

基于 SAKURA-G 实验板的 SM4 硬件电路能量攻击研究

胡文静¹, 王 安¹, 乌力吉¹, 谢鑫君²

(1 清华大学 微电子与纳电子学系, 北京 100084; 2 上海通用识别技术研究所, 上海 201100)

摘 要: 目前, 公开研究 SM4 能量攻击的实验中, 功耗波形都采用计算机仿真、单片机软件实现等手段, 这与实际中使用的 SM4 硬件电路实现的功耗尚有很大差别. 由此分析了一种 SM4 算法硬件电路实现的安全性, 将它的 Verilog 硬件电路下载到 SAKURA-G 开发板中, 并采集其实际运行时的功耗信息. 通过相关性能量分析技术对首轮 S 盒输入、首轮 S 盒输出、首轮输出寄存器多个可能泄露信息的位置进行攻击, 还原了首轮子密钥. 通过相同方法可以还原第 2~4 轮子密钥, 最终推导 128 比特源密钥. 最后分别从系统级和算法级探讨了相应的防御对策.

关键词: SM4 分组密码算法; 相关性能量分析; SAKURA-G

中图分类号: TP309.7

文献标识码: A

文章编号: 1000-7180(2015)04-0015-06

Power Attack of SM4 Hardware
Implementation Based on SAKURA-G BoardHU Wen-jing¹, WANG An¹, WU Li-ji¹, XIE Xin-jun²

(1 Institute of Microelectronics, Tsinghua University, Beijing 100084, China;

2 Shanghai Universal Recognition Technology Institute, Shanghai 201100, China)

Abstract: Currently, in public researches about SM4 power attack, the power traces are generated by computer simulation or software implementation. However, this is different with hardware implementation which is used in actual. A research of a SM4 algorithm hardware implementation is given out, which is applied in market. we download the Verilog code to a SAKURA-G board, and collect the power traces when it actually operates. Correlation Power Analysis method is used to analyze the leakage of the input of the S-box in the first round, the output of the S-box in the first round and the output registers of the first round. We recovered the sub-key of the first round. By the same method, we can recover the sub-keys of round 2-4, and eventually get the 128-bit key.

Key words: SM4; correlation power analysis; SAKURA-G

1 引言

侧信道攻击利用密码硬件设备在实际运行中所泄露的时间、功耗、电磁等信息, 结合设备的输入输出数据来恢复设备中存储的密钥等信息. 自 1999 年 Kocher 提出差分能量攻击^[1]以来, 侧信道分析迅猛发展, 目前已经成为密码设备安全性方面的主要威胁之一.

SM4 分组密码算法是我国广泛应用的商用密码算法, 在无线局域网中保护数据安全传输具有重

要作用. 2009 年, 沈薇在 C 语言仿真平台上完成了 SM4 算法的能量分析实验^[2], 白雪飞等利用晶体管级仿真波形对 SM4 算法实施了差分功耗攻击^[3]; 2010 年李浪等将 SM4 算法硬件电路用 Modelsim 仿真, 并实施差分能量攻击^[4]; 2014 年滕永平等研究了 SM4 算法单片机软件实现的差分功耗攻击^[5]. 然而计算机仿真和单片机软件实现的功耗波形同实际中使用的硬件电路实际^[6-7]的功耗波形有很大差别.

本文给出一个针对 SM4 硬件电路实现实际攻

收稿日期: 2014-04-13; 修回日期: 2014-06-10

基金项目: 核高基重大专项(2014ZX01032205); “十二五”国家密码发展基金密码理论课题(MMJJ201401009)

击的例子,通过将 SM4 硬件电路及相应控制代码下载到 SAKURA-G 开发板中,并采集其实际运行时的功耗信息,对多个可能泄露信息的位置进行相关性能量分析^[8],恢复出首轮密钥,并提出相应的防护对策。

2 相关知识

2.1 SM4 算法及硬件实现

SM4 密码算法和 DES、AES 类似同为分组密码算法,分组长度和密钥长度都是 128 bit,加密算法与密钥扩展算法都采用 32 轮非线性迭代结构.解密算法和加密算法的结构相同,只是密钥使用的顺序相反,解密轮密钥是加密轮密钥的逆序.加密结构如图 1 所示。

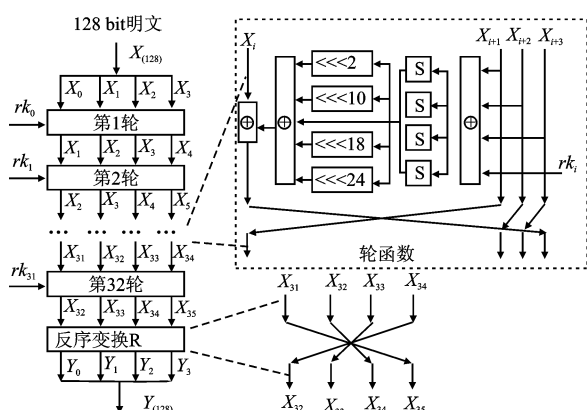


图 1 SM4 算法结构

明文输入为 $X = (X_0, X_1, X_2, X_3) \in (GF(2^{32}))^4$, 密文输出为 $Y = (Y_0, Y_1, Y_2, Y_3) \in (GF(2^{32}))^4$, 第 i 轮运算的输入为 $X_i, X_{i+1}, X_{i+2}, X_{i+3}$, 轮密钥为 $rk_i \in GF(2^{32}), i = 0, 1, \dots, 31$, 则算法的加密变换定义为

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

$$(Y_0, Y_1, Y_2, Y_3) = R(X_{32}, X_{33}, X_{34}, X_{35}) = (X_{35}, X_{34}, X_{33}, X_{32}).$$

式中, T 为合成置换, R 为反序变换. 合成置换 T 由非线性变换(四个并行的 8 位 S 盒)和线性变换构成。

SM4 算法的轮密钥由密钥扩展算法构成,其结构与加密变换类似。

本文中的 SM4 硬件电路首先执行写明文和写密钥操作,随后进行第 1 轮密钥生成、第 1 轮加密、第 2 轮密钥生成、第 2 轮加密……如此交替进行各 32 轮密钥生成和加密.每轮密钥生成和加密均在一个时钟周期内完成. S-box 采用塔域求逆的方法^[9]

实现,4 个 S-box 为并行关系.实验中采用的密钥固定为 01234567 89ABCDEF FEDCBA98 76543210 (对应首轮子密钥为 F12186F9,按字节换算为 10 进制为 241,33,134,249). 本文的攻击目标是通过能量分析技术还原首轮子密钥,进而可以通过相同方法还原第 2~4 轮子密钥,并推导 128 比特源密钥。

2.2 SAKURA 平台搭建

SAKURA-G 是一款专门用于研究硬件电路侧信道安全的开发板,如图 2 所示,它由两块 FPGA 构成,一块主 FPGA 与一块控制 FPGA. 主 FPGA 用于实现密码算法,控制 FPGA 用于实现相应的控制逻辑. LDO 输出的稳压电源经过一个电阻输送到主 FPGA,我们用差分探头来测量电阻两端的电压来得到主 FPGA 的功耗信息。

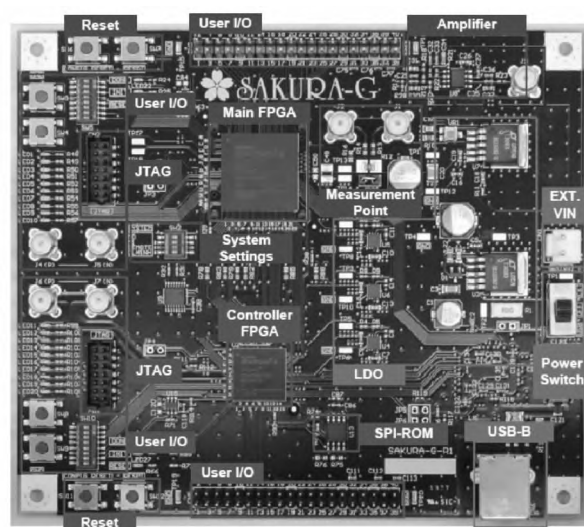


图 2 SAKURA-G 开发板

3 SM4 能量分析攻击

3.1 实验方法简介与波形初步分析

3.1.1 采集设备与原始波形简析

能量分析实验中,首先将 SM4 硬件电路下载到 SAKURA-G 开发板中,用安捷伦 MSO 9000 混合信号示波器完成波形采集. 开发板时钟频率 48 MHz(周期 20.83 ns),采样率选用 10 GSa/s,存储深度占用 30 kSa(实际采到的每条原始波由 30 000 个点构成). 为了降噪和便于高速分析,我们将每条波形进行了重采样,即每 10 个点平均为 1 个点,因此重采样后每条波含有 3 000 个点(每个点代表 1 ns). 一条原始波如图 3 所示。

由仿真结果知,0~354 ns(状态 0~17)是写明文操作,396~687 ns(状态 19~33)是写密钥操作,770~2 125 ns(状态 37~102)是密钥生成与加密操

作, 2 125~2 625 ns(状态 103~126)是在等待读取密文, 2 625 ns 之后开始读取密文. 显然波形特征与该仿真结果完全吻合.

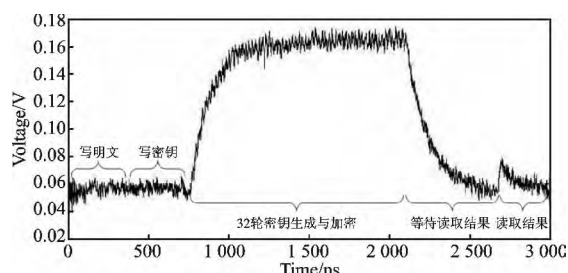


图3 单条原始波形初步分析

可见单次 SM4 加密的波形特征十分明显, SM4 加密区间很容易直接用肉眼分辨, 十分有助于攻击者对 SM4 算法的操作区间进行定位和进一步分析.

3.1.2 平均波形分析

采用 4 000 个随机的不同明文, 加密 4 000 次, 采集到对应的 4 000 条波形, 并将明文和波形分别保存. 对 4 000 条波形进行直接平均得到平均波, 如图 4 所示. 将其 1 000 ns 到 1 500 ns 部分放大, 可得图 5, 500 ns 的时间内刚好是 24 个周期, 与时钟周期 20.83 ns 完全吻合, 这说明每次时钟上升沿带来的寄存器值变化, 以及随后的组合电路发生的翻转, 均能够清晰地反映在功耗波形上.

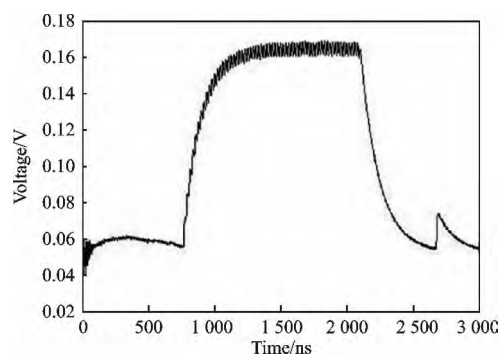


图4 4 000 条波形的平均波形

因此该电路的时钟、整体 SM4, 以及轮函数中均没有添加任何随机延时对策, 导致波形是自然对齐的, 对于攻击者来说易于直接开展统计分析, 即用少量波形完成侧信道攻击.

3.2 对 S-box 处信息泄露的分析

3.2.1 首轮 S-box 输入值泄露点研究

继续使用上一节采到的 4 000 条波形及其对应的 4 000 个明文. 以首轮第一个 S-box 为例, 我们用已知的密钥计算出每个明文对应的首轮第一个 S-box 输入值的汉明重量, 共计 4 000 个, 记作列向量

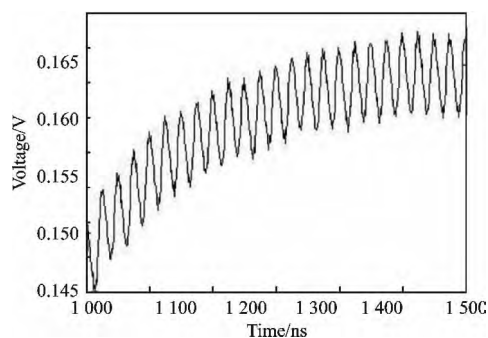


图5 平均波形的放大

V_1 . 而 4 000 条波中第 i 个横坐标对应的 4 000 个点可记作列向量 $V_{2,i}$. 我们用相关系数来计算 V_1 和 $V_{2,i}$ 之间的相关性, 可得到一个值 ρ_i . 当横坐标 i 遍历 1 到 3 000 时, 可构成一条相关系数曲线, 如图 6 所示. 对该曲线进行重采样(每相邻的 10 个点平均为一个点), 可得更清晰的图 7.

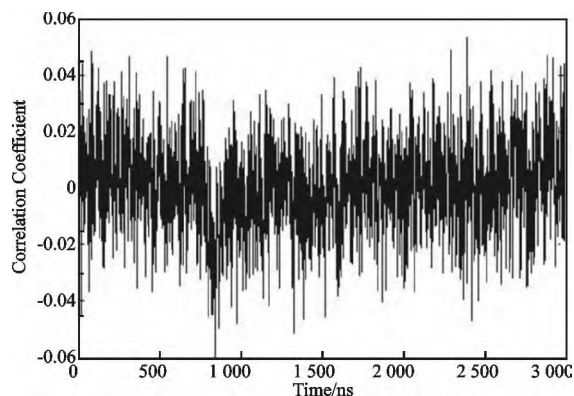


图6 S-box 输入值泄露分析图(1)

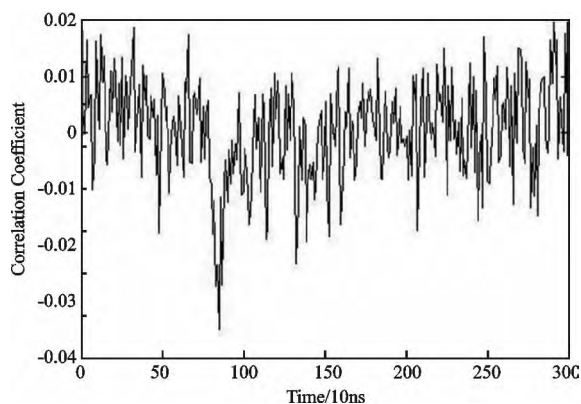


图7 S-box 输入值泄露分析图(2)

从图中可以看出, 在横坐标 781 ns 到 900 ns 之间, 首轮 S-box 输入值处有非常显著的数据泄露.

3.2.2 对 S-box 输入值实施相关能量分析

我们选用首轮第一个 S-box 输入值(设为 x)处

实施能量分析攻击. 具体方法如下:

(1) 采集 4 000 条波形, 并保存 4 000 个对应的随机明文;

(2) 猜测首轮子密钥第一个字节 $k = 0$;

(3) 用 k 和已知 4 000 个明文分别计算 4 000 个 x 值及其汉明重量 $\text{HW}(x)$;

(4) 对于波形的每一个横坐标 i , 用 4 000 个 $\text{HW}(x)$ 和 4 000 个波形点做相关系数, 记作 ρ_i ;

(5) 将 $\rho_{781} \sim \rho_{900}$ 进行平均, 记作 $\hat{\rho}_k$;

(6) $k = k + 1$, 重复 (3), (4), (5), 直到 $k > 255$ 时结束;

(7) 对于 k 从 0 到 255, 画出 $\hat{\rho}_k$ 函数图, 如图 8 所示.

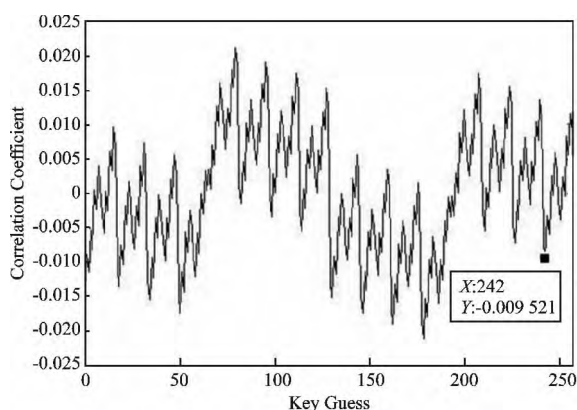


图 8 对 S-box 输入值实施相关能量分析结果图

从图中可以看出, S-box 输入值的泄露严格符合汉明重量模型的规律, 但真实密钥 241 (图中显示为 242 是因为 Matlab 编号从 1 开始而不是从 0 开始, 下同) 并未出现在最低点或最高点. 这说明真实的泄露位置不是 x 处, 而是在密钥加入运算后, 跟 x 有关的一个位置 $f(x)$. 所以, 暂时无法用相关能量分析破解该 SM4, 因为其真实泄露位置不详. 若泄露具体位置搞清楚后, 可以十分容易地按字节恢复出密钥.

3.2.3 S-box 输出值泄露点研究

本节中用类似分析 S-box 输入值的方法来分析 S-box 输出值, 发现泄漏并不明显, 如图 9、图 10 所示. 说明 S-box 输出值处无明显数据泄露.

3.2.4 对 S-box 输出值实施相关能量分析

本节中同样用类似针对 S-box 输入值实施相关能量分析的方法来分析 S-box 输出值, 发现正确密钥并未出现在最高点或最低点, 如图 11 所示.

对 S-box 输出值处实施相关能量分析, 无法得到有用的信息.

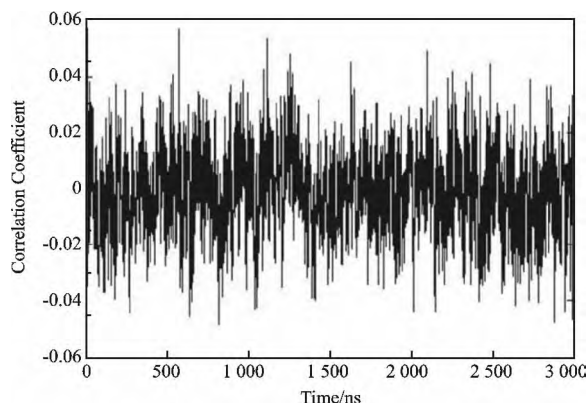


图 9 S-box 输出值泄露分析图 (1)

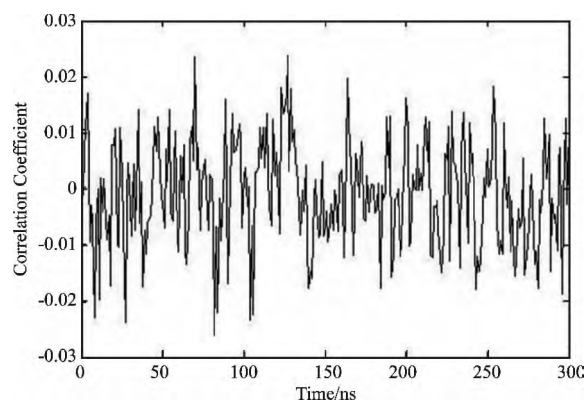


图 10 S-box 输出值泄露分析图 (2)

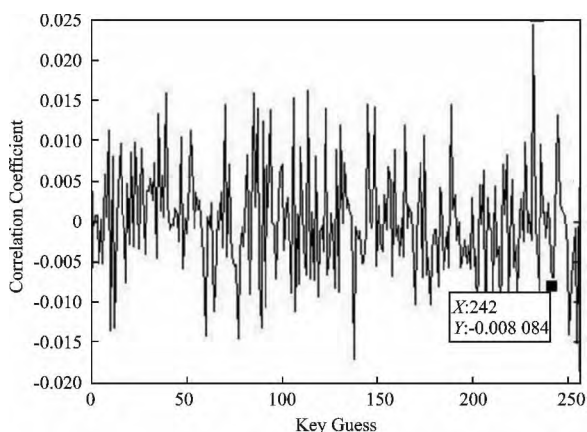


图 11 对 S-box 输出值实施相关能量分析结果图

3.3 对首轮输出寄存器信息泄露的分析

3.3.1 首轮输出寄存器泄露点研究

根据仿真分析得知, 在第一轮结束时, 轮函数寄存器值是从 K_4 变为 X_4 . 因此拟研究从 K_4 到 X_4 的汉明距离. 我们利用明文和密钥计算出 32 比特的首轮输出值 X_4 , 并计算从 K_4 到 X_4 的汉明距离. 4 000 个明文分别对应 4 000 个汉明距离值, 与前述类似, 用这 4 000 个汉明距离值与波形上每一个横坐标对

应的 4 000 个点分别做相关系数,可得到图 12 所示的相关系数曲线。显然,841 ns 到 1 060 ns 之间的相关系数显著高于其余部分。因此,在 841 ns 到 1 060 ns 之间,首轮输出值 X_4 与子密钥 K_4 之间的汉明距离的泄露十分明显。这意味着一定可以在该位置实施相关能量分析。

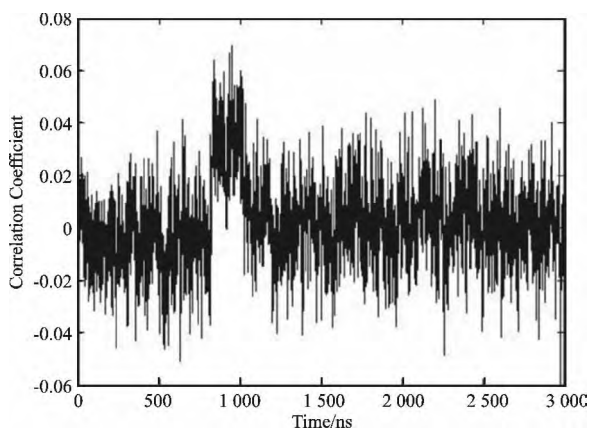


图 12 首轮输出寄存器泄露分析图

3.3.2 对首轮输出寄存器实施相关能量分析

由于直接实施相关能量分析的搜索空间过大 (2^{32}),先尝试在假设 3 字节密钥已知的前提下,恢复第四字节密钥。攻击位置选取首轮输出值 X_4 寄存器处,利用 X_4 与 K_4 之间的汉明距离的泄露来恢复密钥(设二者的第一字节分别为 y 和 k),具体方法如下:

- (1) 采集 4 000 条波形,并保存 4 000 个对应的随机明文;
- (2) 假设首轮子密钥第 2、3、4 字节已知。猜测首轮子密钥第一个字节 $k = 0$;
- (3) 用 k 和已知 4 000 个明文分别计算 4 000 个 y 值及其汉明距离 $HW(y \oplus k)$;
- (4) 对于波形的每一个横坐标 i ,用 4 000 个 $HW(y \oplus k)$ 和 4 000 个波形点做相关系数,记作 ρ_i ;
- (5) 将 $\rho_{841} \sim \rho_{1060}$ 进行平均,记作 ρ_k ;
- (6) $k = k + 1$,重复(3),(4),(5),直到 $k > 255$ 时结束;
- (7) 对于 k 从 0 到 255,画出 $\hat{\rho}_k$ 函数图,如图 13 所示。

显然,可通过该算法恢复出密钥第一个字节 241。通过相同的方法,可恢复出其余三个字节密钥 33、134、249,如图 14、15、16 所示。

此外,我们也尝试了用 4 000 条波形在已知 K_4 第 3、4 字节的情况下,用相关能量分析恢复其第 1、

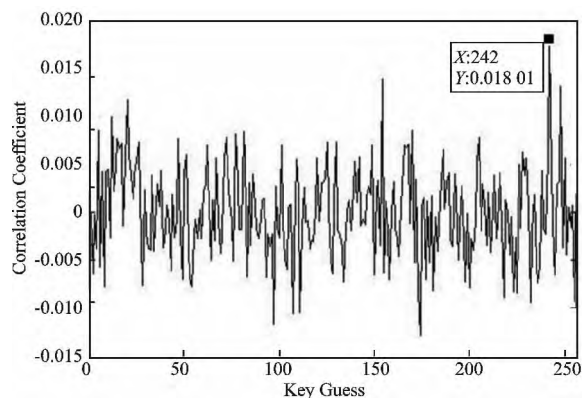


图 13 按字节对首轮输出寄存器实施相关能量分析(1)

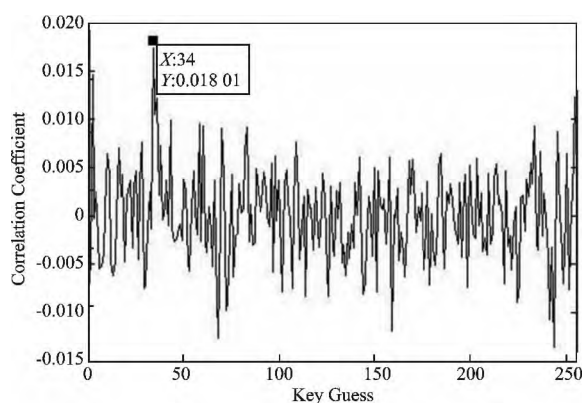


图 14 按字节对首轮输出寄存器实施相关能量分析(2)

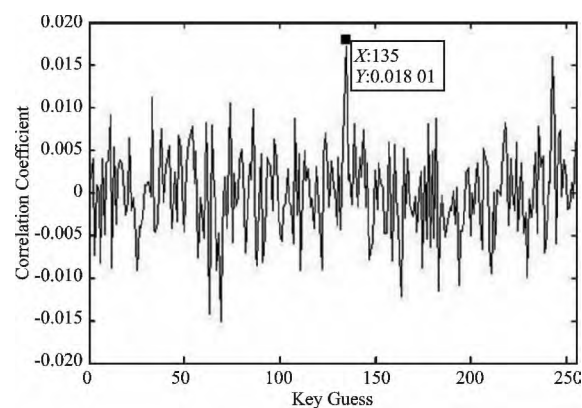


图 15 按字节对首轮输出寄存器实施相关能量分析(3)

2 字节。实验程序在单台 PC 运行约 2 小时,结果如图 17 所示,也正确恢复出了 2 字节密钥 61 729 ($241 \times 256 + 33$)。

以上所有统计分析实验均在 4 000 条波形下完成,我们最后尝试了用 2 500 条波来攻击前两字节的子密钥,发现正确密钥也能出现在最大的几个 ρ 值当中。因此,配合少量的搜索,2 500 条波形也能够破解该 SM4 电路。如图 18 所示。

以上研究说明:

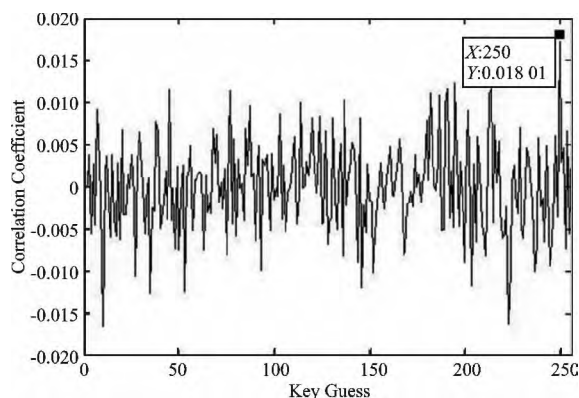


图 16 按字节对首轮输出寄存器实施相关能量分析(4)

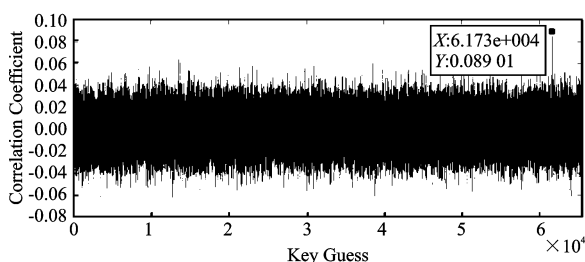


图 17 用 4 000 条波对首轮输出寄存器的前 2 字节实施相关能量分析

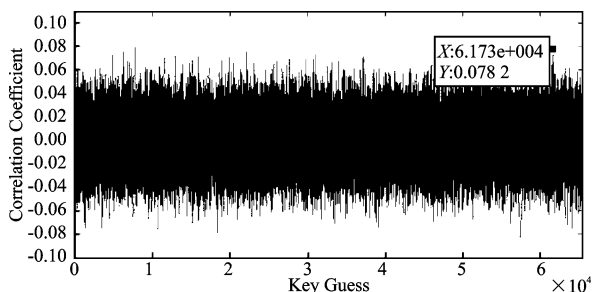


图 18 用 2 500 条波对首轮输出寄存器的前 2 字节实施相关能量分析

(1) 首轮输出寄存器的泄露十分明显, 实验证明 2 500 条波形(采集时间大约仅需 10 分钟)的泄露足以用于恢复 16 比特首轮子密钥。

(2) 根据 16 比特密钥能够清晰恢复出来的情况推断, 32 比特密钥也可以极高的概率正确恢复, 因为并行的比特数越多, 其汉明重量泄露的信噪比越大。

(3) 由于 32 比特搜索空间稍大, 较难在短期内完成完整 32 比特的密钥恢复, 但是用 32 比特的猜测密钥来计算中间值的过程可在攻击之前进行预计算, 其所需时间实际可行。

4 SM4 算法 IP 核抗攻击措施建议

4.1 算法级对策

伪轮: 加入若干个伪轮来让攻击者不确定哪轮

是第一轮或最后一轮。掩码: 将泄露位置全部随机化, 保证理论上无泄露。

4.2 系统级对策

时钟随机化: 将时钟高低电平出现的频率随机化, 使波形难以对齐。伪 SM4: 加入若干个(个数随机)伪 SM4 来让攻击者不确定哪个 SM4 是真实的 SM4。

5 结束语

本文给出了一个对实际应用的 SM4 硬件电路进行功耗攻击的实例, 通过对多个可能泄露信息的位置进行攻击, 成功地找到了信息泄露点, 并成功地恢复出首轮密钥, 运用同样的方法可以恢复出第 2、3、4 轮的密钥, 进而推导出 128 bit 源密钥。由此本文给出了相应的防御对策。

参考文献:

- [1] Kocher P, Jaffe J, Jun B. Differential power analysis [C]// Advances in Cryptology-CRYPTO 99. Berlin, Heidelberg, Springer, 1999: 388-397.
- [2] 沈薇. SM4 算法的能量分析攻击及其防御研究[D]. 西安: 西安电子科技大学, 2009.
- [3] 白雪飞, 郭立, 徐艳华, 等. SMS4 密码算法的差分功耗分析攻击研究[J]. 小型微型计算机系统, 2009, 30(3): 541-544.
- [4] 李浪, 李仁发, 李静, 等. 一种 SM4 加密算法差分功耗攻击[J]. 计算机科学, 2010, 37(7): 39-41.
- [5] 滕永平, 陈运, 陈俊, 等. SM4 算法的差分功耗以及相关功耗分析研究[J]. 成都信息工程学院学报, 2014, 29(1): 14-18.
- [6] 陈吉华. VHDL——一种统一的设计描述语言[J]. 微电子学与计算机, 1993, 10(8): 35-38.
- [7] Brier E, Clavier C, Olivier F. Correlation power analysis with a leakage model[C]// CHES 2004, LNCS 3156. USA: Cambridge, 2004: 16-19.
- [8] Canright D, Batina L. A very compact "perfectly masked" S-Box for AES[C]// ACNS 2008. New York: LNCS 5037, 2008: 446-459.

作者简介:

胡文静 女, (1988-), 硕士研究生. 研究方向为集成电路工程与信息安全. E-mail: hwj12@mails. tsinghua. edu. cn.

王安 男, (1983-), 博士后. 研究方向为信息安全.

乌力吉 男, (1965-), 博士, 博士生导师. 研究方向为信息安全与汽车电子.

谢鑫君 男, (1980-), 硕士, 助理工程师. 研究方向为信息安全与数据处理.