密码学报 ISSN 2095-7025 CN 10-1195/TN
Journal of Cryptologic Research, 2018, 5(6): 651–670
©《密码学报》编辑部版权所有.
E-mail: jcr@cacrnet.org.cn
http://www.jcr.cacrnet.org.cn
Tel/Fax: +86-10-82789618

对两个 SM4 白盒方案的分析*

潘文伦, 秦体红, 贾 音, 张立廷

摩石实验室, 成都卫士通信息产业股份有限公司, 北京 100070
通信作者: 潘文伦, E-mail: pan.wenlun@westone.com.cn

摘 要: 传统密码算法在设计时并未考虑算法运行平台的安全风险. Chow 等在 2002 年提出了白盒攻击模型, 假定攻击者具有完全控制算法运行过程的能力, 可以获取算法的运行状态、更改算法运行的中间值等. 此模型更符合密码设备在失控环境下的应用情况, 因为一个合法的用户也可能变为一个潜在的攻击者. 在这种环境下, 传统攻击模型中设计的密码算法将不再安全. 如何保护密码算法在白盒环境下的安全性, 在数字版权保护、移动终端安全等领域具有强烈的现实需求. Chow 等使用混淆与查找表等方式设计了 AES、DES 白盒方案, 肖雅莹等在 2009 年使用类似方法设计了 SM4 算法的白盒方案 (肖-来方案), 白鲲鹏等进一步通过复杂化内部解码编码过程以及引入更多随机数的方式设计了一个新的 SM4 白盒方案 (白-武方案). 本文分析了这两个 SM4 白盒方案. 首先指出林婷婷等对肖-来方案分析的复杂度计算存在偏差 (林-来分析). 具体来讲, 该分析中唯一确定了编码矩阵及仿射常数, 而实质上根据该分析方法, 编码矩阵与仿射常数存在 $61200 \cdot 2^{32}$ 种可能取值. 进一步地, 我们改进了林-来的分析方法, 通过调整仿射常数的恢复顺序, 大幅降低了计算复杂度. 如恢复查找表外部编码的仿射常数时, 我们通过搜索等价密钥再确定仿射常数的方式只需不超过 2^{10} 次查表运算就可确定该仿射常数, 而林-来分析中获取该仿射常数的计算复杂度为 2^{46}. 同时, 我们提出了首个针对白-武方案的第三方分析, 指出其密钥和外部编码的取值空间大小为 $61200 \cdot 2^{128}$. 我们的分析表明, 肖-来、白-武方案的安全性主要依赖外部编码中仿射常数的安全性. 两个方案的线性变换部分对安全性的影响有限, 且复杂化内部编码解码过程并不能有效提高线性变换的安全性. 另外, 通过对仿射矩阵或仿射常数进行拆分来增大白盒多样性的策略只会增大白盒方案的实现难度, 而对方案的安全性并无明显加强. 这一系列发现将对白盒密码的分析与设计提供借鉴作用.

关键词: 白盒密码; SM4; 查找表; 仿射变换

中图分类号: TP309.7 **文献标识码:** A DOI: 10.13868/j.cnki.jcr.000274

中文引用格式: 潘文伦, 秦体红, 贾音, 张立廷. 对两个 SM4 白盒方案的分析[J]. 密码学报, 2018, 5(6): 651–670.
英文引用格式: PAN W L, QIN T H, JIA Y, ZHANG L T. Cryptanalysis of two white-box SM4 implementations[J]. Journal of Cryptologic Research, 2018, 5(6): 651–670.

Cryptanalysis of Two White-box SM4 Implementations

PAN Wen-Lun, QIN Ti-Hong, JIA Yin, ZHANG Li-Ting

Westone Cryptologic Research Center, Westone Information Industry Inc. Beijing 100070, China
Corresponding author: PAN Wen-Lun, E-mail: pan.wenlun@westone.com.cn

Abstract: Traditional cryptographic algorithms are designed to be secure without considering their

* 基金项目: 国家重点研发计划 (2017YFB0802000); 国家自然科学基金 (61572484)
Foundation: National Key Research and Development Program of China (2017YFB0802000); National Natural Science Foundation of China (61572484)
收稿日期: 2017-10-30 定稿日期: 2018-03-23

platform risks. In 2002, Chow et al. introduced the white-box context, assuming attackers have full control over the execution of a cryptographic algorithm. Attackers can thus observe the algorithm internal states and even modify these values. This white-box model is very practical when evaluating cryptographic devices which are used in untrusted environments, where a legitimate user may be a potential attacker. Then, traditional cryptographic algorithms become no longer secure in such a context, and how to protect their security is strongly required in practice, e.g. digital rights protection and mobile device security. Taking advantage of obfuscation operations and lookup-tables, Chow et al. constructed a white-box implementation of AES and DES. Similarly, Xiao et al. constructed a white-box implementation of SM4 and Bai et al. proposed another white-box SM4 by more complex inner encodings/decodings and random numbers. This paper analyzes these two white-box implementations of SM4. It first points out a flaw in Lin's analysis on Xiao's white-box SM4. In counting encoding matrixes and affine constants, the real value should be $61\ 200 \cdot 2^{32}$ instead of only one in their analysis. Then an improvement of Lin's method is given. By adjusting the order of recovering affine constants, the improved method greatly reduces the attacking complexity. For example, in recovering the affine constants of external encoding of look-up tables, the new method needs no more than 2^{10} look-ups by first searching equivalent keys and then determining affine constants. This is far less than 2^{46} in Lin's analysis. This study also proposes a third-party analysis on Bai's white-box SM4, and points out that the size of keys and external encodings is $61\ 200 \cdot 2^{128}$. Analysis shows that, both of the two white-box SM4 rely on the security of their affine constants in external encodings, the linear transformation parts contribute very little to security, and simply complicating internal encodings/decodings. Furthermore, at the expense of implementation hardness to increase diversity, the method of splitting affine matrixes or constants cannot efficiently improve security. All these findings will be useful to analyze and design white-box ciphers.

Key words: white-box cryptography; SM4; lookup-tables; Affine transformation

1 引言

常用的标准密码算法, 如 DES、AES、SM4 等在设计时并未考虑算法运行终端的安全风险. 然而密码算法在实际使用时可能运行在不安全的平台上, 攻击者通过观察算法运行的过程, 提取算法运行产生的中间值可恢复密钥, 从而可绕开黑盒模型下分析算法整体的难题. 例如在数字版权保护中, 内容提供商向用户提供文本、音频、视频等收费数字内容服务, 用户在客户端解密数字内容. 恶意用户在客户端可监控算法运行过程, 恢复密钥并非法传播从而造成内容提供商的损失. 基于此观察, Chow 等在 2002 年提出白盒攻击模型 [1], 该模型假设攻击者对算法的运行过程具有完全的控制权, 即可以动态观测算法运行过程、任意修改算法运行的中间值等. 在此攻击模型下, 传统密码算法的标准实现方式不堪一击, 构造安全的白盒密码方案具有很大的现实意义.

Chow 等在提出白盒攻击模型时, 通过将若干步骤组合起来所构成的函数用查找表表示 (即遍历函数的输入, 存储函数的输出值构成查找表, 从而可隐藏该函数的内部信息), 将密钥嵌入到查找表并使用随机双射 (Chow 等称随机双射为编码) 来保护查找表, 开创性地构造了首个白盒 AES 方案 [1] 及首个白盒 DES 方案 [2]. 因查找表构造方式的多样性, 局部破解 Chow 等的白盒方案中的单个查找表极其困难. 然而, Billet 等 [3] 在 2004 年提出一种有效的攻击, 能在 2^{30} 复杂度内恢复 Chow 等的白盒 AES 方案中的密钥, 随后, Tolhuizen [4]、Lepoint [5]、Mulder [6] 等逐步将复杂度降低到 2^{22}. AES、DES 白盒方案的设计 [7–11] 与分析 [12–17] 相互促进了白盒密码研究的进步.

SM4 (原 SMS4) 分组密码算法于 2006 年由国家密码管理办公室正式发布, 并于 2012 年成为密码行业标准 GM/T 0002-2012, 于 2016 年成为国家标准 GB/T 32907-2016《信息安全技术 SM4 分组密码算法》 [18]. 肖雅莹等在 2009 年设计了首个 SM4 白盒方案 [19](肖 - 来方案), 该方案使用与 Chow 等的 AES 白盒方案类似的构造方法, 不久被林婷婷等攻破 [20]. 2016 年, 白鲲鹏等构造了另外一个白盒实

现 SM4 的方案 [21](白 - 武方案), 该方案与肖 - 来方案类似, 均使用仿射变换与查找表来保护算法运行的内部信息, 只是将方案内部编码解码过程设计得更为复杂, 增大了分析的难度, 并引入更多的随机数来提高算法的混淆程度. 史扬等通过引入同构变换来进一步增加白盒多样性的方式也构造了一个 SM4 白盒方案 [22]. 本文只分析肖 - 来及白 - 武的白盒方案.

与目前大多数 AES、DES 白盒方案相同, 肖 - 来、白 - 武的 SM4 白盒方案均采用外部编码来保护算法第一轮与最后一轮的信息 (对算法输入信息进行的编码我们称为输入编码, 对算法输出的被编码过的信息的解码称为输出解码). 当方案的输入编码或输出解码不再安全时, 通过分析方案的第一轮或最后一轮, 攻击者可依次恢复 SM4 算法的每一轮密钥.

需要注意的是, 对具有外部编码的白盒密码方案, 攻击者需同时恢复密钥及外部编码才能解密信息, 故具有外部编码的白盒密码方案的密钥空间为密码算法的密钥空间与外部编码的取值空间之积. 肖 - 来与白 - 武 SM4 白盒方案的外部编码均为 4 个 32 比特向量的仿射变换, 仿射变换的取值空间为 4 个 32 阶可逆矩阵及 4 个 32 比特向量的取值空间, 空间大小为 $(\prod_{i=0}^{31}(2^{32}-2^i))^4\cdot(2^{32})^4$, 而 SM4 算法的密钥空间大小为 2^{128}, 故这两个白盒方案的密钥空间大小为 $2^{256}\cdot(\prod_{i=0}^{31}(2^{32}-2^i))^4$.

本文在外部编码受到安全保护的假定下, 分析了肖 - 来、白 - 武的 SM4 白盒方案. 一方面发现了林 - 来分析中估算复杂度的偏差, 另一方面针对两个白盒 SM4 方案给出了具体的恢复密钥攻击, 指出密钥空间大小分别为 $61\ 200\cdot 2^{32}$ 和 $61\ 200\cdot 2^{128}$.

同时, 我们发现两个方案所有编码矩阵的取值空间均为 61 200. 进一步, 肖 - 来方案中所有仿射常数及轮密钥的取值空间为 2^{32}, 白 - 武方案的所有仿射常数及轮密钥的取值空间为 2^{128}. 分析表明, 这两个方案的安全强度主要由外部编码中仿射常数的安全性决定. 因两个方案中线性变换矩阵的取值空间大小相同, 故白 - 武方案中通过复杂化内部编码解码过程并不能提高线性变换矩阵的安全性, 以及通过对矩阵拆分或随机数拆分来增加白盒多样性的策略, 对算法安全性并无作用. 造成两个方案外部编码实际取值空间大小不同的主要原因在于白 - 武方案中将所有仿射常数隐藏至查找表中, 而在肖 - 来方案中泄露了复合变换的仿射常数. 根据泄露的复合变换的仿射常数, 可以获取一些线性关系, 从而只需猜测其中一个 32 比特仿射常数就可以确定其它所有仿射常数.

2 符号约定及 SM4 算法简介

本文符号约定如下:

Z_2^m : m比特向量集合, 本文均将向量看做列向量, 即$x\in Z_2^m$表示x为m行一列的向量

$Z_2^{m\times n}$: $m\times n$维比特矩阵集合

$X_0, X_1, \cdots, X_{35}$: SM4算法的内部状态, $X_i\in Z_2^{32}$

$X_0', X_1', \cdots, X_{35}'$: 被编码过后的内部状态$X_i'\in Z_2^{32}$

Key : SM4算法的主密钥, Key $\in Z_2^{128}$

K_r : SM4算法的轮密钥, $K_r\in Z_2^{32}, r=1,2,\cdots,32$

$+,\cdot$: 有限域GF(2)上的加法与乘法

$||$: 连接符

$<<<$: 循环左移位

$\oplus$: 异或

$\odot$: 点乘, 若$A,B\in Z_2^{m\times n}$, 则$A\odot B$表示矩阵A与矩阵B对应的分量分别相乘构成的$m\times n$矩阵. 若$A\in Z_2^m, B\in Z_2^{m\times n}$, 则$A\odot B$表示$A$分别与$B$的每一列相乘构成的$m\times n$矩阵

$S(\cdot)$: SM4算法的S盒查表运算, 输入8比特输出8比特

τ : $\tau(a)=S(a_0)||S(a_1)||S(a_2)||S(a_3)$, 其中$a=a_0||a_1||a_2||a_3, a\in Z_2^{32}, a_i\in Z_2^8$

$L: L(a) = a \oplus (a <<< 2) \oplus (a <<< 10) \oplus (a <<< 18) \oplus (a <<< 24)$, 其中$a \in Z_2^{32}$, L可用 32 阶矩阵M表示如下:

$$M = \begin{pmatrix} M_1 & M_2 & M_2 & M_3 \\ M_3 & M_1 & M_2 & M_2 \\ M_2 & M_3 & M_1 & M_2 \\ M_2 & M_2 & M_3 & M_1 \end{pmatrix}, \text{其中} M_1 = \begin{pmatrix} 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix},$$

$$M_2 = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}, M_3 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

$M_0, M_1, M_2, M_3: M = M_0||M_1||M_2||M_3, M_j$分别表示矩阵$M$的4个$32 \times 8$分块

轮函数$F: F(X_0, X_1, X_2, X_3, K_r) = X_0 \oplus T(X_1 \oplus X_2 \oplus X_3 \oplus K_r)$, 其中$T = L \circ \tau$

反序变换$R: R(a_0, a_1, a_2, a_3) = (a_3, a_2, a_1, a_0)$, 其中$a_i \in Z_2^{32}, i = 0, 1, 2, 3$

编码：$y = f(x)$称为将x编码为y, 其中f为可逆函数

级联编码：若$x = x_1||x_2||\cdots||x_m, y = y_1||y_2||\cdots||y_m$, 且$y_i = f_i(x_i)$, f_i均为可逆函数, $i = 1, 2, \cdots, m$, 则称$y = f_1(x_1)||f_2(x_2)||\cdots||f_m(x_m)$为对$x$的级联编码.

SM4 算法的密钥长度为 128 比特, 表示为 $\text{Key} = (\text{MK}_0, \text{MK}_1, \text{MK}_2, \text{MK}_3)$, 其中 $\text{MK}_i \in Z_2^{32}, i = 0, 1, 2, 3$. 根据密钥扩展算法, 由主密钥 Key 生成 32 轮轮密钥 $K_1, K_2, \cdots, K_{32}$, 其中 $K_r \in Z_2^{32}, r = 1, 2, \cdots, 32$. 设明文输入为 $(X_0, X_1, X_2, X_3) \in (Z_2^{32})^4$, 密文输出为 $(Y_0, Y_1, Y_2, Y_3) \in (Z_2^{32})^4$, 轮密钥为 $K_1, K_2, \cdots, K_{32}$, 则 SM4 的加密变换为:

$$X_{r+3} = F(X_{r-1}, X_r, X_{r+1}, X_{r+2}, K_r) = X_{r-1} \oplus T(X_r \oplus X_{r+1} \oplus X_{r+2} \oplus K_r), r = 1, 2, \cdots, 32$$

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

解密变换与加密变换过程相同, 只是轮密钥使用顺序相反.

3 肖 - 来、白 - 武 SM4 白盒方案简介

3.1 肖 - 来 SM4 白盒方案

肖 - 来 SM4 白盒方案使用仿射变换 $X_i' = P_i \cdot X_i + p_i$ 来隐藏 SM4 算法的每个内部状态 X_i, 其中 $P_i \in Z_2^{32\times 32}$, $p_i \in Z_2^{32}$, $i = 0, 1, 2, \cdots, 35$. 并使用查找表来实现变换函数 T, 通过对查找表的输入输出值编码的方式来隐藏变换函数 T 中所含有的密钥信息. 为降低查找表的规模, 肖 - 来方案中使用分块矩阵 $E_r = \text{diag}(E_{r,0}, E_{r,1}, E_{r,2}, E_{r,3})$ 来编码查找表的输入信息, 从而可将函数 T 使用 4 个 8 进 32 出的查找表来实现, 其中 $E_{r,i} \in Z_2^{8\times 8}$ 且为可逆矩阵, $r = 1, 2, \cdots, 32, i = 0, 1, 2, 3$.

肖 - 来方案的具体实现过程如下.

(1) 随机生成以下 32 阶可逆矩阵:

$$P_0, P_1, \cdots, P_{35}, Q_1, Q_2, \cdots, Q_{32}, E_1, E_2, \cdots, E_{32}$$

及以下 32×1 随机向量:

$$p_0, p_1, \cdots, p_{35}, p'_4, p'_5, \cdots, p'_{35}, q_1, q_2, \cdots, q_{32}, e_1, e_2, \cdots, e_{32}$$

其中 $E_r = \mathrm{diag}(E_{r,0}, E_{r,1}, E_{r,2}, E_{r,3})$, $E_{r,i}$ 为 8 阶可逆矩阵. 并记 $e_r = e_{r,0}||e_{r,1}||e_{r,2}||e_{r,3}, e_{r,i}$ 为 8×1 向量, $r = 1, 2, \cdots, 32, i = 0, 1, 2, 3$.
根据上述矩阵及向量计算以下复合仿射变换:

$$\begin{aligned} B_r &= P_{r+3} \cdot Q_r^{-1}, \quad b_r = P_{r+3} \cdot Q_r^{-1} \cdot q_r \\ C_r &= P_{r+3} \cdot P_{r-1}^{-1}, \quad c_r = P_{r+3} \cdot P_{r-1}^{-1} \cdot p_{r-1} + p'_{r+3} \\ D_{r,j} &= E_r \cdot P_{r+j}^{-1}, \quad d_{r,j} = E_r \cdot P_{r+j}^{-1} \cdot p_{r+j} + e_r \end{aligned}$$

其中 $r = 1, 2, \cdots, 32, j = 0, 1, 2$.

(2) 输入编码:

$$X'_i = P_i \cdot X_i + p_i, i = 0, 1, 2, 3$$

输出解码:

$$X_i = P_i^{-1} \cdot (X'_i + p_i), i = 32, 33, 34, 35$$

(3) 存储以下矩阵及向量:

$$B_r, b_r, C_r, c_r, D_{r,j}, d_{r,j}, r = 1, 2, \cdots, 32, j = 0, 1, 2$$

以及以下 8 进 32 出的查找表 (遍历 8 比特输入值 x 存储 32 比特输出值 y):

$$\mathrm{Table}_{r,i} : \mathrm{GF}(2^{32}) \leftarrow \mathrm{GF}(2^8)$$

$$y = R_i \cdot S(E_{r,i}^{-1} \cdot (x + e_{r,i}) + K_{r,i}) + \boldsymbol{q}_{r,i}, r = 1, 2, \cdots, 32, i = 0, 1, 2, 3$$

其中 $R_0||R_1||R_2||R_3 = Q_r \cdot M$, R_i 为 32×8 矩阵. $\sum_{i=0}^{3} \boldsymbol{q}_{r,i} = q_r$, 后面分析表明 $\boldsymbol{q}_{r,i}$ 如何取值对方案的安全性没有影响, 肖 - 来方案中取 $\boldsymbol{q}_{r,0} = \boldsymbol{q}_{r,1} = \boldsymbol{q}_{r,2} = 0, \boldsymbol{q}_{r,3} = q_r$.

(4) 第 r 轮, 由输入 $(X'_{r-1}, X'_r, X'_{r+1}, X'_{r+2})$ 计算 X'_{r+3} 的过程如下:

(a) 计算

$$y_r = \sum_{j=0}^{2} (D_{r,j} \cdot X'_{r+j} + d_{r,j})$$

(b) 记 $y_r = y_{r,0}||y_{r,1}||y_{r,2}||y_{r,3}$, 由每个 $y_{r,i}$ 分别查询查询 $\mathrm{Table}_{r,i}$ 得

$$\boldsymbol{z}_{r,i} = \mathrm{Table}_{r,i}(y_{r,i}), i = 0, 1, 2, 3$$

(c) 计算

$$X'_{r+3} = (B_r \cdot \sum_{i=0}^{3} \boldsymbol{z}_{r,i} + b_r) + (C_r \cdot X'_{r-1} + c_r)$$

由第 (4) 步经 32 轮计算后, 可得被编码的密文 $(Y'_0, Y'_1, Y'_2, Y'_3) = R(X'_{32}, X'_{33}, X'_{34}, X'_{35}) = (X'_{35}, X'_{34}, X'_{33}, X'_{32})$.

在白盒攻击环境下, 肖 - 来方案中攻击者可见的信息有: 编码后的内部状态 X'_i, 复合仿射变换 $\{B_r, b_r\}, \{C_r, c_r\}, \{D_{r,j}, d_{r,j}\}$, 以及每轮 4 个 8 进 32 出查找表 $\text{Table}_{r,i}$, 其中 $r = 1, 2, \cdots, 32, i = 0, 1, 2, \cdots, 35, j = 0, 1, 2$.

肖 - 来方案的主要原理如下 (只考察在第 r 轮由 $X'_{r-1}, X'_r, X'_{r+1}, X'_{r+2}$ 计算 X'_{r+3} 过程):

首先对 $X'_{r-1}, X'_r, X'_{r+1}, X'_{r+2}$ 解码并编码. 解码即对 X'_i 做相应的逆仿射变换恢复 X_i 的值 $X_i = P_i^{-1} \cdot (X'_i + p_i)$, $i = r-1, r, r+1, r+2$. 再对解码后的值 X_{r-1} 使用仿射变换 $z_r = P_{r+3} \cdot X_{r-1} + p'_{r+3}$ 编码, 并对 X_r, X_{r+1}, X_{r+2} 使用相同的仿射变换 $\{E_r, e_r\}$ 编码, 其中 $i = 0, 1, 2$. 上述过程可以统一以如下公式表示:

$$\begin{pmatrix} z'_r \\ \boldsymbol{y}_{r,0} \\ \boldsymbol{y}_{r,1} \\ \boldsymbol{y}_{r,2} \end{pmatrix} = \begin{pmatrix} P_{r+3} \\ E_r \\ E_r \\ E_r \end{pmatrix} \odot \begin{pmatrix} P_{r-1}^{-1} \\ P_r^{-1} \\ P_{r+1}^{-1} \\ P_{r+2}^{-1} \end{pmatrix} \odot \begin{pmatrix} X'_{r-1} \\ X'_r \\ X'_{r+1} \\ X'_{r+2} \end{pmatrix} + \begin{pmatrix} p_{r-1} \\ p_r \\ p_{r+1} \\ p_{r+2} \end{pmatrix} + \begin{pmatrix} p'_{r+3} \\ e_r \\ e_r \\ e_r \end{pmatrix}$$

其中, 解码编码过程同时完成, 并不需要计算出 $X_{r-1}, X_r, X_{r+1}, X_{r+2}$ 的值. 对给定的复合仿射变换, 任意选取一个可逆仿射变换, 均可以找到另外一个可逆仿射变换使此两个仿射变换的复合变换为给定的仿射变换, 故根据给定的复合后的仿射变换, 无法将两个仿射变换分离出来, 从而根据上述解码编码过程无法直接恢复内部状态 $X_{r-1}, X_r, X_{r+1}, X_{r+2}$.

其次, 在上述第一步中, 因对 X'_r, X'_{r+1}, X'_{r+2} 解码后使用的同一个仿射变换编码, 易知

$$y_r = \sum_{i=0}^{2} \boldsymbol{y}_{r,i} = E_r \cdot (X_r + X_{r+1} + X_{r+2}) + e_r$$

其中 $E_r = \text{diag}(E_{r,0}, E_{r,1}, E_{r,2}, E_{r,3})$ 为分块矩阵, 即 y_r 为对和值 $X_r + X_{r+1} + X_{r+2}$ 的 4 个 8 比特分量分别独立的做仿射变换得来.

再次, 由 y_r 经 4 个查找表完成了对 y_r 的级联解码以及函数 T 的计算并对结果进行编码, 即

$$z_r = \sum_{i=0}^{3} \boldsymbol{z}_{r,i} = Q_r \cdot T(X_r + X_{r+1} + X_{r+2} + K_r) + q_r$$

最后再对 z_r 解码并使用仿射变换 $\{P_{r+3}, p'_{r+3}\}$ 编码, 然后与 z'_r 异或完成对 X_{r+3} 的计算并将其编码为 $X'_{r+3} = P_{r+3} \cdot X_{r+3} + p_{r+3}$.

在肖 - 来方案中, 轮密钥 K_r 被分散隐藏在 4 个查找表 $\text{Table}_{r,0}$, $\text{Table}_{r,1}$, $\text{Table}_{r,2}$, $\text{Table}_{r,3}$ 中, 每个查找表的输入值分别被仿射变换 $\{E_{r,i}, e_{r,i}\}$ 编码, 输出值被仿射变换 $\{Q_r, q_r\}$ 编码. 同样的, 根据查找表的输入输出值, 无法直接获取变换函数 T 的输入输出值信息, 从而在此局部确保了密钥的安全.

对于该方案, 从被编码的密文 $(X'_{35}, X'_{34}, X'_{33}, X'_{32})$ 中恢复明文 (X_0, X_1, X_2, X_3), 需恢复所有轮密钥 K_r 及外部编码 (输入或输出编码的 4 个仿射变换: $\{P_0, p_0\}$,$\{P_1, p_1\}$,$\{P_2, p_2\}$,$\{P_3, p_3\}$ 或 $\{P_{32}, p_{32}\}$, $\{P_{33}, p_{33}\}$,$\{P_{34}, p_{34}\}$,$\{P_{35}, p_{35}\}$). 由 SM4 算法的特点, 只需恢复连续 4 轮轮密钥就可恢复所有轮密钥, 故轮密钥的取值空间大小为 2^{128}. 二元 32 阶可逆矩阵的数目为 $\prod_{i=0}^{31}(2^{32} - 2^i)$, 故外部编码的取值空间大小为 $(\prod_{i=0}^{31}(2^{32} - 2^i))^4 \cdot (2^{32})^4$. 我们需找到正确的密钥, 同时恢复外部编码才能恢复明文, 所需搜索的空间大小为 $2^{256} \cdot (\prod_{i=0}^{31}(2^{32} - 2^i))^4$.

3.2 白-武白盒 SM4 方案

与肖-来方案类似, 白-武方案同样使用仿射变换 $X_i' = P_i \cdot X_i + p_i$ 来隐藏 SM4 算法的每个内部状态 X_i, 其中 $P_i \in Z_2^{32\times 32}, p_i \in Z_2^{32}, i = 0,1,2,\cdots,35$. 并使用查找表来隐藏变换函数 T 中所含有的密钥信息. 不同之处在于, 白-武方案在降低查找表的规模时使用了两个 8 阶级联编码矩阵并引入了更多的随机数, 以此来增加内部编码解码的复杂度及提高算法的分析难度.

白-武白盒 SM4 方案的具体实现过程如下:

(1) 随机生成 32 阶可逆矩阵:

$$P_0, P_1, \cdots, P_{35}, E_1, E_2, \cdots, E_{32}, F_1, F_2, \cdots, F_{32}$$

及 32× 1 随机向量:

$$p_i, \boldsymbol{p}_{i',j}, \boldsymbol{p}'_{r+3,j}, \boldsymbol{p}''_{r+3,j}, \boldsymbol{e}_{r,0}, \boldsymbol{e}_{r,1}, \boldsymbol{e}_{r,2}, \boldsymbol{e}_{r,3}, \boldsymbol{e}_{r,4}, \boldsymbol{e}_{r,5}, \boldsymbol{f}_{r,0}, \boldsymbol{f}_{r,1}, \boldsymbol{f}_{r,2}, \boldsymbol{f}_{r,3}, \boldsymbol{f}_{r,4}, \boldsymbol{f}_{r,5}$$

其中 $i = 0,1,\cdots,35$, $i' = 0,1,\cdots,34$, $j = 0,1,2,3$, $r = 1,2,\cdots,32$, $\sum_{j=0}^{3} \boldsymbol{p}_{i',j} = p'_i$, $\sum_{j=0}^{3} \boldsymbol{p}'_{r+3,j} = p'_{r+3}, \sum_{j=0}^{3} \boldsymbol{p}''_{r+3,j} = p''_{r+3}, p'_{r+3} + p''_{r+3} = p_{r+3}, E_r = \mathrm{diag}(E_{r,0}, E_{r,1}, E_{r,2}, E_{r,3}), F_r = \mathrm{diag}(F_{r,0}, F_{r,1}, F_{r,2}, F_{r,3})$.
计算

$$e_r = \sum_{i=0}^{5} \boldsymbol{e}_{r,i}, f_r = \sum_{i=0}^{5} \boldsymbol{f}_{r,i}$$

并记

$$e_r = e_{r,0}||e_{r,1}||e_{r,2}||e_{r,3}, f_r = f_{r,0}||f_{r,1}||f_{r,2}||f_{r,3}$$

且将 e_r, f_r 重组为以下 16 比特向量

$$u_{r,i} = e_{r,i}||f_{r,i}$$

其中 $i = 0,1,2,3, r = 1,2,\cdots,32$.

(2) 输入编码:

$$X_i' = P_i \cdot X_i + p_i, i = 0,1,2,3$$

输出解码:

$$X_i = P_i^{-1} \cdot (X_i' + p_i), i = 32,33,34,35$$

(3) 第 r 轮, 存储以下 16 个查找表 (遍历 8 比特输入值 x 存储 32 比特输出值 y):

$$\begin{aligned}
TC_{r,j}: &\quad y = P_{r+3} \cdot P_{r-1,j}^{-1} \cdot x + P_{r-1}^{-1} \cdot \boldsymbol{p}_{r-1,j} + \boldsymbol{p}'_{r+3,j}, && j = 0,1,2,3\\
TA_{r,j}: &\quad y = E_r \cdot {P_{r,j}}^{-1} \cdot x + P_r^{-1} \cdot \boldsymbol{p}_{r,j} + \boldsymbol{e}_{r,j}, && j = 0,1\\
TA_{r,j+2}: &\quad y = E_r \cdot P_{r+1,j}^{-1} \cdot x + P_{r+1}^{-1} \cdot \boldsymbol{p}_{r+1,j} + \boldsymbol{e}_{r,j+2}, && j = 0,1\\
TA_{r,j+4}: &\quad y = E_r \cdot P_{r+2,j}^{-1} \cdot x + P_{r+2}^{-1} \cdot \boldsymbol{p}_{r+2,j} + \boldsymbol{e}_{r,j+4}, && j = 0,1\\
TB_{r,j}: &\quad y = F_r \cdot P_{r,j}^{-1} \cdot x + P_r^{-1} \cdot \boldsymbol{p}_{r,j+2} + \boldsymbol{f}_{r,j}, && j = 0,1\\
TB_{r,j+2}: &\quad y = F_r \cdot P_{r+1,j}^{-1} \cdot x + P_{r+1}^{-1} \cdot \boldsymbol{p}_{r+1,j+2} + \boldsymbol{f}_{r,j+2}, && j = 0,1\\
TB_{r,j+4}: &\quad y = F_r \cdot P_{r+2,j}^{-1} \cdot x + P_{r+2}^{-1} \cdot \boldsymbol{p}_{r+2,j+2} + \boldsymbol{f}_{r,j+4}, && j = 0,1
\end{aligned}$$

以及以下 4 个查找表 (遍历 16 比特的输入值 x 存储 32 比特输出值 y):

$$TD_{r,j}: \quad y = P_{r+3} \cdot \boldsymbol{M}_i \cdot S(E_{r,j}^{-1}||F_{r,j}^{-1} \cdot (x + u_{r,j}) + K_{r,j}) + \boldsymbol{p}''_{r+3,j}, j = 0, 1, 2, 3$$

注, $E_{r,j}^{-1}||F_{r,j}^{-1}$ 为 8×16 阶矩阵, $(E_{r,j}^{-1}||F_{r,j}^{-1})(x + u_{r,j})$ 表示将 16 比特的输入值 x 的两个 8 比特分量分别经过 $E_{r,j}^{-1}$ 与 $F_{r,j}^{-1}$ 解码后再将两解码后的值加起来.

(4) 第 r 轮由 $X'_{r-1}, X'_r, X'_{r+1}, X'_{r+2}$ 计算 X_{r+3} 并编码为 X'_{r+3} 的过程如下:

(a) 由 X'_{r-1} 的 4 个 8 比特分量分别查询查找表 $TC_{r,0}, TC_{r,1}, TC_{r,2}, TC_{r,3}$ 得:

$$\boldsymbol{c}_{r,0}, \boldsymbol{c}_{r,1}, \boldsymbol{c}_{r,2}, \boldsymbol{c}_{r,3}$$

由 X'_r 的 4 个 8 比特分量分别查询查找表 $TA_{r,0}, TA_{r,1}, TB_{r,0}, TB_{r,1}$ 得:

$$\boldsymbol{a}_{r,0}, \boldsymbol{a}_{r,1}, \boldsymbol{b}_{r,0}, \boldsymbol{b}_{r,1}$$

由 X'_{r+1} 的 4 个 8 比特分量分别查询查找表 $TA_{r,2}, TA_{r,3}, TB_{r,2}, TB_{r,3}$ 得:

$$\boldsymbol{a}_{r,2}, \boldsymbol{a}_{r,3}, \boldsymbol{b}_{r,2}, \boldsymbol{b}_{r,3}$$

由 X'_{r+2} 的 4 个 8 比特分量分别查询查找表 $TA_{r,4}, TA_{r,5}, TB_{r,4}, TB_{r,5}$ 得:

$$\boldsymbol{a}_{r,4}, \boldsymbol{a}_{r,5}, \boldsymbol{b}_{r,4}, \boldsymbol{b}_{r,5}$$

(b) 计算

$$a_r = \sum_{i=0}^{5} \boldsymbol{a}_{r,i}, b_r = \sum_{i=0}^{5} \boldsymbol{b}_{r,i}$$

记

$$a_r = a_{r,0}||a_{r,1}||a_{r,2}||a_{r,3}, b_r = b_{r,0}||b_{r,1}||b_{r,2}||b_{r,3}$$

将 a_r, b_r 重组为 4 个 16 比特向量

$$y_{r,0}, y_{r,1}, y_{r,2}, y_{r,3}, 其中 y_{r,i} = a_{r,i}||b_{r,i}, i = 0, 1, 2, 3$$

(c) 根据 $y_{r,0}, y_{r,1}, y_{r,2}, y_{r,3}$ 分别查询查找表 $TD_{r,0}, TD_{r,1}, TD_{r,2}, TD_{r,3}$ 得:

$$\boldsymbol{d}_{r,0}, \boldsymbol{d}_{r,1}, \boldsymbol{d}_{r,2}, \boldsymbol{d}_{r,3}$$

(d) 计算 X'_{r+3}

$$X'_{r+3} = \sum_{i=0}^{3} (\boldsymbol{c}_{r,i} + \boldsymbol{d}_{r,i})$$

在白-武方案中, 内部状态 X_i 被编码成 $X'_i = P_i \cdot X_i + p_i$ 而隐藏, 每一轮中公开可见的信息为上述 20 个查找表. 而轮函数 F 的计算完全由查表及对查表结果重组或者异或完成.

白-武方案的主要原理如下 (只考察在第 r 轮由输入 $X'_{r-1}, X'_r, X'_{r+1}, X'_{r+2}$ 计算 X'_{r+3} 过程):

首先, 16 个查找表 $TA_{r,i}, TB_{r,i}, TC_{r,j} i = 0, 1, \cdots, 5, j = 0, 1, 2, 3$ 完成对 $X'_{r-1}, X'_r, X'_{r+1}, X'_{r+2}$ 的解码及编码, 该过程可用如下公式表示.

$$
\begin{pmatrix}
\boldsymbol{c}_{r,0} & \boldsymbol{c}_{r,1} & \boldsymbol{c}_{r,2} & \boldsymbol{c}_{r,3} \\
\boldsymbol{a}_{r,0} & \boldsymbol{a}_{r,1} & \boldsymbol{b}_{r,0} & \boldsymbol{b}_{r,1} \\
\boldsymbol{a}_{r,2} & \boldsymbol{a}_{r,3} & \boldsymbol{b}_{r,2} & \boldsymbol{b}_{r,3} \\
\boldsymbol{a}_{r,4} & \boldsymbol{a}_{r,5} & \boldsymbol{b}_{r,4} & \boldsymbol{b}_{r,5}
\end{pmatrix}
$$

$$
= \begin{pmatrix}
P_{r+3} & P_{r+3} & P_{r+3} & P_{r+3} \\
E_r & E_r & F_r & F_r \\
E_r & E_r & F_r & F_r \\
E_r & E_r & F_r & F_r
\end{pmatrix} \odot \left(\begin{pmatrix}
P_{r-1,0}^{-1} & P_{r-1,1}^{-1} & P_{r-1,2}^{-1} & P_{r-1,3}^{-1} \\
P_{r,0}^{-1} & P_{r,1}^{-1} & P_{r,2}^{-1} & P_{r,3}^{-1} \\
P_{r+1,0}^{-1} & P_{r+1,1}^{-1} & P_{r+1,2}^{-1} & P_{r+1,2}^{-1} \\
P_{r+2,0}^{-1} & P_{r+2,1}^{-1} & P_{r+2,2}^{-1} & P_{r+2,3}^{-1}
\end{pmatrix} \right.
$$

$$
\odot \begin{pmatrix}
X'_{r-1,0} & X'_{r-1,1} & X'_{r-1,2} & X'_{r-1,3} \\
X'_{r,0} & X'_{r,1} & X'_{r,2} & X'_{r,3} \\
X'_{r+1,0} & X'_{r+1,1} & X'_{r+1,2} & X'_{r+1,3} \\
X'_{r+2,0} & X'_{r+2,1} & X'_{r+2,2} & X'_{r+2,3}
\end{pmatrix} + \begin{pmatrix}
P_{r-1}^{-1} \\ P_r^{-1} \\ P_{r+1}^{-1} \\ P_{r+2}^{-1}
\end{pmatrix} \odot \left. \begin{pmatrix}
\boldsymbol{p}_{r-1,0} & \boldsymbol{p}_{r-1,1} & \boldsymbol{p}_{r-1,2} & \boldsymbol{p}_{r-1,3} \\
\boldsymbol{p}_{r,0} & \boldsymbol{p}_{r,1} & \boldsymbol{p}_{r,2} & \boldsymbol{p}_{r,3} \\
\boldsymbol{p}_{r+1,0} & \boldsymbol{p}_{r+1,1} & \boldsymbol{p}_{r+1,2} & \boldsymbol{p}_{r+1,3} \\
\boldsymbol{p}_{r+2,0} & \boldsymbol{p}_{r+2,1} & \boldsymbol{p}_{r+2,2} & \boldsymbol{p}_{r+2,3}
\end{pmatrix} \right)
$$

$$
+ \begin{pmatrix}
\boldsymbol{p}'_{r+3,0} & \boldsymbol{p}'_{r+3,1} & \boldsymbol{p}'_{r+3,2} & \boldsymbol{p}'_{r+3,3} \\
\boldsymbol{e}_{r,0} & \boldsymbol{e}_{r,1} & \boldsymbol{f}_{r,0} & \boldsymbol{f}_{r,1} \\
\boldsymbol{e}_{r,2} & \boldsymbol{e}_{r,3} & \boldsymbol{f}_{r,2} & \boldsymbol{f}_{r,3} \\
\boldsymbol{e}_{r,4} & \boldsymbol{e}_{r,5} & \boldsymbol{f}_{r,4} & \boldsymbol{f}_{r,5}
\end{pmatrix}
$$

其中 $X'_i = X'_{i,0}||X'_{i,1}||X'_{i,2}||X'_{i,3}, P_i^{-1} = P_{i,0}^{-1}||P_{i,1}^{-1}||P_{i,2}^{-1}||P_{i,3}^{-1}$ 即将矩阵 P_i^{-1} 分成 4 个 32×8 矩阵, $p_i = \sum_{j=0}^{3} \boldsymbol{p}'_{i,j}$ 即将向量 p_i 拆分为 4 个 32 比特向量之和, $i = r-1, r, r+1, r+2$. 易知

$$
\sum_{j=0}^{3} \boldsymbol{p}_{r+3}^{-1}(\boldsymbol{c}_{r,j} + \boldsymbol{p}'_{r+3,j}) = X_{r-1}
$$

$$
E_r^{-1}(\boldsymbol{p}_{r,0} + \boldsymbol{p}_{r,1} + \boldsymbol{e}_{r,0} + \boldsymbol{e}_{r,1}) + F_r^{-1}(\boldsymbol{b}_{r,0} + \boldsymbol{b}_{r,1} + \boldsymbol{f}_{r,0} + \boldsymbol{f}_{r,1}) = X_r
$$

$$
E_r^{-1}(\boldsymbol{p}_{r,2} + \boldsymbol{p}_{r,3} + \boldsymbol{e}_{r,2} + \boldsymbol{e}_{r,3}) + F_r^{-1}(\boldsymbol{b}_{r,2} + \boldsymbol{b}_{r,3} + \boldsymbol{f}_{r,2} + \boldsymbol{f}_{r,3}) = X_{r+1}
$$

$$
E_r^{-1}(\boldsymbol{p}_{r,4} + \boldsymbol{p}_{r,5} + \boldsymbol{e}_{r,4} + \boldsymbol{e}_{r,5}) + F_r^{-1}(\boldsymbol{b}_{r,4} + \boldsymbol{b}_{r,5} + \boldsymbol{f}_{r,4} + \boldsymbol{f}_{r,5}) = X_{r+2}
$$

记 $a_r = \sum_{i=0}^{5} \boldsymbol{a}_i$, $b_r = \sum_{i=0}^{5} \boldsymbol{b}_i$,$e_r = \sum_{i=0}^{5} \boldsymbol{e}_{r,i}$,$f_r = \sum_{i=0}^{5} \boldsymbol{f}_{r,i}$, 则我们有

$$
E_r(a_r + e_r) + F_r(b_r + f_r) = X_r + X_{r+1} + X_{r+2}
$$

记 $a_r = a_{r,0}||a_{r,1}||a_{r,2}||a_{r,3}, b_r = b_{r,0}||b_{r,1}||b_{r,2}||b_{r,3}$, 将 a_r, b_r 重组为 4 个 16 比特向量 $y_{r,0}, y_{r,1}, y_{r,2}, y_{r,3}$, 其中 $y_{r,i} = a_{r,i}||b_{r,i}, i = 0, 1, 2, 3$. 并记

$$
e_r = e_{r,0}||e_{r,1}||e_{r,2}||e_{r,3}, f_r = f_{r,0}||f_{r,1}||f_{r,2}||f_{r,3}
$$

将 e_r, f_r 重组为 4 个 16 比特向量

$$
u_{r,i} = e_{r,i}||f_{r,i}, i = 0, 1, 2, 3
$$

再根据 $y_{r,0}, y_{r,1}, y_{r,2}, y_{r,3}$ 分别查询 4 个 $TD_{r,j}$ 查找表得 $\boldsymbol{d}_{r,0}, \boldsymbol{d}_{r,1}, \boldsymbol{d}_{r,2}, \boldsymbol{d}_{r,3}$, 此过程可用如下公式表示.

$$\begin{pmatrix} d_{r,0} \\ d_{r,1} \\ d_{r,2} \\ d_{r,3} \end{pmatrix} = \begin{pmatrix} P_{r+3} \\ P_{r+3} \\ P_{r+3} \\ P_{r+3} \end{pmatrix} \odot \begin{pmatrix} \boldsymbol{M}_0 \\ \boldsymbol{M}_1 \\ \boldsymbol{M}_2 \\ \boldsymbol{M}_3 \end{pmatrix} \odot \begin{pmatrix} S((E_{r,0}^{-1}||F_{r,0}^{-1})(y_{r,0}+u_{r,0})+K_{r,0}) \\ S((E_{r,1}^{-1}||F_{r,1}^{-1})(y_{r,1}+u_{r,1})+K_{r,1}) \\ S((E_{r,2}^{-1}||F_{r,2}^{-1})(y_{r,2}+u_{r,2})+K_{r,2}) \\ S((E_{r,3}^{-1}||F_{r,3}^{-1})(y_{r,3}+u_{r,3})+K_{r,3}) \end{pmatrix} + \begin{pmatrix} \boldsymbol{p}''_{r+3,0} \\ \boldsymbol{p}''_{r+3,1} \\ \boldsymbol{p}''_{r+3,2} \\ \boldsymbol{p}''_{r+3,3} \end{pmatrix}$$

易知

$$\sum_{j=0}^{3} \boldsymbol{d}_{r,j} = P_{r+3} \cdot T(X_r + X_{r+1} + X_{r+2} + K_r) + p''_{r+3}, \text{ 其中 } p''_{r+3} = \sum_{j=0}^{3} \boldsymbol{p}''_{r+3,j}$$

从而

$$X'_{r+3} = \sum_{i=0}^{3} (\boldsymbol{c}_{r,i} + \boldsymbol{d}_{r,i}) = P_{r+3} \cdot X_{r+3} + p_{r+3}$$

其中 $p_{r+3} = \sum_{i=0}^{3} (\boldsymbol{p}'_{r+3,i} + \boldsymbol{p}''_{r+3,i})$.

由上述可知, 白 - 武方案与肖 - 来方案均使用仿射变换来隐藏 SM4 算法的每个内部状态 X_i, 只是白 - 武方案使用更为复杂的内部编码解码方式. 相比于肖 - 来方案, 白 - 武方案加密过程中减少了仿射变换的操作, 整个加密过程完全由查表及对查表结果重组或异或完成, 白 - 武方案的算法加解密速度快于肖 - 来方案. 但白 - 武方案的查找表的构造相对复杂, 查找表的规模增大到 32.5 MB, 远大于肖 - 来方案所需的存储空间.

白 - 武方案采用与肖 - 来方案相同的外部编码, 故该方案的密钥空间与肖 - 来方案相同, 空间大小也为 $2^{256} \cdot (\prod_{i=0}^{31}(2^{32} - 2^i))^4$.

4 对白盒 SM4 方案的分析

4.1 林 - 来分析的偏差

林婷婷等在文献 [20] 中构造了一种对肖 - 来 SM4 白盒方案的攻击. 通过将前一轮查找表的输出与下一轮的解码编码过程组合起来, 消除 32 阶仿射变换的线性部分, 再通过分析级联编码后的 4 个 8 比特向量之间的线性关系恢复级联编码矩阵, 进而恢复所有仿射变换的线性部分, 并最终恢复密钥. 但文中对仿射矩阵及仿射常数的数目估计有误, 以下做简要介绍.

首先, 文献 [20] 中第 2.2.2 节恢复 E_{i+1}^{-1} 的线性部分中, 作者获取了以下关系式

$$A_{jr} \cdot A'_{rj} = l[E_{i+1,j}^{-1}] \cdot L_{j0} \cdot (L_{r0})^{-1} \cdot L_{r1} \cdot (L_{j1})^{-1} \cdot (l[E_{i+1,j}^{-1}])^{-1}$$

其中 8 阶矩阵 $A_{jr} \cdot A'_{rj}$ 可求出, $L_{j0}, L_{r0}, L_{r1}, L_{j1} \in \{M_1, M_2, M_3\}, l[E_{i+1,j}^{-1}]$ 为未知的 8 阶可逆矩阵. 我们记 $A = A_{jr} \cdot A'_{rj}, L = L_{j0} \cdot (L_{r0})^{-1} \cdot L_{r1} \cdot (L_{j1})^{-1}, \Sigma = l[E_{i+1,j}^{-1}]$, 则上式可表示如下

$$A = \Sigma \cdot L \cdot \Sigma^{-1}$$

其中 8 阶矩阵 L 为 SM4 算法中 32 阶线性变换矩阵 M 中的分块矩阵 M_1, M_2 或 M_3(关于 M_1, M_2, M_3 的取值详见第 2 节中的符号约定). 由矩阵 A 与矩阵 L 相似可知, 存在可逆矩阵 P 满足

$$A = P^{-1} \cdot L \cdot P$$

从而

$$L \cdot (P \cdot \Sigma) = (P \cdot \Sigma) \cdot L$$

记 $X = P \cdot \Sigma$, 由上述方程可得关于 8 阶矩阵 X 中 64 个变元的 64 个齐次线性方程组. 该方程组的 64 阶系数矩阵由矩阵 L 完全决定, 分析显示该系数矩阵的秩为 48, 从而 X 的解的个数为 2^{16}. 进一步, 由计

算机实验可得, X 的 2^{16} 个解中存在 61 200 个可逆矩阵, 故 Σ 的解的个数为 61 200 个. 因此, 林-来分析中对矩阵 Σ 的可能取值数目估计有误.

其次, 文中第 2.2.6 节确定子密钥 k_{ri} 中, 作者获取了关于 32 比特向量 $a_{i+4}, a''_{i+4}, e_{i+1}, a_{i+3}, e_i, k_{ri}$ 的以下 6 个方程

$$\begin{aligned}
c[M_{i+4}^{i+1}] &= l[E_{i+1}^{-1}] \cdot A_{i+4}^{-1} \cdot a_{i+4} \oplus e_{i+1} \\
c[M_{i+3}^{i+1}] &= l[E_{i+1}^{-1}] \cdot A_{i+3}^{-1} \cdot a_{i+3} \oplus e_{i+1} \\
c[M_{i+3}^{i}] &= l[E_{i}^{-1}] \cdot A_{i+3}^{-1} \cdot a_{i+3} \oplus e_{i} \\
g_{i+1} &= l[E_{i+1}^{-1}] \cdot A_{i+4}^{-1} \cdot a''_{i+4} \oplus e_{i+1} \\
\tau^{-1}(L^{-1}(Q_i(Y_0))) &= l[E_i](X_0) \oplus e_i \oplus k_{ri} \\
A_{i+3}^{-1}(X_{i+3}) \oplus A_{i+3}^{-1} \cdot a_{i+3} \oplus k_{ri} &= \tau^{-1}(L^{-1}(Q_i(Y)))
\end{aligned}$$

除上述 6 个变量外, 方程组中其它矩阵或向量均为已知值或可求出. 因作者假设 $E_{i+1}^{-1}(x) = l[E_{i+1}^{-1}](x \oplus e_{i+1})$, 则 $E_{i+1}(x) = l[E_{i+1}](x \oplus e_{i+1})$, 因此上述方程中, 方程

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

实质应为

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

设 $x_0 = A_{i+4}^{-1} \cdot a_{i+4}$,$x_1 = l[E_{i+1}] \cdot e_{i+1}$,$x_2 = A_{i+3}^{-1} \cdot a_{i+3}$,$x_3 = l[E_i] \cdot e_i$,$x_4 = A_{i+4}^{-1} \cdot a''_{i+4}$,$x_5 = k_{ri}$, 则上述 6 个方程等价于

$$\begin{aligned}
x_0 \oplus x_1 &= \alpha_0 \\
x_1 \oplus x_2 &= \alpha_1 \\
x_2 \oplus x_3 &= \alpha_2 \\
x_1 \oplus x_4 &= \alpha_3 \\
x_3 \oplus x_5 &= \alpha_4 \\
x_2 \oplus x_5 &= \alpha_5
\end{aligned}$$

其中 $\alpha_i (i = 0, 1, \cdots, 5)$ 可由上述 6 个方程直接导出. 易知该方程组的系数矩阵的秩为 5, 故该方程组的解不唯一. 考虑到 x_i 均为 32 比特向量, 故 $(x_0, x_1, x_2, x_3, x_4, x_5)$ 解空间大小为 2^{32}. 从而密钥 k_{ri} 及仿射常数 $a_{i+4}, a''_{i+4}, e_{i+1}, a_{i+3}, e_i$ 的可能取值数为 2^{32}. 若 $a_{i+4}, a''_{i+4}, e_{i+1}, a_{i+3}, e_i, k_{ri}$ 已知, 可导出其它所有仿射常数及轮密钥. 从而在确定该方案中所有仿射变换的线性部分后, 轮密钥及所有仿射常数的取值空间大小为 2^{32}. 故林-来分析中确定轮密钥的数目有误.

4.2 肖-来方案的分析

以下我们改进林-来的分析方法, 通过调整恢复仿射常数的顺序来降低计算复杂度. 在白盒攻击环境下, 肖-来方案中公开可见的信息有以下复合仿射变换的变换矩阵、仿射常量

$$B_r, b_r, C_r, c_r, D_{r,j}, d_{r,j}, j = 0, 1, 2, r = 1, 2, \cdots, 32$$

以及查找表 $\mathrm{Table}_{r,i}, i = 0, 1, 2, 3, r = 1, 2, \cdots, 32$. 且我们可控制每一轮的输入值 $X'_r, X'_{r+1}, X'_{r+2}, X'_{r+3}$. 根据这些信息, 我们首先恢复每一轮的级联编码矩阵 E_r, 进而恢复所有仿射变换矩阵 $P_i, i = 0, 1, \cdots, 35$. 然后恢复所有仿射变换的仿射常数 e_r, p_i 等, 从而恢复算法内部状态 $X_i, i = 0, 1, 2, \cdots, 35$, 最后获取每一轮的轮密钥. 分析过程如下.

与林-来 [20] 的分析过程类似, 我们首先恢复第 $r+1$ 轮的级联编码矩阵 E_{r+1}. 设 SM4 算法的内部状态 X'_{r+3} 在第 $r+1$ 轮中解码并经 $\{E_{r+1}, e_{r+1}\}$ 级联编码后的值为 z, 且设 $z = z_0||z_1||z_2||z_3$, 其中 z_i 为 z 的 8 比特分量. 设第 r 轮中 4 个查找表 $\mathrm{Table}_{r,0}, \mathrm{Table}_{r,1}, \mathrm{Table}_{r,2}, \mathrm{Table}_{r,3}$ 的输入值分别为 x_0, x_1, x_2, x_3, 并记 $x = x_0||x_1||x_2||x_3$. 以下我们分析肖 - 来白盒 SM4 方案中, 由 x 与 X'_{r-1} 计算出 z 的过程. 在白盒攻击环境下, x 与 X'_{r-1} 的取值可控制, 我们取 X'_{r-1} 为某一固定值, 则 x 与 z 之间的关系如下.

$$
\begin{aligned}
z &= D_{r+1,2} \cdot X'_{r+3} + d_{r+1,2} \\
&= D_{r+1,2} \cdot B_r \cdot z_r + b_r + C_r \cdot X'_{r-1} + c_r + d_{r+1,2} \\
&= D_{r+1,2} \cdot B_r \cdot z_r + \alpha \quad (\text{记} \alpha = D_{r+1,2} \cdot b_r + C_r \cdot X'_{r-1} + c_r + d_{r+1,2}) \\
&= E_{r+1} \cdot Q_r^{-1} \cdot z_r + \alpha \\
&= E_{r+1} \cdot Q_r^{-1} \cdot Q_r \cdot M \cdot \begin{pmatrix} S(E_{r,0}^{-1}(x_0 + e_{r,0}) + K_{r,0}) \\ S(E_{r,1}^{-1}(x_1 + e_{r,1}) + K_{r,1}) \\ S(E_{r,2}^{-1}(x_2 + e_{r,2}) + K_{r,2}) \\ S(E_{r,3}^{-1}(x_3 + e_{r,3}) + K_{r,3}) \end{pmatrix} + \alpha' (\text{记 } \alpha' = E_{r+1}\, Q_r^{-1}\, q_r + \alpha) \\
&= \begin{pmatrix} E_{r+1,0} & & & \\ & E_{r+1,1} & & \\ & & E_{r+1,2} & \\ & & & E_{r+1,3} \end{pmatrix} \cdot \begin{pmatrix} M_1 & M_2 & M_2 & M_3 \\ M_3 & M_1 & M_2 & M_2 \\ M_2 & M_3 & M_1 & M_2 \\ M_2 & M_2 & M_3 & M_1 \end{pmatrix} \cdot \begin{pmatrix} S(E_{r,0}^{-1}(x_0 + e_{r,0}) + K_{r,0}) \\ S(E_{r,1}^{-1}(x_1 + e_{r,1}) + K_{r,1}) \\ S(E_{r,2}^{-1}(x_2 + e_{r,2}) + K_{r,2}) \\ S(E_{r,3}^{-1}(x_3 + e_{r,3}) + K_{r,3}) \end{pmatrix} + \alpha'
\end{aligned}
$$

记 $\alpha' = \alpha_0||\alpha_1||\alpha_2||\alpha_3$, $f_i(x_i) = S(E_{r,i}^{-1}(x_i + e_{r,i}) + K_{r,i})$, 分别取 $(x_0, x_1, x_2, x_3) = (x_0, 0, 0, 0)$, $(x_0, x_1, x_2, x_3) = (0, x_1, 0, 0)$, $(x_0, x_1, x_2, x_3) = (0, 0, x_2, 0)$, $(x_0, x_1, x_2, x_3) = (0, 0, 0, x_3)$ 得

$$
\begin{pmatrix} z_0 \\ z_1 \\ z_2 \\ z_3 \end{pmatrix} = \begin{pmatrix} E_{r+1,0} \cdot M_1 \cdot f_0(x_0) + \gamma_0 \\ E_{r+1,1} \cdot M_3 \cdot f_0(x_0) + \gamma_1 \\ E_{r+1,2} \cdot M_2 \cdot f_0(x_0) + \gamma_2 \\ E_{r+1,3} \cdot M_2 \cdot f_0(x_0) + \gamma_3 \end{pmatrix}, \quad \begin{pmatrix} z'_0 \\ z'_1 \\ z'_2 \\ z'_3 \end{pmatrix} = \begin{pmatrix} E_{r+1,0} \cdot M_2 \cdot f_1(x_1) + \gamma'_0 \\ E_{r+1,1} \cdot M_1 \cdot f_1(x_1) + \gamma'_1 \\ E_{r+1,2} \cdot M_3 \cdot f_1(x_1) + \gamma'_2 \\ E_{r+1,3} \cdot M_2 \cdot f_1(x_1) + \gamma'_3 \end{pmatrix},
$$

$$
\begin{pmatrix} z''_0 \\ z''_1 \\ z''_2 \\ z''_3 \end{pmatrix} = \begin{pmatrix} E_{r+1,0} \cdot M_2 \cdot f_2(x_2) + \gamma''_0 \\ E_{r+1,1} \cdot M_2 \cdot f_2(x_2) + \gamma''_1 \\ E_{r+1,2} \cdot M_1 \cdot f_2(x_2) + \gamma''_2 \\ E_{r+1,3} \cdot M_3 \cdot f_2(x_2) + \gamma''_3 \end{pmatrix}, \quad \begin{pmatrix} z'''_0 \\ z'''_1 \\ z'''_2 \\ z'''_3 \end{pmatrix} = \begin{pmatrix} E_{r+1,0} \cdot M_3 \cdot f_3(x_3) + \gamma'''_0 \\ E_{r+1,1} \cdot M_2 \cdot f_3(x_3) + \gamma'''_1 \\ E_{r+1,2} \cdot M_2 \cdot f_3(x_3) + \gamma'''_2 \\ E_{r+1,3} \cdot M_1 \cdot f_3(x_3) + \gamma'''_3 \end{pmatrix}
$$

其中

$$
\begin{aligned}
\gamma_0 &= E_{r+1,0} \cdot (M_2 f_1(0) + M_2 f_2(0) + M_3 f_3(0)) + \alpha_0, \\
\gamma_1 &= E_{r+1,1} \cdot (M_1 f_1(0) + M_2 f_2(0) + M_2 f_3(0)) + \alpha_1, \\
\gamma_2 &= E_{r+1,2} \cdot (M_3 f_1(0) + M_1 f_2(0) + M_2 f_3(0)) + \alpha_2, \\
\gamma_3 &= E_{r+1,3} \cdot (M_2 f_1(0) + M_3 f_2(0) + M_1 f_3(0)) + \alpha_3, \\
\gamma'_0 &= E_{r+1,0} \cdot (M_1 f_0(0) + M_2 f_2(0) + M_3 f_3(0)) + \alpha_0, \\
\gamma'_1 &= E_{r+1,1} \cdot (M_3 f_0(0) + M_2 f_2(0) + M_2 f_3(0)) + \alpha_1, \\
\gamma'_2 &= E_{r+1,2} \cdot (M_2 f_0(0) + M_1 f_2(0) + M_2 f_3(0)) + \alpha_2, \\
\gamma'_3 &= E_{r+1,3} \cdot (M_2 f_0(0) + M_3 f_2(0) + M_1 f_3(0)) + \alpha_3, \\
\gamma''_0 &= E_{r+1,0} \cdot (M_1 f_0(0) + M_2 f_1(0) + M_3 f_3(0)) + \alpha_0,
\end{aligned}
$$

$$\begin{aligned}
\gamma_1'' &= E_{r+1,1} \cdot (M_3 f_0(0) + M_1 f_1(0) + M_2 f_3(0)) + \alpha_1,\\
\gamma_2'' &= E_{r+1,2} \cdot (M_2 f_0(0) + M_3 f_1(0) + M_2 f_3(0)) + \alpha_2,\\
\gamma_3'' &= E_{r+1,3} \cdot (M_2 f_0(0) + M_2 f_1(0) + M_1 f_3(0)) + \alpha_3,\\
\gamma_0''' &= E_{r+1,0} \cdot (M_1 f_0(0) + M_2 f_1(0) + M_2 f_2(0)) + \alpha_0,\\
\gamma_1''' &= E_{r+1,1} \cdot (M_3 f_0(0) + M_1 f_1(0) + M_2 f_2(0)) + \alpha_1,\\
\gamma_2''' &= E_{r+1,2} \cdot (M_2 f_0(0) + M_3 f_1(0) + M_1 f_2(0)) + \alpha_2,\\
\gamma_3''' &= E_{r+1,3} \cdot (M_2 f_0(0) + M_2 f_1(0) + M_3 f_2(0)) + \alpha_3
\end{aligned}$$

易知在 X'_{r-1} 取值固定时, $\gamma_i, \gamma_i', \gamma_i'', \gamma_i'''$ 的取值均为固定值.

我们先考察矩阵 $E_{r+1,0}$ 与 $E_{r+1,1}$. 由上述四组方程中每组的前两个方程, 分别消去 $f_i(x_i)$ 得

$$\begin{aligned}
M_1^{-1} \cdot E_{r+1,0}^{-1}(z_0 + \gamma_0) &= M_3^{-1} \cdot E_{r+1,1}^{-1}(z_1 + \gamma_1)\\
M_2^{-1} \cdot E_{r+1,0}^{-1}(z_0' + \gamma_0') &= M_1^{-1} \cdot E_{r+1,1}^{-1}(z_1' + \gamma_1')\\
E_{r+1,0}^{-1}(z_0'' + \gamma_0'') &= E_{r+1,1}^{-1}(z_1'' + \gamma_1'')\\
M_3^{-1} \cdot E_{r+1,0}^{-1}(z_0''' + \gamma_0''') &= M_2^{-1} \cdot E_{r+1,1}^{-1}(z_1''' + \gamma_1''')
\end{aligned}$$

随机选取 x_i, 经过查表及仿射变换获取 z_i, z_i', z_i'', z_i''' 的值, 任选两组 z_i, z_i', z_i'', z_i''', 计算差分 $\Delta_{z_i}, \Delta_{z_i'}, \Delta_{z_i''}, \Delta_{z_i'''}$ 可消除常数项 $\gamma_i, \gamma_i', \gamma_i'', \gamma_i'''$, 从而得

$$\begin{aligned}
M_1^{-1} \cdot E_{r+1,0}^{-1} \cdot \Delta_{z_0} &= M_3^{-1} \cdot E_{r+1,1}^{-1} \cdot \Delta_{z_1}\\
M_2^{-1} \cdot E_{r+1,0}^{-1} \cdot \Delta_{z_0'} &= M_1^{-1} \cdot E_{r+1,1}^{-1} \cdot \Delta_{z_1'}\\
E_{r+1,0}^{-1} \cdot \Delta_{z_0''} &= E_{r+1,1}^{-1} \cdot \Delta_{z_1''}\\
M_3^{-1} \cdot E_{r+1,0}^{-1} \cdot \Delta_{z_0'''} &= M_2^{-1} \cdot E_{r+1,1}^{-1} \cdot \Delta_{z_1'''}
\end{aligned}$$

从差分 $\Delta_{z_i}, \Delta_{z_i'}, \Delta_{z_i''}, \Delta_{z_i'''}$ 中分别选取 8 个线性无关的差分 (每个 $\Delta_{z_i}, \Delta_{z_i'}, \Delta_{z_i''}, \Delta_{z_i'''}$ 的取值都不超过 256 个) 构成矩阵 $\Delta_i, \Delta_i', \Delta_i'', \Delta_i'''$, 整理得

$$\begin{aligned}
E_{r+1,0} \cdot M_1 M_3^{-1} \cdot E_{r+1,1}^{-1} &= \Delta_0 \cdot \Delta_1^{-1}\\
E_{r+1,0} \cdot M_2 M_1^{-1} \cdot E_{r+1,1}^{-1} &= \Delta_0' \cdot \Delta_1'^{-1}\\
E_{r+1,0} \cdot E_{r+1,1}^{-1} &= \Delta_0'' \cdot \Delta_1''^{-1}\\
E_{r+1,0} \cdot M_3 M_2^{-1} \cdot E_{r+1,1}^{-1} &= \Delta_0''' \cdot \Delta_1'''^{-1}
\end{aligned}$$

消去 $E_{r+1,1}$ 得

$$\begin{aligned}
E_{r+1,0} \cdot M_1 M_3^{-1} \cdot E_{r+1,0}^{-1} &= F\\
E_{r+1,0} \cdot M_2 M_1^{-1} \cdot E_{r+1,0}^{-1} &= G\\
E_{r+1,0} \cdot M_3 M_2^{-1} \cdot E_{r+1,0}^{-1} &= H
\end{aligned}$$

其中 $F = \Delta_0 \cdot \Delta_1^{-1} \cdot \Delta_1'' \cdot \Delta_0''^{-1}$, $G = \Delta_0 \cdot \Delta_1'^{-1} \cdot \Delta_1 \cdot \Delta_0''^{-1}$, $H = \Delta_0''' \cdot \Delta_1'''^{-1} \cdot \Delta_1'' \cdot \Delta_0''^{-1}$.

我们需从上述三个方程中确定矩阵 $E_{r+1,0}$. 因 F, G, H 分别与 $M_1 M_3^{-1}, M_2 M_1^{-1}, M_3 M_2^{-1}$ 相似, 故存在可逆矩阵 F', G', H' 满足

$$F = F'^{-1} \cdot M_1 M_3^{-1} \cdot F'$$
$$G = G'^{-1} \cdot M_2 M_1^{-1} \cdot G'$$
$$H = H'^{-1} \cdot M_3 M_2^{-1} \cdot H'$$

则

$$(F' \cdot E_{r+1,0}) \cdot M_1 M_3^{-1} \cdot (F' \cdot E_{r+1,0})^{-1} = M_1 M_3^{-1}$$
$$(G' \cdot E_{r+1,0}) \cdot M_2 M_1^{-1} \cdot (G' \cdot E_{r+1,0})^{-1} = M_2 M_1^{-1}$$
$$(H' \cdot E_{r+1,0}) \cdot M_3 M_2^{-1} \cdot (H' \cdot E_{r+1,0})^{-1} = M_3 M_2^{-1}$$

满足方程

$$\Sigma \cdot M_1 M_3^{-1} = M_1 M_3^{-1} \cdot \Sigma$$
$$\Sigma \cdot M_2 M_1^{-1} = M_2 M_1^{-1} \cdot \Sigma$$
$$\Sigma \cdot M_3 M_2^{-1} = M_3 M_2^{-1} \cdot \Sigma$$

的矩阵 Σ 分别均有 2^{16} 个 (上述每个方程由 Σ 的 64 个变元组成的线性方程组的系数矩阵由矩阵 M_1, M_2, M_3 决定, 系数矩阵的秩均为 48, 故 Σ 的解的数目为 2^{16}. 事实上上述三个方程相互等价, 只需求解其中一个, 另外两个只需做一个线性变换就可得到). 由计算机实验可以确定 Σ 的 2^{16} 个可能的解中可逆矩阵的数目为 61 200 个, 从而可逆矩阵 $E_{r+1,0}$ 的可能取值数目为 61 200. 且由分析过程可知, 当 $E_{r+1,0}$ 确定时, 可进一步确定 $E_{r+1,1}, E_{r+1,2}, E_{r+1,3}$, 即级联编码矩阵 E_{r+1} 的可能取值数目为 61 200 个.

当 E_{r+1} 确定时, 由 $D_{r+1,j} = E_{r+1} \cdot P_{r+1+j}^{-1}$ 可求出矩阵 $P_{r+1}, P_{r+2}, P_{r+3}$, 再由 $C_r = P_{r+3} \cdot P_{r-1}^{-1}$ 可得矩阵 P_{r-1}, 以及由 $B_r = P_{r+3} \cdot Q_r^{-1}$ 可得矩阵 Q_r, 从而可求得肖 - 来白盒 SM4 方案中所使用的所有仿射变换的线性变换矩阵 $P_i, Q_r, E_r (i = 0, 1, \cdots, 35, r = 1, 2, \cdots, 32)$, 也即所有这些变换矩阵的取值空间大小为 61 200.

记 $K'_{r,i} = E_{r,i}^{-1} \cdot e_{r,i} + K_{r,i}$, 当所有变换矩阵确定后, 由查找表

$$y = R_i \cdot S(E_{r,i}^{-1} \cdot x + K'_{r,i}) + \boldsymbol{q}_{r,i}$$

选取两个输入值 x, x' 得

$$y + y' = R_i \cdot (S(E_{r,i}^{-1} \cdot x + K'_{r,i}) + S(E_{r,i}^{-1} \cdot x' + K'_{r,i}))$$

根据上式搜索 $K'_{r,i}$, 空间为 2^8. 又由文献 [23] 知, SM4 算法中的 S 盒采用与 AES 算法中类似的方式构造, 均为有限域上的逆函数与仿射变换复合而生成, S 盒的差分均匀度均为 4, 即上述差分方程的解的个数至多为 4 个, 从而搜索不超过 2^{10} 可确定 $K'_{r,i}$. 记 $K'_r = K'_{r,0} || K'_{r,1} || K'_{r,2} || K'_{r,3}$, 并称 K'_r 为肖 - 来 SM4 白盒方案的等价轮密钥 (由此处可知 $\boldsymbol{q}_{r,i}$ 如何取值对获取等价密钥并无影响).

当 $K'_{r,i}$ 确定后, 根据查找表, 任意选取输入值获取查找表输出值后, 可确定查找表中的仿射常数 $\boldsymbol{q}_{r,i}$, 从而可确定 $q_r = \sum_{i=0}^{3} \boldsymbol{q}_{r,i}$. 由

$$b_r = P_{r+3} \cdot Q_r^{-1} \cdot q_r + p_{r+3} + p'_{r+3}$$
$$c_r = P_{r+3} \cdot P_{r-1}^{-1} \cdot p_{r-1} + p'_{r+3}$$

得

$$\gamma_r = P_{r-1}^{-1} \cdot p_{r-1} + P_{r+3}^{-1} \cdot p_{r+3} = P_{r+3}^{-1} \cdot c_r + P_{r+3}^{-1} \cdot b_r + Q_r^{-1} \cdot q_r$$

以下我们说明可以使用等价轮密钥 K_r' 以及常数 γ_r 完成肖 - 来方案的白盒加密过程.

首先, 对肖 - 来方案白盒加密过程的输入明文信息 X_0', X_1', X_2', X_3' 进行处理

$$X_i'' = P_i^{-1}\ X_i', i = 0, 1, 2, 3$$

其次, 第 r 轮由输入信息 $(X_{r-1}'', X_r'', X_{r+1}'', X_{r+2}'')$ 以及等价轮密钥 K_r' 与常数信息 γ_r 计算 X_{r+3}'' 的过程如下

$$X_{r+3}'' = X_{r-1}'' + T(X_r'' + X_{r+1}'' + X_{r+2}'' + K_r') + \gamma_r$$

经 32 轮计算后, 可得 $(X_{32}'', X_{33}'', X_{34}'', X_{35}'')$. 由以上分析过程可知 $X_{32}'', X_{33}'', X_{34}'', X_{35}''$ 与肖 - 来方案白盒加密结束后得到的信息 $X_{32}', X_{33}', X_{34}', X_{35}'$ 满足如下关系:

$$X_i' = P_i \cdot X_i'', i = 32, 33, 34, 35$$

故我们可以使用等价密钥 K_r' 与常数 γ_r 完成数据的白盒加解密过程.

进一步, 对于肖 - 来方案, 在白盒攻击环境下, 方案中复合仿射变换的仿射常数值 $b_r, c_r, d_{r,j}$ 公开可见, 我们可得以下方程组:

$$\begin{aligned}
b_r &= P_{r+3} \cdot Q_r^{-1} \cdot q_r + p_{r+3} + p_{r+3}' \\
c_r &= P_{r+3} \cdot P_{r-1}^{-1} \cdot p_{r-1} + p_{r+3}' \\
d_{r,0} &= E_r \cdot P_r^{-1} \cdot p_r + e_r \\
d_{r,1} &= E_r \cdot P_{r+1}^{-1} \cdot p_{r+1} + e_r \\
d_{r,2} &= E_r \cdot P_{r+2}^{-1} \cdot p_{r+2} + e_r \\
d_{r+1,0} &= E_{r+1} \cdot P_{r+1}^{-1} \cdot p_{r+1} + e_{r+1} \\
d_{r+1,1} &= E_{r+1} \cdot P_{r+2}^{-1} \cdot p_{r+2} + e_{r+1} \\
d_{r+1,2} &= E_{r+1} \cdot P_{r+3}^{-1} \cdot p_{r+3} + e_{r+1}
\end{aligned}$$

由上述方程可知, 若其中一个仿射常数已知, 则可确定其它仿射常数. 如若 p_{r+3} 已知, 则根据上述方程, 我们可依次确定 $p_{r+3}', p_{r-1}, e_{r+1}, p_{r+2}, p_{r+1}, e_r, p_r$, 进而确定所有仿射常数 $p_i, e_r, i = 0, 1, \cdots, 35, r = 1, 2, \cdots, 32$. 因此, 对于肖 - 来方案, 在所有仿射矩阵确定之后, 所有仿射常数的取值空间为 2^{32}, 即确定其中某一 32 比特仿射常数 p_i 或 e_r 后, 我们可确定所有其它仿射常数. 当所有仿射变换恢复后, 我们可获取每一轮轮密钥, 从而, 肖 - 来白盒方案的密钥搜索空间为 $61\ 200 \cdot 2^{32}$.

由以上分析, 肖 - 来方案中所有仿射变换的线性部分仅提供了不到 2^{16} 的安全强度, 同时由于方案中大量线性变换的影响, 方案的运行效率远低于 SM4 标准算法, 难以满足实际应用. 其次, 泄露两个仿射变换的复合变换的仿射常数严重影响了每个仿射常数的安全性. 方案中的仿射常数虽然独立随机选取的, 但在仿射变换的线性变换矩阵恢复后, 所有仿射常数其实相互依赖, 只需猜测其中一个 32 比特的仿射常数就可以获取所有其它仿射常数. 因此, 方案的设计中应避免泄露复合变换的仿射常数部分.

4.3 白 - 武方案分析

与肖 - 来方案相同, 白 - 武方案也使用仿射变换 $X_i' = P_i \cdot X_i + p_i$ 将 SM4 的内部状态 X_i 变换成 X_i' 从而隐藏 X_i. 相比之下, 白 - 武方案的内部解码编码复杂许多, 但以下我们将说明可用类似的方法来恢复白 - 武方案中所使用的所有仿射变换的线性矩阵部分.

白-武方案中, 公开的信息为以下查找表

$$TA_{r,i}, TB_{r,i}, TC_{r,j}, TD_{r,j}, i = 0,1,2,3,4,5, j = 0,1,2,3, r = 1,2,\cdots,32$$

记第 r 轮中查找表 $TD_{r,0}, TD_{r,1}, TD_{r,2}, TD_{r,3}$ 的 4 个 16 比特输入值分别为 x_0, x_1, x_2, x_3, 并设该 4 个查找表的 4 个 32 比特输出值为 $\boldsymbol{d}_{r,0}, \boldsymbol{d}_{r,1}, \boldsymbol{d}_{r,2}, \boldsymbol{d}_{r,3}$, 并记 $K'_{r,i} = K_{r,i} + (E_{r,i}^{-1}||F_{r,i}^{-1}) \cdot u_{r,i}$, 则

$$\begin{pmatrix} \boldsymbol{d}_{r,0} \\ \boldsymbol{d}_{r,1} \\ \boldsymbol{d}_{r,2} \\ \boldsymbol{d}_{r,3} \end{pmatrix} = \begin{pmatrix} P_{r+3} \\ P_{r+3} \\ P_{r+3} \\ P_{r+3} \end{pmatrix} \odot \begin{pmatrix} \boldsymbol{M}_0 \\ \boldsymbol{M}_1 \\ \boldsymbol{M}_2 \\ \boldsymbol{M}_3 \end{pmatrix} \odot \begin{pmatrix} S((E_{r,0}^{-1}||F_{r,0}^{-1}) \cdot x_0 + K'_{r,0}) \\ S((E_{r,1}^{-1}||F_{r,1}^{-1}) \cdot x_1 + K'_{r,1}) \\ S((E_{r,2}^{-1}||F_{r,2}^{-1}) \cdot x_2 + K'_{r,2}) \\ S((E_{r,3}^{-1}||F_{r,3}^{-1}) \cdot x_3 + K'_{r,3}) \end{pmatrix} + \begin{pmatrix} \boldsymbol{p}''_{r+3,0} \\ \boldsymbol{p}''_{r+3,1} \\ \boldsymbol{p}''_{r+3,2} \\ \boldsymbol{p}''_{r+3,3} \end{pmatrix}$$

设 X'_{r-1} 的 4 个 8 比特分量经过查找表 $TC_{r,0}, TC_{r,1}, TC_{r,2}, TC_{r,3}$ 之后的 32 比特输出值分别为 $\boldsymbol{c}_{r,0}, \boldsymbol{c}_{r,1}, \boldsymbol{c}_{r,2}, \boldsymbol{c}_{r,3}$, 则有

$$X'_{r+3} = \sum_{i=0}^{3} (\boldsymbol{c}_{r,i} + \boldsymbol{d}_{r,i})$$

我们取 X'_{r-1} 为某一固定值, 从而 $c_{r,i}$ 也为固定值, 即有 $X'_{r+3} = \sum_{i=0}^{3} \boldsymbol{d}_{r,i} + \alpha$, 其中 $\alpha = \sum_{i=0}^{3} \boldsymbol{c}_{r,i}$. 从而有

$$X'_{r+3} = P_{r+3} \cdot M \cdot \begin{pmatrix} S((E_{r,0}^{-1}||F_{r,0}^{-1}) \cdot x_0 + K'_{r,0}) \\ S((E_{r,1}^{-1}||F_{r,1}^{-1}) \cdot x_1 + K'_{r,1}) \\ S((E_{r,2}^{-1}||F_{r,2}^{-1}) \cdot x_2 + K'_{r,2}) \\ S((E_{r,3}^{-1}||F_{r,3}^{-1}) \cdot x_3 + K'_{r,3}) \end{pmatrix} + \alpha + p''_{r+3}$$

又

$$X'_{r+3} = P_{r+3} \cdot X_{r+3} + p_{r+3}$$

故

$$X_{r+3} = M \cdot \begin{pmatrix} S((E_{r,0}^{-1}||F_{r,0}^{-1}) \cdot x_0 + K'_{r,0}) \\ S((E_{r,1}^{-1}||F_{r,1}^{-1}) \cdot x_1 + K'_{r,1}) \\ S((E_{r,2}^{-1}||F_{r,2}^{-1}) \cdot x_2 + K'_{r,2}) \\ S((E_{r,3}^{-1}||F_{r,3}^{-1}) \cdot x_3 + K'_{r,3}) \end{pmatrix} + \beta$$

其中 $\beta = P_{r+3}^{-1} \cdot (\alpha + p_{r+3} + p''_{r+3}) = P_{r+3}^{-1} \cdot \alpha + p'_{r+3}$.

设 X'_{r+3} 的四个 8 比特分量经过查找表 $TA_{r+1,4}, TA_{r+1,5}, TB_{r+1,4}, TB_{r+1,5}$ 以后的 4 个 32 比特输出值为 z_0, z_1, z_2, z_3. 根据查找表 $TA_{r+1,4}, TA_{r+1,5}, TB_{r+1,4}, TB_{r+1,5}$ 的构造原理可知

$$E_{r+1}^{-1}(z_0 + z_1 + \boldsymbol{e}_{r+1,4} + \boldsymbol{e}_{r+1,5}) + F_{r+1}^{-1}(z_2 + z_3 + \boldsymbol{f}_{r+1,4} + \boldsymbol{f}_{r+1,5}) = X_{r+3}$$

从而我们有

$$E_{r+1}^{-1}(z_0 + z_1) + F_{r+1}^{-1}(z_2 + z_3) = \begin{pmatrix} M_1 & M_2 & M_2 & M_3 \\ M_3 & M_1 & M_2 & M_2 \\ M_2 & M_3 & M_1 & M_2 \\ M_2 & M_2 & M_3 & M_1 \end{pmatrix} \cdot \begin{pmatrix} f_0(x_0) \\ f_1(x_1) \\ f_2(x_2) \\ f_3(x_3) \end{pmatrix} + \beta'$$

其中 $f_i(x_i) = S((E_{r,i}^{-1}||F_{r,i}^{-1}) \cdot x_i + K'_{r,i})$,$\beta' = \beta + E_{r+1}^{-1}(\boldsymbol{e}_{r+1,4} + \boldsymbol{e}_{r+1,5}) + F_{r+1}^{-1}(\boldsymbol{f}_{r+1,4} + \boldsymbol{f}_{r+1,5})$.

设 $\beta' = \beta_0||\beta_1||\beta_2||\beta_3$, $z_0 + z_1 = \boldsymbol{z}$, $z_2 + z_3 = \mathrm{z}$, 并设 $\boldsymbol{z} = \boldsymbol{z}_0||\boldsymbol{z}_1||\boldsymbol{z}_2||\boldsymbol{z}_3, \mathrm{z} = \mathrm{z}_0||\mathrm{z}_1||\mathrm{z}_2||\mathrm{z}_3$, 其中

$\boldsymbol{z}_i, \mathrm{z}_i$ 分别表示 $\boldsymbol{z}, \mathrm{z}$ 的 8 比特分量. 取 $(x_0, x_1, x_2, x_3) = (x_0, 0, 0, 0)$, 则根据上式有

$$\begin{pmatrix} E_{r+1,0}^{-1} \cdot \boldsymbol{z}_0 + F_{r+1,0}^{-1} \cdot \mathrm{z}_0 \\ E_{r+1,1}^{-1} \cdot \boldsymbol{z}_1 + F_{r+1,1}^{-1} \cdot \mathrm{z}_1 \\ E_{r+1,2}^{-1} \cdot \boldsymbol{z}_2 + F_{r+1,2}^{-1} \cdot \mathrm{z}_2 \\ E_{r+1,3}^{-1} \cdot \boldsymbol{z}_3 + F_{r+1,3}^{-1} \cdot \mathrm{z}_3 \end{pmatrix} = \begin{pmatrix} M_1 \cdot f_0(x_0) + \gamma_0 \\ M_3 \cdot f_0(x_0) + \gamma_1 \\ M_2 \cdot f_0(x_0) + \gamma_2 \\ M_2 \cdot f_0(x_0) + \gamma_3 \end{pmatrix}$$

其中 $\gamma_0, \gamma_1, \gamma_2, \gamma_3$ 为定值 (当 X'_{r-1} 为给定值时).

我们选取上述方程组的前两个方程消去变量 $f_0(x_0)$ 得 $E_{r+1,0}$ 与 $E_{r+1,1}$ 满足如下关系:

$$M_1^{-1} \cdot (E_{r+1,0}^{-1} \cdot \boldsymbol{z}_0 + F_{r+1,0}^{-1} \cdot \mathrm{z}_0 + \gamma_0) = M_3^{-1} \cdot (E_{r+1,1}^{-1} \cdot \boldsymbol{z}_1 + F_{r+1,1}^{-1} \cdot \mathrm{z}_1 + \gamma_1)$$

对任意给定的 $(\boldsymbol{z}_0, \mathrm{z}_0, \boldsymbol{z}_1, \mathrm{z}_1)$, 搜索到 $(*, \mathrm{z}_0, *, \mathrm{z}_1)$ 的概率为 $(2^8)^2/(2^8)^4 = 1/2^{16}$, 即对给定的 $(\boldsymbol{z}_0, \mathrm{z}_0, \boldsymbol{z}_1, \mathrm{z}_1)$ 平均选取 2^{16} 个不同的 x_0 取值, 可搜索到另外一组 $(\boldsymbol{z}_0, \mathrm{z}_0, \boldsymbol{z}_1, \mathrm{z}_1)$ 的取值满足形式 $(\boldsymbol{z}'_0, \mathrm{z}_0, \boldsymbol{z}'_1, \mathrm{z}_1)$, 其中 $(\mathrm{z}'_0, \mathrm{z}'_1) \neq (\mathrm{z}_0, \mathrm{z}_1)$. 则由

$$\begin{aligned} M_1^{-1} \cdot (E_{r+1,0}^{-1} \cdot \boldsymbol{z}_0 + F_{r+1,0}^{-1} \cdot \mathrm{z}_0 + \gamma_0) &= M_3^{-1} \cdot (E_{r+1,1}^{-1} \cdot \boldsymbol{z}_1 + F_{r+1,1}^{-1} \cdot \mathrm{z}_1 + \gamma_1) \\ M_1^{-1} \cdot (E_{r+1,0}^{-1} \cdot \boldsymbol{z}'_0 + F_{r+1,0}^{-1} \cdot \mathrm{z}_0 + \gamma_0) &= M_3^{-1} \cdot (E_{r+1,1}^{-1} \cdot \boldsymbol{z}'_1 + F_{r+1,1}^{-1} \cdot \mathrm{z}_1 + \gamma_1) \end{aligned}$$

可得

$$M_1^{-1} \cdot E_{r+1,0}^{-1} \cdot \Delta_{\boldsymbol{z}_0} = M_3^{-1} \cdot E_{r+1,1}^{-1} \cdot \Delta_{\boldsymbol{z}_1}, \text{其中} \Delta_{z_0} = \boldsymbol{z}_0 + \boldsymbol{z}'_0, \Delta_{\boldsymbol{z}_1} = \boldsymbol{z}_1 + \boldsymbol{z}'_1$$

从而, 我们可用与对肖 - 来方案的分析相同的方法, 恢复矩阵 $E_{r+1,0}$, 且矩阵 $E_{r+1,0}$ 的取值空间大小为 61 200. 类似地, 当矩阵 $E_{r+1,0}$ 确定后, 我们可确定矩阵 $E_{r+1,1}, E_{r+1,2}, E_{r+1,3}$, 即完全确定矩阵 E_{r+1}, 进一步可确定矩阵 F_{r+1}.

设 $X'_{r+3} = x_0 || x_1 || x_2 || x_3$, 由 x_0, x_1, x_2, x_3 分别经过查找表 $TA_{r+1,4}, TA_{r+1,5}, TB_{r+1,4}, TB_{r+1,5}$ 后的输出值分别为 $\boldsymbol{y}_0, \boldsymbol{y}_1, \boldsymbol{y}_2, \boldsymbol{y}_3$, 则

$$\begin{aligned} \boldsymbol{y}_0 &= E_{r+1} \cdot (P_{r+3,0}^{-1} \cdot x_0 + P_{r+3}^{-1} \cdot \boldsymbol{p}_{r+3,0}) + \boldsymbol{e}_{r+1,4} \\ \boldsymbol{y}_1 &= E_{r+1} \cdot (P_{r+3,1}^{-1} \cdot x_1 + P_{r+3}^{-1} \cdot \boldsymbol{p}_{r+3,1}) + \boldsymbol{e}_{r+1,5} \\ \boldsymbol{y}_2 &= F_{r+1} \cdot (P_{r+1,4}^{-1} \cdot x_2 + P_{r+1}^{-1} \cdot \boldsymbol{p}_{r+1,2}) + \boldsymbol{f}_{r+1,4} \\ \boldsymbol{y}_3 &= F_{r+1} \cdot (P_{r+1,5}^{-1} \cdot x_3 + P_{r+1}^{-1} \cdot \boldsymbol{p}_{r+1,3}) + \boldsymbol{f}_{r+1,5} \end{aligned}$$

从而有

$$\begin{aligned} E_{r+1}^{-1} \cdot (\boldsymbol{y}_0 + \boldsymbol{y}_1) + F_{r+1}^{-1} \cdot (\boldsymbol{y}_2 + \boldsymbol{y}_3) =& P_{r+3}^{-1} \cdot x + P_{r+1}^{-1} \cdot p_{r+3} \\ &+ E_{r+1}^{-1} \cdot (\boldsymbol{e}_{r+1,4} + \boldsymbol{e}_{r+1,4}) + F_{r+1}^{-1} \cdot (\boldsymbol{f}_{r+1,4} + \boldsymbol{f}_{r+1,5}) \end{aligned}$$

由上式的差分形式

$$E_{r+1}^{-1} \cdot (\Delta_{\boldsymbol{y}_0} + \Delta_{\boldsymbol{y}_1}) + F_{r+1}^{-1} \cdot (\Delta_{\boldsymbol{y}_2} + \Delta_{\boldsymbol{y}_3}) = P_{r+3}^{-1} \cdot \Delta_x$$

通过选取不同的差分 Δ_x, 获取差分 $\Delta_{\boldsymbol{y}}$, 根据上式可确定矩阵 P_{r+3}. 类似地, 我们可以确定白 - 武方案中所有仿射变换的线性变换矩阵部分 $P_i, E_r, F_r, i = 0, 1, 2, \cdots, 35, r = 1, 2, \cdots, 32$. 所有这些仿射矩阵的取值空间大小为 61 200, 即在白盒攻击环境下, 白 - 武方案与肖 - 来方案相同, 方案中所使用的仿射变换的线性变换矩阵实质上只有 61 200 组可能的取值.

记 $K'_r = K'_{r,0} || K'_{r,1} || K'_{r,2} || K'_{r,3}, K'_{r,j} = v_{r,j} + K_{r,j}$, 其中 $v_{r,j} = (E_{r,j}^{-1} || F_{r,j}^{-1}) \cdot u_{r,j}, j = 0, 1, 2, 3$. 设

$v_r = v_{r,0}||v_{r,1}||v_{r,2}||v_{r,3}$, 易知 $v_r = E_r^{-1} \cdot e_r + F_r^{-1} \cdot f_r$. 若矩阵 $E_{r,j}, F_{r,j}$ 已知, 可由查找表

$$TD_{r,j}: \ y = P_{r+3} \cdot \boldsymbol{M}_i \cdot S((E_{r,j}^{-1}||F_{r,j}^{-1})(x + u_{r,j}) + K_{r,j}) + \boldsymbol{p}''_{r+3,j}$$

搜索确定 $K'_{r,j}$, 并在 $K'_{r,j}$ 确定后获取 $\boldsymbol{p}''_{r+3,j}$. 再由 $p''_{r+3} = \sum_{j=0}^{3} \boldsymbol{p}''_{r+3,j}$ 我们可得到 p''_{r+3}.

当所有变换矩阵 $P_i, E_r, F_r, i = 0, 1, \cdots, 35, r = 1, 2, \cdots, 32$ 均已确定时, 由查找表 $TA_{r,i}, TB_{r,i}, TC_{r,j}, i = 0, 1, 2, 3, 4, 5, j = 0, 1, 2, 3$ 可得仿射常数满足以下方程:

$$\begin{aligned}
P_{r-1}^{-1} \cdot \boldsymbol{p}_{r-1,j} + P_{r+3}^{-1} \cdot \boldsymbol{p}'_{r+3,j} &= \gamma_{r,j}, && j = 0, 1, 2, 3 \\
P_r^{-1} \cdot \boldsymbol{p}_{r,j} + E_r^{-1} \cdot \boldsymbol{e}_{r,j} &= \alpha_{r,j}, && j = 0, 1 \\
P_{r+1}^{-1} \cdot \boldsymbol{p}_{r+1,j} + E_r^{-1} \cdot \boldsymbol{e}_{r,j+2} &= \alpha_{r,j+2}, && j = 0, 1 \\
P_{r+2}^{-1} \cdot \boldsymbol{p}_{r+2,j} + E_r^{-1} \cdot \boldsymbol{e}_{r,j+4} &= \alpha_{r,j+4}, && j = 0, 1 \\
P_r^{-1} \cdot \boldsymbol{p}_{r,j+2} + F_r \cdot \boldsymbol{f}_{r,j} &= \beta_{r,j}, && j = 0, 1 \\
P_{r+1}^{-1} \cdot \boldsymbol{p}_{r+1,j+2} + F_r^{-1} \cdot \boldsymbol{f}_{r,j+2} &= \beta_{r,j+2}, && j = 0, 1 \\
P_{r+2}^{-1} \cdot \boldsymbol{p}_{r+2,j+2} + F_r^{-1} \cdot \boldsymbol{f}_{r,j+4} &= \beta_{r,j+4}, && j = 0, 1
\end{aligned}$$

其中 $\alpha_{r,i}, \beta_{r,i}, \gamma_{r,j} (i = 0, 1, \cdots, 5, j = 0, 1, 2, 3)$ 均可由相应的查找表直接确定. 由 $\sum_{j=0}^{3} \boldsymbol{p}_{i,j} = p_i$, $\sum_{j=0}^{3} \boldsymbol{e}_{r,j} = e_{r,j}, \sum_{j=0}^{3} \boldsymbol{f}_{r,j} = f_{r,j}, v_r = E_r^{-1} \cdot e_r + F_r^{-1} \cdot f_r, K'_r = v_r + K_r$ 及 $p_{r+3} = p'_{r+3} + p''_{r+3}$, 整理上述方程得

$$P_{r-1}^{-1} \cdot p_{r-1} + P_{r+3}^{-1} \cdot p_{r+3} = \gamma_r \tag{1}$$

$$P_r^{-1} \cdot p_r + P_{r+1}^{-1} \cdot p_{r+1} + P_{r+2}^{-1} \cdot p_{r+2} + K_r = K'_r + \alpha_r + \beta_r \tag{2}$$

其中 $\alpha_r = \sum_{j=0}^{5} \alpha_{r,j}, \beta_r = \sum_{j=0}^{5} \beta_{r,j}, \gamma'_r = \sum_{j=0}^{3} \gamma_{r,j}, \gamma_r = \gamma'_r + P_{r+3}^{-1} \cdot p''_{r+3}$.

记 $K''_r = K'_r + \alpha_r + \beta_r$, 并称 K''_r 为白-武方案白盒加密过程的等价轮密钥. 由以上分析, 所有等价轮密钥 $K''_1, K''_2, \cdots, K''_{32}$ 的可能取值空间大小为 61 200. 通过等价轮密钥 K''_r 及常数 γ_r 可以使用以下与肖-来方案完全相同的方式完成白-武方案的白盒加密过程:

首先, 对白-武方案白盒加密过程的输入明文信息 X'_0, X'_1, X'_2, X'_3 进行处理

$$X''_i = P_i^{-1} \cdot X'_i, i = 0, 1, 2, 3$$

其次, 第 r 轮由输入信息 $X''_{r-1}, X''_r, X''_{r+1}, X''_{r+2}$ 以及等价轮密钥 K''_r 与常数信息 γ_r 计算 X''_{r+3} 的过程如下

$$X''_{r+3} = X''_{r-1} + T(X''_r + X''_{r+1} + X''_{r+2} + K''_r) + \gamma_r$$

经 32 轮计算后, 可得 $(X''_{32}, X''_{33}, X''_{34}, X''_{35})$. 且 $X''_{32}, X''_{33}, X''_{34}, X''_{35}$ 与白-武方案白盒加密结束后得到的信息 $X'_{32}, X'_{33}, X'_{34}, X'_{35}$ 满足如下关系

$$X'_i = P_i \cdot X''_i, i = 32, 33, 34, 35$$

进一步, 对于白-武方案, 当获取所有仿射矩阵后, 由式 (1) 和 (2) 可知, 若我们猜测连续 4 个 32 比特向量 $p_{r-1}, p_r, p_{r+1}, p_{r+2}$, 就可获取仿射常数 $p_0, p_1, \cdots, p_{35}$ 以及轮密钥 $K_1, K_2, \cdots, K_{32}$. 从而白-武方案中密钥及外部编码的取值空间大小为 $61\ 200 \cdot 2^{128}$. 由对白-武方案的分析可知, 复杂化内部编码解码过程并不能提高算法的安全性. 其次, 通过矩阵分块或对随机数进行拆分的方法来提高白盒多样性的策略对算法的安全性并无影响. 方案的安全性主要由外部编码中的仿射常数 p_0, p_1, p_2, p_3 的安全性决定, 当仿射常数泄露时, 算法不再具有安全性.

5 总结

作为我国广泛应用的商用分组密码算法, SM4 的白盒化研究对移动终端的安全防护具有重要的实用价值. 本文通过研究两个 SM4 的白盒方案, 优化了现有分析方法, 同时通过一些基本的观察为设计新型白盒化方案提供了参考和建议. 此外, 我们发现当前的白盒化方案普遍效率不高, 远远不能满足实用需求. 在保障合理安全强度的前提下提高白盒方案的效率, 将是白盒密码研究中的重要方向.

致谢

感谢来学嘉老师及其团队在本文写作过程中给予的指导和帮助! 感谢审稿专家提出的意见, 这对本文内容的提高及我们未来的研究都有很大的帮助!

References

[1] CHOW S, EISEN P, JOHNSON H, et al. White-box cryptography and an AES implementation[C]. In: Selected Areas in Cryptography—SAC 2002. Springer Berlin Heidelberg, 2003: 250–270. [DOI: 10.1007/3-540-36492-7_17]

[2] CHOW S, EISEN P, JOHNSON H, et al. A white-box DES implementation for DRM applications[C]. In: Digital Rights Management—DRM 2002. Springer Berlin Heidelberg, 2003: 1–15. [DOI: 10.1007/978-3-540-44993-5_1]

[3] BILLET O, GILBERT H, ECH-CHATBI C. Cryptanalysis of a white box AES implementation[C]. In: Selected Areas in Cryptography—SAC 2004. Springer Berlin Heidelberg, 2004: 227–240. [DOI: 10. 1007/978-3-540-30564-4_16]

[4] TOLHUIZEN L. Improved cryptanalysis of an AES implementation[C]. In: Proceedings of the 33rd WIC Symposium on Information Theory. WIC (Werkgemeenschap voor Informatie-en Communicatietheorie), 2012: 68–70. [DOI: 10.1.1.913.5807]

[5] LEPOINT T, RIVAIN M, MULDER Y D, et al. Two attacks on a white-box AES implementation[C]. In: Selected Areas in Cryptography—SAC 2013. Springer Berlin Heidelberg, 2013: 265–285. [DOI: 10. 1007/978-3-662-43414-7_14]

[6] MULDER Y D, ROELSE P, PRENEEL B. Revisiting the BGE attack on a white-box AES implementation[J]. IACR Cryptology ePrint Archive, 2013: 2013/450. http://eprint.iacr. org/2013/450.

[7] XIAO Y Y, LAI X J. A secure implementation of white-box AES[C]. In: Proceedings of 2009 2nd International Conference on Computer Science and Its Applications—CSA 2009. IEEE, 2009: 1–6. [DOI: 10.1109/CSA. 2009. 5404239]

[8] XIAO Y Y. White-box Cryptography and Implementations of AES and SMS4[D]. [Master Dissertation]. Shanghai: Shanghai Jiao Tong University, 2010.
肖雅莹. 白盒密码及 AES 与 SMS4 算法的实现 [D]. [硕士论文]. 上海: 上海交通大学, 2010.

[9] KARROUMI M. Protecting white-box AES with dual ciphers[C]. In: Information Security and Cryptology—ICISC 2010. Springer Berlin Heidelberg, 2011: 278–291. [DOI: 10. 1007/978-3-642-24209-0_19]

[10] LUO R, LAI X J, YOU R. A new attempt of white-box AES implementation[C]. In: Proceedings of 2014 International Conference on Security, Pattern Analysis, and Cybernetics—SPAC 2014. IEEE, 2014: 423–429. [DOI: 10.1109/SPAC. 2014. 6982727]

[11] LINK H E, NEUMANN W D. Clarifying obfuscation: Improving the security of white-box DES[C]. In: Proceedings of 2005 International Conference on Information Technology: Coding and Computing—ITCC 2005. IEEE, 2005: 679–684. [DOI: 10.1109/ITCC.2005.100]

[12] MICHIELS W, GORISSEN P, HOLLMANN H D L. Cryptanalysis of a generic class of white-box implementations[C]. In: Selected Areas in Cryptography—SAC 2008. Springer Berlin Heidelberg, 2009: 414–428. [DOI: 10.1007/978-3-642-04159-4_27]

[13] MULDER Y D, ROELSE P, PRENEEL B. Cryptanalysis of the Xiao-Lai white-box AES implementation[C]. In: Selected Areas in Cryptography—SAC 2012. Springer Berlin Heidelberg, 2013: 34–49. [DOI: 10.1007/978-3-642-35999-6_3]

[14] LEPOINT T, RIVAIN M. Another nail in the Coffin of white-box AES implementations[J]. IACR Cryptology ePrint Archive, 2013: 2013/455. https://eprint.iacr.org/2013/455.

[15] BAEK C H, CHEON J H, HONG H. White-box AES implementation revisited[J]. IACR Cryptology ePrint Archive, 2014: 2014/688. https://eprint. iacr. org/2014/688.

[16] JACOB M, BONEH D, FELTEN E. Attacking an obfuscated cipher by injecting faults[C]. In: Digital Rights

Management—DRM 2002. Springer Berlin Heidelberg, 2003: 16–31. [DOI: 10.1007/978-3-540-44993-5_2]

[17] WYSEUR B, MICHIELS W, GORISSEN P, et al. Cryptanalysis of white-box DES implementations with arbitrary external encodings[C]. In: Selected Areas in Cryptography—SAC 2007. Springer Berlin Heidelberg, 2007: 264–277. [DOI: 10.1007/978-3-540-77360-3_17]

[18] Standardization Administration. Information Security Technology—SM4 Block Cipher Algorthm[EB/OL]. 2016. http://c.gb688.cn/bzgk/gb/showGb?type=online&hcno=7803DE42D3BC5E80B0C3E5D8E873D56A.
国家标准化管理委员会. 信息安全技术 SM4 分组密码算法 [EB/OL]. 2016. http://c.gb688.cn/bzgk/gb/showGb?type=online&hcno=7803DE42D3BC5E80B0C3E5D8E873D56A.

[19] XIAO Y Y, LAI X J. White-box crytography and implementation of SMS4[C]. In: Proceedings of the 2009 CACR Annual Meeting. Beijing: Science Press, 2009: 24–34.
肖雅莹, 来学嘉. 白盒密码及 SMS4 算法的白盒实现 [C]. 中国密码学会 2009 年会. 北京: 科学出版社, 2009: 24–34.

[20] LIN T T, LAI X J. Efficient attack to white-box SMS4 implementation[J]. Journal of Software, 2013, 24(9): 2238–2249. [DOI: 10.3724/SP.J.1001.2013.04356]
林婷婷, 来学嘉. 对白盒 SMS4 实现的一种有效攻击 [J]. 软件学报, 2013, 24(9): 2238–2249. [DOI: 10.3724/SP.J.1001.2013.04356]

[21] BAI K P, WU C K. A secure white-box SM4 implementation[J]. Security and Communication Networks, 2016, 9(10): 996–1006. [DOI: 10.1002/sec.1394]

[22] SHI Y, WEI W J, HE Z J. A lightweight white-box symmetric encryption algorithm against node capture for WSNs[J]. Sensors, 2015, 15(5): 11928–11952. [DOI: 10. 3390/s150511928]

[23] LIU F, JI W, DING J T, et al. Analysis of the SMS4 block cipher[C]. In: Information Security and Privacy—ACISP 2007. Springer Berlin Heidelberg, 2007: 158–170. [DOI: 10.1007/978-3-540-73458-1_13]

作者信息

潘文伦 (1988–), 湖北天门人, 工程师. 主要研究领域为白盒密码的分析与设计.
pan.wenlun@westone.com.cn

秦体红 (1981–), 安徽安庆人, 工程师. 主要研究领域为密码算法的高速、安全实现.
qthxnjd@163.com

贾音 (1993–), 河南洛阳人, 工程师. 主要研究领域为白盒密码的设计与分析、区块链技术.
jy09091001@163.com

张立廷 (1982–), 山东即墨人, 副研究员. 主要研究领域为可证明安全理论及密码应用.
liting.zhang@hotmail.com