文章编号:2096-1618(2021)05-0499-04

针对SM4密码算法的模板攻击

匡晓云[1], 黄开天[1], 兰 天[2], 杜之波[3], 吴 震[3]
(1. 南方电网科学研究院,广东 广州 510663;2. 华大半导体有限公司,上海 200120;3. 成都信息工程大学网络空间安全学院,四川 成都 610225)

摘要:目前针对SM4密码算法的侧信道攻击研究主要是故障分析攻击和相关性能量分析攻击,为了探索模板攻击针对SM4密码算法分析攻击应用问题,提出了针对SM4密码算法的模板攻击方法,模板构建时,选择SM4密码算法的S盒输出作为能量分析攻击点,汉明重量作为能量分析模型,构建关于S盒输出的模板,模板匹配时,选择多条曲线的联合概率进行匹配,概率最大值对应的密钥即为攻击出的正确密钥。针对SM4密码算法智能卡实测实验,验证了该攻击方法的有效性。该攻击方法不仅可以应用到针对SM4密码算法其他攻击点的模板攻击,而且对其他分组密码算法的模板攻击具有借鉴意义。

关 键 词:侧信道攻击;模板攻击;SM4密码算法;S盒
中图分类号:TP309.1 **文献标志码**:A
doi:10.16836/j.cnki.jcuit.2021.05.004

0 引言

密码技术是网络空间安全的基础和支撑,从底层的物理设备,到上层的系统应用,密码技术在网络空间安全中发挥着保密性、完整性、真实性和不可否认性等安全作用。密码技术中的密码算法,无论实现方式是软实现还是硬实现,都要依附于具体的物理载体,密码物理载体在运行时存在能量、电磁等侧信息的泄露,而这些信息和密钥存在一定的相关性,通过对侧信息的分析可还原密码算法运行时所使用的密钥,所以密码算法本身的安全,并不意味着密码算法实现的安全[1]。侧信道分析攻击[2]就是通过对密码设备的侧信息的采集和分析实现密钥的还原,侧信道分析攻击已经对各类密码算法构成严重威胁[3]。能量分析攻击是侧信道分析攻击常用的方法之一,能量分析攻击包括差分能量分析攻击、相关性能量分析攻击和模板攻击等攻击方法,这些攻击方法也是目前评估密码产品安全的重要测评方法。

SM4密码算法是中国发布的商用分组密码算法,也是目前中国密码行业标准以及国家信息安全技术标准密码算法之一。针对SM4密码算法的侧信道攻击研究,主要是针对SM4密码算法的侧信道故障分析攻击,以及针对SM4密码算法的侧信道相关性能量分析攻击等研究内容。文献[4-5]针对SM4密码算法进行了随机故障注入攻击和改进差分故障攻击研究。文献[6-8]对SM4密码算法相关性能量分析攻击和攻击点展开了研究。文献[9]提出了基于深度学习的S盒逆向分析算法,探索了深度学习在侧信道攻击中应用。而关于SM4密码算法的模板攻击方向研究,国内外公开发表的研究成果较少。模板攻击也是侧信道攻击中常用的攻击方法,用判别的方法来获取能量曲线中隐藏的密钥信息,所以研究针对SM4密码算法的模板攻击,对模板攻击和针对分组密码算法的模板攻击都具有十分重要的意义。

本文根据SM4密码算法结构特点和模板攻击原理,分析了针对SM4密码算法模板攻击的可行性,根据对SM4密码算法信息泄露点的分析,提出和设计了针对SM4密码算法的模板攻击方法,探索了模板攻击方法在SM4密码算法上应用的可行性,通过针对SM4密码算法智能卡的实测攻击实验,验证了该攻击方法的有效性,该攻击方法和实测攻击过程对其他密码算法的模板攻击具有重要参考意义。

1 SM4密码算法描述

SM4密码算法的分组长度是128比特,由32轮非线性迭代结构构成,加解密结构相同,只是轮子密钥使用顺序相反,SM4密码算法加密流程[6]如图1所示。

收稿日期:2021-01-14
基金项目:国家重点研发计划资助项目(2018YFB0904900、2018YFB0904901);"十三五"国家密码发展基金资助项目(MMJJ20180224);四川省重点研发资助项目(2019YFG0096)

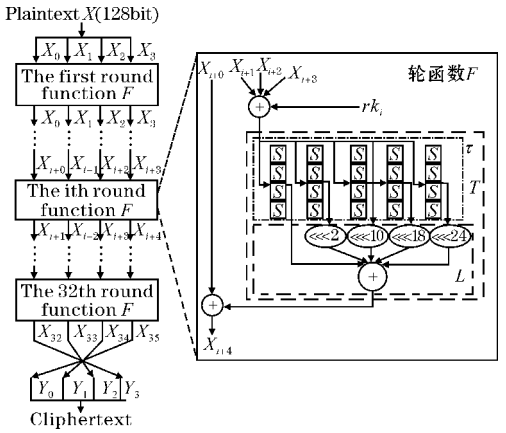

图 1 SM4 密码算法加密流程

设 128 比特的明文输入为 X,轮输入为(X_i,X_{i+1},X_{i+2},X_{i+3}),轮密钥为 rk_i,其中 X_i 和 rk_i 均为 32 比特,加密时,对 $i=0,1,2\cdots31$ 执行:

$$X_{i+4}=F(X_i,X_{i+1},X_{i+2},X_{i+3},rk_i)$$
$$=X_i\oplus \mathrm{T}(X_{i+1}\oplus X_{i+2}\oplus X_{i+3}\oplus rk_i) \quad (1)$$

输出的密文 $Y=(X_{35},X_{34},X_{33},X_{32})$。

在图 1 和式(1)中,轮函数 F 由异或运算和合成置换 T 构成,合成置换 T 又由非线性变换 τ 和线性变换 L 组成[6]。非线性变换 τ,是由 4 个 8 进 8 出的 S 盒构成。设线性变换的输入为 B,输出为 C,则线性变换 L 的运算如式(2)所示。

$$C=L(B)=B\oplus(B<<<2)\oplus(B<<<10)\oplus;(B<<<18)\oplus(B<<<24) \quad (2)$$

SM4 密码算法的密钥扩展运算也是有 32 轮迭代结构构成,设 SM4 密码算法 128 比特加密密钥为 MK,则每轮产生的轮密钥如式(3)和式(4)所示。

$$(K_0,K_1,K_2,K_3)=(MK_0\oplus FK_0,MK_1\oplus FK_1,MK_2\oplus FK_2,MK_3\oplus FK_3) \quad (3)$$

$$rk_i=K_{i+1}=K_i\oplus T'(K_{i+1}\oplus K_{i+2}\oplus K_{i+3}\oplus CK_i) \quad (4)$$

在式(3)和(4)中,$MK=(MK_0,{}_MK_1,MK_2,MK_3)$,$FK_0$,$FK_1$,$FK_2$,$FK_3$ 和 CK_i 为标准规定的 32 比特常数[10]。

式(4)中,函数的运算和合成置换 T 类似,只是线性变换不同,函数 T'的线性变换为

$$L'(B)=B\oplus(B<<<13)\oplus(B<<<23) \quad (5)$$

2 模板攻击

模板攻击通过采集可控的密码芯片的能量曲线来构建模板,最后进行模板匹配,完成对密码芯片的攻击。模板攻击通常由两个阶段构成:第一阶段是模板构建,第二阶段是模板匹配。

2.1 模板构建

采集能量曲线,选择信息泄露点和能量模型,根据能量模型将能量曲线划分 θ 个集合,对每个集合构建模板,模板由能量曲线的均值向量 M_α 和噪声矩阵的协方差矩阵 $\boldsymbol{\Sigma}_\alpha$ 构成,表示为 $T_\alpha=<M_\alpha,\boldsymbol{\Sigma}_\alpha>$,其中 $1\leqslant\alpha\leqslant\theta$。

设模板 T_α 中对应能量曲线条数为 m,每条曲线的采样点数为 n,单条能量曲线 t_λ 可以表示为 $t_{\lambda,1},\cdots,t_{\lambda,l},\cdots,t_{\lambda,n}(0\leqslant\lambda\leqslant m-1,1\leqslant l\leqslant n)$,则模板的均值向量为

$$M_\alpha=[\sum_{i=0}^{m-1}t_{\lambda,1}/m,\cdots,\sum_{i=0}^{m-1}t_{\lambda,l}/m,\cdots,\sum_{i=0}^{m-1}t_{\lambda,n}/m] \quad (6)$$

能量曲线的噪声曲线可以用能量曲线与能量曲线的均值差来计算,用 N_λ 表示,对应的噪声矩阵表示为 $\boldsymbol{N}_{m\times n}$,$\boldsymbol{N}_\lambda$ 和 $\boldsymbol{N}_{m\times n}$ 计算如式(7)和式(8)所示,则模板的协方差均值如式(9)所示。

$$N_i=(t_{\lambda,1}-\sum_{i=0}^{m-1}t_{\lambda,1}/m),\cdots,(t_{\lambda,2}-\sum_{\lambda=0}^{m-1}t_{\lambda,l}/m),\cdots,(t_{\lambda,n}-\sum_{\lambda=0}^{m-1}t_{\lambda,n}/m) \quad (7)$$

$$\boldsymbol{N}_{m\times n}=\begin{bmatrix}N_{0,1}&\cdots&N_{0,n}\\ \cdots&N_{\lambda,l}&\cdots\\ N_{m,1}&\cdots&N_{m,n}\end{bmatrix} \quad (8)$$

$$\boldsymbol{\Sigma}_\alpha=\begin{bmatrix}\mathrm{cov}(N_1,N_1)&\cdots&\mathrm{cov}(N_1,N_n)\\ \cdots&\mathrm{cov}(N_\lambda,N_\lambda)&\cdots\\ \mathrm{cov}(N_n,N_1)&\cdots&\mathrm{cov}(N_n,N_n)\end{bmatrix} \quad (9)$$

2.2 模板匹配

采集被攻击密钥和已知明文 X_i 进行加密运算时对应的能量曲线 t'_i,曲线条数为 j,猜测密钥 k',根据已知明文和密钥 k',以及能量模型,评估其符合特定模板的概率,计算公式为

$$p(t'_i\mid \mathrm{T}_{(k',X_i)})=\frac{1}{\sqrt{(2\pi)^n\mid\boldsymbol{\Sigma}_{(k',X_i)}\mid}}$$
$$\exp(-\frac{1}{2}(t'_i-M_{(k',X_i)})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}_{(k',X_i)}(t'_i-M_{(k',X_i)})) \quad (10)$$

对所有被攻击曲线计算联合概率,概率最大值对应的密钥即为攻击出的正确密钥。

$$P=\prod_{i=1}^{j}p(t'_i\mid \mathrm{T}_{(k',X_i)}) \quad (11)$$

3 针对 SM4 密码算法的模板攻击算法

SM4 密码算法的信息泄漏点比较多,常用的信息泄露点有 F 函数中 S 盒输入、S 盒输出、线性变换输出和轮输出等。针对 SM4 密码算法的模板攻击研究,选择的信息泄露点是 S 盒输出,能量模型为汉明重量模型,以攻击第一轮其中一个 S 盒对应的密钥为例,针对

SM4 密码算法的模板攻击算法如下:

(1)在相同采集条件下采集已知密钥 k 和随机明文 X_λ 进行加密运算时对应的能量曲线 t_λ,曲线条数为 n,$1\leqslant\lambda\leqslant n$,和被攻击密钥和随机明文 X'_i 进行加密运算时的能量曲线 t'_i,曲线条数为 j,$1\leqslant i\leqslant j$;

(2) 数据预处理。为了提高能量曲线的信噪比,采用滤波和静态对齐对采集的能量曲线进行预处理。

(3) 特征点选择。计算模板的多元高斯分布时,一般不选择能量曲线上所有的点。这是因为从效率方面考虑,计算维度非常大的协方差矩阵效率低,内存占用过大。此外,使用过高的维度计算协方差矩阵时,将产生奇异矩阵。所以特征点的选择对模板攻击能否成功非常关键。兴趣点选择的基本原则是:选择能量曲线上对不同信息表现出最大能耗差异的那些能量点。

(4) 模板构建。根据信息泄露点和能量模型,构建泄露函数,如(12)式所示,其中 $\beta\in[1,4]$。

$$d=f(X_\lambda,k)=HW(\tau(X_\lambda,k)_\beta) \tag{12}$$

由于 S 盒输出是 8 比特,所以 d 的范围是[0,8],根据 d 将步骤(3)中能量曲线特征点划分 9 个集合,根据文中所述对每个集合构建模板 $T_\alpha<M_\alpha,\sum_\alpha>$,其中 $a=d$。

(5) 模板匹配。采集 j 条被攻击曲线,进行模板匹配,根据式 (11)计算 j 条曲线的联合概率,概率最大值对应的密钥即为攻击出的正确密钥。

4　针对 SM4 密码算法的模板攻击实验

针对 SM4 密码算法的模板攻击实验,模板构建选择的密钥为 0x123456789ABCDEFFEDCBA9876543210,明文随机。实验条件是 Inspector 侧信道攻击平台,测试对象为 SM4 密码算法智能卡,被攻击密钥为 128 比特 0,采集到的能量曲线如图 2所示。

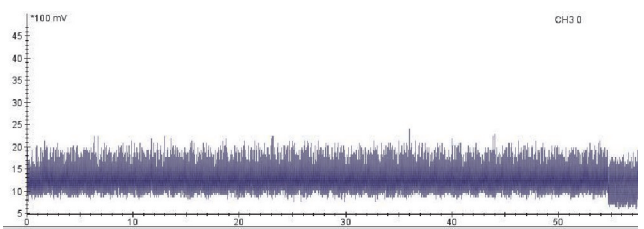

图 2　SM4 密码算法智能卡能量曲线

4.1　数据预处理

低通滤波:采用 Inspector 软件提供滤波工具对能量曲线做低通滤波处理,滤波参数为 30,滤波后的能量曲线的部分信号数据如图 3 所示。

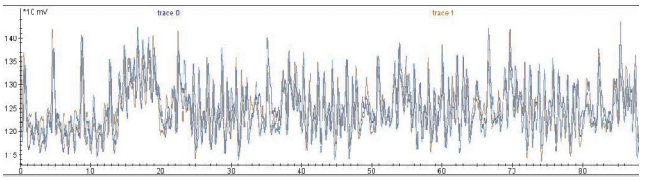

图 3　滤波后的曲线

数据对齐:对滤波后的数据进行静态对齐,对齐之后的曲线如图 4 所示。

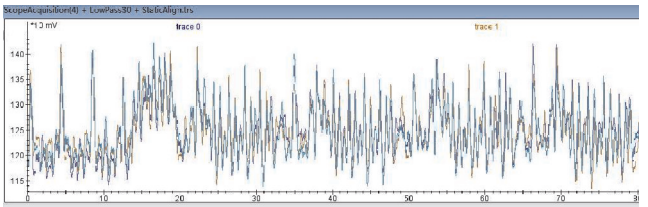

图 4　静态对齐后的曲线

4.2　特征点选择

对第一轮 4 个 S 盒输出对应的特征点进行选择,所采用的方法是 SOST(sum of squared t-test)。形成每个 S 盒对应的特征点集合,共 4 个特点集合。

4.3　模板构建

根据模板构建方法,针对第一轮 4 个 S 盒输出,对选择的特征点集合,分别构建模板。由于每个 S 盒输出为 8 比特,根据汉明重量模型,每个 S 盒可构造汉明重量为 0、1、2、3、4、5、6、7 和 8 对应的模板,共 9 个模板,训练所用能量曲线条数为 4000 条。

4.4　模板攻击

根据文中所述模板攻击方法进行模板攻击,攻击所用能量曲线条数为 1000 条,攻击结果如图 5所示。

```
Results after 1000 traces
Best correlation Round 1: Key 4:
rank: 1, candidate: 35 (0x23), confidence: 40.6114
rank: 2, candidate: 201 (0xC9), confidence: 38.3975
rank: 3, candidate: 29 (0x1D), confidence: 38.3874
rank: 4, candidate: 177 (0xB1), confidence: 38.3750
Best correlation Round 1: Key 3:
rank: 1, candidate: 59 (0x3B), confidence: 39.7740
rank: 2, candidate: 249 (0xF9), confidence: 38.1309
rank: 3, candidate: 72 (0x48), confidence: 37.8412
rank: 4, candidate: 11 (0x0B), confidence: 37.7907
Best correlation Round 1: Key 2:
rank: 1, candidate: 96 (0x60), confidence: 39.6464
rank: 2, candidate: 165 (0xA5), confidence: 38.0331
rank: 3, candidate: 236 (0xEC), confidence: 37.9778
rank: 4, candidate: 21 (0x15), confidence: 37.9623
Best correlation Round 1: Key 1:
rank: 1, candidate: 69 (0x45), confidence: 40.1417
rank: 2, candidate: 34 (0x22), confidence: 39.0312
rank: 3, candidate: 135 (0x87), confidence: 38.9870
rank: 4, candidate: 10 (0x0A), confidence: 38.9493
the best key:
Key 4 is: 0x23; Key 3 is: 0x3B; Key 2 is: 0x60; Key 1 is: 0x45;
```

图 5　模板攻击结果

从高字节到低字节排序,第一轮轮密钥最终的攻击结果为:0x45603B23。

同理,第二轮轮密钥最终的攻击结果为0x26440963,第三轮轮密钥最终的攻击结果为0xC5C2E19C,第四轮轮密钥最终的攻击结果为0xE1E0DDAA。根据前四轮轮密钥,反推出最终的加密密钥为128比特0,该结果和被攻击测试密钥相同,验证了攻击结果的正确性。

5 结束语

对SM4密码算法进行模板攻击研究,提出了针对SM4密码算法的模板攻击方法,探索和实践了该攻击方法针对SM4密码算法智能卡的应用。该攻击方法不仅对SM4密码算法产品的安全研究具有实际的应用意义,而且对其他密码算法的模板攻击安全研究具有重要的借鉴意义。

参考文献:

[1] Paul K, Joshua J, Benjamin J. Differential power analysis[A]. Proceedings of the 19th Annual International Cryptology Conference on Advances in Cryptology[C]. 1999:388-397.

[2] 杜之波, 孙元华, 王燚. 针对AES密码算法的多点联合能量分析攻击[J]. 通信学报, 2016, 37(S1):78-84.

[3] Schramm K, Wollinger T J, Paar C. A new class of collision attacks and its application to DES[C]. Fast Software Encryption-FSE 2003, LNCS2779, 2003:2-16.

[4] 金雨璇,杨宏志,王相宾,等. 对SM4算法的改进差分故障攻击[J]. 密码学报,2020,7(4):453-464.

[5] 荣雪芳,吴震,王敏,等. 基于随机故障注入的SM4差分故障攻击方法[J]. 计算机工程,2016(7).

[6] 杜之波,吴震,王敏,等. 针对SM4密码算法的多点联合能量分析攻击[J]. 计算机研究与发展,2016,53(10):2224-2229.

[7] 杜之波,吴震,王敏,等. 针对SM4轮输出的改进型选择明文功耗分析攻击[J]. 通信学报,2015,36(10):85-91.

[8] 王敏,饶金涛,吴震,等. SM4密码算法的频域能量分析攻击[J]. 信息网络安全,2015(8):14-19.

[9] 马向亮,李冰,杨丹,等. 基于深度学习的类SM4算法S盒逆向分析[J/OL]. 北京邮电大学学报:https://doi.org/10.13190/j.jbupt.2020-034,[2020-12-31]:1-7.

[10] Office of State Commercial Cipher Administration. Block cipher for WLAN products-SMS4[EB/OL]. http://www.oscca.go-v.cn/UpFile/200621016423197990.pdf,2006-02-10.

Template Attack Against SM4 Cryptographic Algorithm

KUANG Xiaoyun[1], HUANG Kaitian[1], LAN Tian[2], DU Zhibo[3], WU Zhen[3]

(1. Electric Power Research Institute, CSG, Guangzhou 510663, China; 2. Huada Semiconductor Co., Ltd., Shanghai 200120, China; 3. College of Cybersecurity, Chengdu University of Information Technolgy, Chengdu 610225, China)

Abstract: At present, the research of the side channel attack against the SM4 cryptographic algorithm is mostly about the fault analysis attack and correlation power analysis attack. A method of the template attack against SM4 cryptographic algorithm was proposed to explore the application of the template attack against SM4 cryptographic algorithm. When constructing the template, the S-box output of the SM4 cryptographic algorithm was selected as the attack point. And Hamming weight model was selected as power analysis model. The template about S-box output was constructed during the template attack. The joint probability of multiple power traces was selected for template matching. The key corresponding to the maximum probability was the correct key. The effectiveness of this method is demonstrated by the experiment of the SM4 cryptographic algorithm smart card. The method can not only be applied to the other attack points of the SM4 cryptographic algorithm, but also be used for reference to other block cipher algorithms.

Keywords: side-channel analysis attack; template attack; SM4 cryptographic algorithm; S-box