and Zero Correlation Linear Cryptanalysis," in *CRYPTO 2015*.

[26] A. Bogdanov, L. R. Knudsen, G. Leander, C. Paar, A. Poschmann, M. J. Robshaw, …, C. Vikkielsoe, "PRESENT: An Ultra-lightweight Block Cipher," *Springer Berlin Heidelberg Press*, 2007.

[27] G. Lender, "On Linear Hulls, Statistical Saturation Attacks, Present and a Cryptanalysis of PUFFIN," EUROCRYPT 2011 (EUROCRYPT'11), pp. 303-322, 2011.

[28] L. Li, K. Jia, X. Wang, and X. Dong, "Meet-in-the-Middle Techniques for Truncated Differential and Its Applications to CLEFIA and Camellia," in *Fast Software Encryption*, pp. 48-70, 2015.

[29] X. Lai, and J. L. Massey, "Markov Ciphers and Differential Cryptanalysis," in *EUROCRYPT 1991 (EUROCRYPT'91)*, pp. 17-38, 1991.

[30] K. Nyberg, "Linear Approximations of Block Ciphers," in *Advances in Cryptology - EUROCRYPT 1994 (EUROCRYPT'94)*, pp.439-444, 1994.

[31] J. Y. Cho, "Linear Cryptanalysis of Reduced-round PRESENT," in *Topics in Cryptology - CT - RSA 2010*, pp. 302-317, 2010.

[32] M. Z'aba, H. Raddum, M. Henricksen, and E. Dawson, "Bit-pattern Based Integral Attack," in *Fast Software Encryption - FSE 2008*, pp. 363-381, 2008.

[33] L. Wen, and M. Wang, "Integral Zero-Correlation Distinguisher for ARX Block Cipher, with Application to SHACAL-2," in *Information Security and Privacy*, pp. 454-461, 2014.

[34] T. Shirai, K. Shibutani, T. Akishita, S. Moriai, and T. Iwata, "The 128-bit Block Cipher CLEFIA," in Fast Software Encryption – FSE2007, pp. 181-195, 2007.

[35] A. Kato, S. Moriai, M. Kanda, "The Camellia Cipher Algorithm and its use with IPSec," https://tools.ietf.org/html/rfc4312, 2005.

王美琴 于2007年在山东大学信息安全专业获得博士学位。现任山东大学密码技术与信息安全教育部重点实验室副主任。研究领域为对称密码分析与设计。研究兴趣包括：零相关线性分析、积分攻击等。
Email: mqwang@sdu.edu.cn

孙玲 于 2014 年在山东大学大学数学与应用数学专业获得理学学士学位。现在山东大学信息安全专业攻读硕士学位。研究领域为分组密码的分析与设计。研究兴趣包括: 差分分析与线性分析等。
Email: lingsun@mail.sdu.edu.cn

陈怀风 于2012年在山东大学信息安全专业获得理学学士学位。研究领域为分组密码的分析与设计。研究兴趣包括: 差分分析和线性分析等。
Email: hfchen@mail.sdu.edu.cn

刘瑜 于 2008 年在山东大学信息安全专业获得硕士学位。现在山东大学信息安全专业攻读博士学位。研究领域为分组密码。研究兴趣包括: 差分分析和线性分析等。
Email: liuyu01@mail.sdu.edu.cn