Cryptanalysis of a Type of White-Box Implementations of the SM4 Block Cipher

Jiqiang Lu, Jingyu Li

***Abstract*—The SM4 block cipher was first released in 2006 as SMS4 used in the Chinese national standard WAPI, and became a Chinese national standard in 2016 and an ISO international standard in 2021. White-box cryptography aims primarily to protect the secret key used in a cryptographic software implementation in the white-box scenario that assumes an attacker to have full access to the execution environment and execution details of an implementation. Since white-box cryptography has many real-life applications nowadays, a few white-box implementations of the SM4 block cipher has been proposed with its increasingly wide use, among which a type of constructions is dominated, that use an affine (or extremely even linear) diagonal block encoding to protect the original output of an SM4 round function and use the encoding or its inverse to protect the original input of the S-box layer of the next round, such as Xiao and Lai's implementation in 2009, Shang's implementation in 2016, Yao and Chen's and Wu et al.'s implementations in 2020. In this paper, we show that this type of white-box SM4 constructions is rather insecure against collision-based attacks, by devising attacks on Xiao and Lai's, Shang's, Yao and Chen's and Wu et al.'s implementations with a time complexity of respectively about $2^{19.4}$, $2^{35.6}$, $2^{19.4}$ and $2^{17.1}$ to recover a round key, and thus their security is much lower than previously published or expected. Thus, such white-box SM4 constructions should be avoided unless being enhanced somehow.**

***Index Terms*—White-box cryptography, SM4 (SMS4) block cipher, collision attack.**

I. Introduction

In 2002, Chow et al. [9], [10] introduced white-box cryptography and proposed white-box implementations of the AES [28] and DES [29] block ciphers. White-box cryptography works under the white-box security model, which assumes an attacker has full access to the execution environment and execution details (such as intermediate values, CPU calls, memory registers, etc) of a software implementation, giving the attacker more power than the black-box and grey-box security models. Nowadays, white-box cryptography has many real-life application scenarios like TV boxes, mobile phones and game consoles, and some white-box cryptography solutions have been in use.

The primary security threat for white-box cryptography is key extraction attack, which aims to extract the key used in white-box implementation. Chow et al.'s white-box AES implementation has been cryptanalysed extensively [6], [19], [25], [32], and the main attack results are as follows. In 2004, Billet et al. [6] presented an attack with a time complexity of 2^{30} (referred to below as BGE attack). In 2013, Lepoint et al. [19] improved the BGE attack to have a time complexity of 2^{22}, and presented a collision-based attack with a time complexity of 2^{22}. There are also a few attacks [15], [16], [22], [35] on Chow et al.'s white-box DES implementation. On the other hand, a number of different white-box implementation designs have been proposed [1], [3], [8], [17], [24], [36], but almost all of them have been broken with a practical or semi-practical time complexity [3], [11], [19], [26], [27]. Generally speaking, it has been well understood that the line of white-box implementation for an existing cryptographic algorithm is hardly possible to achieve the full security under the black-box model, but it is expected that it can still provide some protection with realistic significance.

The SM4 block cipher was first released in 2006 as the SMS4 [12] block cipher used in the Chinese national standard WAPI (WLAN Authentication and Privacy Infrastructure), which has a 128-bit block length and a 128-bit user key with a total of 32 rounds. SMS4 became a Chinese cryptographic industry standard in 2012, labeled with SM4, which then became a Chinese national standard [13] in 2016 and an ISO international standard in 2021 [14]. The main white-box implementation results of SMS4/SM4 are as follows. In 2009, Xiao and Lai [37] proposed the first white-box SM4 implementation in a relatively traditional way with a series of lookup tables and affine transformations. In 2013, Lin and Lai [20] attacked Xiao and Lai's white-box SM4 implementation with a time complexity of around 2^{47} by combining the BGE attack with other techniques like differential cryptanalysis [5]. In 2015, Shi et al. [31] proposed a lightweight white-box SM4 implementation based on the idea of dual cipher [4]. In 2016, Shang [30] improved Xiao and Lai's white-box SM4 implementation mainly by merging two individual lookup tables for two S-boxes into a larger whole, and got a security complexity of around 2^{48}; and Bai and Wu [2] proposed a white-box SM4 implementation with an S-box input being divided into two shares. In 2018, Lin et al. [21] applied Biryukov et al.'s affine equivalence technique [7] to

Manuscript received MONTH DAY, YEAR; revised MONTH DAY, YEAR. This work was supported by National Natural Science Foundation of China (No. 61972018) and Guangxi Key Laboratory of Cryptography and Information Security (No. GCIS202102). This is an extended version of the paper appeared in Proceedings of ISC 2021 — The 24th Information Security Conference [23]. In this extended version, we corrected and revised the phase of how to recover the round key and the part of time complexity analysis for Yao and Chen's and Xiao and Lai's implementations, and cryptanalysed two other white-box SM4 implementations, namely Shang's and Wu et al.'s implementations. (Corresponding author: Jiqiang Lu.)

Jiqiang Lu is with School of Cyber Science and Technology, Beihang University, Beijing 100191, China, with Guangxi Key Laboratory of Cryptography and Information Security, Guilin 541004, China, and also with Hangzhou Innovation Institute, Beihang University, Hangzhou 310053, China (e-mail: lvjiqiang@buaa.edu.cn)

Jingyu Li is with School of Cyber Science and Technology, Beihang University, Beijing 100191, China (e-mail: lijingyu98@buaa.edu.cn))

Digital Object Identifier XX.XXXX/TIFS.XXXX.XXXXXXX

attack Shi et al.'s white-box SM4 implementation with a time complexity of 2^{49}. In 2020, Yao and Chen [38] proposed a white-box SM4 implementation with some original internal states expanded by dummy states under the control of a secret random number, and got the lowest attack complexity of about 2^{51} among a variety of attack techniques; and Wu et al. [34] proposed a white-box SM4 implementation with lookup tables and linear transformations, and showed it was resistant against BGE attack. In 2021, Wang et al. [33] applied Lepoint et al.'s collision-based idea to attack Shi et al.'s white-box SM4 implementation with a time complexity of around 2^{23}.

In this paper, we are concerned with Xiao and Lai's, Shang's, Yao and Chen's and Wu et al.'s white-box SM4 implementations, which are more or less different one another from a structural view but fundamentally all employ the construction method that uses an affine (or extremely even linear) diagonal block encoding to protect the original output of an SM4 round function and uses the inverse of the encoding to protect the original input of the S-box layer of the next round. Especially, we focus on Yao and Chen's white-box SM4 implementation due to its representativeness, and apply Lepoint et al.'s collision-based idea to devise an attack with a total time complexity of about $2^{19.4}$; in particular, we first find that the effect of those dummy states can be bypassed without any workload by devising an appropriate collision function, then we find a trick to recover the linear parts of the concerned affine output encodings at ease, and finally we use another trick to the collision function to recover the constant parts of the affine output encodings and the round key. The attack significantly reduces the security of Yao and Chen's white-box SM4 implementation, from the designers' estimated semi-practical level 2^{51} to a very practical level. The attack is likewise applied to Xiao and Lai's white-box SM4 implementation with a time complexity of about $2^{19.4}$ too, reducing much the best previously published attack complexity of 2^{32} based on affine equivalence technique, and is applied to Shang's and Wu et al.'s white-box SM4 implementations with a time complexity of about $2^{35.6}$ and $2^{17.1}$, respectively, with more or less modifications due to their respective specifications. These suggest that their security is much lower than previously published or expected, their realistic significance is reduced, and such white-box SM4 constructions should be avoided unless being enhanced somehow.

The remainder of the paper is organised as follows. We describe the notation and the SM4 block cipher in the next section, and present our attacks on Yao and Chen's, Xiao and Lai's, Shang's and Wu et al.'s white-box SM4 implementations in Sections III to VI, respectively. Section VII concludes this paper.

II. Preliminaries

In this section, we give the notation used throughout this paper, and briefly describe the SM4 block cipher.

A. Notation

We use the following notation throughout this paper.

$\oplus$ bitwise exclusive OR (XOR)
$\gg$ right shift of a bit string
$\lll$ left rotation of a bit string
$||$ bit string concatenation
$\circ$ functional composition

B. The SM4 Block Cipher

SM4 [12], [13] is a generalised Feistel cipher with 32 rounds, a 128-bit block size and a 128-bit key length. Denote by $(X_i, X_{i+1}, X_{i+2}, X_{i+3})$ the 128-bit input to the i-th round, by rk_i the 32-bit i-th round key, where $X_i \in \mathrm{GF}(2)^{32}$ and $i = 0, 1, \ldots, 31$.

Define the nonlinear function $\tau : \mathrm{GF}(2)^{32} \to \mathrm{GF}(2)^{32}$ that applies the same 8-bit S-box $\mathbf{S}$ four times in parallel as

$$x \mapsto \big(\mathbf{S}(x_{[31\ldots24]}), \mathbf{S}(x_{[23\ldots16]}), \mathbf{S}(x_{[15\ldots8]}), \mathbf{S}(x_{[7\ldots0]})\big);$$

and define the linear function $\mathbf{L} : \mathrm{GF}(2)^{32} \to \mathrm{GF}(2)^{32}$ as

$$x \mapsto x \oplus (x \lll 2) \oplus (x \lll 10) \oplus (x \lll 18) \oplus (x \lll 24). \quad (1)$$

Then, the invertible transformation $\mathbf{T} : \mathrm{GF}(2)^{32} \times \mathrm{GF}(2)^{32} \to \mathrm{GF}(2)^{32}$ is defined to be

$$(x, rk_i) \to \mathbf{L}(\tau(x \oplus rk_i)),$$

and the round function $\mathbf{F} : \mathrm{GF}(2)^{128} \times \mathrm{GF}(2)^{32} \to \mathrm{GF}(2)^{128}$ under round key rk_i is

$$\begin{aligned}&((X_i, X_{i+1}, X_{i+2}, X_{i+3}), rk_i) \mapsto (X_{i+1}, X_{i+2}, X_{i+3},\\ &X_i \oplus \mathbf{T}\big(X_{i+1} \oplus X_{i+2} \oplus X_{i+3}, rk_i\big)\big). \quad (2)\end{aligned}$$

The encryption procedure of SM4, as depicted in Fig. 1, consists of the 32 round functions $\mathbf{F}$'s and finally a reverse transformation $R : \mathrm{GF}(2)^{128} \to \mathrm{GF}(2)^{128}$ defined as

$$(X_{32}, X_{33}, X_{34}, X_{35}) \mapsto (X_{35}, X_{34}, X_{33}, X_{32}).$$

The decryption process of SM4 is the same as the encryption process, except that the round keys are used in the reverse order. We refer the reader to [12], [13] for detailed specifications.

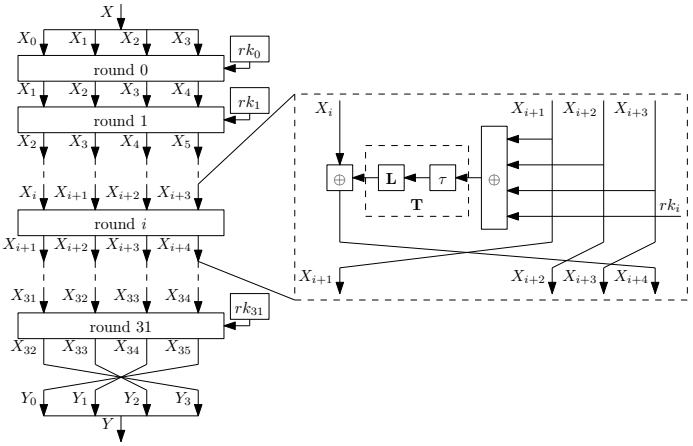

Figure 1. SM4 encryption procedure

Particularly, it is easy and worthy to note that the linear transformation $\mathbf{L}$ (as described in Eq. (1)) of SM4 can also be represented as an invertible 32×32-bit matrix

$$\begin{bmatrix} B_1 & B_2 & B_2 & B_3 \\ B_3 & B_1 & B_2 & B_2 \\ B_2 & B_3 & B_1 & B_2 \\ B_2 & B_2 & B_3 & B_1 \end{bmatrix}, \tag{3}$$

with B_1, B_2 and B_3 being invertible 8×8-bit block matrices.

Let x_0, x_1, x_2, x_3 be four byte variables, represent $\mathbf{L}$ as four 32×8-bit matrices $\begin{bmatrix} \mathbf{L}_0 & \mathbf{L}_1 & \mathbf{L}_2 & \mathbf{L}_3 \end{bmatrix}$, and define

$$\begin{aligned} \mathbf{L}_0(x) &= x \cdot \begin{bmatrix} B_1 & B_3 & B_2 & B_2 \end{bmatrix}^T, \\ \mathbf{L}_1(x) &= x \cdot \begin{bmatrix} B_2 & B_1 & B_3 & B_2 \end{bmatrix}^T, \\ \mathbf{L}_2(x) &= x \cdot \begin{bmatrix} B_2 & B_2 & B_1 & B_3 \end{bmatrix}^T, \\ \mathbf{L}_3(x) &= x \cdot \begin{bmatrix} B_3 & B_2 & B_2 & B_1 \end{bmatrix}^T, \end{aligned}$$

then $\mathbf{L}(x_0||x_1||x_2||x_3) = \mathbf{L}_0(x_0) \oplus \mathbf{L}_1(x_1) \oplus \mathbf{L}_2(x_2) \oplus \mathbf{L}_3(x_3)$.

III. Collision-Based Attack on Yao and Chen's White-Box SM4 Implementation

In this section, we first describe Yao and Chen's white-box SM4 implementation, and then present our attack on it.

A. Yao and Chen's White-Box SM4 Implementation

Yao and Chen's white-box SM4 implementation [38] is based on internal state expansion, particularly, the 32×32-bit matrix representation described in Eq. (3) of the linear transformation $\mathbf{L}$ is expanded to the following 64×64-bit matrix $\widehat{\mathbf{L}}$ with the 8×8-bit zero matrix $\mathbf{0}$:

$$\widehat{\mathbf{L}} = \begin{bmatrix} B_1 & \mathbf{0} & B_2 & \mathbf{0} & B_2 & \mathbf{0} & B_3 & \mathbf{0} \\ \mathbf{0} & B_1 & \mathbf{0} & B_2 & \mathbf{0} & B_2 & \mathbf{0} & B_3 \\ B_3 & \mathbf{0} & B_1 & \mathbf{0} & B_2 & \mathbf{0} & B_2 & \mathbf{0} \\ \mathbf{0} & B_3 & \mathbf{0} & B_1 & \mathbf{0} & B_2 & \mathbf{0} & B_2 \\ B_2 & \mathbf{0} & B_3 & \mathbf{0} & B_1 & \mathbf{0} & B_2 & \mathbf{0} \\ \mathbf{0} & B_2 & \mathbf{0} & B_3 & \mathbf{0} & B_1 & \mathbf{0} & B_2 \\ B_2 & \mathbf{0} & B_2 & \mathbf{0} & B_3 & \mathbf{0} & B_1 & \mathbf{0} \\ \mathbf{0} & B_2 & \mathbf{0} & B_2 & \mathbf{0} & B_3 & \mathbf{0} & B_1 \end{bmatrix}.$$

Represent the matrix $\widehat{\mathbf{L}}$ as four 64×16-bit matrices, that is, $\widehat{\mathbf{L}} = \begin{bmatrix} \widehat{\mathbf{L}}_0 & \widehat{\mathbf{L}}_1 & \widehat{\mathbf{L}}_2 & \widehat{\mathbf{L}}_3 \end{bmatrix}$. Then, an encryption round of Yao and Chen's white-box SM4 implementation consists of the following three parts according to Eq. (2), as depicted in Fig. 2. Note first that X_l is the corresponding original value protected with an affine output encoding $P_l(x) = A_l \cdot x \oplus a_l$, where x is a 32-bit variable, the linear part A_l is a secret (randomly generated) general invertible 32×32-bit matrix, the constant part a_l is a secret (randomly generated) 32-bit vector, and $l = 0, 1, \cdots, 35$.

1) Part 1 – Implement $X_{i+1} \oplus X_{i+2} \oplus X_{i+3} \mapsto X$: In order to obtain the original value of $X_{i+1} \oplus X_{i+2} \oplus X_{i+3}$ from the protected forms X_{i+1}, X_{i+2} and X_{i+3}, apply first the inverses P_{i+1}^{-1}, P_{i+2}^{-1} and P_{i+3}^{-1} of the three output encodings respectively to X_{i+1}, X_{i+2} and X_{i+3}, followed by an identical diagonal output encoding $E_i = \text{diag}(E_{i,0}, E_{i,1}, E_{i,2}, E_{i,3})$, where $E_{i,0}, E_{i,1}, E_{i,2}, E_{i,3}$ are four general invertible 8×8-bit affine transformations ($i = 0, 1, \ldots, 31$).

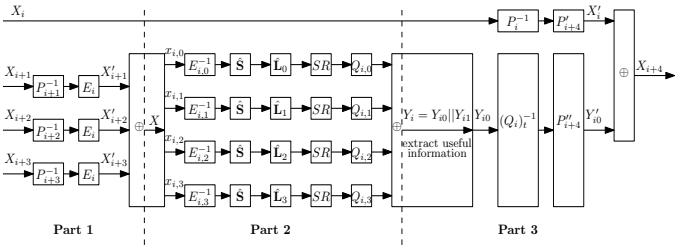

Figure 2. An encryption round of Yao and Chen's white-box SM4 implementation

This part can be summarised as

$$\begin{aligned} X'_{i+j} &= E_i \circ P_{i+j}^{-1}(X_{i+j}), \quad j = 1, 2, 3; \\ X &= X'_{i+1} \oplus X'_{i+2} \oplus X'_{i+3}, \end{aligned}$$

where X is a 32-bit variable. Observe that the final result of this part $X = E_i \circ (P_{i+1}^{-1}(X_{i+1}) \oplus P_{i+2}^{-1}(X_{i+2}) \oplus P_{i+3}^{-1}(X_{i+3}))$ is the original value of $X_{i+1} \oplus X_{i+2} \oplus X_{i+3}$ protected with the output encoding E_i in such a way that its four bytes are protected respectively with the four 8-bit encodings $E_{i,0}, E_{i,1}, E_{i,2}, E_{i,3}$.

2) Part 2 – Implement $\mathbf{T}(X, rk_i) \mapsto Y_i (= Y_{i0}||Y_{i1})$: The input X of the second part is the output of the first part, represent X as 4 bytes $X = (x_{i,0}, x_{i,1}, x_{i,2}, x_{i,3})$, and represent the round key rk_i as 4 bytes $rk_i = (rk_{i,0}, rk_{i,1}, rk_{i,2}, rk_{i,3})$, where $i = 0, 1, \ldots, 31$. Next, construct four lookup tables that map from 8-bit input to 64-bit output each, as follow:

$$\begin{aligned} Table_{i,0} &= G_{i,0} \circ \widehat{\mathbf{L}}_0[\widehat{\mathbf{S}}(E_{i,0}^{-1}(x_{i,0}), rk_{i,0}, \alpha_{i,0})_{t_{i,0}}], \\ Table_{i,1} &= G_{i,1} \circ \widehat{\mathbf{L}}_1[\widehat{\mathbf{S}}(E_{i,1}^{-1}(x_{i,1}), rk_{i,1}, \alpha_{i,1})_{t_{i,1}}], \\ Table_{i,2} &= G_{i,2} \circ \widehat{\mathbf{L}}_2[\widehat{\mathbf{S}}(E_{i,2}^{-1}(x_{i,2}), rk_{i,2}, \alpha_{i,2})_{t_{i,2}}], \\ Table_{i,3} &= G_{i,3} \circ \widehat{\mathbf{L}}_3[\widehat{\mathbf{S}}(E_{i,3}^{-1}(x_{i,3}), rk_{i,3}, \alpha_{i,3})_{t_{i,3}}], \end{aligned}$$

where

- $\alpha_{i,j}$ is an 8-bit random number ($j = 0, 1, 2, 3$);
- $\widehat{\mathbf{L}}_j$ is the corresponding j-th 64×16-bit part of $\widehat{\mathbf{L}}$;
- $(t_{i,0}, t_{i,1}, t_{i,2}, t_{i,3})$ $(t_{i,j} \in \{0, 1\})$ is a 4-bit random vector, and

$$\widehat{\mathbf{S}}(E_{i,j}^{-1}(x_{i,j}), rk_{i,j}, \alpha_{i,j})_{t_{i,j}} = \begin{cases} \mathbf{S}(E_{i,j}^{-1}(x_{i,j}) \oplus rk_{i,j}) \;||\; \mathbf{S}(E_{i,j}^{-1}(x_{i,j}) \oplus \alpha_{i,j}), & t_{i,j} = 0; \\ \mathbf{S}(E_{i,j}^{-1}(x_{i,j}) \oplus \alpha_{i,j}) \;||\; \mathbf{S}(E_{i,j}^{-1}(x_{i,j}) \oplus rk_{i,j}), & t_{i,j} = 1. \end{cases}$$

  That is, the $\widehat{\mathbf{S}}$ operation is constructed by expanding the original $\mathbf{S}$ operation with a dummy $\mathbf{S}$ operation under the control of the 1-bit $t_{i,j}$ parameter.
- $G_{i,j}$ is the composition of a shift matrix SR and an output encoding $Q_{i,j}$. The shift matrix SR transforms the expanded 64-bit value after $\widehat{\mathbf{L}}_j$ into such a 64-bit value that the former half is the original 32-bit part (without expansion) and the latter half consists only of some dummy bits. $Q_{i,j}$ is of the affine form $Q_{i,j}(x) = L_Q \cdot x \oplus C_{Q_{i,j}}$, here x is a 64-bit variable, the linear part L_Q is a block diagonal matrix being composed of eight 8×8-bit matrices, and the constant part $C_{Q_{i,j}}$ consists of eight concatenated 8-bit vectors.

The final output of this part is the XOR of the four 64-bit outputs of the four lookup tables, which is denoted by $Y_i = Y_{i0}||Y_{i1}$ with Y_{i0} being supposed to be the original useful 32-bit value.

3) Part 3 – Implement $Y_{i0} \oplus X_i \mapsto X_{i+4}$: This part first extracts the original useful 32-bit value from the 64-bit expanded output of the second part, and then calculates X_{i+4}, as follows.

$$\begin{aligned} Y'_{i0} &= P''_{i+4} \circ (Q_i)_t^{-1}(Y_{i0}), \\ X'_i &= P'_{i+4} \circ P_i^{-1}(X_i), \\ X_{i+4} &= Y'_{i0} \oplus X'_i, \end{aligned}$$

where $(Q_i)_t^{-1}$ represents the corresponding part of the inverse of the encodings $L_Q \cdot x \oplus (C_{Q_{i,0}} \oplus C_{Q_{i,1}} \oplus C_{Q_{i,2}} \oplus C_{Q_{i,3}})$ of the second part, and P'_{i+4} and P''_{i+4} are new affine output encodings of the forms $P'_{i+4}(x) = P_{i+4} \oplus a'_{i+4}$ and $P''_{i+4}(x) = P_{i+4} \oplus a''_{i+4}$, respectively, so that X_{i+4} is a protected form with an affine output encoding $P_{i+4}(x) = A_{i+4} \cdot x \oplus a_{i+4}$, like X_i.

As a result, the whole white-box SM4 implementation can be obtained by iterating the above process for all the 32 rounds with possibly independent encodings.

Yao and Chen analysed its security against a variety of attack techniques like BGE, and got that the attack complexity using affine equivalence technique was 2^{97}, and the lowest attack complexity was 2^{51} among all used attack techniques.

B. Attacking Yao and Chen's White-Box SM4 Implementation

In this subsection, we apply Lepoint et al.'s collision-based idea to attack Yao and Chen's white-box SM4 implementation with a time complexity of about $2^{19.4}$. AES and SM4 have different structures, and Yao and Chen's white-box SM4 implementation is distinct from Chow et al.'s white-box AES implementation: there are dummy states with indeterminate positions and the encoding used in X_{i+4} involves a general 32×32-bit matrix, which does not allow us to apply Lepoint et al.'s attack idea efficiently within one round, as for Chow et al.'s white-box AES implementation. However, after a detailed investigation we find an appropriate collision function by considering two consecutive rounds in Yao and Chen's white-box SM4 implementation, plus a trick that can recover the linear parts of the concerned encodings, to bypass the effects due to the dummy states and etc.

1) Devising a Collision Function: As illustrated in Fig. 3 at a high level, the collision function used in our attack takes as input the two 32-bit input parameters $(x_{i,0}||x_{i,1}||x_{i,2}||x_{i,3}, X_i)$ in the second part of an encryption round of Yao and Chen's white-box SM4 implementation, and ends with the output of an $E_{i+1,j}$ operation of the X_{i+4} branch in the first part of the next encryption round $(j = 0, 1, 2, 3)$. Observe that E_i and E_{i+1} are diagonal affine transformations, $E_{i,j}$ and $E_{i+1,j}$ are invertible 8×8-bit affine transformations, and $x_{i,j}$ is the original input byte to the j-th original S-box of the i-th encryption round in a protected form with $E_{i,j}$.

The collision function is functionally equivalent and can be simplified to the one depicted in Fig. 4. In our attack and all subsequent descriptions, we set X_i such that $P'_{i+4} \circ P_i^{-1}(X_i) = 0$, and denote the constant $A_{i+4}^{-1} \cdot a''_{i+4} \oplus A_{i+4}^{-1} \circ P'_{i+4} \circ P_i^{-1}(X_i) = A_{i+4}^{-1} \cdot a''_{i+4}$ by ε_i. We now explain where the value ε_i comes from. Let $\widehat{X}$ denotes the original 32-bit value immediately after the $\mathbf{L}$ operation under the input $X = (x_{i,0}, x_{i,1}, x_{i,2}, x_{i,3})$, then we have

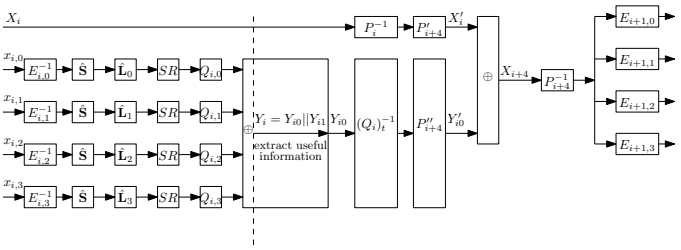

Figure 3. Our collision function on Yao and Chen's white-box SM4 implementation

$$\begin{aligned} & P_{i+4}^{-1} \circ (Y'_{i0} \oplus P'_{i+4} \circ P_i^{-1}(X_i)) \\ &= P_{i+4}^{-1}(Y'_{i0}) \oplus A_{i+4} \circ P'_{i+4} \circ P_i^{-1}(X_i) \\ &= P_{i+4}^{-1} \circ P''_{i+4}(\widehat{X}) \\ &= P_{i+4}^{-1} \circ (P_{i+4}(\widehat{X}) \oplus a''_{i+4}) \\ &= P_{i+4}^{-1} \circ (A_{i+4}(\widehat{X}) \oplus a_{i+4} \oplus a''_{i+4}) \\ &= A_{i+4}^{-1} \circ (A_{i+4}(\widehat{X}) \oplus a_{i+4} \oplus a''_{i+4} \oplus a_{i+4}) \\ &= \widehat{X} \oplus A_{i+4}^{-1} \cdot a''_{i+4}, \end{aligned}$$

which is equal to $\widehat{X} \oplus \varepsilon_i$ under $P'_{i+4} \circ P_i^{-1}(X_i) = 0$.

As a consequence, the collision function denoted by $f(x_{i,0}, x_{i,1}, x_{i,2}, x_{i,3}, X_i)$, or simply $f(x_{i,0}, x_{i,1}, x_{i,2}, x_{i,3})$ under $P'_{i+4} \circ P_i^{-1}(X_i) = 0$, is

$$\begin{aligned} & f(x_{i,0}, x_{i,1}, x_{i,2}, x_{i,3}) \\ &= \begin{bmatrix} E_{i+1,0} \\ E_{i+1,1} \\ E_{i+1,2} \\ E_{i+1,3} \end{bmatrix} \circ \oplus_{\varepsilon_i} \circ \mathbf{L} \circ \begin{bmatrix} \mathbf{S} \circ \oplus_{rk_{i,0}} \circ E_{i,0}^{-1}(x_{i,0}) \\ \mathbf{S} \circ \oplus_{rk_{i,1}} \circ E_{i,1}^{-1}(x_{i,1}) \\ \mathbf{S} \circ \oplus_{rk_{i,2}} \circ E_{i,2}^{-1}(x_{i,2}) \\ \mathbf{S} \circ \oplus_{rk_{i,3}} \circ E_{i,3}^{-1}(x_{i,3}) \end{bmatrix}. \end{aligned}$$

Furthermore, we express f as a concatenation of four byte functions f_0, f_1, f_2 and f_3:

$$\begin{aligned} & f(x_{i,0}, x_{i,1}, x_{i,2}, x_{i,3}) \\ =& [f_0(x_{i,0}, x_{i,1}, x_{i,2}, x_{i,3}), f_1(x_{i,0}, x_{i,1}, x_{i,2}, x_{i,3}), \\ & f_2(x_{i,0}, x_{i,1}, x_{i,2}, x_{i,3}), f_3(x_{i,0}, x_{i,1}, x_{i,2}, x_{i,3})]^T; \end{aligned}$$

and define $\mathbf{S}_j$ function as

$$\begin{aligned} \mathbf{S}_j(\cdot) &= \mathbf{S} \circ \oplus_{rk_{i,j}} \circ E_{i,j}^{-1}(\cdot) \\ &= \mathbf{S}(rk_{i,j} \oplus E_{i,j}^{-1}(\cdot)), \quad j = 0, 1, 2, 3. \end{aligned} \tag{4}$$

2) Recovering $\mathbf{S}_j$ Functions: Next we try to recover the functions $\mathbf{S}_0$, $\mathbf{S}_1$, $\mathbf{S}_2$ and $\mathbf{S}_3$ by exploiting collisions on the output of the functions f_j. We first use the following collision to recover $\mathbf{S}_0$ and $\mathbf{S}_1$:

$$f_0(\alpha, 0, 0, 0) = f_0(0, \beta, 0, 0), \tag{5}$$

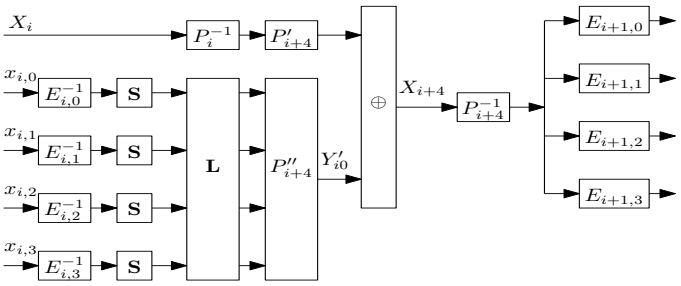

Figure 4. Equivalent of collision function on Yao and Chen's white-box SM4 implementation

where $\alpha, \beta \in \mathrm{GF}(2)^8$. By the linear transformation $\mathbf{L}$ in Eq. (3), Eq. (5) immediately means the following equation:

$$\begin{aligned} & E_{i+1,0} \circ \oplus_{\varepsilon_{i,0}} \circ \big(B_1 \circ \mathbf{S}_0(\alpha) \oplus B_2 \circ \mathbf{S}_1(0) \oplus \\ & B_2 \circ \mathbf{S}_2(0) \oplus B_3 \circ \mathbf{S}_3(0)\big) \\ = \quad & E_{i+1,0} \circ \oplus_{\varepsilon_{i,0}} \circ \big(B_1 \circ \mathbf{S}_0(0) \oplus B_2 \circ \mathbf{S}_1(\beta) \oplus \\ & B_2 \circ \mathbf{S}_2(0) \oplus B_3 \circ \mathbf{S}_3(0)\big), \end{aligned}$$

where $\varepsilon_{i,0}$ is the corresponding byte of the constant ε_i. Since $E_{i+1,0}$ is a bijection, we have the following equation:

$$B_1 \circ \mathbf{S}_0(\alpha) \oplus B_2 \circ \mathbf{S}_1(0) = B_1 \circ \mathbf{S}_0(0) \oplus B_2 \circ \mathbf{S}_1(\beta).$$

For convenience, define $u_m = \mathbf{S}_0(m)$ and $v_m = \mathbf{S}_1(m)$, then we have

$$B_1 \circ (u_0 \oplus u_\alpha) = B_2 \circ (v_0 \oplus v_\beta). \tag{6}$$

Since $\alpha \mapsto f_0(\alpha, 0, 0, 0)$ and $\beta \mapsto f_0(0, \beta, 0, 0)$ are bijections, we can find 256 collisions. After removing $(\alpha, \beta) = (0, 0)$, we get 255 pairs (α, β) satisfying Eq. (5), each providing an equation of the form of Eq. (6). In the same way, we use other f_j functions ($j \in \{1, 2, 3\}$) to generate similar equations with different coefficients in $\{B_1, B_2, B_3\}$. Finally, we get 4×255 linear equations with all 512 unknowns, as follows:

$$\begin{cases} B_1 \circ (u_0 \oplus u_\alpha) = B_2 \circ (v_0 \oplus v_\beta); \\ B_3 \circ (u_0 \oplus u_\alpha) = B_1 \circ (v_0 \oplus v_\beta); \\ B_2 \circ (u_0 \oplus u_\alpha) = B_3 \circ (v_0 \oplus v_\beta); \\ B_2 \circ (u_0 \oplus u_\alpha) = B_2 \circ (v_0 \oplus v_\beta). \end{cases} \tag{7}$$

Define $u'_m = u_0 \oplus u_m$ and $v'_m = v_0 \oplus v_m$, with $m \in \{1, 2, \ldots, 255\}$, so that the number of unknowns is reduced to $2 \times 255 = 510$. Thus, Eq. (6) can be rewritten as

$$B_1 \circ u'_\alpha = B_2 \circ v'_\beta,$$

meaning that the linear system of Eq. (7) can be represented with 510 unknowns as

$$\begin{cases} B_1 \circ u'_\alpha = B_2 \circ v'_\beta, \\ B_3 \circ u'_\alpha = B_1 \circ v'_\beta, \\ B_2 \circ u'_\alpha = B_3 \circ v'_\beta, \\ B_2 \circ u'_\alpha = B_2 \circ v'_\beta. \end{cases}$$

The 4×255 equations yield a linear system of rank 509; and in such a linear equation system, all other unknowns can be expressed as a function of one of them, say u'_1, that is, there exist coefficients a_i and b_i such that $u'_m = a_m \cdot u'_1$ and $v'_m = b_m \cdot u'_1$. That is,

$$\begin{aligned} u_m &= a_m \cdot (u_0 \oplus u_1) \oplus u_0, \\ v_m &= b_m \cdot (u_0 \oplus u_1) \oplus v_0. \end{aligned} \tag{8}$$

Next we can recover the $\mathbf{S}_0$ function by exhaustive search on the pair (u_0, u_1), and at last we use the following equation from the definition of the $\mathbf{S}_0$ function to verify whether the obtained $\mathbf{S}_0$ function is right or not:

$$\mathbf{S}^{-1} \circ \mathbf{S}_0(\cdot) = rk_{i,0} \oplus E_{i,0}^{-1}(\cdot).$$

Since $E_{i,0}^{-1}$ is an 8×8-bit invertible affine transformation, the above function has an algebraic degree of at most 1. For a wrong pair (u_0, u_1), a wrong candidate function $\mathbf{S}_0^*$ would be got which is an affine equivalent to $\mathbf{S}_0$, namely there exists an 8×8-bit matrix a and an 8-bit vector b such that $\mathbf{S}_0^*(\cdot) = a \cdot \mathbf{S}_0(\cdot) \oplus b$, with $a \neq 0$ and $(a, b) \neq (0, 1)$. The function $\mathbf{S}^{-1} \circ \mathbf{S}_0^*(\cdot)$ satisfies

$$\mathbf{S}^{-1} \circ \mathbf{S}_0^*(\cdot) = \mathbf{S}^{-1}\big(a \cdot \mathbf{S}\big(rk_{i,0} \oplus E_{i,0}^{-1}(\cdot)\big) \oplus b\big).$$

In this case, $\mathbf{S}^{-1} \circ \mathbf{S}_0^*(\cdot)$ has an algebraic degree greater than 1 with an overwhelming probability. More specifically, we set the function $\hat{g}(\cdot) = \mathbf{S}^{-1} \circ \mathbf{S}^*(\cdot)$, used Lai's higher-order derivative concept [18] to calculate the first-order derivative of $\hat{g}$, and finally ran ten thousand tests without obtaining a function with an algebraic degree of 1 or less. For instance, the first-order derivative $\hat{\varphi}$ at point $(\mathtt{01})$ is set to

$$\hat{\varphi}(x) = \hat{g}(x \oplus \mathtt{01}) \oplus \hat{g}(x),$$

and we verify whether $\hat{\varphi}(x)$ is constant with at most 2^7 inputs of x, since $\hat{\varphi}(x) = \hat{\varphi}(x \oplus \mathtt{01})$. For each wrong pair, the probability that $\hat{\varphi}(x)$ is constant is roughly 2^{-8}, so wrong guesses can be quickly removed.

After recovering $\mathbf{S}_0$, we can use Eq. (8) to recover $\mathbf{S}_1$ by exhaustive search on v_0, and similarly recover $\mathbf{S}_2$ and $\mathbf{S}_3$ with other equations finally.

3) Recovering the Linear Parts of Output Encodings $E_{i+1,j}$: After the $\mathbf{S}_j$ functions have been recovered ($j = 0, 1, 2, 3$), however it is not as easy to recover the output encodings $E_{i+1,j}$ as Lepoint et al.'s attack on Chow et al.'s white-box AES implementation, because of the existence of the unknown constant ε_i, which is partially due to the different structures of Feistel and SPN ciphers and the design of Yao and Chen's white-box SM4 implementation. Anyway, we find a trick to recover the linear part of the output encodings $E_{i+1,j}$. Since $E_{i+1,j}$ is an invertible affine transformation, we write $E_{i+1,j}(\cdot) = C_{i+1,j}(\cdot) \oplus c_{i+1,j}$, where the linear part $C_{i+1,j}$ is a general invertible 8×8-bit matrix and $c_{i+1,j}$ is an 8-bit constant.

Given a 32-bit input $X = (x_{i,0}, x_{i,1}, x_{i,2}, x_{i,3})$ to the f collision function, denote the original 32-bit value immediately after the $\widehat{\mathbf{L}}$ operation as follows:

$$\begin{aligned} Y &= \begin{bmatrix} Y_0 & Y_1 & Y_2 & Y_3 \end{bmatrix}^T \\ &= \mathbf{L}_0 \circ \mathbf{S}_0(x_{i,0}) \oplus \mathbf{L}_1 \circ \mathbf{S}_1(x_{i,1}) \oplus \\ &\quad \mathbf{L}_2 \circ \mathbf{S}_2(x_{i,2}) \oplus \mathbf{L}_3 \circ \mathbf{S}_3(x_{i,3}). \end{aligned}$$

As $\mathbf{L}$ is public and we have recovered $\mathbf{S}_j$ above ($j = 0, 1, 2, 3$), we can compute Y_j. The output of the f collision function is

$$f = \begin{bmatrix} f_0 & f_1 & f_2 & f_3 \end{bmatrix}^T = \begin{bmatrix} E_{i+1,0}(Y_0 \oplus \varepsilon_{i,0}) \\ E_{i+1,1}(Y_1 \oplus \varepsilon_{i,1}) \\ E_{i+1,2}(Y_2 \oplus \varepsilon_{i,2}) \\ E_{i+1,3}(Y_3 \oplus \varepsilon_{i,3}) \end{bmatrix},$$

where $(\varepsilon_{i,0}, \varepsilon_{i,1}, \varepsilon_{i,2}, \varepsilon_{i,3}) = \varepsilon_i$.

Subsequently, to recover $E_{i+1,j}$, we need to know the 8-bit unknown constant $\varepsilon_{i,j}$. A straightforward way is to try by exhaustive search, which would cause an additional complexity factor of 2^8. However, we find we can recover the linear part $C_{i+1,j}$ at ease with a negligible time complexity, as follows.

First, we consider the output of the arbitrary 32-bit input $X = (x_{i,0}, x_{i,1}, x_{i,2}, x_{i,3})$ under the f_0 collision function,

$$\begin{aligned} f_0(X) &= E_{i+1,0}(Y_0 \oplus \varepsilon_{i,0}) \\ &= C_{i+1,0}(Y_0) \oplus C_{i+1,0}(\varepsilon_{i,0}) \oplus c_{i+1,0}, \end{aligned} \tag{9}$$

where Y_0 is defined above, which denotes the corresponding original 8-bit value immediately after the $\mathbf{L}$ operation under the input X.

Next, we choose the 32-bit input $X0 = (\hat{x}_{i,0}, \hat{x}_{i,1}, \hat{x}_{i,2}, \hat{x}_{i,3})$ to the f collision function, so that the original 32-bit value immediately after the $\mathbf{L}$ operation is 0; this can be done easily by choosing $X0$ such that

$$(\mathbf{S}_0(\hat{x}_{i,0}), \mathbf{S}_1(\hat{x}_{i,1}), \mathbf{S}_2(\hat{x}_{i,2}), \mathbf{S}_3(\hat{x}_{i,3})) = \mathbf{L}^{-1}(0) = 0.$$

Thus, its corresponding output under the f_0 collision function is

$$f_0(X0) = C_{i+1,0}(\varepsilon_{i,0}) \oplus c_{i+1,0}. \tag{10}$$

At last, XORing Eq. (9) and Eq. (10), we get $f_0(X) \oplus f_0(X0) = C_{i+1,0}(Y_0)$. As a consequence, we can recover the linear part $C_{i+1,0}$ of the output encodings $E_{i+1,0}$. The linear parts of other output encodings $E_{i+1,j}$ can be recovered similarly.

4) Recovering Round Key rk_i: Subsequently, we cannot recover a round key byte from the above collision function in a way similar to Lepoint et al.'s attack on Chow et al.'s white-box AES implementation, because there is an unknown constant $\varepsilon_{i,0}$ and $E_{i,j}$ is also unknown, although its linear part $C_{i,j}$ can be recovered as above. We find these problems can be solved by modifying the collision function, as follows.

Suppose that $E_{i,j}(\cdot) = C_{i,j}(\cdot) \oplus c_{i,j}$ and the linear part $C_{i,j}$ has been recovered as above, where $C_{i,j}$ is a general invertible 8×8-bit matrix and $c_{i,j}$ is an 8-bit constant ($j = 0, 1, 2, 3$). We first show how to recover the two unknown 8-bit constants $\varepsilon_{i,j}$ and $C_{i,j}(rk_{i,j}) \oplus c_{i,j}$, that is, $\varepsilon_{i,j}$ and $E_{i,j}(rk_{i,j})$, rather than to recover $rk_{i,j}$ directly. Compared with Lepoint et al.'s attack, this increases an additional complexity factor of 2^8, since we need to guess two 8-bit unknowns here, instead of one, but it is comparable to the time complexity of the above phase of recovering $\mathbf{S}_j$'s.

According to Eq. (4) and Eq. (5), we have

$$\begin{aligned} & f_0\big(E_{i,0}(\mathbf{S}^{-1}(x) \oplus rk_{i,0}), 0, 0, 0\big) \\ = \; & f_0\big(C_{i,0}(\mathbf{S}^{-1}(x)) \oplus E_{i,0}(rk_{i,0}), 0, 0, 0\big) \\ = \; & E_{i+1,0}(B_1(x) \oplus \delta \oplus \varepsilon_{i,0}) \\ = \; & E_{i+1,0}(B_1(x \oplus B_1^{-1}(\varepsilon_{i,0})) \oplus \delta), \end{aligned} \tag{11}$$

where $\delta = B_2 \circ \mathbf{S}_1(0) \oplus B_2 \circ \mathbf{S}_2(0) \oplus B_3 \circ \mathbf{S}_3(0)$ is a constant that can be easily computed.

We replace x with $x \oplus B_1^{-1}(\varepsilon_{i,0})$ in Eq. (11), and define the function g as

$$\begin{aligned} g(x) &= f_0\big(C_{i,0}(\mathbf{S}^{-1}(x \oplus B_1^{-1}(\varepsilon_{i,0}))) \oplus E_{i,0}(rk_{i,0}), 0, 0, 0\big) \\ &= E_{i+1,0}(B_1(x) \oplus \delta). \end{aligned}$$

Because of the 8×8-bit invertible affine transformation $E_{i+1,0}$, the function g has an algebraic degree of at most 1. For a wrong guess $\hat{rk}_{i,0} \neq rk_{i,0}$, the function $\hat{g}$ is defined as

$$\begin{aligned} \hat{g}(x) &= f_0\big(C_{i,0}(\mathbf{S}^{-1}(x \oplus B_1^{-1}(\varepsilon_{i,0}))) \oplus E_{i,0}(\hat{rk}_{i,0}), 0, 0, 0\big) \\ &= E_{i+1,0}(B_1 \circ \mathbf{S}(\mathbf{S}^{-1}(x) \oplus \hat{rk}_{i,0} \oplus rk_{i,0}) \oplus \delta). \end{aligned}$$

In this case, with a similar test, $\hat{g}$ has an algebraic degree of more than 1 with an overwhelming probability. We extract $(\varepsilon_{i,0}, E_{i,0}(rk_{i,0}))$ by exhaustive search, that is, similarly we verify whether the first-order derivative $\hat{\varphi}(x) = \hat{g}(x \oplus \texttt{01}) \oplus \hat{g}(x)$ of $\hat{g}(x)$ at point $\texttt{01}$ is constant for each guess $(\varepsilon_{i,0}, E_{i,0}(\hat{rk}_{i,0}))$. For a wrong guess $(\varepsilon_{i,0}, E_{i,0}(\hat{rk}_{i,0}))$, the probability that $\hat{\varphi}(x)$ is constant is roughly 2^{-8}, so wrong guesses can be quickly removed.

As a result, we can also recover $(\varepsilon_{i,j}, E_{i,j}(rk_{i,j}))$ for $j = 1, 2, 3$, by changing the definition of the function g. Thus, we recover $\varepsilon_i = (\varepsilon_{i,0}, \varepsilon_{i,1}, \varepsilon_{i,2}, \varepsilon_{i,3})$, and further we can recover the encodings $E_{i,j}$ by $E_{i,j}(Y_j \oplus \varepsilon_{i,j})$ (or by deducing $c_{i,j}$ under those equations like Eq. (9)). Finally, we get the round key bytes $rk_{i,j}$ from the recovered $E_{i,j}(rk_{i,j})$. Four round keys enables us to determine the full secret key of SM4 in principle.

5) Time Complexity: In the phase of recovering $\mathbf{S}_0$, there are 2^{16} candidates (u_0, u_1) for exhaustive search, and to verify whether $\hat{\varphi}(x)$ is constant we need to calculate $\hat{\varphi}(x)$ for at most 2^7 inputs. For a wrong guess (u_0, u_1), the probability that $\hat{\varphi}(x)$ is constant is 2^{-8} roughly. Thus, the expected value of the test is $1+1/256+\cdots+1/(256^{127}) \approx 1$. The expected time complexity of recovering $\mathbf{S}_0$ is hence about $2^{16} \cdot 1 \cdot 2 = 2^{17}$ (dominated by $\mathbf{S}/\mathbf{S}^{-1}$ computations).

We recover $\mathbf{S}_1$, $\mathbf{S}_2$ and $\mathbf{S}_3$ by exhaustive search on v_0 and produce an expected time complexity of $3 \cdot (2^8 \cdot 1 \cdot 2) = 3 \cdot 2^9$. Thus, the expected time complexity of recovering all the four $\mathbf{S}_j$'s is about $2^{17} + 3 \cdot 2^9 = 259 \cdot 2^9$.

The time complexity for recovering the linear part of output encoding $E_{i+1,j}$ is negligible. The expected time complexity of recovering $(\varepsilon_{i,0}, E_{i,0}(rk_{i,0}))$ is about $2^{16} \cdot 1 \cdot 2 = 2^{17}$, so the expected time complexity of recovering a round key is $4 \cdot (2^{16} \cdot 1 \cdot 2) = 2^{19}$. To sum up, the expected total time complexity of recovering one round key is about $259 \cdot 2^9 + 2^{19} \approx 2^{19.4}$.

IV. Collision-Based Attack on Xiao and Lai's White-Box SM4 Implementation

Xiao and Lai's white-box SM4 implementation [37] is similar to Yao and Chen's white-box SM4 implementation at a high level, except that there is no state expansion to the S-box layer and thus the original $\mathbf{L}$ operation is used. Fig. 5 depicts an encryption round of Xiao and Lai's white-box SM4 implementation, where Q_i is a general invertible affine output encoding. Therefore, we can apply our above collision-based attack to Xiao and Lai's white-box SM4 implementation in the same way, and the attack's time complexity is also about $2^{19.4}$ for recovering a round key.

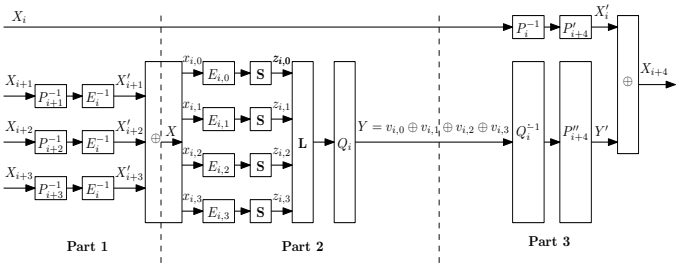

Figure 5. An encryption round of Xiao and Lai's white-box SM4 implementation

V. Collision-Based Attack on Shang's White-Box SM4 Implementation

Shang's white-box SM4 implementation [30] is based on Xiao and Lai's white-box SM4 implementation, mainly by applying two general 16-bit affine encodings $E_{i,0}$ and $E_{i,1}$ to the input of the S-box layer in parallel, each corresponding to two S-boxes and subsequently a 32×16-bit component $\mathbf{L}_0$ or $\mathbf{L}_1$ of the $\mathbf{L}$ matrix, and thus two 16×32-bit tables. Fig. 6 depicts an encryption round of Shang's white-box SM4 implementation.

We can similarly exploit our above collision-based attack to Shang's white-box SM4 implementation after a few modifications. Specifically, we represent $\mathbf{L}$ by Eq. (3) with 16×16-bit blocks as

$$\mathbf{L} = \begin{bmatrix} L_{0,0} & L_{0,1} \\ L_{1,0} & L_{1,1} \end{bmatrix} = \begin{bmatrix} L_{0,0} & L_{0,1} \\ L_{0,1} & L_{0,0} \end{bmatrix},$$

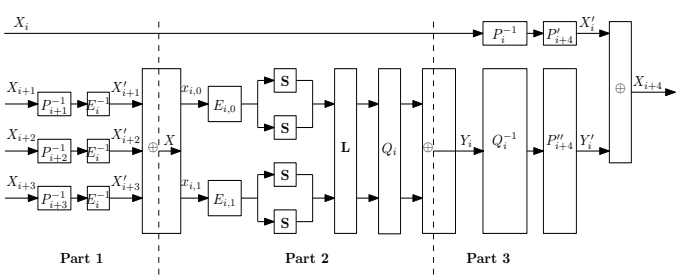

Figure 6. An encryption round of Shang's white-box SM4 implementation

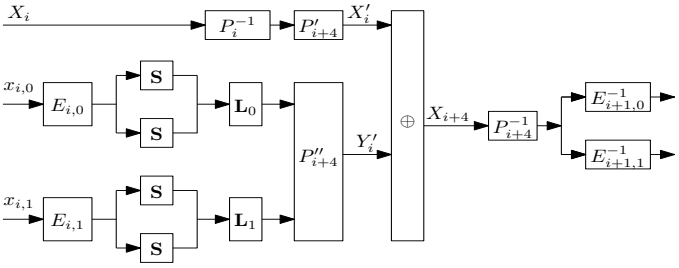

Figure 7. Equivalent of collision function on Shang's white-box SM4 implementation

define

$$\mathbf{S}_0(x) = \begin{pmatrix} \mathbf{S} \\ \mathbf{S} \end{pmatrix} \big((rk_{i,0}||rk_{i,1}) \oplus E_{i,0}(x)\big),$$
$$\mathbf{S}_1(x) = \begin{pmatrix} \mathbf{S} \\ \mathbf{S} \end{pmatrix} \big((rk_{i,2}||rk_{i,3}) \oplus E_{i,1}(x)\big),$$

where $x \in \text{GF}(2)^{16}$, and define a collision function on Shang's white-box SM4 implementation whose equivalent f is depicted in Fig. 7, as follows:

$$\begin{aligned} &f(x_{i,0}, x_{i,1}) = [f_0(x_{i,0}, x_{i,1}), f_1(x_{i,0}, x_{i,1})]^T \\ &= \begin{bmatrix} E_{i+1,0}^{-1} \\ E_{i+1,1}^{-1} \end{bmatrix} \oplus_{\epsilon_i} \circ \mathbf{L} \circ \begin{bmatrix} \begin{pmatrix} \mathbf{S} \\ \mathbf{S} \end{pmatrix} \big((rk_{i,0}||rk_{i,1}) \oplus E_{i,0}(x_{i,0})\big) \\ \begin{pmatrix} \mathbf{S} \\ \mathbf{S} \end{pmatrix} \big((rk_{i,2}||rk_{i,3}) \oplus E_{i,1}(x_{i,1})\big) \end{bmatrix} \\ &= \begin{bmatrix} E_{i+1,0}^{-1} \\ E_{i+1,1}^{-1} \end{bmatrix} \oplus_{\epsilon_i} \circ \mathbf{L} \circ \begin{bmatrix} \mathbf{S}_0(x_{i,0}) \\ \mathbf{S}_1(x_{i,1}) \end{bmatrix}. \end{aligned}$$

Then, we consider the collision $f_h(\alpha, 0) = f_h(0, \beta)$, where $\alpha, \beta \in \text{GF}(2)^{16}$ and $h = 0, 1$. At last, we get the following linear system of $2 \times (2^{16} - 1)$ equations with $2 \times (2^{16} - 1)$ unknowns:

$$L_{0,0} \circ u'_\alpha \oplus L_{0,1} \circ v'_\beta = 0,$$
$$L_{0,1} \circ u'_\alpha \oplus L_{0,0} \circ v'_\beta = 0,$$

where $u'_\alpha = \mathbf{S}_0(\alpha) \oplus \mathbf{S}_0(0)$ and $v'_\beta = \mathbf{S}_1(\beta) \oplus \mathbf{S}_1(0)$ and $(\alpha, \beta) \neq (0, 0)$. Subsequently, by a similar process, we can recover the $\mathbf{S}_h$ function with an expected time complexity of about $2^{34} + 2^{18}$, and recover the linear part of $E_{i+1,h}$ with a negligible time complexity. Note that here each $\mathbf{S}_h$ computation involves two $\mathbf{S}$ computations.

At last, suppose that $E_{i,h}(\cdot) = C_{i,h}(\cdot) \oplus c_{i,h}$ and the linear part $C_{i,h}$ has been recovered as above, where $C_{i,h}$ is a general invertible 16×16-bit matrix and $c_{i,h}$ is a 16-bit constant. Similarly, we depend on the following revised functions to recover the key bytes $rk_{i,0}||rk_{i,1}$ with an expected time complexity of about 2^{34}:

$$\begin{aligned} &f_0\big(E_{i,0}^{-1}\big(\begin{pmatrix} \mathbf{S}^{-1} \\ \mathbf{S}^{-1} \end{pmatrix}(x) \oplus (rk_{i,0}||rk_{i,1})\big), 0\big) \\ &= f_0\big(C_{i,0}^{-1}\big(\begin{pmatrix} \mathbf{S}^{-1} \\ \mathbf{S}^{-1} \end{pmatrix}(x)\big) \oplus E_{i,0}^{-1}(rk_{i,0}||rk_{i,1}), 0\big) \\ &= E_{i+1,0}^{-1}(L_{0,0}(x) \oplus L_{0,1} \circ \mathbf{S}_1(0) \oplus \varepsilon_{i,0}) \\ &= E_{i+1,0}^{-1}\big(L_{0,0}(x \oplus L_{0,0}^{-1} \cdot \varepsilon_{i,0}) \oplus L_{0,1} \circ \mathbf{S}_1(0)\big), \end{aligned}$$

and

$$g(x) = f_0\big(C_{i,0}^{-1}\big(\begin{pmatrix}\mathbf{S}^{-1}\\ \mathbf{S}^{-1}\end{pmatrix}(x \oplus L_{0,0}^{-1}\cdot\varepsilon_{i,0})\big)\oplus E_{i,0}^{-1}(rk_{i,0}||rk_{i,1}),0\big)$$
$$= E_{i+1,0}^{-1}\big(L_{0,0}(x)\oplus L_{0,1}\circ\mathbf{S}_1(0)\big),$$

where $\varepsilon_{i,0}$ is the corresponding 16-bit part of ε_i.

Thus, the total expected time complexity of recovering one round key rk_i from Shang's white-box SM4 implementation is about $2^{34}+2^{18}+2\cdot 2^{34}\approx 2^{35.6}$.

VI. Collision-Based Attack on Wu et al.'s White-Box SM4 Implementation

In this section, we briefly describe Wu et al.'s white-box SM4 implementation and our attack.

A. Wu et al.'s White-Box SM4 Implementation

An encryption round of Wu et al.'s white-box SM4 implementation [34] is made up of three parts, as depicted in Fig. 8, but it is expanded to 36 rounds to produce the original output (without protection), and there are respectively two types of lookup tables in the second and third parts, especially, the second type of lookup tables uses three different construction methods for different rounds.

1) Part 1: The first part is processed as follows, and the 32-bit output X is protected by a diagonal invertible matrix E_i:

$$X'_{i+j} = A_{i,j}(X_{i+j}), j = 1,2,3;$$
$$X = X'_{i+1}\oplus X'_{i+2}\oplus X'_{i+3},$$

where $A_{i,j}$ is a composite 32-bit invertible matrix with different encodings ($i = 0,1,\ldots,35$), as follows,

$$\begin{array}{ccc}
A_{0,1}=E_0P, & A_{0,2}=E_0P, & A_{0,3}=E_0P,\\
A_{1,1}=E_1P, & A_{1,2}=E_1P, & A_{1,3}=E_1R_0^{-1},\\
A_{2,1}=E_2P, & A_{2,2}=E_2R_0^{-1}, & A_{2,3}=E_2R_1^{-1},\\
A_{3,1}=E_3R_0^{-1}, & A_{3,2}=E_3R_1^{-1}, & A_{3,3}=E_3R_2^{-1},\\
A_{4,1}=E_4R_1^{-1}, & A_{4,2}=E_4R_2^{-1}, & A_{4,3}=E_4R_3^{-1},\\
\vdots & \vdots & \vdots\\
A_{31,1}=E_{31}R_{28}^{-1}, & A_{31,2}=E_{31}R_{29}^{-1}, & A_{31,3}=E_{31}R_{30}^{-1},\\
A_{32,1}=E_{32}R_{25}^{-1}, & A_{32,2}=E_{32}R_{26}^{-1}, & A_{32,3}=E_{32}R_{27}^{-1},\\
A_{33,1}=E_{33}R_{26}^{-1}, & A_{33,2}=E_{33}R_{27}^{-1}, & A_{33,3}=E_{33}R_{28}^{-1},\\
A_{34,1}=E_{34}R_{27}^{-1}, & A_{34,2}=E_{34}R_{28}^{-1}, & A_{34,3}=E_{34}R_{29}^{-1},\\
A_{35,1}=E_{35}R_{28}^{-1}, & A_{35,2}=E_{35}R_{29}^{-1}, & A_{35,3}=E_{35}R_{30}^{-1},
\end{array}$$

with $P=\mathrm{diag}(P_0,P_1,P_2,P_3)$ being a diagonal invertible matrix and $R_0,R_1,\cdots,R_{30}$ being general invertible matrices.

2) Part 2: Construct four 8-bit to 32-bit lookup tables of the first type $Table_{i,j}$ ($j=0,1,2,3$), and XOR the outputs of the four tables into a 32-bit Y_i, as follows:

$$Y_i = \bigoplus_{j=0}^{3} Table_{i,j}$$
$$= Q_i\circ P\circ\mathbf{L}\circ\begin{bmatrix}\mathbf{S}\circ\oplus_{rk_{i,0}}\circ P_0^{-1}\circ E_{i,0}^{-1}(x_{i,0})\\ \mathbf{S}\circ\oplus_{rk_{i,1}}\circ P_1^{-1}\circ E_{i,1}^{-1}(x_{i,1})\\ \mathbf{S}\circ\oplus_{rk_{i,2}}\circ P_2^{-1}\circ E_{i,2}^{-1}(x_{i,2})\\ \mathbf{S}\circ\oplus_{rk_{i,3}}\circ P_3^{-1}\circ E_{i,3}^{-1}(x_{i,3})\end{bmatrix}.$$

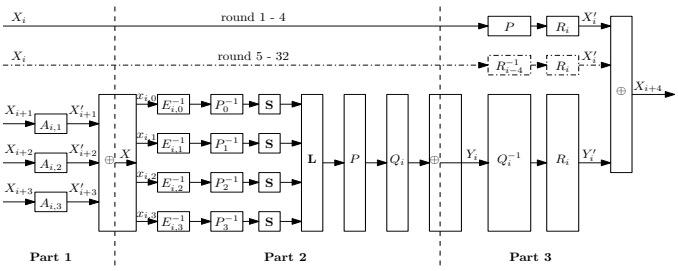

Figure 8. An encryption round of Wu et al.'s white-box SM4 implementation

3) Part 3: Construct four 16-bit to 8-bit lookup tables of the second type. Let (X_i,Y_i) be the input of the four tables, and X_{i+4} be the output, then

- Rounds 1-4:
$$X'_i = R_i\circ P(X_i), Y'_i = R_i\circ Q_i^{-1}(Y_i),$$
$$X_{i+4} = Y'_i\oplus X'_i;$$
- Rounds 5-32:
$$X'_i = R_i\circ R_{i-4}(X_i), Y'_i = R_i\circ Q_i^{-1}(Y_i),$$
$$X_{i+4} = Y'_i\oplus X'_i;$$
- Rounds 33-36:
$$X'_i = R_i\circ R_{i-8}(X_i), Y'_i = P^{-1}\circ Q_i^{-1}(Y_i),$$
$$X_{i+4} = Y'_i\oplus X'_i.$$

B. Attacking Wu et al.'s White-Box SM4 Implementation

Note that all the encodings are linear (invertible matrices) in Wu et al.'s white-box SM4 implementation, rather than affine encodings as in the above three white-box SM4 implementations. As a consequence, we can devise a collision function in a similar way as above, but much easier, since there is no effect of unknown constant parts associated with the encodings. Fig. 9 (top) depicts a collision function on Wu et al.'s white-box SM4 implementation, where Q_i and Q_i^{-1} are cancelled with each other.

Further, the collision function can be simplified. First, we set $X_i=0$, and thus $R_i\circ P(X_i)=0$, since R_i and P are invertible matrices. Second, R_i and R_i^{-1} are cancelled with each other. Thus, after we adjust the position of encoding P, we get a simplified collision function as depicted in Fig. 9 (bottom), that is

$$f(x_{i,0},x_{i,1},x_{i,2},x_{i,3})$$
$$=[f_0(x_{i,0},x_{i,1},x_{i,2},x_{i,3}),f_1(x_{i,0},x_{i,1},x_{i,2},x_{i,3}),$$
$$f_2(x_{i,0},x_{i,1},x_{i,2},x_{i,3}),f_3(x_{i,0},x_{i,1},x_{i,2},x_{i,3})]^T$$
$$=\begin{bmatrix}\hat{E}_{i+1,0}\\ \hat{E}_{i+1,1}\\ \hat{E}_{i+1,2}\\ \hat{E}_{i+1,3}\end{bmatrix}\circ\mathbf{L}\circ\begin{bmatrix}\mathbf{S}\circ\oplus_{rk_{i,0}}\circ\hat{E}_{i,0}^{-1}(x_{i,0})\\ \mathbf{S}\circ\oplus_{rk_{i,1}}\circ\hat{E}_{i,1}^{-1}(x_{i,1})\\ \mathbf{S}\circ\oplus_{rk_{i,2}}\circ\hat{E}_{i,2}^{-1}(x_{i,2})\\ \mathbf{S}\circ\oplus_{rk_{i,3}}\circ\hat{E}_{i,3}^{-1}(x_{i,3})\end{bmatrix},$$

where $\hat{E}_{i,j}^{-1}(\cdot)=P_j^{-1}\circ E_{i,j}^{-1}(\cdot)$ and $\hat{E}_{i+1,j}(\cdot)=E_{i+1,j}\circ P_j(\cdot)$ are invertible 8×8-bit matrices ($j=0,1,2,3$). Note that $\hat{E}_{i,j}^{-1}$ and $\hat{E}_{i+1,j}$ are matrixes simply, not affine transformations, and thus this attack is a simplified case.

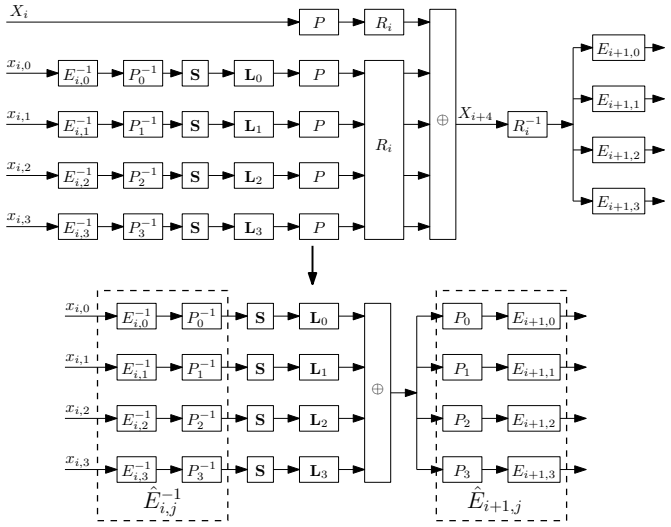

Figure 9. Collision function and its equivalent on Wu et al's white-box SM4 implementation

Define $\mathbf{S}_j$ function as

$$\begin{aligned}\mathbf{S}_j(\cdot) &= \mathbf{S} \circ \oplus_{rk_{i,j}} \circ \hat{E}_{i,j}^{-1}(\cdot) \\ &= \mathbf{S}(rk_{i,j} \oplus \hat{E}_{i,j}^{-1}(\cdot)), \quad j = 0, 1, 2, 3.\end{aligned}$$

Then, we can similarly recover the four functions $\mathbf{S}_j$'s with an expected time complexity of about $2^{16} \cdot 1 \cdot 2 + 3 \cdot (2^8 \cdot 1 \cdot 2) = 259 \cdot 2^9$. Since $\hat{E}_{i+1,0}$ is an invertible 8×8-bit matrix, we can recover it immediately by calculating $\hat{E}_{i+1,0}(\cdot) = f_0(\psi^{-1}(\cdot), 0, 0, 0)$, where $\psi : \alpha \mapsto B_1 \circ \mathbf{S}_0(\alpha) \oplus B_2 \circ \mathbf{S}_1(0) \oplus B_2 \circ \mathbf{S}_2(0) \oplus B_3 \circ \mathbf{S}_3(0)$. Similarly for recovering $\hat{E}_{i+1,1}, \hat{E}_{i+1,2}$ and $\hat{E}_{i+1,3}$.

At last, we set function g as

$$\begin{aligned}g(x) =& f_j\big(\hat{E}_{i,0}(\mathbf{S}^{-1}(x) \oplus rk_{i,0}), 0, 0, 0\big) \\ =& \hat{E}_{i+1,0}(B_1 \circ x \oplus B_2 \circ \mathbf{S}_1(0) \oplus B_2 \circ \mathbf{S}_2(0) \oplus \\ & B_3 \circ \mathbf{S}_3(0)),\end{aligned}$$

and we can similarly recover the round key byte $rk_{i,0}$ and finally the whole round key rk_i with an expected time complexity of about $4 \cdot (2^8 \cdot 1 \cdot 2) = 2^{11}$. Therefore, the total expected time complexity for recovering a round key is about $259 \cdot 2^9 + 2^{11} \approx 2^{17.1}$.

VII. Concluding Remarks

The SM4 block cipher is a Chinese national standard and an ISO international standard, formerly known as SMS4. A few white-box SM4 implementations have been proposed since 2009, with an increasingly wide use of SM4. In this paper, we have analysed the security against collision-based attacks of four white-box SM4 implementations with the construction method that uses an affine (or linear) diagonal block encoding to protect the original output of an SM4 round function and uses the inverse of the encoding to protect the original input of the S-box layer of the next round, and have presented attacks with a practical time complexity. Thus, their security is much lower than previously published and their realistic significance is reduced. Our attacks indicate that a white-box SM4 implementation with this construction method is hardly practically secure generally, and such white-box SM4 constructions should be avoided unless improved somehow.

References

[1] C. H. Baek, J. H. Cheon, and H. Hong, "White-box AES implementation revisited," *Journal of Communications and Networks*, vol. 18, no. 3, pp. 273–287, 2016.

[2] K. P. Bai and C. K. Wu, "A secure white-box SM4 implementation," *Security and Communication Networks*, vol. 9, no. 10, pp. 996–1006, 2016.

[3] K. P. Bai, C. K. Wu, and Z. F. Zhang, "Protect white-box AES to resist table composition attacks," *IET Information Security*, vol. 12, no. 4, pp. 305–313, 2018.

[4] E. Barkan and E. Biham, "In how many ways can you write Rijndael?," in *International Conference on the Theory and Application of Cryptology and Information Security (ASIACRYPT 2002)*, vol. 2501, pp. 160–175, Springer, 2002.

[5] E. Biham, A. Shamir, Differential cryptanalysis of the Data Encryption Standard. Springer-Verlag, 1993.

[6] O. Billet, H. Gilbert, and C. Ech-Chatbi, "Cryptanalysis of a White Box AES Implementation," in *International Conference on Selected Areas in Cryptography (SAC 2004)*, vol. 3357, pp. 227–240, Springer, 2004.

[7] A. Biryukov, C. De Cannière, A. Braeken, and B. Preneel, "A toolbox for cryptanalysis: Linear and affine equivalence algorithms," in *International Conference on the Theory and Applications of Cryptographic Techniques (EUROCRYPT 2003)*, vol. 2656, pp. 33–50, Springer, 2003.

[8] J. Bringer, H. Chabanne, and E. Dottax, "White box cryptography: Another attempt," *IACR Cryptology ePrint Archive*, no. 2006, p. 468, 2006.

[9] S. Chow, P. Eisen, H. Johnson, and P. C. Van Oorschot, "White-box cryptography and an AES implementation," in *International Conference on Selected Areas in Cryptography (SAC 2002)*, vol. 2595, pp. 250–270, Springer, 2002.

[10] S. Chow, P. Eisen, H. Johnson, and P. C. Van Oorschot, "A white-box DES implementation for DRM applications," in *ACM Workshop on Digital Rights Management (DRM 2002)*, vol. 2696, pp. 1–15, Springer, 2002.

[11] P. Derbez, P. A. Fouque, B. Lambin, and B. Minaud, "On recovering affine encodings in white-box implementations," *IACR Transactions on Cryptographic Hardware and Embedded Systems*, vol. 2018, no. 3, pp. 121–149, 2018.

[12] *The SMS4 Block Cipher (in Chinese)*. Office of State Commercial Cryptography Administration of China, 2006.

[13] GB/T 32907-2016, *Information Security Technology — SM4 Block Cipher Algorithm*, Standardization Administration of China, 2016.

[14] International Standard, *ISO/IEC 18033-3:2010/AMD1:2021, Amendment 1 - Information technology - Security techniques - Encryption algorithms - Part 3: Block ciphers - SM4*, International Standardization of Organization (ISO), 2021.

[15] L. Goubin, J. M. Masereel, and M. Quisquater, "Cryptanalysis of white box DES implementations," in *International Conference on Selected Areas in Cryptography (SAC 2007)*, vol. 4876, pp. 278–295, Springer, 2007.

[16] M. Jacob, D. Boneh, and E. Felten, "Attacking an obfuscated cipher by injecting faults," in *ACM Workshop on Digital Rights Management (DRM 2002)*, vol. 2696, pp. 16–31, Springer, 2003.

[17] M. Karroumi, "Protecting white-box AES with dual ciphers," in *International Conference on Information Security and Cryptology (ICISC 2010)*, vol. 6829, pp. 278–291, Springer, 2011.

[18] X. J. Lai, "Higher order derivatives and differential cryptanalysis," Communications and Cryptography: Two Sides of One Tapestry, no. 1994, pp. 227–233, Springer Science & Business Media, 1994.

[19] T. Lepoint, M. Rivain, Y. De Mulder, P. Roelse, and B. Preneel, "Two attacks on a white-box AES implementation," in *International Conference on Selected Areas in Cryptography (SAC 2013)*, vol. 8282, pp. 265–285, Springer, 2014.

[20] T. T. Lin and X. J. Lai, "Efficient attack to white-box SMS4 implementation (in Chinese)," *Journal of Software*, vol. 24, no. 9, pp. 2238–2249, 2013.

[21] T. T. Lin, H. L. Yan, X. J. Lai, Y. X. Zhong, and Y. Jia, "Security evaluation and improvement of a white-box SMS4 implementation based on affine equivalence algorithm," *The Computer Journal*, vol. 61, no. 12, pp. 1783–1790, 2018.

[22] H. E. Link and W. D. Neumann, "Clarifying obfuscation: Improving the security of white-box DES," in *International Conference on Information Technology: Coding and Computing (ITCC 2005)*, vol. 1, pp. 679–684, IEEE, 2005.

[23] J. Q. Lu, J. Y. Li, "Cryptanalysis of two white-box implementations of the SM4 block cipher," in *International Conference on Information Security (ISC 2021)*, vol. 13118. pp. 54–69, Springer, 2021.

[24] R. Luo, X. J. Lai and R. You, "A new attempt of white-box AES implementation," in *Proceedings 2014 IEEE International Conference on Security, Pattern Analysis, and Cybernetics (SPAC 2014)*, IEEE, pp. 423-429, 2014.

[25] W. Michiels, P. Gorissen, and H. D. L. Hollmann, "Cryptanalysis of a generic class of white-box implementations," in *International Conference on Selected Areas in Cryptography (SAC 2008)*, vol. 5381, pp. 414–428, Springer, 2008.

[26] Y. De Mulder, P. Roelse, and B. Preneel, "Cryptanalysis of the Xiao - Lai white-box AES implementation," in *International Conference on Selected Areas in Cryptography (SAC 2012)*, vol. 7707, no. 1, pp. 34–49, Springer, 2013.

[27] Y. De Mulder, B. Wyseur, and B. Preneel, "Cryptanalysis of a perturbated white-box AES implementation," in *International Conference on Cryptology in India (INDOCRYPT 2010)*, vol. 6498, pp. 292–310, Springer, 2010.

[28] FIPS-197, *Advanced Encryption Standard (AES)*, National Institute of Standards and Technology (NIST) , 2001.

[29] FIPS-46, *Data Encryption Standard (DES)*, National Bureau of Standards (NBS), 1977.

[30] P. Shang, "White-box cryptography algorithm design and implementation of SMS4 (in Chinese)," Master's thesis, University of Electronic Science and Technology of China, 2016.

[31] Y. Shi, W. J. Wei, and Z. J. He, "A lightweight white-box symmetric encryption algorithm against node capture for WSNs," *Sensors*, vol. 15, no. 5, pp. 11928–11952, 2015.

[32] L. Tolhuizen, "Improved cryptanalysis of an AES implementation," in *Proceedings of the 33rd WIC Symposium on Information Theory in the Benelux and The 3rd Joint WIC / IEEE Symposium on Information Theory and Signal Processing in the Benelux*. pp. 68–71, Werkgemeenschap voor Informatie- en Communicatietheorie (WIC), 2012.

[33] R. S. Wang, H. Guo, J. Q. Lu, and J. W. Liu, "Cryptanalysis of a white-box SM4 implementation based on collision attack," *IET Information Security*, [Online]. Available: https://doi.org/10.1049/ise2.12045

[34] Z. Wu, J. Bai, D. S. Li, B. Li, B. Zeng, and Z. Q. Zhang, "White-box cryptographic video data sharing system based on SM4 algorithm (in Chinese)," *Journal of Beijing University of Aeronautics and Astronautics*, vol. 46, no. 9, pp. 1660–1669, 2020.

[35] B. Wyseur, W. Michiels, P. Gorisseii, and B. Preneel, "Cryptanalysis of white-box DES implementations with arbitrary external encodings," in *International Conference on Selected Areas in Cryptography (SAC 2007)*, vol. 4876, pp. 264–277, Springer, 2007.

[36] Y. Y. Xiao and X. J. Lai, "A secure implementation of White-Box AES," in *Proceedings of the Second International Conference on Computer Science and its Applications (CSA 2009)*, pp. 1–6, IEEE, 2009.

[37] Y. Y. Xiao and X. J. Lai, "White-box cryptography and a SMS4 implementation (in Chinese)," in *Proceedings of 2009 Annual Conference of the Chinese Association of Cryptologic Research*, pp. 24–34, 2009.

[38] S. Yao and J. Chen, "A new method for white-box implementation of SM4 algorithm (in Chinese)," *Journal of Cryptologic Research*, vol. 7, no. 3, pp. 358–374, 2020.