Hindawi
Security and Communication Networks
Volume 2017, Article ID 1461520, 10 pages
https://doi.org/10.1155/2017/1461520

Research Article

New Linear Cryptanalysis of Chinese Commercial Block Cipher Standard SM4

Yu Liu,[1,2] Huicong Liang,[1] Wei Wang,[1] and Meiqin Wang[1]

[1]*Key Laboratory of Cryptologic Technology and Information Security, Ministry of Education, Shandong University, Jinan, China*
[2]*Weifang University, Weifang, China*

Correspondence should be addressed to Meiqin Wang; mqwang@sdu.edu.cn

Received 8 June 2017; Accepted 1 August 2017; Published 6 September 2017

Academic Editor: Jesús Díaz-Verdejo

SM4 is a Chinese commercial block cipher standard used for wireless communication in China. In this paper, we use the partial linear approximation table of S-box to search for three rounds of iterative linear approximations of SM4, based on which the linear approximation for 20-round SM4 has been constructed. However, the best previous identified linear approximation only covers 19 rounds. At the same time, a linear approximation for 19-round SM4 is obtained, which is better than the known results. Furthermore, we show the key recovery attack on 24-round SM4 which is the best attack according to the number of rounds.

1. Introduction

SMS4 [1], issued in 2006 by Chinese government, serves the WAPI (WLAN Authentication and Privacy Infrastructure) as the underling block cipher for the security of wireless LANs. In 2012, SMS4 was announced as the Chinese commercial block cipher standard, renamed SM4 [2].

SM4 receives more attention from the cryptographic community and a lot of cryptanalytic results for SM4 have been produced. In [3], the rectangle and boomerang attacks on 18-round SM4 and the linear and differential attacks on 22-round SM4 have been presented. Using multiple linear attack, Etrog and Robshaw gave an attack on 23-round SM4 in [4]. Besides these, the differential attack and the multiple linear attack on 22-round SM4 have been introduced in [5, 6]. Till now, the best differential attack for 23-round SM4 is given in [7]. Cho and Nyberg proposed a multidimensional linear attack on 23-round SM4 in [8]. The best linear attack on 23-round SM4 is provided by Liu and Chen in [9]. Bai and Wu proposed a new lookup-table-based white-box implementation for SM4 which could protect the large linear encodings from being cancelled out in [10]. Moreover, related-key differential attack on SM4 has been given in [11] and the lower bound of the number of linear active S-boxes for SMS4-like ciphers has been analyzed in [12].

Linear cryptanalysis [13] is one of the most important techniques in the analysis of symmetric-key cryptographic primitives. The linear cryptanalysis focuses on the linear approximation between plaintext, ciphertext, and key. If a cipher behaves differently from a random permutation for linear cryptanalysis, this can be used to build a distinguisher or even a key recovery attack through adding some rounds. The subkeys of appended rounds are guessed and the ciphertexts are decrypted and/or plaintexts are encrypted using these subkeys to calculate intermediate state at the ends of distinguisher. If the subkeys are correctly guessed, then the distinguisher should hold. Otherwise, it will fail. Linear cryptanalysis has been used to analyze many ciphers such as [14–17].

Our Contributions. In terms of the number of rounds that all the previous attacks for SM4 can work, the best key recovery attacks on SM4 are linear cryptanalysis and differential cryptanalysis, and both of them are based on 19-round distinguishers. Whether we can get a better distinguisher is our first motivation to improve the attacks on SM4. Therefore, we focus on searching the linear approximation for SM4 to improve the attacks on SM4. The contributions of this paper are summarized as follows.

The best previous linear attacks work on the 19-round linear approximations. We design a new search algorithm for the

Table 1: Summary of linear approximations of SM4.

Rounds	Bias	Reference
19	$2^{-62.27}$	[9]
19	2^{-58}	Section 3
20	2^{-61}	Section 3

iterative linear approximations for small rounds of SM4 by gradually expanding the partial linear approximation table of S-box. Firstly, it is proved that there is no one-round or two-round iterative linear approximation for SM4, and then some properties are obtained for the iterative linear approximations of 3-round SM4. Based on these properties, we utilize our searching algorithm to get an 19-round linear approximation with bias 2^{-58} and a 20-round linear approximation with bias 2^{-61}. The results about our identified linear approximations with the previous ones are depicted in Table 1. It can be seen that our linear approximations are the best ones so far.

The best previous attacks can work on 23-round SM4. Utilizing our identified 20-round linear approximation of SM4, we give a key recovery attack on 24-round SM4, which is the best attack according to the number of rounds for SM4. Moreover, the new 19-round linear approximation is used to attack 23-round SM4. As a result, the best previous linear attack on 23-round SM4 is improved. A summary of our attacks and the previous attacks on SM4 is listed in Table 2.

The paper is organized as follows. Section 2 briefly describes the notations used in this paper and introduces the SM4 block cipher. Section 3 shows how to search the better linear approximations for SM4. In Section 4, we use the 19-round and 20-round linear approximations to attack 23-round and 24-round SM4, respectively. Section 5 concludes this paper.

2. Preliminaries

2.1. Notations. In this subsection, we will present the notations used in this paper as follows:

(i) $\oplus$: a bitwise XOR operation

(ii) $\|$: concatenation of two words

(iii) $\lll$: left cyclic shift operation

(iv) $\circ$: multiplication of two vectors, matrix and vector, or two matrices

(v) $\cdot$: bitwise inner product

(vi) &&: logical AND operation

(vii) $X[i]$: the ith bit of X

(viii) $X[i-j]$: a bit string starting from the ith bit to the jth bit of X.

2.2. Brief Description of SM4. SM4 is a Chinese national standard block cipher used in WAPI for WLAN. It has 128-bit block size and the key size is also 128 bits. The design of SM4 is based on the unbalanced generalized Feistel structure and the number of rounds is 32. We denote the plaintext

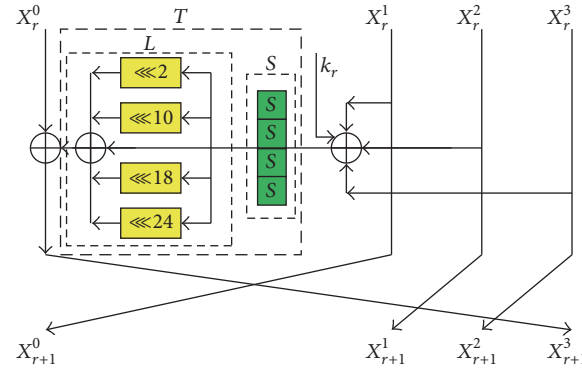

Figure 1: The round function of SM4.

as $(P_0, P_1, P_2, P_3) = (X_0^0, X_0^1, X_0^2, X_0^3) \in (\mathbb{F}_2^{32})^4$, and the encryption procedure is described as follows:

$$\begin{aligned} X_{r+1}^3 &= X_r^0 \oplus T\left(X_r^1 \oplus X_r^2 \oplus X_r^3 \oplus k_r\right) \\ &= X_r^0 \oplus L \circ S\left(X_r^1 \oplus X_r^2 \oplus X_r^3 \oplus k_r\right), \\ &\qquad \text{for } r = 0, \ldots, 31, \end{aligned} \tag{1}$$

where k_r is the rth round's subkey ($0 \le r \le 31$). The ciphertext $(C_0, C_1, C_2, C_3) = (X_{32}^3, X_{32}^2, X_{32}^1, X_{32}^0)$. The decryption procedure is the same as the encryption procedure with the reverse order of subkeys.

One round of SM4 is shown in Figure 1. It can be known from Figure 1 that T is composed of the nonlinear layer S and the linear transformation L. Layer S has four 8×8 S-boxes used in parallel. The specification of the S-box could be referred to [1]. Let $X \in \mathbb{F}_2^{32}$ and $Y \in \mathbb{F}_2^{32}$ be the 32-bit input and output words of the linear transformation L. Then

$$\begin{aligned} Y &= L(X) \\ &= X \oplus (X \lll 2) \oplus (X \lll 10) \oplus (X \lll 18) \\ &\quad \oplus (X \lll 24). \end{aligned} \tag{2}$$

The key schedule of SM4 is similar to the encryption procedure but the only difference between them is that the linear transformation in the key schedule is

$$L'(X) = X \oplus (X \lll 13) \oplus (X \lll 23). \tag{3}$$

The 128-bit master key $(\mathrm{MK}_0, \mathrm{MK}_1, \mathrm{MK}_2, \mathrm{MK}_3)$ is first masked with the constants $\mathrm{FK}_0, \mathrm{FK}_1, \mathrm{FK}_2, \mathrm{FK}_3$ and then input to the key schedule function.

$$\begin{aligned} (K_0, K_1, K_2, K_3) &= (\mathrm{MK}_0 \oplus \mathrm{FK}_0, \mathrm{MK}_1 \oplus \mathrm{FK}_1, \mathrm{MK}_2 \\ &\quad \oplus \mathrm{FK}_2, \mathrm{MK}_3 \oplus \mathrm{FK}_3), \end{aligned} \tag{4}$$

where $\mathrm{FK}_0 = 0xa3b1bac6$, $\mathrm{FK}_1 = 0x56aa3350$, $\mathrm{FK}_3 = 0x677d9197$, and $\mathrm{FK}_3 = 0xb27022dc$. And then k_r is computed as follows:

$$k_r = K_{r+4} = K_r \oplus L' \circ S\left(K_{r+1} \oplus K_{r+2} \oplus K_{r+3} \oplus \mathrm{CK}_r\right), \tag{5}$$

where CK_r, $r = 0, 1, \ldots, 31$, is the constant.

Table 2: Summary of attacks on SM4.

Rounds	Attack type	Data	Time	Memory	Source
13	Integral	2^{16}	2^{114}	—	[6]
16	Impossible differential	$2^{107.9}$	$2^{87.7}$	$2^{107.9}$	[18]
18	Rectangle	2^{124}	$2^{112.8}$	2^{128}	[3]
18	Boomerang	2^{120}	$2^{116.8}$	2^{123}	[3]
22	Multiple linear	2^{112}	$2^{119.8}$	$2^{118.8}$	[19]
23	Differential	2^{118}	$2^{126.7}$	—	[7]
23	Multidimensional linear	$2^{122.6}$	$2^{122.7}$	$2^{120.6}$	[9]
23	Linear	$2^{126.5}$	2^{122}	2^{116}	[9]
23	Linear	$2^{120.3}$	$2^{121.7}$	2^{85}	Section 4
24	Linear	$2^{126.6}$	$2^{126.6}$	2^{85}	Section 4

3. Search for the Linear Approximations of SM4

In terms of the number of rounds, all previous attacks for SM4 can work. One of the best key recovery attacks on SM4 is linear and differential cryptanalysis, and both of them are based on 19-round distinguishers. Whether we can get a better distinguisher is our first motivation to improve the attacks on SM4. Therefore, the key point is to search for the linear approximation of SM4. As far as we know, some methods to search for linear approximations of SM4 have been considered in [3, 4, 9, 19].

The search method in [3] is to construct linear approximations for reduced-round SM4 by identifying a one-round linear approximation with the same input and output masks for the T function. In this way, the number of active T functions can be minimized. As a result, an 18-round linear approximation with bias $2^{-57.28}$ for SM4 has been found.

In [4], Etrog and Robshaw derived a 5-round iterative linear approximation where only the last two rounds are active, and then they concatenated three five-round iterative linear approximations to construct an 18-round linear approximation with bias $2^{-56.2}$.

In [19], Liu et al. used the branch-and-bound algorithm in [20] to obtain a series of 5-round iterative linear approximations, which are utilized to construct an 18-round linear approximation with bias $2^{-56.14}$.

In order to get a better linear approximation for SM4, Liu and Chen gave a more dedicated search algorithm in [9]. They firstly used an MILP-based method to search the mode for the linear approximation with the minimum number of active S-boxes for reduced-round SM4; then based on the identified mode they found the 19-round linear approximation with bias $2^{-62.27}$.

It is obvious that even if the number of active S-boxes for a linear approximation is minimized, the absolute of its bias might not be maximum. From this point, we focus on searching for better linear approximations with a few more active S-boxes.

At CT-RSA 2014, Biryukov and Velichkov extended the branch-and-bound algorithm to search for the differential characteristics of ARX ciphers where the partial differential distribution table for modular addition is used in order to improve the search efficiency [21]. Inspired from this idea, we will use the partial linear approximation table to search for linear approximations of SM4.

At first, some properties for basic operations such as the XOR operation, the three-forked branching operation, and the linear map will be introduced.

Lemma 1 (XOR operation [22]). *Let $f(x_0, x_1) = x_0 \oplus x_1$; the input mask vector and output mask are $\Gamma = (\Gamma_0, \Gamma_1)$ and Λ, respectively. Then $\Pr(\Lambda \cdot f(x_0, x_1) = \Gamma_0 \cdot x_0 \oplus \Gamma_1 \cdot x_1) \neq 1/2$ if and only if $\Gamma_0 = \Gamma_1 = \Lambda$.*

Lemma 2 (three-forked branching operation [22]). *Let $f(x) = (x, x)$; the input mask and output linear mask vector are Γ and $\Lambda = (\Lambda_0, \Lambda_1)$, respectively. Then $\Pr(\Gamma \cdot x = \Lambda \cdot f(x)) \neq 1/2$ if and only if $\Gamma = \Lambda_0 \oplus \Lambda_1$.*

Lemma 3 (linear map [23]). *Let $f(x) = M \circ x$ with the input mask vector $\Gamma = (\Gamma_0, \Gamma_1, \ldots, \Gamma_{n-1})$ and output mask vector $\Lambda = (\Lambda_0, \Lambda_1, \ldots, \Lambda_{n-1})$; then $\Pr(\Gamma \cdot x = \Lambda \cdot f(x)) \neq 1/2$ if and only if $\Gamma = M^T \circ \Lambda$, where M^T is the transposed matrix of M and M is an $n \times n$ invertible binary matrix.*

Biases in the linear approximation table for S-box of SM4 take the values $l/2^8$ ($2 \leq l \leq 16$). If we put all the linear approximation table into the search program, the program will be too slow to get a better linear approximation. Thus, the partial linear approximation table is used in the search algorithm. The basic idea is that linear approximations of S-box with higher bias are utilized first. If no better linear approximation is output, then we can expand the partial linear approximation table by appending more linear approximations of S-box with less bias successively till a better linear approximation is output.

In order to get a better linear approximation, one common method is to find iterative linear approximations for short rounds first based on which long rounds of linear approximations could be produced directly. Thus, we will focus on searching for iterative linear approximations of SM4.

Now three properties for iterative linear approximations of SM4 are shown as follows.

Property 4. There is no one-round iterative linear approximation with active S-boxes on SM4.

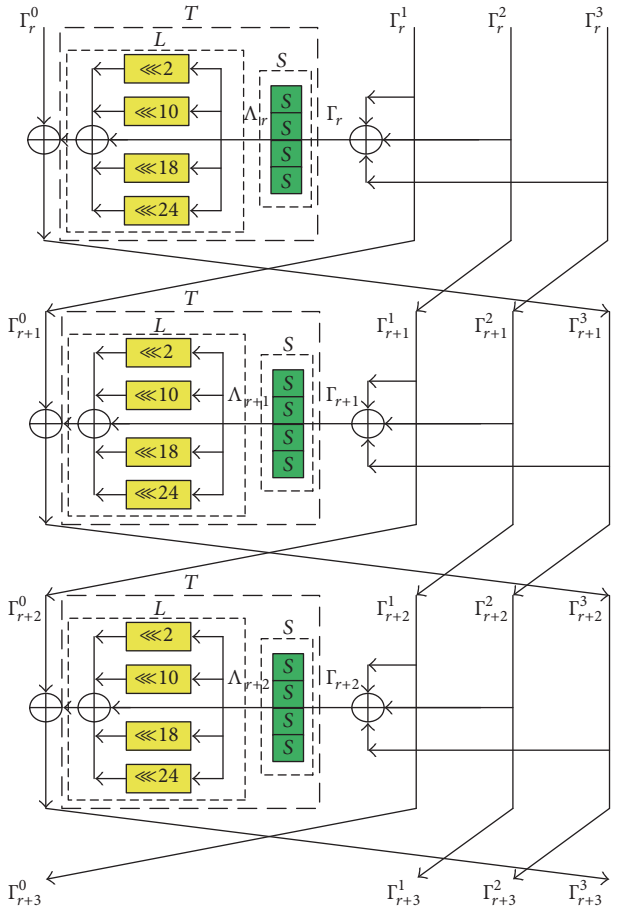

Figure 2: Linear approximation of 3-round SM4.

Proof. From Figure 2, if there is an iterative linear approximation for the first round, we have

$$\begin{aligned}\Gamma_r^0 &= \Gamma_{r+1}^0,\\ \Gamma_r^1 &= \Gamma_{r+1}^1,\\ \Gamma_r^2 &= \Gamma_{r+1}^2,\\ \Gamma_r^3 &= \Gamma_r^0.\end{aligned} \tag{6}$$

Using the property of three-forked branch, we have

$$\begin{aligned}\Gamma_r^1 &= \Gamma_{r+1}^0 \oplus \Gamma_r,\\ \Gamma_r^2 &= \Gamma_{r+1}^1 \oplus \Gamma_r,\\ \Gamma_r^3 &= \Gamma_{r+1}^2 \oplus \Gamma_r.\end{aligned} \tag{7}$$

From (6) and (7), we get

$$\begin{aligned}\Gamma_{r+1}^1 &= \Gamma_r^0 \oplus \Gamma_r,\\ \Gamma_r^0 \oplus \Gamma_r \oplus \Gamma_r &= \Gamma_{r+1}^2,\\ \Gamma_r^3 &= \Gamma_r^0 \oplus \Gamma_r \oplus \Gamma_r \oplus \Gamma_r = \Gamma_r^0 \oplus \Gamma_r = \Gamma_r^0,\end{aligned} \tag{8}$$

which implies $\Gamma_r = 0$ and all the S-boxes in this round are passive. Thus, there is no one-round iterative linear approximation for SM4. □

Property 5. The iterative linear approximation for two rounds of SM4 does not exist.

Proof. If there is an iterative linear approximation for the first two rounds in Figure 2, then we have

$$\begin{aligned}\Gamma_r^0 &= \Gamma_{r+2}^0,\\ \Gamma_r^1 &= \Gamma_{r+2}^1,\\ \Gamma_r^2 &= \Gamma_{r+2}^2,\\ \Gamma_r^3 &= \Gamma_{r+1}^0.\end{aligned} \tag{9}$$

With the property of three-forked branch, we have

$$\begin{aligned}\Gamma_r^1 &= \Gamma_{r+1}^0 \oplus \Gamma_r,\\ \Gamma_r^2 &= \Gamma_{r+2}^0 \oplus \Gamma_r \oplus \Gamma_{r+1},\\ \Gamma_r^3 &= \Gamma_{r+2}^1 \oplus \Gamma_r \oplus \Gamma_{r+1},\\ \Gamma_{r+2}^2 &= \Gamma_r^0 \oplus \Gamma_{r+1}.\end{aligned} \tag{10}$$

According to (9) and (10), we derive

$$\begin{aligned}\Gamma_r^2 &= \Gamma_r^0 \oplus \Gamma_r \oplus \Gamma_{r+1},\\ \Gamma_r^3 &= \Gamma_r^1 \oplus \Gamma_r \oplus \Gamma_{r+1},\\ \Gamma_r^2 &= \Gamma_r^0 \oplus \Gamma_{r+1}.\end{aligned} \tag{11}$$

Thus,

$$\Gamma_r^0 \oplus \Gamma_{r+1} = \Gamma_r^0 \oplus \Gamma_r \oplus \Gamma_{r+1} \Longrightarrow \Gamma_r = 0, \tag{12}$$

which means that $\Gamma_r^0 = \Gamma_{r+2}^0 = 0$. Substitute the terms Γ_r in the above formulas and we have

$$\begin{aligned}\Gamma_r^1 &= \Gamma_{r+1}^0 = \Gamma_{r+2}^1,\\ \Gamma_r^2 &= \Gamma_{r+1},\\ \Gamma_r^3 &= \Gamma_{r+2}^1 \oplus \Gamma_{r+1} = \Gamma_{r+1}^0 \oplus \Gamma_{r+1} = \Gamma_{r+1}^0,\end{aligned} \tag{13}$$

so $\Gamma_{r+1} = 0$, which means that all S-boxes in the first two rounds are passive. Therefore, 2-round iterative linear approximation for SM4 does not exist. □

Property 6. For the iterative linear approximation of 3-round SM4, the minimum number of active S-boxes is 3. Meanwhile, each round has one active S-box and the active S-boxes are located in the same positions of three rounds.

Table 3: Linear approximation for 19-round SM4.

Round	i	Γ_i^0	Λ_i	Γ_i	Bias	Γ_i^1	Γ_i^2	Γ_i^3
1	0	028*A*0828	002*A*0000	00950000	2^{-4}	021*F*0828	028*B*0828	028*A*0828
2	1	028*A*0828	002*A*0000	00940000	2^{-4}	021*E*0828	021*F*0828	028*A*0828
3	2	028*A*0828	002*A*0000	00010000	2^{-4}	028*B*0828	021*E*0828	028*A*0828
4	3	028*A*0828	002*A*0000	00950000	2^{-4}	021*F*0828	028*B*0828	028*A*0828
5	4	028*A*0828	002*A*0000	00940000	2^{-4}	021*E*0828	021*F*0828	028*A*0828
6	5	028*A*0828	002*A*0000	00010000	2^{-4}	028*B*0828	021*E*0828	028*A*0828
7	6	028*A*0828	002*A*0000	00950000	2^{-4}	021*F*0828	028*B*0828	028*A*0828
8	7	028*A*0828	002*A*0000	00940000	2^{-4}	021*E*0828	021*F*0828	028*A*0828
9	8	028*A*0828	002*A*0000	00010000	2^{-4}	028*B*0828	021*E*0828	028*A*0828
10	9	028*A*0828	002*A*0000	00950000	2^{-4}	021*F*0828	028*B*0828	028*A*0828
11	10	028*A*0828	002*A*0000	00940000	2^{-4}	021*E*0828	021*F*0828	028*A*0828
12	11	028*A*0828	002*A*0000	00010000	2^{-4}	028*B*0828	021*E*0828	028*A*0828
13	12	028*A*0828	002*A*0000	00950000	2^{-4}	021*F*0828	028*B*0828	028*A*0828
14	13	028*A*0828	002*A*0000	00940000	2^{-4}	021*E*0828	021*F*0828	028*A*0828
15	14	028*A*0828	002*A*0000	00010000	2^{-4}	028*B*0828	021*E*0828	028*A*0828
16	15	028*A*0828	002*A*0000	00950000	2^{-4}	021*F*0828	028*B*0828	028*A*0828
17	16	028*A*0828	002*A*0000	00940000	2^{-4}	021*E*0828	021*F*0828	028*A*0828
18	17	028*A*0828	002*A*0000	00010000	2^{-4}	028*B*0828	021*E*0828	028*A*0828
19	18	028*A*0828	002*A*0000	00950000	2^{-4}	021*F*0828	028*B*0828	028*A*0828
20	19	028*A*0828	002*A*0000	*	*	021*E*0828	021*F*0828	028*A*0828

Proof. If there is an iterative linear approximation for three rounds in Figure 2, then we have

$$\begin{aligned} \Gamma_r^1 \oplus \Gamma_r \oplus \Gamma_{r+2} &= \Gamma_r^2, \\ \Gamma_r^3 &= \Gamma_r^2 \oplus \Gamma_r \oplus \Gamma_{r+1}, \\ \Gamma_r^3 \oplus \Gamma_r \oplus \Gamma_{r+1} \oplus \Gamma_{r+2} &= \Gamma_r^0, \\ \Gamma_r^0 \oplus \Gamma_{r+1} \oplus \Gamma_{r+2} &= \Gamma_r^1. \end{aligned} \tag{14}$$

So

$$\begin{aligned} \Gamma_{r+1} &= \Gamma_r \oplus \Gamma_{r+2}, \\ \Gamma_r^0 &= \Gamma_r^3. \end{aligned} \tag{15}$$

We focus on the linear approximation with less active S-boxes. From (15), it is impossible for a three-round iterative linear approximation to have only one active S-box. If there are two active S-boxes, then $\Gamma_r = 0$ or $\Gamma_{r+1} = 0$ or $\Gamma_{r+2} = 0$. Hence, all S-boxes in the three-round linear approximation are passive. Take $\Gamma_r = 0$ as an example.

If $\Gamma_r = 0$, then $\Lambda_r = 0$, which implies $\Gamma_r^0 = 0$. Then

$$\begin{aligned} \Gamma_r^3 = \Gamma_{r+3}^0 = \Gamma_{r+3}^3 = \Gamma_r^0 = 0, \\ \because \Gamma_{r+2}^0 = \Gamma_{r+3}^3 \\ \therefore \Gamma_{r+2}^0 = 0. \end{aligned} \tag{16}$$

Thus, $\Lambda_{r+2} = 0$; we have $\Gamma_{r+2} = 0$.

$$\begin{aligned} 0 = \Gamma_r^3 \oplus \Gamma_r = \Gamma_{r+1}^2 = \Gamma_{r+1} \oplus \Gamma_{r+2}^1 = \Gamma_{r+1} \oplus \Gamma_{r+2} \oplus \Gamma_{r+3}^0 \\ = \Gamma_{r+1}. \end{aligned} \tag{17}$$

In the cases $\Gamma_{r+1} = 0$ and $\Gamma_{r+2} = 0$, we can also obtain that there is no active S-box in the three-round linear approximation by the similar way of the case $\Gamma_r = 0$. Therefore, the iterative linear approximation for three-round SM4 has at least three active S-boxes. From (15), it is clear that each round has one active S-box and these active S-boxes are located in the same positions of three rounds. □

From Property 6, we will try to search for the iterative linear approximation of 3-round SM4 where each round has only one active S-box. The search algorithm is listed in Algorithm 1. In Algorithm 1, the following notations are used. $\Gamma_r = \Gamma_{r,0} \| \Gamma_{r,1} \| \Gamma_{r,2} \| \Gamma_{r,3}$ and $\Lambda_r = \Lambda_{r,0} \| \Lambda_{r,1} \| \Lambda_{r,2} \| \Lambda_{r,3}$ are input and output masks of S-layer in the rth round. $\Gamma_{r,j}$ and $\Lambda_{r,j}$ are input and output mask of the jth S-box of the rth round. T^l is a partial linear approximation table of S-box which consists of linear approximations with bias no less than $l/2^8$ $(2 \le l \le 16)$.

After proceeding the search algorithm, we identify 12240 3-round iterative linear approximations with bias 2^{-10}. With any 3-round iterative linear approximation, we can construct linear approximations for 19-round and 20-round SM4 with bias 2^{-58} and 2^{-61}, respectively. Compared with the best previous 19-round linear approximation in [9], the bias has been improved from $2^{-62.27}$ to 2^{-58}. In Tables 3 and 4, we give linear approximations for 19-round and 20-round SM4, respectively, where all masks are denoted as hexadecimal values and "*" is undecided.

4. Key Recovery Attacks for SM4

4.1. Linear Attack on 24-Round SM4. We append two rounds to the bottom and the top of the 20-round linear

(1) **for** $l \leftarrow 16$ to 2 **do**
(2) **for** $j \leftarrow 0$ to 3 **do**
(3) **for all** $\Lambda_{0,j} \neq 0,\ \Lambda_{1,j} \neq 0,\ \Lambda_{2,j} \neq 0,\ \Lambda_{0,j'} = \Lambda_{1,j'} = \Lambda_{2,j'} = 0,\ (0 \leq j' \neq j \leq 3)$ **do**
(4) $\Gamma_{0,j'} = \Gamma_{1,j'} = \Gamma_{2,j'} = 0,\ (0 \leq j' \neq j \leq 3)$
(5) find all m_0 input masks indexed by $\Lambda_{0,j}$ from T^l, and store in $R_j[m_0]$
(6) **for** $k_0 \leftarrow 0$ to $m_0 - 1$ **do**
(7) $\Gamma_{0,j} \leftarrow R_j[k_0]$
(8) find all m_1 input masks indexed by $\Lambda_{1,j}$ from T^l, and store in $R_j[m_1]$
(9) **for** $k_1 \leftarrow 0$ to $m_1 - 1$ **do**
(10) $\Gamma_{1,j} \leftarrow R_j[k_1]$
(11) find all m_2 input masks indexed by $\Lambda_{2,j}$ from T^l, and store in $R_j[m_2]$
(12) **for** $k_2 \leftarrow 0$ to $m_2 - 1$ **do**
(13) $\Gamma_{2,j} \leftarrow R_j[k_2]$
(14) **for** $i \leftarrow 0$ to 2 **do**
(15) $\Gamma_i^0 = L^T \circ (\Lambda_{i,0} \parallel \Lambda_{i,1} \parallel \Lambda_{i,2} \parallel \Lambda_{i,3}) = L^T \circ (\Lambda_i)$
(16) **end for**
(17) **for** $i \leftarrow 0$ to 2 **do**
(18) $\Gamma_i = \Gamma_{i,0} \parallel \Gamma_{i,1} \parallel \Gamma_{i,2} \parallel \Gamma_{i,3}$
(19) **end for**
(20) $\Gamma_3^0 = \Gamma_0^0$
(21) $\Gamma_3^1 = \Gamma_0^0 \oplus \Gamma_1 \oplus \Gamma_2$
(22) $\Gamma_3^2 = \Gamma_1^0 \oplus \Gamma_2$
(23) $\Gamma_0^3 = \Gamma_0^0 \oplus \Gamma_0 \oplus \Gamma_1 \oplus \Gamma_2$
(24) $\Gamma_0^1 = \Gamma_1^0 \oplus \Gamma_0$
(25) $\Gamma_0^2 = \Gamma_2^0 \oplus \Gamma_0 \oplus \Gamma_1$
(26) **if** $\Gamma_0^3 = \Gamma_2^0$ && $\Gamma_0^1 = \Gamma_3^1$ && $\Gamma_0^2 = \Gamma_3^2$ **then**
(27) **return** $\Gamma_0^0, \Gamma_1^0, \Gamma_2^0, \Gamma_3^0, \Gamma_0, \Gamma_1, \Gamma_2$
(28) // 3-round iterative linear approximation
(29) **else**
(30) **continue**.
(31) **end if**
(32) **end for**
(33) **end for**
(34) **end for**
(35) **end for**
(36) **end for**
(37) $l = l - 2$
(38) **end for**

Algorithm 1: Search algorithm for iterative linear approximation of 3-round SM4.

approximation in Table 4, respectively. Then a linear attack on 24-round SM4 is presented. The partial sum technique [24] is used in the partial encryption and decryption procedures. See Figure 3.

According to the linear approximation in Figure 3, we denote $\Gamma_2^0 = \Gamma_2^3 = \Gamma_{22}^0 = \Gamma_{22}^3 = 0x8808A228$, $\Gamma_2^1 = 0x8808C228$, $\Gamma_2^2 = \Gamma_{22}^1 = 0x88080828$, $\Gamma_{22}^2 = 0x88086828$, $\Lambda_1 = \Lambda_{22} = 0x00008200$, and $\Lambda_{1,2} = \Lambda_{22,2} = 0x82$. From the linear approximation, we have

$$\Gamma_2^0 \cdot X_2^0 \oplus \Gamma_2^1 \cdot X_2^1 \oplus \Gamma_2^2 \cdot X_2^2 \oplus \Gamma_2^3 \cdot X_2^3 \oplus \Gamma_{22}^0 \cdot X_{22}^0 \oplus \Gamma_{22}^1 \cdot X_{22}^1 \oplus \Gamma_{22}^2 \cdot X_{22}^2 \oplus \Gamma_{22}^3 \cdot X_{22}^3 = \kappa. \tag{18}$$

Consider the partial encryption and decryption; the left side of the above equation can be written as follows:

$$\begin{aligned}
&\Gamma_2^0 \cdot P_2 \oplus \Gamma_2^1 \cdot P_3 \oplus \Gamma_2^2 \cdot (P_0 \oplus XX_0) \oplus \Gamma_2^0 \cdot (P_1 \oplus XX_1) \\
&\quad \oplus \Gamma_2^0 \cdot (C_2 \oplus XX_{22}) \oplus \Gamma_2^2 \cdot (C_3 \oplus XX_{23}) \oplus \Gamma_{22}^2 \\
&\quad \cdot C_0 \oplus \Gamma_2^0 \cdot C_1 \\
&= \Gamma_2^0 \cdot P_2 \oplus \Gamma_2^1 \cdot P_3 \oplus \Gamma_2^2 \cdot P_0 \oplus \Gamma_2^0 \cdot P_1 \oplus \Gamma_2^0 \cdot C_2 \oplus \Gamma_2^2 \\
&\quad \cdot C_3 \oplus \Gamma_{22}^2 \cdot C_0 \oplus \Gamma_2^0 \cdot C_1 \oplus \Gamma_2^2 \cdot XX_0 \oplus \Gamma_2^0 \cdot XX_1 \\
&\quad \oplus \Gamma_2^0 \cdot XX_{22} \oplus \Gamma_2^2 \cdot XX_{23},
\end{aligned} \tag{19}$$

where XX_r is the state after transformation T and XS_r is the state after layer S in the rth round.

Since

$$\begin{aligned}
\Gamma_2^0 \cdot XX_{22} &= \Lambda_{22} \cdot XS_{22} = \Lambda_{22,2} \cdot XS_{22}\ [16\text{–}23] \\
&= \Lambda_{1,2} \cdot S\left(\left(X_{22}^1 \oplus X_{22}^2 \oplus X_{22}^3 \oplus k_{22}\right)[16\text{–}23]\right) \\
&= \Lambda_{1,2} \cdot S\left((C_3 \oplus C_0 \oplus C_1)\,[16\text{–}23\right]
\end{aligned}$$

Table 4: Linear approximation for 20-round SM4.

Round	i	Γ_i^0	Λ_i	Γ_i	Bias	Γ_i^1	Γ_i^2	Γ_i^3
1	0	8808*A*228	00008200	00006000	2^{-4}	8808*C*228	88080828	8808*A*228
2	1	8808*A*228	00008200	0000*CA*00	2^{-4}	88086828	8808*C*228	8808*A*228
3	2	8808*A*228	00008200	0000*AA*00	2^{-4}	88080828	88086828	8808*A*228
4	3	8808*A*228	00008200	00006000	2^{-4}	8808*C*228	88080828	8808*A*228
5	4	8808*A*228	00008200	0000*CA*00	2^{-4}	88086828	8808*C*228	8808*A*228
6	5	8808*A*228	00008200	0000*AA*00	2^{-4}	88080828	88086828	8808*A*228
7	6	8808*A*228	00008200	00006000	2^{-4}	8808*C*228	88080828	8808*A*228
8	7	8808*A*228	00008200	0000*CA*00	2^{-4}	88086828	8808*C*228	8808*A*228
9	8	8808*A*228	00008200	0000*AA*00	2^{-4}	88080828	88086828	8808*A*228
10	9	8808*A*228	00008200	00006000	2^{-4}	8808*C*228	88080828	8808*A*228
11	10	8808*A*228	00008200	0000*CA*00	2^{-4}	88086828	8808*C*228	8808*A*228
12	11	8808*A*228	00008200	0000*AA*00	2^{-4}	88080828	88086828	8808*A*228
13	12	8808*A*228	00008200	00006000	2^{-4}	8808*C*228	88080828	8808*A*228
14	13	8808*A*228	00008200	0000*CA*00	2^{-4}	88086828	8808*C*228	8808*A*228
15	14	8808*A*228	00008200	0000*AA*00	2^{-4}	88080828	88086828	8808*A*228
16	15	8808*A*228	00008200	00006000	2^{-4}	8808*C*228	88080828	8808*A*228
17	16	8808*A*228	00008200	0000*CA*00	2^{-4}	88086828	8808*C*228	8808*A*228
18	17	8808*A*228	00008200	0000*AA*00	2^{-4}	88080828	88086828	8808*A*228
19	18	8808*A*228	00008200	00006000	2^{-4}	8808*C*228	88080828	8808*A*228
20	19	8808*A*228	00008200	0000*CA*00	2^{-4}	88086828	8808*C*228	8808*A*228
21	20	8808*A*228	00008200	*	*	88080828	88086828	8808*A*228

$$\begin{aligned}
&\oplus XX_{23}[16\text{–}23] \oplus k_{22}[16\text{–}23]),\\
&\Gamma_2^0 \cdot XX_1 = \Lambda_1 \cdot XS_1 = \Lambda_{1,2} \cdot XS_1[16\text{–}23] = \Lambda_{1,2}\\
&\cdot S\left(\left(X_1^1 \oplus X_1^2 \oplus X_1^3 \oplus k_1\right)[16\text{–}23]\right) = \Lambda_{1,2}\\
&\cdot S\left(\left(P_0 \oplus P_2 \oplus P_3\right)[16\text{–}23] \oplus XX_0[16\text{–}23]\right.\\
&\left.\oplus k_1[16\text{–}23]\right),
\end{aligned} \tag{20}$$

then, (19) can be transformed into

$$\begin{aligned}
&\Gamma_2^0 \cdot P_2 \oplus \Gamma_2^1 \cdot P_3 \oplus \Gamma_2^2 \cdot P_0 \oplus \Gamma_2^0 \cdot P_1 \oplus \Gamma_2^0 \cdot C_2 \oplus \Gamma_2^2 \cdot C_3\\
&\oplus \Gamma_{22}^2 \cdot C_0 \oplus \Gamma_2^0 \cdot C_1 \oplus \Gamma_2^2 \cdot T\left(P_1 \oplus P_2 \oplus P_3 \oplus k_0\right)\\
&\oplus \Lambda_{1,2} \cdot S\left(\left(P_0 \oplus P_2 \oplus P_3\right)[16\text{–}23] \oplus XX_0[16\text{–}23]\right.\\
&\left.\oplus k_1[16\text{–}23]\right) \oplus \Lambda_{1,2} \cdot S\left(\left(C_3 \oplus C_0 \oplus C_1\right)[16\text{–}23]\right.\\
&\left.\oplus XX_{23}[16\text{–}23] \oplus k_{22}[16\text{–}23]\right) \oplus \Gamma_2^2 \cdot T\left(C_0 \oplus C_1\right.\\
&\left.\oplus C_2 \oplus k_{23}\right).
\end{aligned} \tag{21}$$

Let $\mathcal{O} = \Gamma_2^0 \cdot P_2 \oplus \Gamma_2^1 \cdot P_3 \oplus \Gamma_2^2 \cdot P_0 \oplus \Gamma_2^0 \cdot P_1 \oplus \Gamma_2^0 \cdot C_2 \oplus \Gamma_2^2 \cdot C_3 \oplus \Gamma_{22}^2 \cdot C_0 \oplus \Gamma_2^0 \cdot C_1$. The attack process is given as follows:

(1) Collect N plaintext/ciphertext pairs.

(2) Initialize 2^{81} counters $\mathcal{V}_0[0], \ldots, \mathcal{V}_0[2^{81}-1]$ to zero.

(3) For every plaintext/ciphertext pair, calculate $w = \mathcal{O} \parallel (P_0 \oplus P_2 \oplus P_3)[16\text{–}23] \parallel (C_3 \oplus C_0 \oplus C_1)[16\text{–}23] \parallel C_0 \oplus C_1 \oplus C_2 \parallel P_1 \oplus P_2 \oplus P_3$. Then increase the counter $\mathcal{V}_0[w]$ by one.

(4) Guess the 32-bit k_{23}. Allocate 2^{49} counters $\mathcal{V}_1[0], \ldots, \mathcal{V}_1[2^{49}-1]$ to zero.

(5) For every $0 \le w \le 2^{81}-1$, calculate $\mathcal{O} \Leftarrow \mathcal{O} \oplus \Gamma_2^2 \cdot T(C_0 \oplus C_1 \oplus C_2 \oplus k_{23})$, XX_{23} and $x = \mathcal{O} \parallel (P_0 \oplus P_2 \oplus P_3)[16\text{–}23] \parallel (C_3 \oplus C_0 \oplus C_1)[16\text{–}23] \oplus XX_{23}[16\text{–}23] \parallel P_1 \oplus P_2 \oplus P_3$. $\mathcal{V}_1[x] \mathrel{+}= \mathcal{V}_0[w]$.

(6) Guess the 8-bit $k_{22}[16\text{–}23]$. Allocate 2^{41} counters $\mathcal{V}_2[0], \ldots, \mathcal{V}_2[2^{41}-1]$ to zero.

(7) For every $0 \le x \le 2^{49}-1$, calculate $\mathcal{O} \Leftarrow \mathcal{O} \oplus \Lambda_{1,2} \cdot S((C_3 \oplus C_0 \oplus C_1)[16\text{–}23] \oplus XX_{23}[16\text{–}23] \oplus k_{22}[16\text{–}23])$ and $y = \mathcal{O} \parallel (P_0 \oplus P_2 \oplus P_3)[16\text{–}23] \parallel P_1 \oplus P_2 \oplus P_3$. $\mathcal{V}_2[y] \mathrel{+}= \mathcal{V}_1[x]$.

(8) Guess the 32-bit k_0. Allocate 2^9 counters $\mathcal{V}_3[0], \ldots, \mathcal{V}_3[2^9-1]$ to zero.

(9) For every $0 \le y \le 2^{41}-1$, calculate $\mathcal{O} \Leftarrow \mathcal{O} \oplus \Gamma_2^2 \cdot T(P_1 \oplus P_2 \oplus P_3 \oplus k_0)$, XX_0 and $z = \mathcal{O} \parallel (P_0 \oplus P_2 \oplus P_3)[16\text{–}23] \oplus XX_0[16\text{–}23]$. $\mathcal{V}_3[z] \mathrel{+}= \mathcal{V}_2[y]$.

(10) Guess the 8-bit $k_1[16\text{–}23]$. Initialize 2^{80} counters $\mathcal{V}_{\text{key}}[0], \ldots, \mathcal{V}_{\text{key}}[2^{80}-1]$ to zero.

(11) For every $0 \le z \le 2^9-1$, calculate $\mathcal{O} \Leftarrow \mathcal{O} \oplus \Lambda_{1,2} \cdot S((P_0 \oplus P_2 \oplus P_3)[16\text{–}23] \oplus XX_0[16\text{–}23] \oplus k_1[16\text{–}23])$. If $\mathcal{O} = 0$, increase the counter $\mathcal{V}_{\text{key}}[k_0 \parallel k_1[16\text{–}23] \parallel k_{22}[16\text{–}23] \parallel k_{23}]$ by $\mathcal{V}_3[z]$; otherwise, decrease it by $\mathcal{V}_3[z]$.

(12) We set the advantage a to be 47 which implies that the top 2^{33} absolute values in $\mathcal{V}_{\text{key}}$ are kept. For each

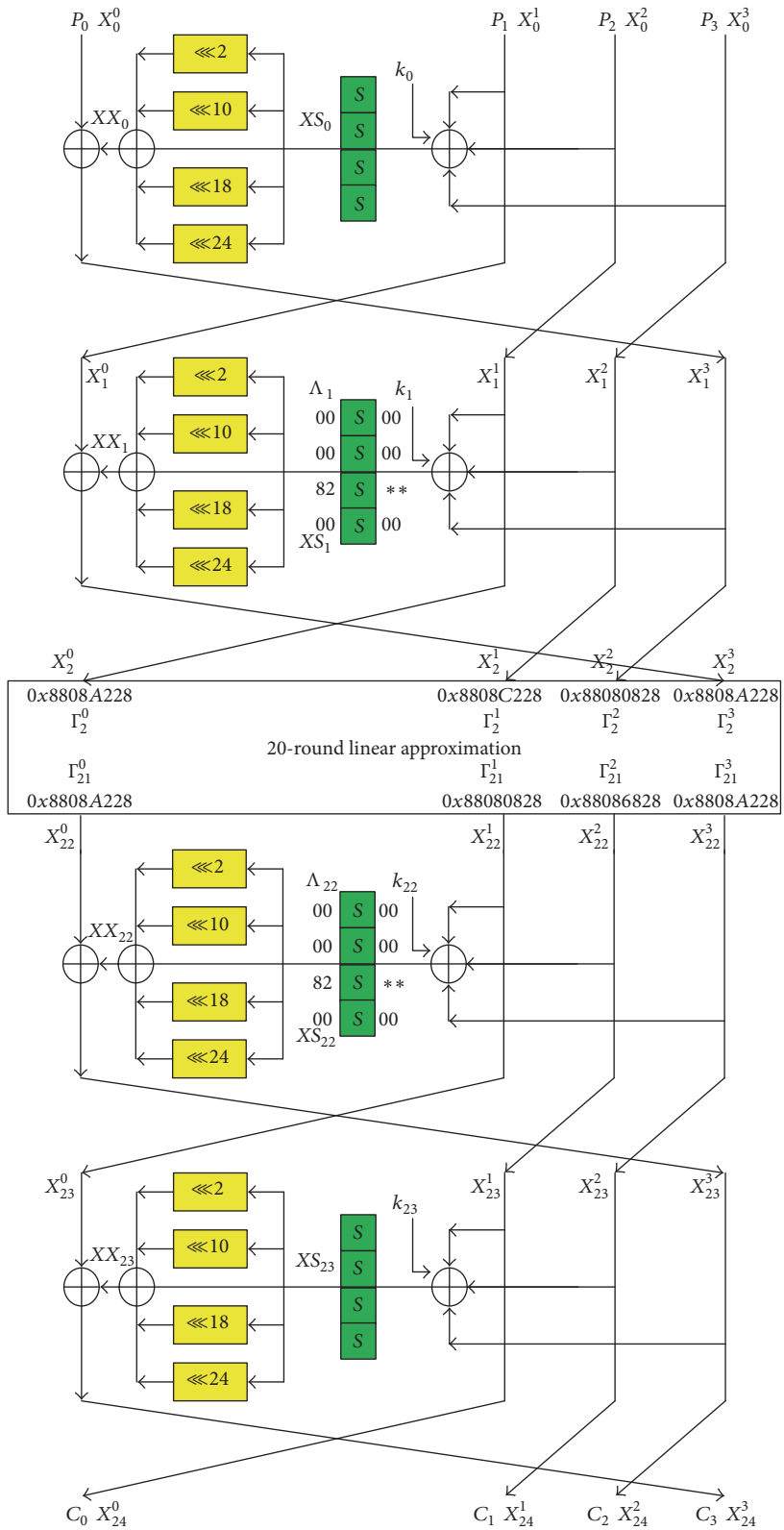

Figure 3: Key recovery attack on 24-round SM4.

kept subkey value, we guess the remaining 88 bits of $k_1[0\text{–}15, 24\text{–}31] \parallel k_2 \parallel k_3$ (the master key can be gotten from the key schedule) and test the key by trail encryptions.

The time complexity of Step (3) is about $2^{126.6}$ operations which is equivalent to $2^{126.6}/24 = 2^{122.0}$ 24-round encryptions. Both Steps (5) and (9) need 2^{113} one-round decryptions or encryptions. Steps (7) and (11) take 2^{89} one-round decryptions or encryptions. The complexity of Step (12) is 2^{121} 24-round encryptions. So the total time complexity is about $2^{122.6}$ encryptions.

The memory complexity of Step (2) is about $2^{81} \times 8 = 2^{84}$ bytes and the counter $\mathscr{V}_{\text{key}}$ requires $2^{80} \times 16 = 2^{84}$ bytes, so the total memory complexity is about $2^{84} + 2^{84} = 2^{85}$ bytes.

If we set the data complexity $N = 2^{126.6}$, the success rate $P_S = \Phi(2\sqrt{N}|\epsilon| - \sqrt{1 + N/2^n}\Phi^{-1}(1 - 2^{-a-1})) = 76.1\%$ by [25]. The time complexity is $2^{126.6}$ 24-round encryptions.

4.2. Linear Attack on 23-Round SM4. Two rounds are added to the bottom and the top of the 19-round linear approximation in Table 3, respectively. The key recovery attack on 23-round SM4 is similar to the attack procedure of 24-round SM4, so we omit details of the process.

If we set the data complexity $N = 2^{120.3}$ and the advantage a to be 47, the time complexity is $2^{120.3} + 2^{121} = 2^{121.7}$ 23-round encryptions, and the memory complexity is 2^{85} bytes. The success rate $P_S = 85.9\%$ is computed with the method in [25].

5. Conclusions

In this paper, it is firstly shown that there is no one-round or two-round iterative linear approximation for SM4 and the property for the 3-round iterative linear approximation. On the basis of the property, we search for the iterative linear approximation of 3-round SM4 by the partial linear approximation table. Next the 20-round linear approximation is constructed by 3-round iterative linear approximations. The best previous distinguishers only cover 19 rounds. Then the key recovery attack on 24-round SM4 is provided, which is the best known attack on SM4 so far. Moreover, we also get a better 19-round linear approximation, used to improve the linear attack on 23-round SM4. As for future work, we hope to use the similar technique to search for a better differential characteristic for SM4.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work is supported by 973 Program (no. 2013CB834205), NSFC Projects (nos. 61133013 and 61572293), and Program for New Century Excellent Talents in University of China (NCET-13-0350).

References

[1] W. Diffie and G. Ledin, "SMS4 Encryption Algorithm for Wireless Networks," Cryptology ePrint Archive 2008/329, 2014, http://eprint.iacr.org/2008/329.pdf.

[2] "Office of State Commercial Cryptography Administration: Specification of SMS4, block cipher for WLAN products-SMS4" (Chinese), http://www.oscca.gov.cn/UpFile/20062101642319799O.pdf.

[3] T. Kim, J. Kim, S. Hong, and J. Sung, "Linear and Differential Cryptanalysis of Reduced SMS4 Block Cipher," IACR Cryptology ePrint Archive 2008/281, 2008, https://eprint.iacr.org/2008/2811.pdf.

[4] J. Etrog and M. J. B. Robshaw, "The Cryptanalysis of Reduced-Round SMS4," in *Selected Areas in Cryptography*, vol. 5381 of *Lecture Notes in Computer Science*, pp. 51–65, Springer, Berlin, Germany, 2008.

[5] W. Zhang, W. Wu, D. Feng, and B. Su, "Some New Observations on the SMS4 Block Cipher in the Chinese WAPI Standard," in *Information Security Practice and Experience*, vol. 5451 of *Lecture Notes in Computer Science*, pp. 324–335, Springer, Berlin, Germany, 2009.

[6] F. Liu, W. Ji, L. Hu et al., "Analysis of the SMS4 Block Cipher," in *Information Security and Privacy*, vol. 4586 of *Lecture Notes in Computer Science*, pp. 158–170, Springer, Berlin, Germany, 2007.

[7] B.-Z. Su, W.-L. Wu, and W.-T. Zhang, "Security of the SMS4 block cipher against differential cryptanalysis," *Journal of Computer Science and Technology*, vol. 26, no. 1, pp. 130–138, 2011.

[8] J. Cho and K. Nyberg, "Improved Linear Cryptanalysis of SMS4 Block Cipher," *Symmetric Key Encryption Workshop*, pp. 1–14, 2011.

[9] M.-J. Liu and J.-Z. Chen, "Improved linear attacks on the Chinese block cipher standard," *Journal of Computer Science and Technology*, vol. 29, no. 6, pp. 1123–1133, 2014.

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

[11] J. Zhang, W. Wu, and Y. Zheng, "Security of SM4 Against (Related-Key) Differential Cryptanalysis," in *Proceedings of the International Conference on Information Security Practice and Experience*, vol. 10060 of *Lecture Notes in Computer Science*, pp. 65–78, Springer, Berlin, Germany, November 2016.

[12] B. Zhang and C. Jin, "Practical security against linear cryptanalysis for SMS4-like ciphers with SP round function," *Science China Information Sciences*, vol. 55, no. 9, pp. 2161–2170, 2012.

[13] T. Helleseth, "Linear cryptanalysis method for des cipher," in *Advances in Cryptology—EUROCRYPT*, vol. 765 of *Lecture Notes in Computer Science*, pp. 386–397, Springer, Berlin, Germany, 1993.

[14] F. Sano, K. Ohkuma, H. Shimizu, and S. Kawamura, "On the security of nested SPN cipher against the differential and linear cryptanalysis," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E86-A, no. 1, pp. 37–46, 2003.

[15] G. Jakimoski and L. Kocarev, "Differential and linear probabilities of a block-encryption cipher," *IEEE Transactions on Circuits and Systems. I. Fundamental Theory and Applications*, vol. 50, no. 1, pp. 121–123, 2003.

[16] Y. Sun, "Linear Cryptanalysis of Light-Weight Block Cipher ICEBERG," in *Advances in Electronic Commerce, Web Application and Communication*, vol. 149, pp. 529–532, Springer Berlin Heidelberg, Berlin, Germany, 2012.

[17] Y. Liu, K. Fu, W. Wang, L. Sun, and M. Wang, "Linear cryptanalysis of reduced-round SPECK," *Information Processing Letters*, vol. 116, no. 3, pp. 259–266, 2016.

[18] D. Toz and O. Dunkelman, "Analysis of two attacks on reduced-round versions of the SMS_4," in *Information and Communications Security*, vol. 5308 of *Lecture Notes in Computer Science*, pp. 141–156, Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.

[19] Z. Liu, D. Gu, and J. Zhang, "Multiple linear cryptanalysis of reduced-round SMS4 block cipher," *Chinese Journal of Electronics*, vol. 19, no. 3, pp. 389–393, 2010.

[20] M. Matsui, "On correlation between the order of *S*-boxes and the strength of DES," in *Advances in cryptology—EUROCRYPT*, vol. 950 of *Lecture Notes in Comput. Sci.*, pp. 366–375, Springer, Berlin, Germany, 1994.

[21] A. Biryukov and V. Velichkov, "Automatic search for differential trails in ARX ciphers," in *Topics in Cryptology—CT-RSA 2014*, vol. 8366 of *Lecture Notes in Comput. Sci.*, pp. 227–250, Springer, Berlin, Germany, 2014.

[22] E. Biham, "On Matsui's linear cryptanalysis," in *Advances in Cryptology*, vol. 950 of *Lecture Notes in Comput. Sci.*, pp. 341–355, Springer, Berlin, Germany, 1994.

[23] J. Daemen, R. Govaerts, and J. Vandewalle, "Correlation matrices," in *Fast Software Encryption*, vol. 1008 of *Lecture Notes in Computer Science*, pp. 275–285, Springer, Berlin, Germany, 1994.

[24] N. Ferguson, J. Kelsey, S. Lucks et al., "Improved Cryptanalysis of Rijndael," in *Fast Software Encryption*, vol. 1978 of *Lecture Notes in Computer Science*, pp. 213–230, Springer, Berlin, Germany, 2000.

[25] A. Bogdanov and E. Tischhauser, "On the Wrong Key Randomisation and Key Equivalence Hypotheses in Matsui's Algorithm 2," in *Fast Software Encryption*, vol. 8424 of *Lecture Notes in Computer Science*, pp. 19–38, Springer, Berlin, Germany, 2013.