引用格式:闫守礼,郭丽敏,王立辉,等. 针对有掩码防护DES的代数侧信道攻击[J].微电子学与计算机,2019,36(10):10-14.[YAN S L, GUO L M, WANG L H,et al. Algebraic side channel attack against DES with mask countermeasure[J]. Microeletronics &Computer,2019,36(10):10-14.]

针对有掩码防护DES的代数侧信道攻击

闫守礼[1],郭丽敏[1],王立辉[1],李清[1,2],俞军[1,2]
(1 上海复旦微电子集团股份有限公司,上海 200433; 2 复旦大学 微电子学院,上海 201203)

摘 要: 基于汉明重量泄漏模型,对带掩码防护的软件DES抗代数侧信道攻击能力进行了评估.首先研究了代数侧信道攻击的攻击原理,然后基于模板攻击得到了DES中间无防护轮次S盒输出的汉明重信息,将其作为可配置参数,利用脚本语言及BAT工具自动生成DES的合取范式表示,最后利用求解器进行密钥求解.结果表明:对仅掩码防护首两轮及尾两轮的软件DES,利用中间连续3轮S盒输出汉明重泄漏即可恢复56比特DES根密钥.

关键词: DES;模板攻击;代数侧信道攻击
中图分类号: TP393.08 **文献标识码**: A **文章编号**: 1000-7180(2019)10-0010-05
DOI:10.19304/j.cnki.issn1000-7180.2019.10.003

Algebraic side channel attack against DES with mask countermeasure

YAN Shou-li[1], GUO Li-min[1], WANG Li-hui[1], LI Qing[1,2], YU Jun[1,2]
(1 Shanghai Fudan Microelectronics Group Company Limited, Shanghai 200433, China;
2 Institute of Microelectronics,Fudan University, Shanghai 201203, China)

Abstract: On the basis of the Hamming Weight leakage model, the anti algebraic side channel attack capability of a masked software DES is evaluated. The algebraic side channel attack principle is studied, and then the Hamming weight information of Sbox output of DES without protection is got based on the template attack, it is as the configurable parameters to generate conjunctive normal form of DES using script language and BAT tools, the key is retrieved by solver finally. The results show that the 56 bit DES root key can be recovered by using the Hamming weight of Sbox output from the middle 3 successive rounds, while the mask is only used to protect the first two rounds and the last two rounds of DES.

Key words: DES; Template Attack; Algebra Side Channel Attack

1 引言

电路加解密过程中会泄漏时间、功耗、电磁、声音等旁路信息,由于电路的旁路信息与加解密数据及加解密操作都存在相关性,因此通过旁路信息可以分析加解密电路的特征,进一步利用数学手段对加解密电路进行攻击得到密钥.模板攻击的概念在1996年被首次提出,这是一种结合概率统计的非常有效的侧信道攻击方法.

对密码算法的代数分析主要是从数学角度研究密码算法的代数结构,建立关于明密文、中间状态与密钥的代数方程组,通过求解方程组破解密钥.对于现代的对称密码算法,如AES、DES等,在设计中已经充分考虑了其数学特性,因此一般的代数分析只能将破解密码算法的难度稍微降低,但总体上仍然超出当前计算机的计算能力.

为提高攻击效率,将传统密码分析与侧信道攻击结合已成为近年来密码分析的一个重要研究方

收稿日期: 2018-12-09; 修回日期: 2019-01-18
基金项目: 十三五装备预先研究项目(3110105-09)

向. Mathieu Renauld 等人在 2009 年提出的代数侧信道攻击(Algebraic side-Channel Attacks)[1]，将侧信道攻击与代数分析相结合，可以充分利用旁路信息，降低代数计算的复杂度，大大减少破解密钥需要的功耗曲线条数.

现有的代数侧信道攻击研究正处于起步阶段，研究对象主要针对轻量级算法 RESESNT[2]、AES[3]及 SMS4[4]. 结果表明：1 条功耗曲线即可得到 PRESENT 和 AES 的完整密钥. DES(Data Encryption Standard)是使用最广泛的对称密码，安全性广受关注，但目前在对 DES 进行代数侧信道攻击方面，尚未发现国内外有公开发表的结果.

本文针对 DES 算法，基于汉明重模型，将模板攻击与代数攻击结合，先利用模板攻击得到带掩码防护软件 DES 运行时中间轮次泄漏的 S 盒输出汉明权重，然后将其作为可配置参数，利用脚本及 BAT 工具自动生成 DES 的 CNF(合取范式)表示，最后利用 MiniSAT2 软件求解得到其根密钥. 文献[1][2][3][4]中提出先建立相应密码算法的代数方程组，再将其转化为 SAT 问题求解；本文则将中间状态信息作为可配置的参数，利用脚本语言及 BAT 工具直接生成 DES 算法的 CNF 表示，提高了攻击效率.

2　代数侧信道攻击

代数侧信道攻击在代数攻击的基础上，利用侧信道攻击方法得到密码加解密过程中泄漏的中间状态信息，将其转化为关于中间状态变量的代数方程，联合密码算法的代数方程组求解得到密钥. 可以说侧信道攻击是利用多条曲线共同作用，以确定密钥信息；而代数侧信道攻击是利用单条曲线，结合一次加解密过程中不同时期的信息进行计算得到密钥信息.

代数侧信道攻击具有下述特性：(1)可以利用全轮旁路信息；(2)在理想情况下，仅利用一条功耗曲线即可得到完整密钥；(3)在未知明密文的情境下，也可成功.

传统的代数侧信道攻击主要包括侧信道攻击、建立密码算法方程组、求解代数方程组 3 个阶段.

DES(Data Encryption Standard)采用了 64 位的分组长度和 56 位的密钥长度，其算法流程如图 1 所示：

其中 IP 是初始置换；IP^{-1}是与初始置换互逆的置换；子密钥 $k_1, k_2, \cdots, k_{16}$ 长度均为 48 位.

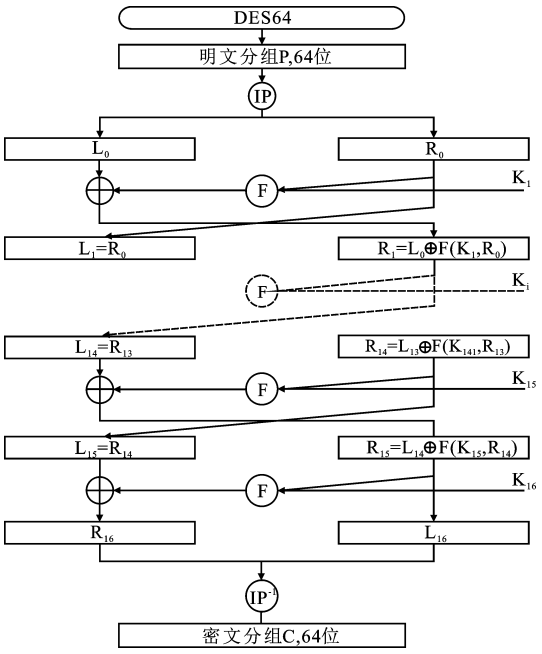

图 1　DES 加密流程

轮函数 F 如图 2 所示，其中 E 为 32 bit 到 48 比特的扩展操作，Sbox1，…，Sbox8 表示查表操作，P 表示 32 比特的置换操作.

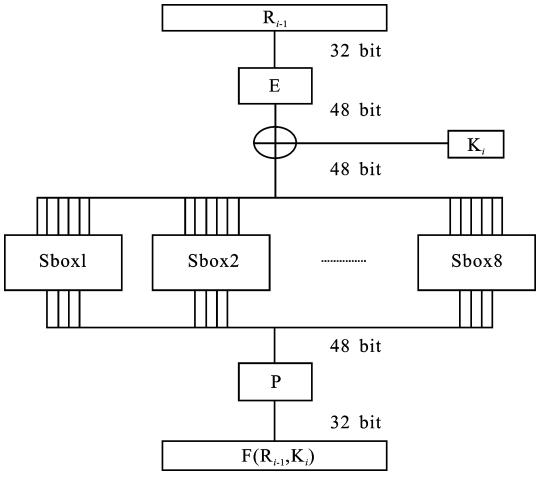

图 2　F 函数结构

3　建立 DES 的 CNF 表达式

直接求解对 DES 建立的庞大的代数方程组是很困难的，本文将其转化为 SAT 问题求解. SAT 问题全称为布尔可满足性问题，在国际上有着十分广泛的研究. SATLive[5]网站详细介绍了 SAT 研究的历史和现状，并一直跟踪最前沿的研究方向.

目前主流的 SAT 求解器都是采用 CNF(合取范式)作为输入的. 若想利用 SAT 求解器求解代数方程组，必须将原来的布尔方程转换为 CNF 形式.

这个转换过程是十分复杂的，有很多文献[6-11]专门讨论了这类问题，也有专门的转化程序. 但仅仅完成将代数方程组转化为 CNF 表示，ASCA 攻击仍然很难实施. 因为根据算法获得代数方程组的步骤，对于一个复杂的算法，例如 DES、AES，仍然是十分复杂的.

本文使用脚本语言及 BAT(Bit Level Analysis Tool)工具[11]实现自动建立 DES 的 CNF 表示功能. BAT 工具的目的是利用程序化的语言对电路进行建模并直接转化成 CNF 形式的逻辑表达式，可直接送入求解器进行求解. BAT 工具使用 Lisp-like[12]脚本进行建模，可以支持三种模型，分别是 machine，exist，forall. machine 模型中采用了初值-next 状态-终值的形式，与时序电路工作方式十分类似；exist 模型可以通过计算得到满足模型方程的第一个解；forall 模型遍历指定输入变量的所有输入情况，检查是否满足. 本文的求解模型中，把密钥作为未知变量，从而使用 exist 模型进行建模.

4 对 DES 的模板攻击

4.1 模板及攻击算法

本文模板攻击的对象为软件实现的 DES 加密算法，其第 1 轮、第 2 轮、第 15 轮及第 16 轮的轮运算使用掩码的 S 盒，在进行 S 盒查表前先进行造表运算. 第 3 轮直到第 14 轮使用标准 S 盒进行直接查表运算. 从而对第 3 轮直到第 14 轮，其 S 盒的输入和输出均为真实数据，存在汉明重的泄漏. 下文为第 3 轮至第 5 轮的 S 盒输出的汉明权重 H 建立模板. H 定义为一个二进制数中 1 的个数，因此，对于 DES 的每个 S 盒来说，其输出汉明权重的取值范围为[0,4]. 需要为第 3 轮至第 5 轮的 8 个 S 盒分别建立模板，那么共需为 24 个 S 盒建立模板.

令$\{W(t,i)\mid 1\leqslant i\leqslant n\}\{W(t,i)\mid 1\leqslant i\leqslant n\}$表示 n 条侧信道信息曲线(下面简称曲线)，n 表示曲线的总条数，假设 N 为每条曲线的点的数目.

$\{P(i)\mid 1\leqslant i\leqslant n\}\{P(i)\mid 1\leqslant i\leqslant n\}$表示 n 个随机明文. 假定 K 为 DES 的 56 比特根密钥.

DES 的模板攻击可分为两个步骤.

(1)模板的建立. 利用随机产生的 100 000 组明文和密钥进行加密并采集 100 000 条功耗曲线，利用这些曲线为第 3，4，5 轮的 S 盒输出建立模板.

(2)记 P 为一个随机明文，利用 P 及待破解的 K 进行 1 000 次加密，采集到 1 000 条功耗曲线，利用这些曲线进行模板匹配(模板匹配的过程请见第 3.3 节). 可以得到第 3 轮至第 5 轮连续 3 轮每轮 8 个 S 盒输出的汉明权重.

4.2 模板建立

模板攻击中考虑的是多维高斯噪声模型. 模板的建立共有 3 个步骤. 步骤 1，数据的预处理. 这一步主要目的是让所有的曲线对齐，并且尽量降低数据的存储用量. 步骤 2，选取有效的信息点，从而提高模板攻击效率. 本文采用计算数据相关性的方法. 步骤 3，计算数据的平均值及相关性噪声矩阵.

对于第 3 轮至第 5 轮的每个 S 盒，按照其输出的汉明权重将所有的曲线分成 5 个集合$\{W_h\mid 0\leqslant h\leqslant 4\}$. 令$\{n_h\mid 0\leqslant h\leqslant 4\}$表示每个集合中曲线的条数. 每个集合的曲线平均值计算如下：

$$\overline{W_h}=\frac{1}{n_h}\sum\nolimits_{W\in W_h}W \tag{1}$$

假设 n_h 条曲线数据构成的矩阵为 $n_h\times N$，令$\{X_i\mid 0\leqslant i\leqslant N\}$为这个矩阵的 N 个列向量. 选择其中的两个元素 X,Y，二者的协方差计算如下

$$COV(X,Y)=\frac{1}{N-1}\sum_{j=1}^{n_h}(X_j-\overline{X})(Y_j-\overline{Y}) \tag{2}$$

那么 SBOX 的协方差矩阵计算如下

$$CM_h(i,j)=COV(X_i,X_j) \tag{3}$$

对于该 S 盒输出可以得到一个$(\overline{W_h},CM_h)$的组合，这便是该 S 盒输出汉明权重为 h 的模板.

数据的预处理对模板攻击有着非常大的作用. 首先对采集到的功耗曲线做重采样，尽量降低数据的存储用量. 然后，再对曲线做静态对齐，保证所有曲线是严格对齐的. 从而，经过预处理之后可以保证有效信息点处在同一位置.

以为第 3 轮第 1 个 S 盒输出选取感兴趣的点为例说明. 根据随机明文及密钥，可以计算得到第 i 条曲线第 3 轮第 1 个 S 盒的输出，其汉明权重记为 s_i. 计算$s=\{s_1,\cdots,s_n\}^{\mathrm{T}}$与经过预处理的功耗曲线中每一点$o^j=\{o_1^j,\cdots,o_n^j\}^{\mathrm{T}}$的相关性$\rho(s,o^j)$. 再从$\{\rho(s,o^1),\cdots,\rho(s,o^N)\}$中选取值最大的若干个点作为有效信息点.

4.3 模板匹配

模板的匹配实际是计算多变量高斯分布的概率. 待匹配的曲线 W 服从模板$(\overline{W_h},CM_h)$的概率为

$$\begin{aligned}&p(W,\overline{W_i},CM_i)\\&=\frac{1}{(2\pi)^{l/2}\mid CM_i\mid^{0.5}}e^{-0.5(W-\overline{W_i})^{\mathrm{T}}CM_i^{-1}(W-\overline{W_i})}\\&=e^{-0.5(W-\overline{W_i})^{\mathrm{T}}CM_i^{-1}(W-\overline{W_i})-l/2\ln(2\pi)-0.5\ln(\mid CM_i\mid)}\\&=e^{p1}\end{aligned} \tag{4}$$

实际上只需要计算 $p1$ 即可. 根据式(4) 可以计算出来每条曲线符合某个模板的概率，在理想情况下，一条曲线便可以匹配成功. 但是由于一条曲线的匹配存在一定判断失误的概率，需要采集多条模板匹配曲线，接着有三种处理方法：第一种方法为采集多条模板匹配曲线，对其做平均，再对得到的平均功耗曲线进行匹配；第二种方法为增强的模板攻击；第三种方法为利用贝叶斯概率公式对多条模板匹配曲线进行模板匹配. 下面说明如何采用贝叶斯概率公式计算多条曲线进行模板匹配的概率.

令 $p_{i,j}$ 表示曲线 W_i 服从汉明权重为 j 的模板的概率，p_j 为 N_W 条模板匹配曲线服从汉明权重为 j 的模板的概率，p_j 的计算公式为

$$p_j = \sum_{i=0}^{N_W} p_{i,j}\, p_j \tag{5}$$

根据上式可以得到 p_j 的值，那么 $\max\{p_j\}$ 对应于正确的汉明权重的值.

令 $p(H_i \mid W_i)$ 表示曲线 W_j 服从第 i 个模板的条件概率，记先验概率为 $p(H_i)$ 和 $p(W_j \mid H_i)$. 对多条曲线的模板匹配来说，令 $p(H_i)$ 为 $p(H_i \mid W_{j-1})$，其中 $p(H_i)$ 的初始值为 $C_4^i/16$，N_W 记为模板匹配的曲线条数. $p(H_i \mid N_W)$ 的计算公式可以表示为

$$p(H_i \mid N_W) = \frac{\left(\prod_{j=1}^{N_W} p(W_j \mid H_i)\right) p(H_i)}{\sum_{i=1}^{5}\left(\prod_{j=1}^{N_W} p(W_j \mid H_i)\right) p(H_i)} \tag{6}$$

根据上式可以得到 $p(H_i \mid H_W)$ 的值，那么 $\max\{p(H_i \mid N_W)\}$ 对应于正确的汉明权重的值.

5　攻击实验

先利用模板攻击对第 3 轮至第 5 轮每轮 8 个 S 盒输出的汉明重进行攻击. 下面以对第 3 轮进行攻击为例.

图 3 是对第 3 轮的 8 个 S 盒输出汉明重进行模板匹配的成功率图，其中红色曲线对应于使用贝叶斯公式进行模板匹配的成功率，绿色曲线对应于使用平均功耗曲线进行模板匹配的成功率，蓝色曲线对应于使用增强的模板攻击进行模板匹配的成功率.

利用模板攻击可以依次得到第 3 轮、第 4 轮及第 5 轮连续 3 轮每轮 8 个 S 盒输出的汉明权重.

接下来利用 perl 脚本实现 DES 模型的参数配置及自动模型生成. 可以配置的参数主要包括：

(1)加密曲线的条数，即可以多条曲线(采用相同密钥)并行计算.

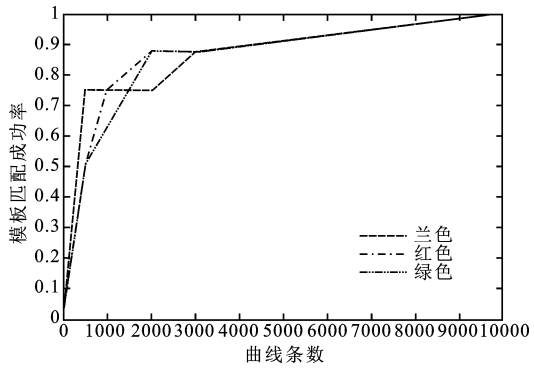

图 3　模板匹配成功率曲线

(2)求解约束条件，即已知的中间变量信息，例如 S 盒输出的汉明重量.

由于算法模型比较复杂，涉及到参数配置，因此将算法模型分割为不同的部分，采用多个脚本各自生成最后组合而成. 脚本的组织结构如图 4 所示：

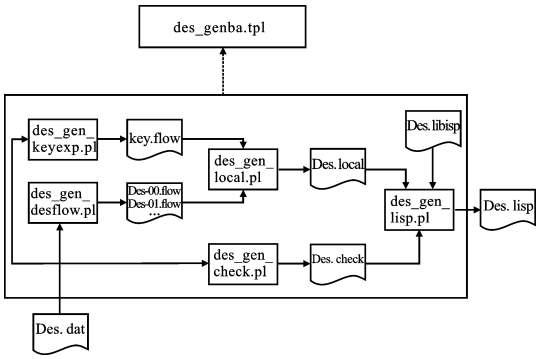

图 4　脚本组织结构图

脚本说明如下：

1. Des. dat 为数据配置文件，包括了每条曲线对应的明密文，已知的中间变量信息，密钥信息.

2. Des_gen_keyflow. pl 为生成密钥扩展部分模型的脚本，从 des. dat 文件中读取密钥信息，生成结果文件存于 Key. flow 中.

3. Des_gen_desflow. pl 为生成轮运算部分模型的脚本，从 des. dat 文件中读取明密文信息生成结果存于 Des-00. flow，Des-01. flow，其中编号 00，01 为曲线编号，根据设定的曲线条数自动设置.

4. Des_gen_check. pl 为生成算法模型校验部分代码的脚本，从 des. dat 文件中读取已知的算法中间状态信息，生成校验模型. 生成的结果存于 Des. check 文件.

5. Des_gen_local. pl 整合密钥扩展和轮运算的模型，生成 lisp 模型中的 local 代码部分，结果存于 Des. local 文件.

6. Des_gen_lisp. pl 整合 local 部分代码和 check 部分代码，同时加入库函数(DES 的常规操作，如 P 置换，S 盒查表运算等，由 deslib. lisp 文件提供，一般不用修改)生成最终的 Des 算法模型存于 Des. lisp.

最终生成的 Des. lisp 文件可以直接传递给 BAT 软件进行运算.

整个运算过程用户只需配置 des. dat 文件即可,实际生成过程,即上述脚本的调用过程,已经整合在 des_genbat. pl 脚本中. 用户配置好 des. dat 文件后运行 des_genbat. pl 脚本可以生成最终结果 des. lisp.

再利用 BAT 工具及 MiniSAT2 求解器[13],对仅掩码防护首两轮及尾两轮的软件 DES,利用普通计算机在 188 秒内成功恢复了 DES 的 56 比特根密钥.

6 结束语

本文基于汉明重量模型,首先对带掩码防护的 DES 进行模板攻击,成功得到其中间无防护轮次 S 盒输出的汉明权重. 进一步结合代数攻击,将由模板攻击得到的 3 轮连续 S 盒输出汉明重信息作为可配置参数,利用脚本语言及 BAT 工具实现自动生成 DES 算法的合取范式表示,提高了攻击效率. 最后利用求解器进行求解,对仅掩码防护首两轮及尾两轮的软件 DES,成功恢复了其 56 比特的根密钥 K.

从实验结果可以看出,为了抵抗代数侧信道攻击,DES 的中间轮次也需要进行掩码防护.

参考文献:

[1] M RENAULD, F.-X STANDAERT. Algebraic Side-Channel Attacks [C]. In Proceedings of the INSCRYPT 2009, LNCS 6151. Beijing, China, 2009: 393-410.

[2] 吴克辉,王韬,赵新杰,等. 基于汉明重的 PRESENT 密码代数旁路攻击[J]. 计算机科学,2011, 38(12): 53-56.

[3] M RENAULD, F.-X STANDAERT, N VEYRAT-CHARVILLON. Algebraic Side-Channel Attacks on the AES: Why Time also Matters in DPA[C]. In proceedings of CHES2009, Lecture Notes in Computer Science. Lausanne, 2009: 97-111.

[4] 刘会英,赵新杰,王韬,等. 基于汉明重的 SMS4 密码代数旁路攻击研究[J]. 计算机学报,2013, 36(6): 1183-1193.

[5] DANIEI LE BERRE. SAT Live. http://www.satlive.org/

[6] P MANOLIOS, D VROON. Efficient circuit to CNF conversion[C]. Theory and Application of Satisfiability Testing, SAT 2007:4-9.

[7] MN VELEV. Efficient translation of boolean formulas to CNF in formal verification of microprocessors[C]. In Proceedings of the 2004 Asia and South Pacific Design Automation Conference: 310-315.

[8] B CHAMBERS, P MANOLIOS, D VROON. Faster SAT solving with better CNF generation[C]. In Proceedings of the Conference on Design, Automation and Test in Europe 2009: 1590-1595.

[9] GV BARD, NT COURTOIS, C JEFFERSON. Efficient methods for conversion and solution of sparse systems of low-degree multivariate polynomials over GF(2) via SAT-solvers. http://eprint.iacr.org.

[10] D TILLE, R KRENZ-BAATH, J SCHIOEFFEL, et al. Improved Circuit-to-CNF Transformation for SAT-based ATPG. http://researchgate.net.

[11] P MANOLIOS, S K. SRINIVASAN, et al. BAT: The Bit-Level Analysis Tool. http://www.ccs.neu.edu/home/pete/research/cav-bat.html.

[12] JOHN MCCARTHY. Lisp (programming language). http://en.wikipedia.org/wiki/Lisp_(programming_language).

[13] NIKLAS EEN, NIKLAS SORENSSON. The Minisat Page. http://minisat.se/

作者简介:

闫守礼 男,(1974-),硕士,工程师. 研究方向为密码芯片安全.

郭丽敏 女,(1986-),硕士,工程师. 研究方向为密码芯片安全.

王立辉(通讯作者) 男,(1982-),博士,工程师. 研究方向为密码芯片安全. E-mail: wanglihui@fmzh.com.cn.

李 清 女,(1968-),硕士,高级工程师. 研究方向为集成电路设计开发.

俞 军 男,(1968-),硕士,高级工程师. 研究方向为集成电路设计开发.