软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software,2013,24(9):2238−2249 [doi: 10.3724/SP.J.1001.2013.04356]

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel/Fax: +86-10-62562563

对白盒 SMS4 实现的一种有效攻击*

林婷婷, 来学嘉

(上海交通大学 计算机科学与工程系,上海 200240)

通讯作者: 林婷婷, E-mail: lintingting00@163.com

摘 要: 传统的密码模型都假设密码系统的运行终端和计算环境是可信任的,但是,随着攻击方式的发展,这样的模型显得越来越脆弱.而白盒攻击环境是指攻击者除了能够获得与传统密码模型同样的资源以外,还对密码系统的内部运行完全可见,并完全掌控执行环境.因此,能够抵抗白盒攻击的密码算法具有更高层次的安全意义.2009 年提出的 SMS4 算法的白盒实现,其目标是在白盒攻击环境下能够防止 SMS4 算法的密钥被恢复.在回顾已有研究的基础上,针对该 SMS4 算法的白盒实现提出了一种有效的攻击,并详细解释了如何以低于 2^{47} 的时间复杂度找出嵌入其中的轮密钥,说明了该白盒设计方法的不可靠性,并为设计安全的白盒实现提供了一种参考.

关键词: 白盒;SMS4;攻击

中图法分类号: TP309 **文献标识码**: A

中文引用格式: 林婷婷,来学嘉.对白盒 SMS4 实现的一种有效攻击.软件学报,2013,24(9):2238−2249. http://www.jos.org.cn/1000-9825/4356.htm

英文引用格式: Lin TT, Lai XJ. Efficient attack to white-box SMS4 implementation. Ruan Jian Xue Bao/Journal of Software, 2013,24(9):2238−2249 (in Chinese). http://www.jos.org.cn/1000-9825/4356.htm

Efficient Attack to White-Box SMS4 Implementation

LIN Ting-Ting, LAI Xue-Jia

(Department of Computer Science & Engineering, Shanghai Jiaotong University, Shanghai 200240, China)

Corresponding author: LIN Ting-Ting, E-mail: lintingting00@163.com

Abstract: In traditional cryptographic model, it is assumed that the communication end points and computing environments of a cryptosystem are trusted. But this model becomes increasingly frailer with the development of the attack method. In the white-box attack model, the adversary can get not only access to the same resources as in the traditional cryptographic model but also total visibility of the internal implementation of the cryptosystem and full control over its execution environment, so it has the higher level of secure significance. The white-box SMS4 implementation, which was proposed in 2009, is aimed at protecting SMS4 operated in the white box context against key exposure. In this paper, based on the review of previous research, we propose an efficient attack and explain in detail how to extract the round key embedded in such a white box SMS4 implementation, with worst time complexity 2^{47}. As a result, we show that the white-box method is unreliable and provide reference for the secure white-box implementation.

Key words: white-box; SMS4; attack

随着网络和计算机技术应用的快速发展,信息安全问题已经成为关乎国计民生的大事.传统密码学建立在黑盒模型下,该模型假设算法运行在一个可信的终端,其运行子结果、内存等均处在不可查看和不可更改的可信任环境中——即假设攻击者不知道密钥,但是知道密码算法,能够控制明文并获得密文输出,但是加密的动态过程是隐藏的,攻击者无法观察算法的运行.

然而,密码学的应用已经遍及到人们日常生活的各个角落,例如电子邮件、网页访问、数字内容分发(电子

* 基金项目: 国家自然科学基金(61073149)
收稿时间: 2012-03-07; 修改时间: 2012-06-29; 定稿时间: 2012-12-03

书、音乐、音像等)、网络交易、电子政务等等.这些应用虽然在一个标准的终端上执行,像 PC、手机、智能卡等,但却使用在一个不可信任的环境中.例如,用户在自己的机器上运行一个数字媒体的播放软件,这个软件对加密过的数字内容进行解密后播放,那么这些软件的运行环境很有可能是不安全的,因为软件解密的动态过程对于攻击者(甚至可能就是用户本身)是可见的,他们很容易就获得密钥信息.此外,近年来提出的时间分析、功耗分析、插入错误分析等旁路攻击[1−4],也称为灰盒攻击(gray-box attack),使得许多密码系统变得不堪一击.而传统的黑盒模型则较少考虑这些问题,它已经不再满足越来越高的安全要求.

白盒密码正是在这样的背景下提出的:2002 年,Chow 等人提出白盒攻击环境(white-box attack context,简称 WABC)[5,6]的概念,允许攻击者对加密软件以及其运行的环境拥有完全的控制能力、与软件的执行者拥有同等的权利.例如,对程序运行的二进制追踪、读取内存中的密钥、观察程序执行的中间结果、任意的静态分析以及改变子计算的结果等,传统的密码算法在这个攻击环境中显得极度脆弱.因此,构造出能够在白盒攻击环境下保证安全性的密码算法迫在眉睫.

Chow 等人提出的 AES 白盒实现[5]和 DES 白盒实现(white-box implementation)[6],其主要目的是在白盒攻击环境下开发出一套保护技术,使得执行密码算法时密钥提取变得困难.他们的方法是:将已有的密码算法分割为若干步骤;将取定的密钥隐藏在最复杂的某个步骤中;对每个步骤使用可逆双射前后编码,并针对每个步骤的所有可能输入计算输出,做成一个查找表;执行密码算法时,就通过顺序地查找这一系列查找表来完成.这样的过程称为密码算法的白盒实现,可以说是设计安全白盒算法的第一步.在 Chow 等人的文章中,还给出了白盒多样性(white-box diversity)和白盒含混度(white-box ambiguity)来刻画白盒实现的安全性,并说明他们的方案满足这两个白盒安全的要求.

但是,在 Chow 等人的白盒 DES 方案发表后不久,2002 年,Jacob 等人指出,文献[6]的白盒实现方法并不安全,他们针对无外部编码(external encoding)的 DES 白盒实现构造了一种注入错误攻击(fault-injection attack)[7],可以以较低的复杂度找出密钥.2005 年,Link 等人[8]在这个攻击的基础上对 DES 的白盒实现进行改进,得出了一种新的 DES 白盒实现方案,这个方案在当时能抵抗已知的几个攻击.2007 年,Wyseur 等人[9]在不考虑外部编码的情况下,基于内部信息,利用截断差分分析对 Chow 的白盒 DES 进行攻击;Goubin 等人[10]也同样基于截断差分分析,针对白盒 DES 的第 1 轮提出了一种攻击.

而对于 AES 的白盒实现方案,2004 年,Billet 等人[11]提出了一种非常有效的攻击方法,我们称之为 BGE 攻击方法,通过选择某些特定的查找表,使用代数的方法去掉其中的非线性部分,并获得密钥.后来,Michiels[12]将其改进为一种通用攻击方法,可以对类似算法的白盒实现进行攻击.

在 Chow 等人提出的白盒 AES 和 DES 算法被证明是不安全的之后,虽然白盒算法一直是学术界讨论的热点,但却一直未有安全的白盒密码算法出现.2009 年,肖雅莹等人[13]提出了 SMS4 白盒实现算法,为了避免占用过多内存,他们使用查找表和仿射变换相结合的方式.从白盒多样性和白盒含混度的角度来看,此方案达到了所需的安全性,其作者也针对 Billet 等人提出的 BGE 攻击方法说明了其方案的安全性.

总的来说,白盒密码的目的是为了保证密码算法在不可信任主机上运行时,密钥不会被窃取.虽然有两种方式可以达到这个目的:一是使用动态的密钥,每次加、解密使用的密钥都不相同;二是将密钥固定,并嵌入到密码算法的执行中,将其与其他数据相混淆,使敌手无法提取出密钥.但是,第 1 种方式实现代价较之固定密钥的方式来说太高,如何降低实现代价将是未来的研究方向;而现有的白盒密码算法均是采用第 2 种方式,并以 Chow 等人使用查找表的方式最为典型,肖雅莹等人采用查找表和仿射变换相结合的方式,而其余实现白盒密码的方式并未多见.我们认为,将密钥隐藏在查找表中的方式是实现白盒密码的一个较好的选择,但从以上描述可知,Chow 等人所提出的两个白盒密码实现方案[5,6]均无法承受各种密码分析[7−12],无法达到所需的安全要求,其中,以 Billet 等人[11]所提出的攻击方法最为典型和有效;而到目前为止,据笔者所知,还未有其他有关肖雅莹等人方案[13]的安全性的评价.然而我们发现,将 BGE 攻击方法与差分分析法、求解方程组等方法相结合,可以恢复出 SMS4 的轮密钥,从而攻破白盒 SMS4 实现.

本文中,我们针对文献[13]中的方案构造了一种攻击方法,能以较低的复杂度找出 SMS4 的轮密钥.其基本

思想是:选定白盒 SMS4 的某一轮,通过将其方案中 Part 2 和 Part 3 与下一轮的 Part 1 合成起来,消去中间网络化编码的部分,得出关于某个仿射变换的一系列代数式,通过求解代数式,获得该仿射变换的线性部分,从而获得输入输出的差分,再完全恢复该仿射变换,进而获得轮密钥.

1 SMS4 白盒方法

1.1 SMS4算法简介

SMS4 算法[14]是国内官方公布的首个商用密码算法,即无线局域网产品使用的 SMS4 密码算法.该算法的分组长度和密钥长度均为 128bit,采用 32 轮非线性迭代结构.解密过程和加密过程的结构相似,但是轮密钥的使用顺序相反.其具体的算法流程如图 1 所示.

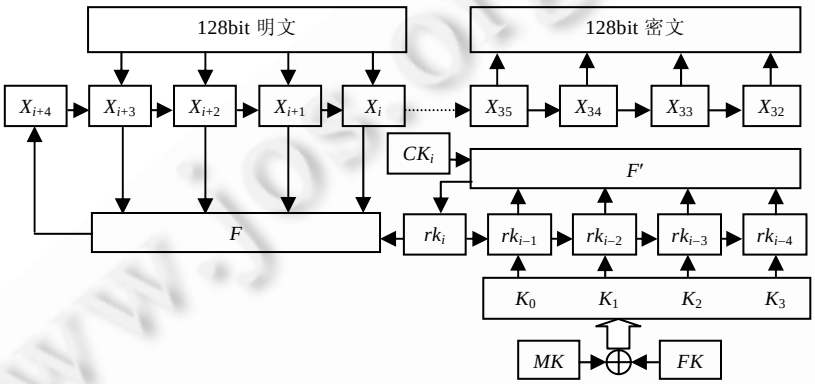

Fig.1 Flow diagrams of algorithm for SMS4

图 1 SMS4 算法流程

这里,明文$(X_{i+1},X_{i+2},X_{i+3},X_{i+4})$为 4 个 32bit,$X_{i+4}=F(X_i,X_{i+1},X_{i+2},X_{i+3},rk_i)=X_i\oplus T(X_{i+1}\oplus X_{i+2}\oplus X_{i+3}\oplus rk_i)$,$T$是一个可逆变换,由非线性变换$\tau$和线性变换 L 复合而成,即 $T(x)=L(\tau(x))$.而非线性变换τ由 4 个并行的 S 盒构成:

$$A=(a_0,a_1,a_2,a_3),\tau(A)=(S_{box}(a_0),S_{box}(a_1),S_{box}(a_2),S_{box}(a_3)).$$

线性变换 L 定义为 $L(X)=X\oplus(X<<<2)\oplus(X<<<10)\oplus(X<<<18)\oplus(X<<<24)$.SMS4 具体细节见文献[14].

1.2 SMS4白盒实现

下面描述 SMS4 的白盒实现方法,其基本思想是:将 SMS4 的每一轮中的某些步骤合成一个查找表,再利用可逆仿射变换作为输入编码(input-encoding)和输出编码(output-encoding)将其混淆.其中,上一变换的输出编码将与下一变换的输入编码互相抵消,称为网络化编码.具体步骤如下:

Part 1:首先计算 $X=X_{i+1}\oplus X_{i+2}\oplus X_{i+3}$,正如前面所述,每一步变换前后都要分别进行输入编码和输出编码:

$$X'_{i+j}=E_i^{-1}\circ P_{i+j}^{-1}(X_{i+j}),i=0,1,\ldots,31,j=0,1,2,3.$$

与 Chow 等人使用双射(bijection)不同的是,这里的输入/输出编码使用仿射变换的形式:$P_{i+j}(x)=A_{i+j}(x)\oplus a_{i+j}$.其中,$A_{i+j}$ 为 $GF(2)$上的 32×32 可逆矩阵,a_{i+j} 为 32bit 列向量,$E_i=diag(E_{i0},E_{i1},E_{i2},E_{i3})$,每个 E_{ij} 均为 $GF(2)$上的 8bit~8bit 的可逆仿射变换.P_{i+j} 和 E_i 随机选择并保密,最后只需要保存 $M_{i+j}^i=E_i^{-1}\circ P_{i+j}^{-1}$,它可以由 $GF(2)$上的 32×32 可逆矩阵和一个 32bit 列向量表示.$X=X_{i+1}\oplus X_{i+2}\oplus X_{i+3}$就由 3 个仿射变换和两个异或构成,如图 2 所示(直角边框表示数据,弧角边框表示变换,下同).

Part 2:接下来计算 $T(X\oplus rk_i)$.在这一步将密钥隐藏在 S 盒中:

$$T(X\oplus rk_i)=L(\tau(X\oplus rk_i))=L(S_{box}(x_{i0}\oplus rk_{i0}),S_{box}(x_{i1}\oplus rk_{i1}),S_{box}(x_{i2}\oplus rk_{i2}),S_{box}(x_{i3}\oplus rk_{i3})),$$

其中,$X=X_{i+1}\oplus X_{i+2}\oplus X_{i+3}=(x_{i0},x_{i1},x_{i2},x_{i3}),rk_i=(rk_{i0},rk_{i1},rk_{i2},rk_{i3})$.由于 S 盒是公开的,所以若是已知 $S_{box}(x_{ij}\oplus rk_{ij})$和 x_{ij},

rk_{ij},j=0,1,2,3 就很容易获得.因此,使用 8bit 到 8bit 的可逆仿射变换 E_{ij} 作为输入编码,32bit~32bit 的可逆仿射变换 Q_i 作为输出编码,记 $S_{ij}(*)=S_{box}(*\oplus rk_{ij})$,如图 3 所示.

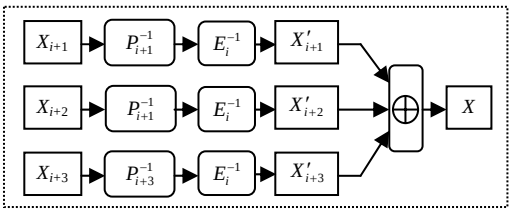

Fig.2 Part 1

图 2 Part 1

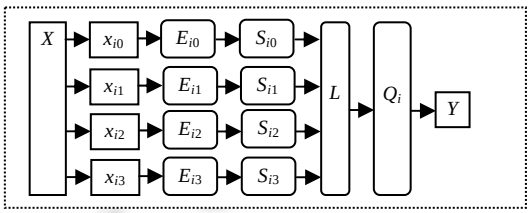

Fig.3 Part 2

图 3 Part 2

这一部分用查找表来实现.由于这一部分是 32bit~32bit 的变换,若直接做成查找表,其大小将是 $2^{32}\times 32=16$GB,虽然达到了安全性,但实际中并不实用.因此,将整个变换分解为 4 个查找表.

将 $X=(x_{i0},x_{i1},x_{i2},x_{i3})$经 E_{ij} 和 S_{ij} 变换后的值记为$(z_{i0},z_{i1},z_{i2},z_{i3})$,将仿射变换 $Q_i\cdot L$ 拆分得:

$$(Q_i\cdot L)\begin{pmatrix}z_{i0}\\ z_{i1}\\ z_{i2}\\ z_{i3}\end{pmatrix}=(R_{i0}\ R_{i1}\ R_{i2}\ R_{i3})\cdot\begin{pmatrix}z_{i0}\\ z_{i1}\\ z_{i2}\\ z_{i3}\end{pmatrix}\oplus r_i=(R_{i0}\cdot z_{i0})\oplus(R_{i1}\cdot z_{i1})\oplus(R_{i2}\cdot z_{i2})\oplus(R_{i3}\cdot z_{i3}\oplus r_i)=v_{i0}\oplus v_{i1}\oplus v_{i2}\oplus v_{i3},$$

其中,R_i 为 32×8 的矩阵,r_i 为 32bit 做成的常数列向量.这样,即可获得 4 个 8bit~32bit 的查找表:$x_{ij}(\rightarrow z_{ij})\rightarrow v_{ij}$,$j$=0,1,2,3.每次计算时,查找对应的 4 个查找表,再将这 4 个结果异或即可获得本次变换的输出 Y.

Part 3:最后计算 X_{i+4}.由于 $X_{i+4}=X_i\oplus T(X_{i+1}\oplus X_{i+2}\oplus X_{i+3}\oplus rk_i)$,$T(X_{i+1}\oplus X_{i+2}\oplus X_{i+3}\oplus rk_i)$的值已经在 Part 2 中计算,并将其看做是本次变换中的输入.同理,仍然要对本次变换的输入/输出进行编码.由于采用了网络化编码的形式,所以输入编码要与上一个变换的输出编码相抵消,如图 4 所示.

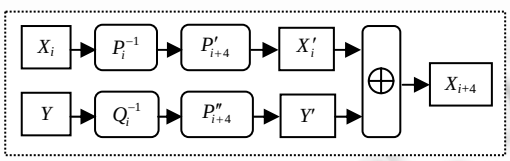

Fig.4 Part 3

图 4 Part 3

其中,$P'_{i+4}(x)=P_{i+4}(x)\oplus a'_{i+4}$,$P''_{i+4}(x)=P_{i+4}(x)\oplus a''_{i+4}$,它们均为 $GF(2)$上的 32bit~32bit 的仿射变换,将会与下一轮中对 X_{i+4} 的编码抵消一部分;而 Q_i^{-1} 则与 Part 2 中的 Q_i 相抵消.

保存 $P'_{i+4}\circ P_i^{-1}=D_i$,$P''_{i+4}\circ Q_i^{-1}=C_i$,则 X_{i+4} 的计算就由两个仿射变换和一个异或构成.

综上所述,白盒 SMS4 算法每一轮首先使用 3 个仿射变换 M_{i+j}^{i} 和两个异或计算,再查找 4 个 8bit~32bit 的查找表并计算 3 个异或,最后使用两个仿射变换 D_i,C_i 和一个异或计算即可.每个变换都经由仿射变换进行输入/输出编码,输入/输出编码采用网络化编码的形式.这样,密钥隐藏在查找表中,即使敌手可以读取内存,也无法获得密钥信息.

2 攻击方法

2.1 白盒攻击环境

Chow 在 AES 白盒实现[5]中提出了白盒攻击环境(white-box attack context,简称 WBAC)的概念,它假设:

(I) 充分享有特权的攻击软件与密码学软件共享一个主机,攻击软件对密码算法的执行完全可以访问;

(II) 动态执行(与某个固定的密钥一起)是可以被观测的;

(III) 密码算法的内部细节是完全可见和可任意更改的.

可以说,白盒攻击环境包括了逆向工程、旁信道攻击和恶意主机攻击等场景.在这个攻击环境中,攻击者除了可以像传统的黑盒攻击一样对整个算法的输入输出进行分析,还可以对程序运行进行二进制追踪;观察程序执行的中间结果;窃听对 CPU 的调用;任意的静态分析;注入数据改变子计算的结果等.在传统的 SMS4 算法中,整个客户端上运行的 SMS4 算法的程序都可以被攻击者观察并分析,因此,密钥很容易被直接观测到.而在文献[13]中的 SMS4 白盒实现方案,攻击者可见的仅仅只有 $P'_{i+4}\circ P_i^{-1}=D_i, P''_{i+4}\circ Q_i^{-1}=C_i, M_{i+j}^{i}=E_i^{-1}\circ P_{i+j}^{-1}$ 和若干个查找表, $E_i^{-1}, P_{i+j}^{-1}, P'_{i+4}, P_i^{-1}, P''_{i+4}, Q_i^{-1}$ 等均不可见.密钥被隐藏在查找表中,无法直接获得.

2.2 攻击具体步骤

我们利用类似 Billet[8]等人的分析方法对白盒 SMS4 进行分析,说明文献[13]中的白盒 SMS4 算法在白盒环境下无法达到安全性.基本思想是,通过将 part 2 和 part 3 与下一轮的 part 1 合成起来,消去中间网络化编码的部分,并经过差分分析、构造性证明、求解方程组等方法,逐一恢复用来混淆的仿射变换 E_{i+1}^{-1} 和 Q_i 的线性部分和常数项部分,从而进一步获得轮密钥.

首先,由于仿射变换由线性部分和常数项组成,为了方便起见,我们记仿射变换 P 的线性部分为 $l[P]$,常数项为 $c[P]$,则仿射变换 $P(x)=l[P](x)\oplus c[P]$.例如仿射变换 $P_{i+j}^{-1}(x)=l[P_{i+j}^{-1}](x)\oplus c[P_{i+j}^{-1}]$,其中, $l[P_{i+j}^{-1}]$ 为 32×32 的可逆矩阵, $c[P_{i+j}^{-1}]$ 为 32bit 的列向量.

其次,我们记 n 阶矩阵相乘的时间复杂度为 n^{ω}.在一般情况下,ω=3;而事实上,还有更好的算法可以使得 ω=2.376.因此,我们下面计算复杂度时取 ω=2.4.而 n 阶矩阵求逆的时间复杂度记为 n^3.

2.2.1 合成变换

由于每个 part 都使用随机仿射变换或者查找表来完成,其白盒含混度和白盒多样性决定了从单个的 Part 恢复密钥是困难的.但是,由于采用了网络化编码的方法,每个 Part 的输入编码都与其上一变换的输出编码完全抵消或者抵消线性的部分,所以,将前后有关联的 Part 2, Part 3 和下一轮的 Part 1 合并起来,可以获得 Part 1 的输出关于 Part 2 的输入的一个整齐的函数表达式,这一表达式将有助于我们求出 E_{i+1}^{-1} 的线性部分.

合并 Part 2, Part 3 和下一轮的 Part 1,如图 5 所示.

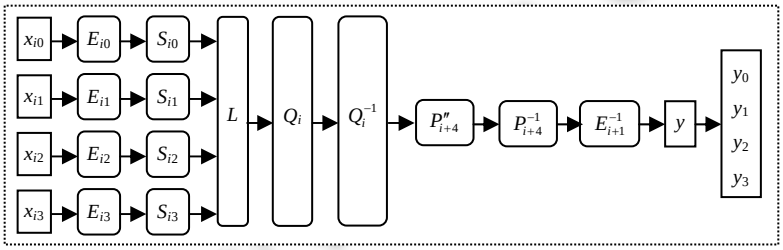

Fig.5 Combination transformation

图 5 合成变换

这里,为了方便分析,我们将 Part 3 中与 X'_i 的异或去掉.这是完全可以实现的,因为 X'_i 是由仿射变换 D_i 独立计算.

由图 5 可以看出, Q_i 与 Q_i^{-1} 互相抵消, P''_{i+4} 与 P_{i+4}^{-1} 抵消只剩下常数 $A_{i+4}^{-1}\cdot a''_{i+4}$.因此,经 L 输出数据的 x,事实上只需与常数 a''_{i+4} 异或后,再经过 E_{i+1}^{-1} 变换即得到 y,我们记为 $y=\tilde{E}_{i+1}(x)=E_{i+1}^{-1}(x\oplus A_{i+4}^{-1}\cdot a''_{i+4})$.

而 $E_{i+1}^{-1}=diag(E_{i+1,0}^{-1}, E_{i+1,1}^{-1}, E_{i+1,2}^{-1}, E_{i+1,3}^{-1})$,其中, $E_{i+1,j}^{-1}$ 为 $GF(2)$上的 8bit~8bit 的仿射变换,j=0,1,2,3.假设 $E_{i+1,j}^{-1}$ 的

线性部分为 $l[E_{i+1,j}^{-1}],\tilde{E}_{i+1}(x)$ 又可记为

$\tilde{E}_{i+1}(x)=E_{i+1}^{-1}(x\oplus A_{i+4}^{-1}\cdot a''_{i+4})=diag(l[E_{i+1,0}^{-1}](x_0)\oplus g_{i+1,0},l[E_{i+1,1}^{-1}](x_1)\oplus g_{i+1,1},l[E_{i+1,2}^{-1}](x_2)\oplus g_{i+1,2},l[E_{i+1,3}^{-1}](x_3)\oplus g_{i+1,3})$,

其中,$x=(x_0,x_1,x_2,x_3)$.

由于 L 可以看做是 32×32 的矩阵,所以可将 L 分成 16 个 8×8 的子块

$$L=\begin{pmatrix}L_{00} & L_{01} & L_{02} & L_{03}\\ L_{10} & L_{11} & L_{12} & L_{13}\\ L_{20} & L_{21} & L_{22} & L_{23}\\ L_{30} & L_{31} & L_{32} & L_{33}\end{pmatrix}.$$

那么,将图 5 的合成变换的输出记为 $y=(y_0\ y_1\ y_2\ y_3)$,则每个 y_j 均可以看做是输入$(x_{i0}\ x_{i1}\ x_{i2}\ x_{i3})$的函数

$$y_j(x_{i0},x_{i1},x_{i2},x_{i3})=l[E_{i+1,j}^{-1}]\left(\left\{\bigoplus_{t=0}^{3}L_{jt}\cdot S_{it}\cdot E_{it}(x_{it})\right\}\right)\oplus g_{i+1,j} \tag{1}$$

注:本节只需要理论推导,因此无计算复杂度.

2.2.2 恢复 E_{i+1}^{-1} 的线性部分

本节,我们利用上一节获得的输入输出关系式,以求解线性方程组为中间步骤,恢复出仿射变换 E_{i+1}^{-1} 的线性部分.在这之前,我们要先证明一个命题,说明任意的两个(y_j,y_r)之间存在仿射的关系,该命题对于恢复 E_{i+1}^{-1} 的线性部分具有重要的作用.

命题 1. 对于上述任意的$(y_j,y_r)(j,r\in 0,1,2,3)$,存在唯一的线性映射 A_{jr} 和常数 con_{jr},使得:

$\forall x_{i0}\in GF(2^8)$,有

$$y_j(x_{i0},0,0,0)=A_{jr}(y_r(x_{i0},0,0,0))\oplus con_{jr} \tag{2}$$

证明:由公式(1)式可知:

$$y_j(x_{i0},0,0,0)=l[E_{i+1,j}^{-1}](\{L_{j0}\cdot S_{i0}\cdot E_{i0}(x_{i0})\}\oplus\beta_j)\oplus g_{i+1,j},\text{其中},\beta_j=\bigoplus_{t=1}^{3}L_{jt}\cdot S_{it}\cdot E_{it}(0),$$

$$y_r(x_{i0},0,0,0)=l[E_{i+1,r}^{-1}](\{L_{r0}\cdot S_{i0}\cdot E_{i0}(x_{i0})\}\oplus\beta_r)\oplus g_{i+1,r},\text{其中},\beta_r=\bigoplus_{t=1}^{3}L_{rt}\cdot S_{it}\cdot E_{it}(0).$$

因此,取

$$con_{jr}=g_{i+1,j}\oplus l[E_{i+1,j}^{-1}](\beta_j)\oplus l[E_{i+1,j}^{-1}]\cdot L_{j0}\cdot L_{r0}^{-1}(\beta_r)\oplus l[E_{i+1,j}^{-1}]\cdot L_{j0}\cdot L_{r0}^{-1}\cdot(l[E_{i+1,r}^{-1}])^{-1}(g_{i+1,r}),$$

$$A_{jr}=l[E_{i+1,j}^{-1}]\cdot L_{j0}\cdot(L_{r0})^{-1}\cdot(l[E_{i+1,r}^{-1}])^{-1},$$

则等式(2)成立,这证明了 A_{jr} 和常数 con_{jr} 的存在性.

反之,假设有一个线性映射 A_{jr} 和常数 con_{jr} 使得等式(2)成立,则有

$$\left.\begin{aligned}
&y_j(x_{i0},0,0,0)=A_{jr}\left(y_r(x_{i0},0,0,0)\right)\oplus con_{jr}\\
&\Rightarrow l[E_{i+1,j}^{-1}](\{L_{j0}\cdot S_{i0}\cdot E_{i0}(x_{i0})\}\oplus\beta_j)\oplus g_{i+1,j}=A_{jr}(l[E_{i+1,r}^{-1}](\{L_{r0}\cdot S_{i0}\cdot E_{i0}(x_{i0})\}\oplus\beta_r)\oplus g_{i+1,r})\oplus con_{jr}\\
&\Rightarrow l[E_{i+1,j}^{-1}]\cdot L_{j0}\cdot S_{i0}\cdot E_{i0}(x_{i0})\oplus l[E_{i+1,j}^{-1}](\beta_j)\oplus g_{i+1,j}=A_{jr}\cdot l[E_{i+1,r}^{-1}]\cdot L_{r0}\cdot S_{i0}\cdot E_{i0}(x_{i0})\oplus\\
&\qquad A_{jr}\cdot l[E_{i+1,r}^{-1}](\beta_r)\oplus A_{jr}(g_{i+1,r})\oplus con_{jr}\\
&\Rightarrow l[E_{i+1,j}^{-1}]\cdot L_{j0}\cdot S_{i0}\cdot E_{i0}(x_{i0})\oplus A_{jr}\cdot l[E_{i+1,r}^{-1}]\cdot L_{r0}\cdot S_{i0}\cdot E_{i0}(x_{i0})=l[E_{i+1,j}^{-1}](\beta_j)\oplus g_{i+1,j}\oplus\\
&\qquad A_{jr}\cdot l[E_{i+1,r}^{-1}](\beta_r)\oplus A_{jr}(g_{i+1,r})\oplus con_{jr}\\
&\Rightarrow(l[E_{i+1,j}^{-1}]\cdot L_{j0}\oplus A_{jr}\cdot l[E_{i+1,r}^{-1}]\cdot L_{r0})\cdot(S_{i0}\cdot E_{i0})(x_{i0})=l[E_{i+1,j}^{-1}](\beta_j)\oplus g_{i+1,j}\oplus A_{jr}\cdot l[E_{i+1,r}^{-1}](\beta_r)\oplus A_{jr}(g_{i+1,r})\oplus con_{jr}
\end{aligned}\right\} \tag{3}$$

等式的右边是常数,所以 $(l[E_{i+1,j}^{-1}]\cdot L_{j0}\oplus A_{jr}\cdot l[E_{i+1,r}^{-1}]\cdot L_{r0})\cdot(S_{i0}\cdot E_{i0})$ 是个常数映射.由于 $S_{i0}\cdot E_{i0}$ 是一一映射,而 $(l[E_{i+1,j}^{-1}]\cdot L_{j0}\oplus A_{jr}\cdot l[E_{i+1,r}^{-1}]\cdot L_{r0})$ 是个线性映射,因此这个常数一定是 0.即

$$l[E_{i+1,j}^{-1}]\cdot L_{j0}\cdot S_{i0}\cdot E_{i0}\oplus A_{jr}\cdot l[E_{i+1,r}^{-1}]\cdot L_{r0}\cdot S_{i0}\cdot E_{i0}=0.$$

故 $A_{jr}=l[E_{i+1,j}^{-1}]\cdot L_{j0}\cdot(L_{r0})^{-1}\cdot(l[E_{i+1,r}^{-1}])^{-1}$.

将 $A_{jr}=l[E_{i+1,j}^{-1}]\cdot L_{j0}\cdot(L_{r0})^{-1}\cdot(l[E_{i+1,r}^{-1}])^{-1}$ 带入等式(3),即可得到

$$con_{jr}=g_{i+1,j}\oplus l[E_{i+1,j}^{-1}](\beta_j)\oplus l[E_{i+1,j}^{-1}]\cdot L_{j0}\cdot L_{r0}^{-1}(\beta_r)\oplus l[E_{i+1,j}^{-1}]\cdot L_{j0}\cdot L_{r0}^{-1}\cdot(l[E_{i+1,r}^{-1}])^{-1}(g_{i+1,r}).$$

从而证明了 A_{jr} 和常数 con_{jr} 的唯一性. □

确定 A_{jr}:事实上,线性映射 A_{jr} 可以看做 8×8 的矩阵,有 64 个未知项;常数 con_{jr} 有 8 个未知项;对于一对(y_j,y_r),将等式 $y_j(x_{i0},0,0,0)=A_{jr}(y_r(x_{i0},0,0,0))\oplus con_{jr}$ 展开可获得 8 个线性方程,则当 x_{i0} 取遍所有值时,共有 2^8 对(y_j,y_r),可获得 $2^8\times8$ 个方程.所以,A_{jr} 和 con_{jr} 中共 72 个未知项,可以组成 8 个含有 9 个未知数的方程组(每个方程组有 9 个方程),因此能以远远低于 8×9^3 的时间复杂度求出 A_{jr} 和 con_{jr}.

确定 $l[E_{i+1,r}^{-1}]$:再次应用命题 1,可得

$$y_r'(0,x_{i1},0,0)=A_{rj}'(y_j'(0,x_{i1},0,0))\oplus con_{rj}',$$

其中, $A_{ij}'=l[E_{i+1,r}^{-1}]\cdot L_{r1}\cdot(L_{j1})^{-1}\cdot(l[E_{i+1,j}^{-1}])^{-1}$,那么

$$\begin{aligned}A_{jr}\cdot A_{rj}'&=l[E_{i+1,j}^{-1}]\cdot L_{j0}\cdot(L_{r0})^{-1}\cdot(l[E_{i+1,r}^{-1}])^{-1}\cdot l[E_{i+1,r}^{-1}]\cdot L_{r1}\cdot(L_{j1})^{-1}\cdot(l[E_{i+1,j}^{-1}])^{-1}\\&=l[E_{i+1,j}^{-1}]\cdot L_{j0}\cdot(L_{r0})^{-1}\cdot L_{r1}\cdot(L_{j1})^{-1}\cdot(l[E_{i+1,j}^{-1}])^{-1}.\end{aligned}$$

A_{jr} 和 A_{rj}' 可以求出,而各分块矩阵 L_{ij} 均已知,因此可记为

$$A=\Sigma\cdot B\cdot\Sigma^{-1}\text{ 或 }A\cdot\Sigma=\Sigma\cdot B,$$

其中, $A=A_{jr}\cdot A_{rj}'$和$B=L_{j0}\cdot(L_{r0})^{-1}\cdot L_{r1}\cdot(L_{j1})^{-1}$均为 8×8 的已知矩阵,求出这些矩阵乘积的复杂度为 4 个 8 阶矩阵乘积的复杂度:$4\times8^{\omega}$; $\Sigma=l[E_{i+1,r}^{-1}]$ 为 8×8 的未知矩阵.由此,方程可获得关于 $\Sigma=l[E_{i+1,r}^{-1}]$ 的 64 个未知数的 64 个方程,因此, $\Sigma=l[E_{i+1,r}^{-1}]$ 能以低于 64^3 的时间复杂度求出(注意,这个方程与求解 A_{jr} 的方程不同,它只能以 64 阶的矩阵求解).因此,这里总的复杂度将远低于 $4\times8^{\omega}+64^3$.

$l[E_{i+1,j}^{-1}]$ 已经求出,即 $E_{i+1,j}^{-1}$ 的线性部分可以求出,经过 4 次相同的计算(j=0,1,2,3),即可以将 E_{i+1}^{-1} 的线性部分求出.

求出 A_{jr} 和 A_{rj}',时间复杂度将低于 $2\times8\times9^3$,因此, $l[E_{i+1,r}^{-1}]$ 能以低于 $2\times8\times9^3+4\times8^{\omega}+64^3$ 的时间复杂度而求出;而经 4 次计算才可获得 E_{i+1}^{-1} 的线性部分,故总的时间复杂度为 $4(2\times8\times9^3+4\times8^{\omega}+64^3)<2^{20}$.

2.2.3 获取 S 盒的输入差分

在白盒 SMS4 的设计中,S 盒中嵌入了密钥,并被仿射变换 E_i 混淆,因此无法直接获得 S 盒的输入差分.但是由于仿射变换后的差分与仿射变换的常数项无关,只要知道其线性部分即可获取差分,因此,我们根据已知的仿射变换和求出的 E_{i+1}^{-1} 的线性部分,分别计算出 E_i^{-1} 的线性部分和 Q_i 的线性部分,从而获得图 3 中 S 盒的输入差分.

按照第 2.2 节的方法,分别求出 $l[E_i^{-1}]$ 和 $l[E_{i+1}^{-1}]$.此外,我们可以获得以下结论:

结论 1. 若仿射变换 M_{i+4}^{i+1} 和 C_i 已知,那么 P_{i+4} 和 Q_i 的线性部分即已知.

证明:由第 1.2 节的 Part 1 所述,仿射变换 M_{i+4}^{i+1} 为已知,故 $l[M_{i+4}^{i+1}]$ 已知.

因 $M_{i+4}^{i+1}=E_{i+1}^{-1}\circ P_{i+4}^{-1}$,从而 $l[P_{i+4}^{-1}]=(l[E_{i+1}^{-1}])^{-1}\cdot l[M_{i+4}^{i+1}]$ 可求出,则 $l[P_{i+4}]=(l[P_{i+4}^{-1}])^{-1}$ 也可求出.

因 $P_{i+4}''(x)=P_{i+4}(x)\oplus a_i''$,则 $l[P_{i+4}'']=l[P_{i+4}]$ 也可求出.

又因为 C_i 已知,而 $P_{i+4}''\circ Q_i^{-1}=C_i$,从而 $l[Q_i^{-1}]=(l[P_{i+4}''])^{-1}\cdot l[C_i]$ 亦可求出,则 $l[Q_i]=(l[Q_i^{-1}])^{-1}$ 也可求出.

注:求出 $l[P_{i+4}^{-1}]$ 的复杂度为 8×8 矩阵求逆和 8×8 矩阵乘法的复杂度之和 8^3+8^{ω};同理,求出 $l[P_{i+4}]$的复杂度为 8^3,而求出 $l[Q_i^{-1}]$ 的复杂度为 8^{ω},求出 $l[Q_i]$的复杂度为 8^3. □

结论 2. 若 $l[E_i^{-1}]$ 已知,那么 $l[E_{i0}]$,$l[E_{i1}]$,$l[E_{i2}]$和 $l[E_{i3}]$均可求出.

证明:由于 $l[E_i^{-1}]$ 已经求出,故 $l[E_i]=(l[E_i^{-1}])^{-1}$ 也可求出,那么 $l[E_{i0}]$,$l[E_{i1}]$,$l[E_{i2}]$和 $l[E_{i3}]$均可获知. □

注:由于 $l[E_i^{-1}]$ 需要按照上一节的方法重新求出,其复杂度为 2^{20},求出 $l[E_{i0}]$,$l[E_{i1}]$,$l[E_{i2}]$和 $l[E_{i3}]$的复杂度为一

次 8×8 矩阵求逆的复杂度加上求 $l[E_i^{-1}]$ 的复杂度,为 $2^{20}+8^3\approx 2^{20}$.

此外,由于 $E_i=diag(E_{i0},E_{i1},E_{i2},E_{i3})$,那么根据第 1.2 节所述,图 3 的 Part 2 变换可以描述为如图 6 所示.

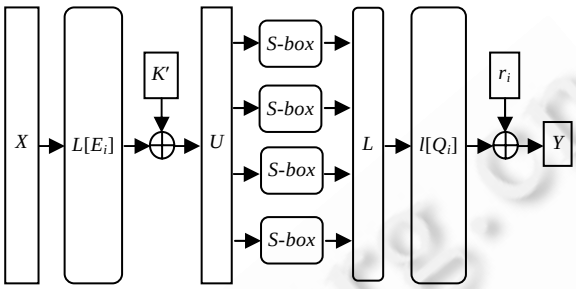

Fig.6 Variant of Part 2

图 6 Part 2 变形

其中,$K'=rk_i\oplus c[E_i]$,U 为 S 盒的输入.

那么,给定 X_1 和 X_2,根据 Part 2 所述,经查表和异或可得结果为 Y_1 和 Y_2.

由于 $l[E_i]$已知,故可计算 S 盒的输入差分

$$\Delta U=U_1\oplus U_2=l[E_i](X_1)\oplus l[E_i](X_2),$$

其对应的整个变换的输出差分为 $Y_1\oplus Y_2=\Delta Y$.此处为理论推导,无计算复杂度.

2.2.4 获取 Q_i 的常数项 r_i

由于上一节已经求出 Q_i 的线性部分 $l[Q_i]$、S 盒的输入差分和整个变换的输出差分的表达式,则可以根据差分分析法,首先获得 Q_i 的常数项 r_i,以便于在最后一节求出 E_i 的常数项.步骤如下:

第 1 步:在查找表中均匀随机地选出$(v_{i0},v_{i1},v_{i2},v_{i3})$和 $(v_{i0}^*,v_{i1}^*,v_{i2}^*,v_{i3}^*)$,则

$$Y=v_{i0}\oplus v_{i1}\oplus v_{i2}\oplus v_{i3},Y^*=v_{i0}^*\oplus v_{i1}^*\oplus v_{i2}^*\oplus v_{i3}^*.$$

对$(v_{i0},v_{i1},v_{i2},v_{i3})$和 $(v_{i0}^*,v_{i1}^*,v_{i2}^*,v_{i3}^*)$ 反向查表得对应的明文$(x_{i0},x_{i1},x_{i2},x_{i3})$和 $(x_{i0}^*,x_{i1}^*,x_{i2}^*,x_{i3}^*)$,则

$$\begin{aligned}\Delta U&=U\oplus U^*\\&=l[E_i](X)\oplus l[E_i](X^*)\\&=(l[E_{i0}](x_{i0})\|l[E_{i1}](x_{i1})\|l[E_{i2}](x_{i2})\|l[E_{i3}](x_{i3}))\oplus(l[E_{i0}](x_{i0}^*)\|l[E_{i1}](x_{i1}^*)\|l[E_{i2}](x_{i2}^*)\|l[E_{i3}](x_{i3}^*)).\end{aligned}$$

这里,$l[E_{ij}](x_{ij})(l[E_{ij}](x_{ij}^*))$ 是 8×8 矩阵与 8×1 矩阵的乘积,共有 8 个这样的计算,故时间复杂度上界为 $8\times 8^2=8^3$.

第 2 步:对 2^{32} 个可能值 r_i 设置 2^{32} 个计数器Λ_j,$1\leqslant j\leqslant 2^{32}$,用每个 r_i 解密 Y 和 Y^*,并计算:

$$\begin{aligned}U'=&S_{box}^{-1}\{L^{-1}\cdot(l[Q_i])^{-1}(Y\oplus r_i)\}_0\|S_{box}^{-1}\{L^{-1}\cdot(l[Q_i])^{-1}(Y\oplus r_i)\}_1\|\\&S_{box}^{-1}\{L^{-1}\cdot(l[Q_i])^{-1}(Y\oplus r_i)\}_2\|S_{box}^{-1}\{L^{-1}\cdot(l[Q_i])^{-1}(Y\oplus r_i)\}_3,\\U'^*=&S_{box}^{-1}\{L^{-1}\cdot(l[Q_i])^{-1}(Y^*\oplus r_i)\}_0\|S_{box}^{-1}\{L^{-1}\cdot(l[Q_i])^{-1}(Y^*\oplus r_i)\}_1\|\\&S_{box}^{-1}\{L^{-1}\cdot(l[Q_i])^{-1}(Y^*\oplus r_i)\}_2\|S_{box}^{-1}\{L^{-1}\cdot(l[Q_i])^{-1}(Y^*\oplus r_i)\}_3,\end{aligned}$$

其中,$\{L^{-1}\cdot(l[Q_i])^{-1}(Y\oplus r_i)\}_0$ 表示$\{L^{-1}\cdot(l[Q_i])^{-1}(Y\oplus r_i)\}$的第 1 个字节,以此类推;$S_{box}^{-1}$ 表示 S 盒逆向查表.

验证 $\Delta U\overset{?}{=}U'\oplus U'^*$,如果相等,则计数器$\Lambda_j$ 加 1.

这里,可以先将$\{L^{-1}\cdot(l[Q_i])^{-1}(Y\oplus r_i)\}$中 1 个 32 阶矩阵乘法 $L^{-1}\cdot(l[Q_i])^{-1}$ 计算出来,其复杂度为 32^{ω},再计算$\{L^{-1}\cdot(l[Q_i])^{-1}(Y\oplus r_i)\}$和$\{L^{-1}\cdot(l[Q_i])^{-1}(Y^*\oplus r_i)\}$,其复杂度为 2×32^2.忽略异或运算和 S 盒查表,因此,总的复杂度为 $32^{\omega}+2\times 32^2$.

第 3 步:重复第 1 步、第 2 步,直到某个计数器的值明显高于其他计数器.那么,这个计数器所对应的 r_i 即为所求.

总的来说,本节找出 r_i 的计算复杂度为 $2^{32}(8^3+32^{\omega}+2\times 32^2)+2^{20}<2^{46}$,其中,$2^{20}$ 为求出 $l[E_{i0}]$,$l[E_{i1}]$,$l[E_{i2}]$和 $l[E_{i3}]$

的复杂度.所以,这里总的复杂度上界为 2^{46}.

2.2.5 恢复 $\tilde{E}_{i+1}$ 的常数项 g_{i+1}

在第 2.1 节,合并 Part 2, Part 3 和下一轮的 Part 1 后,我们获得新的仿射变换 $\tilde{E}_{i+1}(x)$,记它的常数项为 g_{i+1}.本节求出此常数项,它将有助于我们在最后一节计算 E_i 的常数项.在此之前,需要先证明一个命题,此命题是计算 g_{i+1} 的重要依据.

由于 $S_{ij}(*)=S_{box}(*\oplus rk_{ij})$,所以映射 $x \to S_{box}^{-1} \circ S_{ij} \circ E_{ij}(x) = E_{ij}(x) \oplus rk_{ij}$ 是一个仿射变换,其中,rk_{ij} 是第 i 轮密钥的第 j 个字节.由此,我们得到以下命题:

命题 2. 存在唯一的 $GF(2^8)$中的元素对$(\delta_i,\gamma_i)_{i=0,\ldots,3}$,$\delta_i$ 不为 0(Λ_δ表示乘以δ),使得

$$\tilde{P}_0 : x \to (S^{-1} \circ \Lambda_{\delta_0} \circ l[E_{i+1,0}]) \ (y_0(x,\text{‘00’},\text{‘00’},\text{‘00’}) \oplus \gamma_0)$$

$$\tilde{P}_1 : x \to (S^{-1} \circ \Lambda_{\delta_1} \circ l[E_{i+1,0}]) \ (y_0(\text{‘00’},x,\text{‘00’},\text{‘00’}) \oplus \gamma_1)$$

$$\tilde{P}_2 : x \to (S^{-1} \circ \Lambda_{\delta_2} \circ l[E_{i+1,0}]) \ (y_0(\text{‘00’},\text{‘00’},x,\text{‘00’}) \oplus \gamma_2)$$

$$\tilde{P}_3 : x \to (S^{-1} \circ \Lambda_{\delta_3} \circ l[E_{i+1,0}]) \ (y_0(\text{‘00’},\text{‘00’},\text{‘00’},x) \oplus \gamma_3)$$

均为仿射变换.更进一步,这些仿射变换实际上即为 $\tilde{P}_j(x) = E_{ij}(x) \oplus rk_{ij}$.

证明:以 $\tilde{P}_0$ 为例,将 $y_0(x,\text{‘00’},\text{‘00’},\text{‘00’})$带入,化简即得 $x \to S^{-1}\{\delta \cdot S(E_{i0}(x) \oplus rk_{i0})\} \oplus \gamma$,其中,

$\delta=\delta_0 \cdot L_{00}$,$\gamma=\delta_0 \cdot \{\beta_0 \oplus l[E_{i+1,0}](g_{i+1,0}) \oplus l[E_{i+1,0}](\gamma_0)\}$,$\beta_0=L_{01} \cdot S_{i1} \circ E_{i1}(\text{‘00’}) \oplus L_{02} \cdot S_{i2} \circ E_{i2}(\text{‘00’}) \oplus L_{03} \cdot S_{i3} \circ E_{i3}(\text{‘00’})$.

由于 δ_i 不为 0,故 $\delta=\delta_0 \cdot L_{00}$ 不为 0,而 S 表示 SMS4 的 S 盒,要使得 $x \to S^{-1}\{\delta \cdot S(E_{i0}(x) \oplus rk_{i0})\} \oplus \gamma$ 是一个非常数的仿射变换,唯一的可能即是$(\delta,\gamma)=(\text{‘01’},\text{‘00’})$.即证明了$(\delta_0,\gamma_0)$的存在性和唯一性.

同时,由于$(\delta,\gamma)=(\text{‘01’},\text{‘00’})$,即知 $\tilde{P}_0(x) = E_{i0}(x) \oplus rk_{i0}$.同理可证 $\tilde{P}_1,\tilde{P}_2,\tilde{P}_3$. □

确定 δ_i:因为 $\delta=\delta_0 \cdot L_{00}=\text{‘01’}$,而 L_{00} 已知,故 δ_0 可以确定,其复杂度为求解一个 8 元线性方程组的复杂度 8^3.同理可求出 $\delta_1,\delta_2,\delta_3$.

确定 γ_i:而对于所有可能的 2^8 个 γ_0,我们检验其对应的映射

$$\tilde{P}_0 : x \to (S^{-1} \circ \Lambda_{\delta_0} \circ l[E_{i+1,0}]) \ (y_0(x,\text{‘00’},\text{‘00’},\text{‘00’}) \oplus \gamma_0)$$

是否为仿射变换,即可找出唯一的那个 γ_0.这需要利用查找表对所有可能的 x 依次查询,在取定 γ_0 的情况下验证 $(x,\tilde{P}_0(x))$ 是否对应着仿射变换,或者说求出此仿射变换的线性部分和常数项,依然是 8 个含有 9 个未知数的方程组.因此,获取 γ_0 的总的时间复杂度不超过 $2^8 \times 8 \times 9^3 + 8^3 < 2^{21}$,其中,$8^3$ 为求解其中 δ_0 的复杂度.同理可求出 $\gamma_1,\gamma_2,\gamma_3$.

恢复 g_{i+1}:由于 $\gamma=\text{‘00’}$,因此,根据 $\gamma=\delta_0 \cdot \{\beta_0 \oplus l[E_{i+1,0}](g_{i+1,0}) \oplus l[E_{i+1,0}](\gamma_0)\}$,即

$$\beta_0 \oplus l[E_{i+1,0}](g_{i+1,0}) \oplus l[E_{i+1,0}](\gamma_0) = \text{‘00’}$$

$$\Rightarrow \gamma_0 = l[E_{i+1,0}^{-1}]\{L_{00} \cdot S_{i0} \circ E_{i0}(x) \oplus \beta_0\} \oplus l[E_{i+1,0}^{-1}](L_{00} \cdot S_{i0} \circ E_{i0}(x)) \oplus g_{i+1,0}$$

$$\Rightarrow \gamma_0 = y_0(x,\text{‘00’},\text{‘00’},\text{‘00’}) \oplus l[E_{i+1,0}^{-1}](L_{00} \cdot S_{i0} \circ E_{i0}(x)).$$

同理可得:

$$\gamma_1 = y_0(\text{‘00’},x,\text{‘00’},\text{‘00’}) \oplus l[E_{i+1,0}^{-1}](L_{01} \cdot S_{i1} \circ E_{i1}(x)),$$

$$\gamma_2 = y_0(\text{‘00’},\text{‘00’},x,\text{‘00’}) \oplus l[E_{i+1,0}^{-1}](L_{02} \cdot S_{i2} \circ E_{i2}(x)),$$

$$\gamma_3 = y_0(\text{‘00’},\text{‘00’},\text{‘00’},x) \oplus l[E_{i+1,0}^{-1}](L_{03} \cdot S_{i3} \circ E_{i3}(x)).$$

我们令 $\gamma_4=y_0(\text{‘00’},\text{‘00’},\text{‘00’},\text{‘00’})$,亦可记为

$$\gamma_4 = l[E_{i+1,0}^{-1}]\left(\bigoplus_{t=0}^{3} L_{0t} \cdot S_{it} \circ E_{it}(\text{‘00’})\right) \oplus g_{i+1,0},$$

则 $g_{i+1,0}=\gamma_0 \oplus \gamma_1 \oplus \gamma_2 \oplus \gamma_3 \oplus \gamma_4$,从而恢复出 $g_{i+1,0}$.由于 γ_4 仅需要查表,不需要计算,因此,计算 $g_{i+1,0}$ 的复杂度为求出 $\gamma_0,\gamma_1,\gamma_2,\gamma_3$ 的复杂度 4×2^{21};同理可得 $g_{i+1,1},g_{i+1,2}$ 和 $g_{i+1,3}$,从而全部恢复 $\tilde{E}_{i+1}$ 的常数项 g_{i+1},其时间复杂度上界为 $4 \times 4 \times 2^{21} = 2^{25}$.

2.2.6 确定子密钥 rk_i

在图 3 的 Part 2 中,L 已知、S 盒是公开的、X 和 Y 的对应关系可以通过查表获得;而同时,E_i 的线性部分 $l[E_{i0}]$, $l[E_{i1}]$,$l[E_{i2}]$和 $l[E_{i3}]$以及 Q_i 的线性部分已在第 2.3 节求出,Q_i 的常数项 r_i 在第 2.4 节求出,即只需要再求出 E_i 的常数项,即可完全恢复图 3 的 Part 2 中的两个仿射变换 Q_i 和 E_i,从而找到这一轮所对应的轮密钥 rk_i.

首先,由第 1.2 节的 Part 1 可知,$P_{i+4}(x)=A_{i+4}(x)\oplus a_{i+4}$,故 $P_{i+4}^{-1}(x)=A_{i+4}^{-1}(x)\oplus A_{i+4}^{-1}\cdot a_{i+4}$.而由第 2.3 节可知, $l[P_{i+4}^{-1}]$ 可求出,即 A_{i+4}^{-1} 已知,则 A_{i+4} 亦已知.由于 $Q_i(x)=l[Q_i](x)\oplus r_i$($l[Q_i]$和 $l[Q_i]^{-1}$ 亦已在第 2.3 节求出,r_i 在第 2.4 节求出), 则 $Q_i^{-1}(x)=l[Q_i]^{-1}(x)\oplus l[Q_i]^{-1}\cdot r_i$,而 $P''_{i+4}(x)=P_{i+4}(x)\oplus a''_{i+4}=A_{i+4}\oplus a_{i+4}\oplus a''_{i+4}$.

下面,我们假设 $E_{i+1}^{-1}(x)=l[E_{i+1}^{-1}](x)\oplus e_{i+1}$:

(I) 由于仿射变换 $M_{i+4}^{i+1}=E_{i+1}^{-1}\circ P_{i+4}^{-1}$,$M_{i+3}^{i+1}=E_{i+1}^{-1}\circ P_{i+3}^{-1}$ 和 $M_{i+3}^{i}=E_{i}^{-1}\circ P_{i+3}^{-1}$ 已知,故 M_{i+4}^{i+1},M_{i+3}^{i+1} 和 M_{i+3}^{i} 的常数项已知,故有:

$$c[M_{i+4}^{i+1}]=l[E_{i+1}^{-1}]\cdot A_{i+4}^{-1}\cdot a_{i+4}\oplus e_{i+1} \quad ①$$

$$c[M_{i+3}^{i+1}]=l[E_{i+1}^{-1}]\cdot A_{i+3}^{-1}\cdot a_{i+3}\oplus e_{i+1} \quad ②$$

$$c[M_{i+3}^{i}]=l[E_{i}^{-1}]\cdot A_{i+3}^{-1}\cdot a_{i+3}\oplus e_{i} \quad ③$$

(II) 同时,$\tilde{E}_{i+1}(x)=E_{i+1}^{-1}(x\oplus a''_{i+4})$ 的常数项 g_{i+1} 已知,故有

$$g_{i+1}=l[E_{i+1}^{-1}]\cdot A_{i+4}^{-1}\cdot a''_{i+4}\oplus e_{i+1} \quad ④$$

(III) 此外,由于 Q_i 已经完全恢复,所以在 Part 2,我们选定一个 32bit 的 X_0,通过查找表和计算可以获得一个 32bit 的 Y_0,由此可以获得等式

$$\tau^{-1}(L^{-1}(Q_i(Y_0)))=l[E_i](X_0)\oplus e_i\oplus rk_i \quad ⑤$$

其中,τ^{-1} 表示反查 SMS4 的 S 盒.

(IV) 最后,由于 Part 1 中的 3 个分支是可以分别计算的,所以取其中一个分支与 Part 2 合并,如图 7 所示.

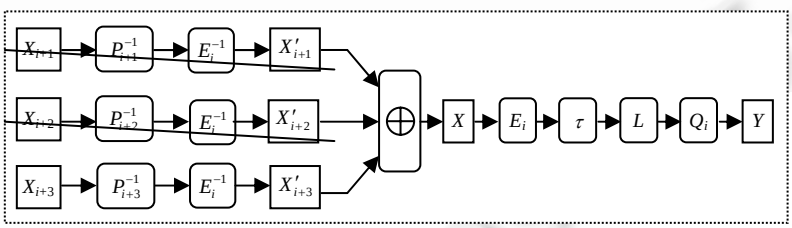

Fig.7 Combination of Part 1 and Part 2

图 7 合并 Part 1 和 Part 2

我们选定一个 32bit 的 $\tilde{X}_{i+3}$,通过 Part 1 和 Part 2 的计算可以获得一个 32bit 的 $\tilde{Y}$,由此可以获得等式:

$$A_{i+3}^{-1}(\tilde{X}_{i+3})\oplus A_{i+3}^{-1}\cdot a_{i+3}\oplus rk_i=\tau^{-1}(L^{-1}(Q_i(\tilde{Y}))) \quad ⑥$$

将以上 6 个等式联立,有 6 个未知数 $a_{i+4}, a''_{i+4}, e_{i+1}, a_{i+3}, e_i, rk_i$,其系数矩阵为非奇异矩阵.故通过求解 32 阶矩阵方程组即可解得 rk_i,从而恢复出第 i 轮的轮密钥,其时间复杂度远远低于$(32^{\omega}+32^3+4(32^3+32^{\omega}))<2^{18}$.

3 攻击复杂度

纵观整个获取密钥的过程,第 2.2.1 节将两轮中的 3 个部分合成起来,获得 y_j 关于输入$(x_{i0}\ x_{i1}\ x_{i2}\ x_{i3})$的代数式,本节主要是理论推导.第 2.2.2 节通过以上代数式证明了命题 1,并以最多 2^{20} 的时间复杂度求出 E_{i+1}^{-1} 的线性部分 $l[E_{i+1}^{-1}]$.第 2.2.3 节以复杂度 8^3+8^{ω}求出 $l[P_{i+4}^{-1}]$,即是第 2.2.6 节所述的 A_{i+4}^{-1} ;以复杂度 8^3 求出 $l[P_{i+4}]$,即是第 2.2.6 节所述的 A_{i+4};以复杂度 8^{ω}和复杂度 8^3 分别求出 $l[Q_i^{-1}]$ 和 $l[Q_i]$;并通过已经求出的线性部分获取 S 盒的输入差

分.第 2.2.4 节根据 S 盒的输入差分,以时间复杂度 2^{46} 求出 Q_i 的常数项 r_i.第 2.2.5 节以时间复杂度 2^{25} 和相关理论推导求出 $\tilde{E}_{i+1}$ 的常数项 g_{i+1}.第 2.2.6 节最终以时间复杂度 2^{18} 恢复出密钥 rk_i.

因此,整个攻击过程的时间复杂度上界为

$$2^{20}+(8^{\omega}+8^3)+8^{\omega}+(8^{\omega}+8^3)+2^{46}+2^{25}+2^{18}<2^{47}.$$

而事实上,在整个攻击中,参与运算的均为 $GF(2)$上的矩阵,其元素为 0 或者 1,因此在实际乘法操作时,都只执行的是逻辑“与”运算;同时,我们在计算复杂度时考虑的仅仅是普通算法,其时间复杂度都并不是最优的.因此,实际的时间复杂度将远远低于 2^{47}.

4 结束语

本文针对文献[13]中的白盒 SMS4 实现进行分析,将 BGE 攻击法、差分分析法以及其他方法相结合,以较低的复杂度恢复出轮密钥,证明了该白盒实现方案并不满足所需的安全性.文献[13]中所提出的白盒 SMS4 实现方案,采用了 Chow 等人所提出的查找表的方法,并对 part 1,part 2 和 part 3 的输入输出采用了网络化编码.从目前看来,网络化编码是本文攻击成功的首要条件(事实上也是 Chow 等人的方案被攻击的首要条件),它可以在合并两个 Part 的时候消去部分仿射变换,减小插入的随机仿射对原 SMS4 算法的影响,并为列出若干等式提供条件.此外,方案中的仿射变换也是本文攻击成功的前提之一,由于仿射变换可以用矩阵和向量来表示,因此可以通过观察输入/输出建立线性方程组,从而解出所需的未知量.因此可以说,仿射变换是威胁方案安全性的最重要的原因.在白盒密码实现算法的设计中,应该要避免以上两种方式同时出现.而只避免其中一种方式是否能保证安全性,以及是否有其他方法可以构造出安全的白盒实现,将是我们下一步探讨的内容.

SMS4 是国家商用密码算法,其目标是保证我国电子商务的安全性.但是对于越来越恶劣的网络使用环境,SMS4 算法的传统使用方式将会逐渐显现出劣势,容易遭受到各种多元化的攻击,这将不利于我国信息安全的稳定和发展.因此,探讨实用的白盒 SMS4 能够防止在最坏环境下对 SMS4 算法的破坏,对于更高层次上保证电子商务的安全性具有长远的意义.

References:

[1] Anderson R, Kuhn M. Low cost attacks on tamper-resistant devices. In: Proc. of the 5th Int'l Workshop on Security Protocols. LNCS 1361, Springer-Verlag, 1997. 125−136. [doi: 10.1007/BFb0028165]

[2] Biham E, Shamir A. Differential fault analysis of secret key cryptosystems. In: Proc. of the 17th Annual Int'l Cryptology Conf. on Advances in Cryptology. New York, 1997. 513−525. [doi: 10.1007/BFb0052259]

[3] Biham E, Shamir A. Power analysis of the key scheduling of the AES candidates. In: Proc. of the 2nd AES Candidate Conf. Rome, 1999. 22−23.

[4] Boneh D, DeMillo RA, Lipton RJ. On the importance of eliminating errors in cryptographic computations. Journal of Cryptology, 2001,14(2):101−119. [doi: 10.1007/s001450010016]

[5] Chow S, Eisen P, Johnson H, Van Oorschot PC. White-Box cryptography and an AES implementation. In: Proc. of the Selected Areas in Cryptography. LNCS 2595, Newfoundland: Springer-Verlag, 2002. 250−270. [doi: 10.1007/3-540-36492-7_17]

[6] Chow S, Eisen P, Johnson H, Van Oorschot PC. A white-box DES implementation for DRM applications. In: Proc. of the ACM Workshop on Security and Privacy in Digital Rights Management. LNCS 2692, Heidelberg: Springer-Verlag, 2002. 1−15. [doi: 10.1007/978-3-540-44993-5_1]

[7] Jacob M, Boneh D, Felten E. Attacking an obfuscated cipher by injecting faults. In: Proc. of the ACM Workshop on Security and Privacy in Digital Rights Management. LNCS 2696, Heidelberg: Springer-Verlag, 2002. 16−31. [doi: 10.1007/978-3-540-44993-5_2]

[8] Link HE, Neumann WD. Clarifying obfuscation: Improving the security of white-box DES. In: Proc. of the Int'l Conf. on Information Technology: Coding and Computing. Washington: IEEE Computer Society, 2005. 679−684. [doi: 10.1109/ITCC.2005.100]

[9] Wyseur B, Michiels W, Gorissen P, Preneel B. Cryptanalysis of white-box DES implementations with arbitrary external encodings. In: Proc. of the Selected Areas in Cryptography. LNCS 4876, Ottawa: Springer-Verlag, 2007. 264–277. [doi: 10.1007/978-3-540-77360-3_17]

[10] Goubin L, Masereel JM, Quisquater M. Cryptanalysis of white box DES implementations. In: Proc. of the Selected Areas in Cryptography. LNCS 4876, Ottawa: Springer-Verlag, 2007. 278–295. [doi: 10.1007/978-3-540-77360-3_18]

[11] Billet.O, Gilbert.H, Ech-Chatbi C. Cryptanalysis of a white box AES implementation. In: Proc. of the Selected Areas in Cryptography. LNCS 3357, Berlin, Heidelberg: Springer-Verlag, 2005. 227–240. [doi: 10.1007/978-3-540-30564-4_16]

[12] Michiels W, Gorissen P, Hollmann HDL. Cryptanalysis of a generic class of white-box implementations. In: Proc. of the Selected Areas in Cryptography. LNCS 5381, Berlin, Heidelberg: Springer-Verlag, 2009. 414–428. [doi: 10.1007/978-3-642-04159-4_27]

[13] Xiao YY, Lai XJ. White-Box cryptography and implementations of SMS4. In: Proc. of the 2009 CACR Annual Meeting. Beijing: Science Press, 2009. 24–34 (in Chinese with English abstract).

[14] The Office of Security Commercial Code Administration (OSCCA). SMS4 cryptographic algorithms used by wireless LAN products. Beijing, 2006 (in Chinese).

附中文参考文献:

[13] 肖雅莹,来学嘉.白盒密码及 SMS4 算法的白盒实现.见:中国密码学会 2009 年会论文集.北京:科学出版社,2009.24–34.

[14] 国家商用密码管理办公室.无线局域网产品使用的 SMS4 密码算法.北京,2006.

林婷婷(1982－),女,四川雅安人,博士,讲师,主要研究领域为密码学与信息安全,密码技术应用.

E-mail: lintingting00@163.com

来学嘉(1954－),男,博士,教授,博士生导师,主要研究领域为密码学与信息安全,密码技术应用.

E-mail: lai-xj@cs.sjtu.edu.cn