Security of SM4 Against (Related-Key) Differential Cryptanalysis

Jian Zhang[1,2], Wenling Wu[1(✉)], and Yafei Zheng[1]

[1] Institute of Software, Chinese Academy of Sciences, Beijing 100190, China
{zhangjian,wwl,zhengyafei}@tca.iscas.ac.cn
[2] State Key Laboratory of Cryptology, Beijing 100190, China

Abstract. In this paper, we study the security of SM4 block cipher against (related-key) differential cryptanalysis by making use of the Mixed Integer Linear Programming (MILP) method.

SM4 is the first commercial block cipher standard of China, which attracts lots of attentions in cryptography. To analyze the security of SM4 against differential attack, we exploit a highly automatic MILP method to determine the minimum number of active S-boxes for consecutive rounds of SM4. We try to dig out the underlying relationships in different rounds, and convert them to the constraints trickily to extend the MILP model, in order to cut off the invalid differential modes as many as possible. We obtain tighter lower bounds on the number of active S-boxes by solving the extended MILP model with optimizer Gurobi. Moreover, we consider the security of SM4 against related-key differential analysis. We construct the extended MILP model by adding more helpful constraints, and get the lower bounds on the number of active S-boxes, which proves the intuition of stronger differential security of SM4 in the related-key setting. Our results shows that there exists no differential characteristic with probability larger than 2^{-128} for 23 rounds of SM4 in the single-key setting and 19 rounds in the related-key setting.

Keywords: Block cipher · SM4 · Differential attack · Active S-box · Related-key differential attack · Mixed-integer linear programming

1 Introduction

SMS4 is the underlying block cipher used in the WAPI (WLAN Authentication and Privacy Infrastructure), which is the Chinese national standard for protecting the wireless LANs. It is declassified by Chinese government in January 2006 ([5] gives an English translation) and becomes the first Chinese commercial block cipher standard in 2012 with a new name "SM4". Therefore, it has been wildly used in Chinese industry and many international corporations, such as Sony, supporting SM4 in relevant products.

SM4 employs an unbalanced generalized Feistel network, with a 128-bit block size, a 128-bit key and a total of 32 rounds. Its simplicity and Chinese standard prominence have encouraged a lot of analysis on the round-reduced SM4 [7,10,

F. Bao et al. (Eds.): ISPEC 2016, LNCS 10060, pp. 65–78, 2016.
DOI: 10.1007/978-3-319-49151-6_5

11,15,18,22]. Among these analysis, we focus on the works of Zhang [23] and Wu [19], which give lower bounds on the number of active S-boxes for consecutive rounds of SM4. Zhang et al. estimate the minimum number of active S-boxes by enumerating possible cases of differential propagation, and conclude that 31 rounds of SM4 is believed to be secure against differential attack after considering 4 rounds of guessing subkeys, which shows poor security margin with just one round for SM4. Wu et al. present a algorithm based on Integer programming to search the lower bound on the number of active S-boxes for various structures, including the structure of SM4, but with a limited number of rounds.

The lower bound on the number of active S-boxes can directly give the upper bound for the probability of the best differential characteristic of the cipher, which can be used to prove the security against differential cryptanalysis. Therefore, determining the lower bound on the number of active S-boxes is of great interest and lots of works have been done. Generally, the lower bound can be get mainly with two methods, proved mathematically [9,13,20,23] and counted with certain algorithm automatically [3,12,19]. In this paper, we employ the highly automatic MILP method to determine the minimum number of active S-boxes for consecutive rounds of SM4 in the single-key setting and related-key setting, respectively.

MILP and Differential Characteristic Search. MILP (Mixed Integer Linear Programming) aims at minimizing or maximizing a linear objective function subject to some linear equalities and inequalities. Using MILP method, what an analyst needs to do is just to write a simple program to generate the MILP model with suitable objective function and constraints resulted from the differential propagation of the cipher. The remaining work can be done by the highly optimized solver such as CPLEX [14] and Gurobi [1]. Because of the simplicity and highly automatic property, MILP method has been wildly applied in cryptography [2,8,12,16,17].

Contributions of this Paper. In this paper, we focus on how to extend Mouha et al.'s MILP method to get a tighter bound on the number of active S-boxes for SM4. When searching the number of active S-boxes, the biggest problem is that two active S-boxes can always cancel out with each other in the XOR operations to produce many invalid differential modes. It is mainly resulted from loss of the equality information by truncating the difference in words. We try to dig out the implicit relationships of the differences in different rounds to reduce the invalid differential modes. We also show how to convert the relationships to the equalities and inequalities trickily, which bring about new constraints into the basic MILP model. Then the optimizer Gurobi is exploited to solve the MILP model, and tighter lower bounds on the number of active S-boxes for SM4 are obtained. The method in this paper can also be applied to other unbalanced generalized Feistel structures to determine the lower bound on the number of active S-boxes. Furthermore, we consider the security of SM4 in the related-key setting by determining the minimum number of active S-boxes with MILP method. According to our knowledge, no results on the security of SM4 in the related-key setting have been published because of the intricate key schedule algorithm.

To prevent the difference in encryption procedure to be cancelled out by the key difference all the time, we present more useful relationships and constraints to extend the MILP model. Our results show that there exists no differential characteristic with probability larger than 2^{-128} in 23 rounds of SM4 cipher in the single-key setting. And 19 rounds are enough to prevent the valid related-key differential characteristic. The intricate key schedule algorithm indeed strengthens the security of SM4 against related-key differential attack.

This paper is organized as follows. We first give the notions which will be used throughout the paper and describe the SM4 cipher in Sect. 2. In Sect. 3, we construct the basic MILP model with Mouha et al.'s method. And then we present the relationships among the differences in different rounds, and show how to convert them to the constraints in Sect. 4. In Sect. 5, we introduce more helpful relationships and constraints to extend the MILP model to determine the minimal number of active S-boxes for SM4 in the related-key setting.

2 Preliminaries

2.1 Notation

In this subsection, we will give the notations and definitions which will be used throughout this paper.

- $\oplus$ denotes bitwise logical exclusive OR operation.
- $\lll i$ denotes left rotation by i bits.
- Z_2^{32} denotes the set of 32-bit words, and Z_2^8 denotes the set of 8-bit bytes. We equally treat the elements in Z_2^{32} and in $(Z_2^8)^4$ in this paper.
- We add a line above the difference variable to denote whether the difference is zero. For $X \in Z_2^{32}$, the new variable $\overline{X} \in \{0, 1\}$ and if $X \neq 0$, $\overline{X} = 1$, otherwise $\overline{X} = 0$. For $x \in Z_2^8$, if $x \neq 0$, $\overline{x} = 1$, otherwise $\overline{x} = 0$.
- $wt(X), X \in Z_2^{32}$ denotes the number of nonzero bytes of X.
- The branch number of a linear transformation $L : Z_2^{32} \to Z_2^{32}$ is defined by

$$\mathcal{B}(L) = \min_{X \neq 0,\, X \in Z_2^{32}} (wt(X) + wt(L(X)))$$

- We use d to stand for all the relevant variables d_j for convenience.

2.2 Description of SM4

SM4 is a block cipher with a 128-bit block size and a 128-bit key size. It consists of 32 rounds, each of which modifies one of the four 32-bit words that make up the block by XORing it with a keyed function of the other three words, as showed in Fig. 1.

Let $(S_i, S_{i+1}, S_{i+2}, S_{i+3}) \in (Z_2^{32})^4$ be the input of round $i (i = 0, 1, \cdots, 31)$ and $RK_i \in Z_2^{32}$ denotes the corresponding subkey in round i. Note that round 0 is referred to the first round. Then the encryption of SM4 is as follows,

$$S_{i+4} = S_i \oplus T(S_{i+1} \oplus S_{i+2} \oplus S_{i+3} \oplus RK_i),$$

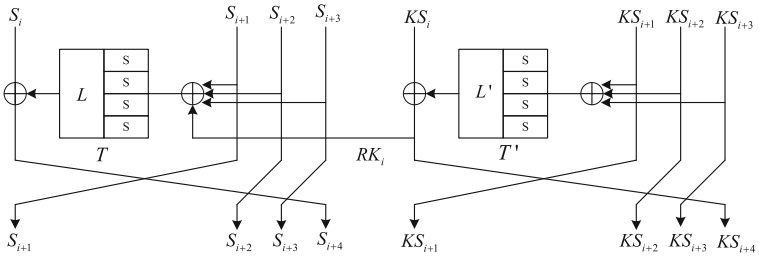

Fig. 1. Encryption and key schedule of round i of SM4 cipher.

for $i = 0, 1, \cdots, 31$. Then the ciphertext is generated by applying a reverse transformation R,

$$(C_1, C_2, C_3, C_4) = R(S_{32}, S_{33}, S_{34}, S_{35}) = (S_{35}, S_{34}, S_{33}, S_{32})$$

The transformation R aims at making the decryption procedure of SM4 be identical to the encryption procedure with the subkey used in reverse order. The round function T is composed of non-linear substitution layer and linear transformation L. The substitution layer is made up of four 8×8 bijective S-boxes in parallel and L is defined by,

$$L(S) = S \oplus (S \lll 2) \oplus (S \lll 10) \oplus (S \lll 18) \oplus (S \lll 24),$$

where $S \in Z_2^{32}$.

The key schedule algorithm is quite similar to the encryption procedure as showed in Fig. 1. The subkey in round $i (i = 0, 1, \cdots, 31)$ is got by,

$$RK_i = KS_{i+4} = KS_i \oplus T'(KS_{i+1} \oplus KS_{i+2} \oplus KS_{i+3} \oplus CK_i),$$

where $\{CK_i | i = 0, 1, \cdots, 31\}$ are some constants and (KS_0, KS_1, KS_2, KS_3) can be get from the main key. The round function T' is almost the same as T with a different linear transformation defined as,

$$L'(S) = S \oplus (S \lll 13) \oplus (S \lll 23),$$

where $S \in Z_2^{32}$. The value of constants and more details can be found in [5].

Note that it can be easily verified by a computer experiment that the branch number of L and L' are 5 and 4, respectively.

3 Basic MILP Model

In this section, we construct the basic MILP model using the constraints resulted from the operations of SM4 round function.

For better understanding of the (in)equalities in the following, we clarify the definitions of different variables. For round i $(i = 0, \cdots, 31)$, we

use $(X_i, X_{i+1}, X_{i+2}, X_{i+3}) \in (Z_2^{32})^4$ to denote the input difference, and X_i consists of four byte differences, i.e. $X_i = (x_{4i+1}, x_{4i+2}, x_{4i+3}, x_{4i+4}), x_{4i+k} \in Z_2^8, k = 1, 2, 3, 4$. We introduce some intermediate variables, $F_i = (f_{4i+1}, f_{4i+2}, f_{4i+3}, f_{4i+4})$ and $Y_i = (y_{4i+1}, y_{4i+2}, y_{4i+3}, y_{4i+4})$, which are computed by $F_i = X_{i+1} \oplus X_{i+3}$ and $Y_i = X_{i+2} \oplus X_{i+3}$. From byte perspective, $f_{4i+k} = x_{4i+4+k} \oplus x_{4i+12+k}$ and $y_{4i+k} = x_{4i+8+k} \oplus x_{4i+12+k}$. Particularly, we denote that $Y_{-1} = X_1 \oplus X_2$. The input and output difference of T function are denoted respectively by $In_i = (z_{4i+1}, z_{4i+2}, z_{4i+3}, z_{4i+4})$ and $Out_i = (w_{4i+1}, w_{4i+2}, w_{4i+3}, w_{4i+4})$. Note that all the variables f, y, z, w, sz are in Z_2^8, and we use $\overline{f}, \overline{y}, \overline{z}, \overline{w}, \overline{sz} \in \{0, 1\}$ to denote whether the corresponding difference is zero or not. In the rest of the paper, $i \in [0, 31]$ denotes the round number, $k \in [1, 4]$ denotes the byte number in one word.

Constraints Imposed by Linear Transformation. Because the branch number of L is 5, there are at least 5 active bytes in the input and output differences of L. Furthermore, the output difference of S-box is active only and if only the input difference is also active. Thus, we have,

$$\begin{cases} \sum_{k=1}^{4} \overline{z}_{4i+k} + \sum_{k=1}^{4} \overline{w}_{4i+k} \geq 5d_i \\ \overline{z}_{4i+k} \leq d_i, k = 1, 2, 3, 4 \\ \overline{w}_{4i+k} \leq d_i, k = 1, 2, 3, 4 \end{cases}$$

where d is the dummy variable taking values in $\{0, 1\}$.

Constraints Imposed by XOR Operations. For $a \oplus b = c, a, b, c \in Z_2^8$, the constraints are introduced by,

$$\begin{cases} \overline{a} + \overline{b} + \overline{c} \geq 2r \\ \overline{a} \leq r, \overline{b} \leq r, \overline{c} \leq r \end{cases}$$

where r is the dummy variable taking values in $\{0, 1\}$.

To reduce the number of invalid differential mode, we tackle the XORing of three words carefully from different perspective, distinguished by which two words are XORed firstly, i.e.

$$In_i = Y_{i-1} \oplus X_{i+3}, \quad In_i = X_{i+1} \oplus Y_i, \quad In_i = F_i \oplus X_{i+2}$$

Thus, we can obtain three sets of constraints.

Finally, we set up the objective function for r rounds of SM4 to the sum of all variables representing the input of the S-boxes, as follows,

$$Minimize : \sum_{i=0}^{r-1} \sum_{k=1}^{4} \overline{z}_{4i+k}.$$

The basic model is far from achieving a tighter lower bound on the number of active S-boxes. In the following, we will extend the model by adding more constraints.

4 Relationships Among Different Rounds

In this section, we mainly try to explore the relationships among different rounds to cut off the invalid differential modes, and thus obtain a tighter lower bound on the number of active S-boxes.

We notice that although the equality information among the difference variables gets lost because of truncation, we can still catch some equality information from the "zero" values. For example, we can know $x_9 = x_{13}$ from either $\overline{x}_9 = \overline{x}_{13} = 0$ or $\overline{y}_1 = 0$. However, we can not judge if $x_9 = x_{13}$ or not from the equation $\overline{x}_9 = \overline{x}_{13} = 1$. Because of this property, we try to explore the equality relationships and then introduce the corresponding constraints where equality information can only be caught from "zero" values.

We firstly give the relationships in three consecutive rounds. Note that the proof of all the theorems in the following is omitted and can be found in the full version of the paper [21].

Theorem 1. *For any three consecutive rounds (from round i to round $i+2$), it holds that,*

$$f_{4i+k} = f_{4i+8+k} \Leftrightarrow w_{4i+4+k} = 0$$

where $i \in [0, 31], k \in [1, 4]$.

According to Theorem 1, we should remove the differential modes with weight 1, i.e. $(\overline{f}_{4i+k}, \overline{f}_{4i+8+k}, \overline{w}_{4i+4+k}) \in \{(0,0,1),(1,0,0),(0,1,0)\}$. Therefore, we introduce the constraints,

$$\begin{cases} \overline{w}_{4i+4+k} \leq \overline{f}_{4i+k} + \overline{f}_{4i+8+k} \\ \overline{f}_{4i+k} \leq \overline{w}_{4i+4+k} + \overline{f}_{4i+8+k} \\ \overline{f}_{4i+8+k} \leq \overline{w}_{4i+4+k} + \overline{f}_{4i+k} \end{cases} \tag{1}$$

Theorem 2. *From round i to round $i+2$ of SM4 cipher, among the following three conditions: $x_{4i+4+k} = x_{4i+16+k}$, $w_{4i+4+k} = 0$, $y_{4i+8+k} = 0$, any two conditions can lead to the rest one.*

Moreover, it can be easily found that the condition $x_{4i+4+k} = x_{4i+16+k}$ is equivalent to each one of the following three conditions,

- $z_{4i+k} = z_{4i+4+k}$
- $f_{4i+k} = y_{4i+4+k}$
- $y_{4(i-1)+k} = f_{4i+4+k}$

According to Theorem 2, any two of $z_{4i+k} = z_{4i+4+k}$, $w_{4i+4+k} = 0$, $y_{4i+8+k} = 0$ can lead to the rest one, then we should remove the differential modes with weight 1, i.e. $(\overline{z}_{4i+k}, \overline{z}_{4i+4+k}, \overline{w}_{4i+4+k}, \overline{y}_{4i+8+k}) \in \{(1,0,0,0),(0,1,0,0),(0,0,1,0),(0,0,0,1)\}$. We can get the similar constraints to (1). This also goes for $(\overline{f}_{4i+k}, \overline{y}_{4i+4+k}, \overline{w}_{4i+4+k}, \overline{y}_{4i+8+k})$ and $(\overline{y}_{4i-4+k}, \overline{f}_{4i+4+k}, \overline{w}_{4i+4+k}, \overline{y}_{4i+8+k})$.

We then study the relationships among more rounds, some of which are given by Su et al. [15].

Theorem 3. *For any four consecutive rounds, there is at least one active S-box if the plaintext difference is nonzero.*

Theorem 3 can be easily converted to the constraints,

$$\sum_{j=i}^{i+3}\sum_{k=1}^{4}\overline{z}_{4j+k}\geq 1$$

where $i\in[0,31]$.

Theorem 4. [15] *For any 5 consecutive rounds (from round i to round $i+4$) of SM4 cipher, it holds that*

$$In_i\oplus In_{i+4}=Out_{i+1}\oplus Out_{i+2}\oplus Out_{i+3}.$$

From byte perspective, we have,

$$z_{4i+k}\oplus z_{4i+16+k}=w_{4i+4+k}\oplus w_{4i+8+k}\oplus w_{4i+12+k}$$

It brings about the constraints,

$$\begin{cases}\overline{z}_{4i+k}+\overline{z}_{4i+16+k}+\overline{w}_{4i+4+k}+\overline{w}_{4i+8+k}+\overline{w}_{4i+12+k}\geq 2u_{4i+k}\\ \overline{z}_{4i+k}\leq u_{4i+k},\overline{z}_{4i+16+k}\leq u_{4i+k}\\ \overline{w}_{4i+4+k}\leq u_{4i+k},\overline{w}_{4i+8+k}\leq u_{4i+k},\overline{w}_{4i+12+k}\leq u_{4i+k}\end{cases}$$

where u is the dummy variable taking values in $\{0,1\}$.

We denote the input difference of S-box of round i by $SIn_i=(sz_{4i+1},sz_{4i+2},sz_{4i+3},sz_{4i+4})\in Z_2^{32}$, then we have the following important corollaries,

Corollary 1. *For any 5 consecutive rounds (from round i to round $i+4$), if $(\overline{I}n_i,\overline{I}n_{i+1},\overline{I}n_{i+2},\overline{I}n_{i+3},\overline{I}n_{i+4})=(0,1,1,0,0)$, then it holds that $\overline{z}_{4i+4+k}=\overline{z}_{4i+8+k}$ for any $k\in[1,4]$.*

The constraints generated by corollary 1 are given by,

$$\begin{cases}\overline{z}_{4i+4+k}\leq\overline{z}_{4i+8+k}+\sum_{j=1}^{4}\overline{z}_{4i+j}+\sum_{j=13}^{20}\overline{z}_{4i+j},k=1,2,3,4\\ \overline{z}_{4i+8+k}\leq\overline{z}_{4i+4+k}+\sum_{j=1}^{4}\overline{z}_{4i+j}+\sum_{j=13}^{20}\overline{z}_{4i+j},k=1,2,3,4\end{cases}$$

Similar property and constraints can be obtained for the differential modes $(0,1,0,1,0)$ and $(0,0,1,1,0)$.

We can also find more useful corollaries according to Theorem 4 and present them in the full version of this paper [21].

Theorem 5. *For any 5 consecutive rounds (from round i to round $i+4$), if $w_{4i+8+k}=w_{4i+12+k}=0$, then it holds that $y_{4i+k}=y_{4i+16+k}$.*

The constraints resulted from Theorem 5 can be easily given by,

$$\begin{cases} \overline{y}_{4i+k} \leq \overline{y}_{4i+16+k} + \overline{w}_{4i+8+k} + \overline{w}_{4i+12+k}, k = 1, 2, 3, 4 \\ \overline{y}_{4i+16+k} \leq \overline{y}_{4i+k} + \overline{w}_{4i+8+k} + \overline{w}_{4i+12+k}, k = 1, 2, 3, 4 \end{cases}$$

Theorem 6. *For any 6 consecutive rounds (from round i to round $i+5$), if $z_{4i+k} = z_{4i+4+k} = w_{4i+4+k} = 0$, then it holds that*

$$z_{4i+16+k} \oplus w_{4i+16+k} = z_{4i+20+k}$$

Then we have the constraints,

$$\begin{cases} \overline{z}_{4i+16+k} + \overline{w}_{4i+16+k} + \overline{z}_{4i+20+k} + 2(\overline{z}_{4i+k} + \overline{z}_{4i+4+k} + \overline{w}_{4i+4+k}) \geq 2p_{4i+k} \\ \overline{z}_{4i+16+k} \leq p_{4i+k} + \overline{z}_{4i+k} + \overline{z}_{4i+4+k} + \overline{w}_{4i+4+k} \\ \overline{w}_{4i+16+k} \leq p_{4i+k} + \overline{z}_{4i+k} + \overline{z}_{4i+4+k} + \overline{w}_{4i+4+k} \\ \overline{z}_{4i+20+k} \leq p_{4i+k} + \overline{z}_{4i+k} + \overline{z}_{4i+4+k} + \overline{w}_{4i+4+k} \end{cases}$$

If $\overline{z}_{4i+k} + \overline{z}_{4i+4+k} + \overline{w}_{4i+4+k} \neq 0$, the constraints are trivial. Otherwise, they are just the constraints resulted from the XOR operation.

Theorem 7. *For any 6 consecutive rounds (from round i to round $i+5$), if $z_{4i+16+k} = z_{4i+20+k} = w_{4i+16+k} = 0$, then it holds that*

$$z_{4i+4+k} \oplus w_{4i+4+k} = z_{4i+k}$$

Theorem 8. *For any 6 consecutive rounds (from round i to round $i+5$), if $In_i \oplus In_{i+1} \oplus In_{i+4} \oplus In_{i+5} \neq 0$, then it holds that*

$$wt(SIn_{i+1} \oplus SIn_{i+4}) + wt(In_i \oplus In_{i+1} \oplus In_{i+4} \oplus In_{i+5}) \geq 5.$$

The constraints resulted from Theorems 7 and 8 are omitted here because of similarity.

Experimental Results. We extend the basic MILP model by introducing the constraints resulted from the relationships of difference among different rounds. Although we have explored so many relationships, the resulted differential mode may still be invalid mainly because of loss of equality information. Once we have found an invalid differential mode, We can add new constraints to cut it off, which can make the lower bound get from the MILP model tighter and tighter. For example, for 19 rounds of SM4 cipher, the model returns a lower bound with 17 actives S-boxes only when $(In_0, In_1, \cdots, In_{18}) \in U = \{(1,0,0,1,1,0,0,1,1,0,0,1,1,0,0,1,1,0,0), (0,0,1,1,0,0,1,1,0,0,1,1,0,0,1,1,0,0,1)\}$, but we find none valid actual differential mode after we have searched all cases. Then we can know there are at least 18 active S-boxes if the differential mode belongs to U, which have also been found by Su et al. [15]. After adding this constraint into our model, we get the lower bounds for different rounds of SM4 by solving the extended MILP model, which are concluded in Table 1.

Table 1. Lower bounds on the number of active S-boxes for different rounds of SM4 in the single-key setting and related-key setting.

Rounds	Single-key				Related-key	
	Time(seconds)	This paper	In [23]	In [19]	Time(seconds)	This paper
1	0	0	-	0	0	0
2	0	0	-	0	0	0
3	0	0	-	0	0	0
4	0	1	-	1	0	1
5	0	2	-	2	0	2
6	0	2	-	2	1	4
7	3	5	5	5	10	6
8	6	6	6	6	89	8
9	16	7	7	7	237	9
10	23	8	8	8	317	10
11	24	9	9	8	757	11
12	22	10	10	10	1345	13
13	69	10	11	10	5883	14
14	75	10	11	10	27420	14
15	410	13	12	12	44492	16
16	395	14	13	13	60017	18
17	696	15	14	15	1.5 days	19
18	1381	16	15	15	12 days	20
19	8156	18	16	16	<30 days	22
20	12771	18	16	18	–	-
21	18038	19	17	18	–	-
22	24691	20	18	-	–	-
23	36470	22	19	-	–	-
24	82857	23	20	-	–	-
25	102451	23	21	-	–	-
26	117849	24	22	-	–	-

From Table 1, we can find that our results always give a tighter lower bound on the number of active S-boxes for different rounds of SM4 cipher. Moreover, we can try to search the actual differential characteristic corresponding to each lower bound, and we find that the lower bounds given in Table 1 are almost tight especially when $r < 20$. Particularly, for 14 rounds of SM4, we can find the actual differential characteristic with probability 2^{-68} following a differential mode with weight $(0, 0, 1, 1, 0, 0, 3, 3, 0, 0, 1, 1, 0, 0)$. It has 10 active S-boxes, which indicates that some errors exist in the results of Zhang et al. Since the

maximal probability of the S-box is 2^{-6}, 23 rounds of SM4 with at least 22 active S-boxes are enough to prevent the differential characteristic with probability larger than 2^{-128}. Even if we add 4 more rounds of guessing the subkeys, 27 rounds of SM4 are believed to be secure against the differential attack based on certain differential characteristic, with 5 rounds as enough security margin.

5 Security Against Related-Key Differential Analysis

In this section, we study the security of SM4 against related-key differential analysis. General speaking, the key schedule algorithm of a block cipher is always much simpler than encryption procedure, which always results in a weaker security in the related-key setting, such as AES [4] and Present [6]. However, SM4 adopts a key schedule algorithm similar to encryption procedure, which makes it difficult to analyze the security against related-key differential analysis. Thanks to the automatic MILP method, we will show the lower bounds on the number of active S-boxes. But one biggest problem is that the differences in encryption procedure can be cancelled out by the subkey differences all the time because of the strong nonlinearity of the key schedule algorithm. We will present some helpful relationships to cut off the invalid differential modes.

We first describe the notations which will be used to construct the MILP model. For the encryption procedure, We introduce one more variable $TIn_i = (tz_{4i+1}, tz_{4i+2}, tz_{4i+3}, tz_{4i+4}), i \in [0, 31]$, which is computed by $TIn_i = X_{i+1} \oplus X_{i+2} \oplus X_{i+3}$. And we change the definition of In_i as $In_i = TIn_i \oplus KX_{i+4}$, where KX_{i+4} denotes the subkey of round i. Definitions of the other variables (x, f, y, z, w, Out) stay unchanged. For the key schedule algorithm which is almost the same as the encryption procedure, just with a different linear permutation L' whose branch number is 4, we add a character "k" as the prefix of all the variables in Sect. 3 to get the new variables, such as $kx, kf, ky, kz, kw, kIn, kOut$.

The basic model is quite similar to the basic model in Sect. 3 with a few differences. The objective function should be changed to,

$$Minimize : \sum_{i=0}^{r-1} \sum_{k=1}^{4} (\overline{z}_{4i+k} + \overline{kz}_{4i+k})$$

The constraints resulted from L' in key schedule are given by,

$$\begin{cases} \sum_{k=1}^{4} \overline{kz}_{4i+k} + \sum_{k=1}^{4} \overline{kw}_{4i+k} \geq 4kd_i \\ \overline{kz}_{4i+1} + \overline{kz}_{4i+2} + \overline{kz}_{4i+3} + \overline{kz}_{4i+4} \geq kd_i \\ \overline{kw}_{4i+1} + \overline{kw}_{4i+2} + \overline{kw}_{4i+3} + \overline{kw}_{4i+4} \geq kd_i \\ \overline{kz}_{4i+k} \leq kd_i, k = 1, 2, 3, 4 \\ \overline{kw}_{4i+k} \leq kd_i, k = 1, 2, 3, 4 \end{cases}$$

where kd is the dummy variable taking values in $\{0, 1\}$. Here, two more inequalities are introduced to prevent the occurrence of nonzero input but zero output,

which is different from the constraints of L. Other constraints can be obtained by imitating the ones in Sect. 3, which we omit here.

The relationships among different rounds and corresponding constraints can also be obtained similarly. We just present some relationships here as examples.

For the key schedule algorithm, we have,

Theorem 9. *For any 5 consecutive rounds (from round i to round $i+4$), it holds that*

$$kIn_i \oplus kIn_{i+4} = kOut_{i+1} \oplus kOut_{i+2} \oplus kOut_{i+3}.$$

Theorem 10. *For any 5 consecutive rounds (from round i to round $i+4$), if $kIn_i \oplus kIn_{i+4} \neq 0$, then there are at least four active S-box in the 5 consecutive rounds.*

For the encryption procedure, we have,

Theorem 11. *For any 5 consecutive rounds (from round i to round $i+4$) of SM4 cipher, it holds that*

$$TIn_i \oplus TIn_{i+4} = Out_{i+1} \oplus Out_{i+2} \oplus Out_{i+3}.$$

Theorem 12. *For any 5 consecutive rounds (from round i to round $i+4$) of SM4 cipher, if $TIn_i \oplus TIn_{i+4} \neq 0$, then we have,*

$$wt(TIn_i) + wt(TIn_{i+4}) + wt(In_i) + wt(In_{i+1}) + wt(In_{i+2}) \geq 5$$

Other relationships can be obtained similarly. We just note that the branch number of L' is 4 for key schedule algorithm, the input difference of encryption procedure can be zero due to the effects of the key differences, and the variable tz plays an important role in describing the relationships.

In the following, we will present some significant relationships to link the key schedule and encryption process.

Theorem 13. *For any 3 consecutive rounds (from round i to round $i+2$) of SM4 cipher in related-key setting, among the three conditions: $z_{4i+k} = z_{4i+4+k}$, $w_{4i+k} = 0$, $ky_{4i+8+k} = y_{4i+8+k}$, any two conditions can bring about the remaining one.*

Because we can only catch the equality information from the "zero" values, Theorem 13 can only cut off the differential modes $(\overline{z}_{4i+k}, \overline{z}_{4i+4+k}, \overline{w}_{4i+k}, \overline{ky}_{4i+8+k}, \overline{y}_{4i+8+k})$ with wight 1, such as $(1,0,0,0,0)$. Therefore, we can get the similar constraints to (1).

Furthermore, when $z_{4i+k} = z_{4i+4+k}$, $w_{4i+k} = 0$, we know $ky_{4i+8+k} = y_{4i+8+k}$ according to Theorem 13. This equality information can be propagated forward as illustrated in Fig. 2. Thus, if $kz_{4i+12+k} = 0$, we can deduce the checksum of the red variables in Fig. 2 is zero because it is equal to the checksum of green variables, then we can know $z_{4i+8+k} = x_{4i+12+k}$. Therefore, we can also cut off the differential modes $(\overline{z}_{4i+k}, \overline{z}_{4i+4+k}, \overline{w}_{4i+k}, \overline{kz}_{4i+12+k}, \overline{z}_{4i+8+k}, \overline{x}_{4i+12+k})$ with weight 1.

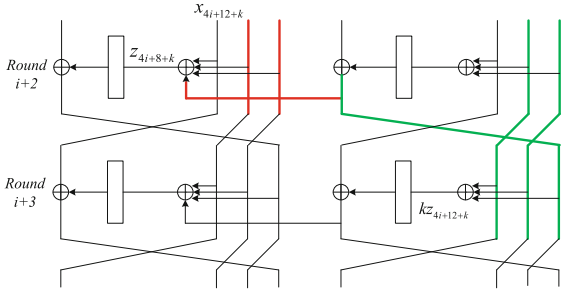

Fig. 2. Propagate the equality information forward. The checksum of red variables is equal to the checksum of green variables, s.t. $kz_{4i+12+k} = 0 \Leftrightarrow z_{4i+8+k} = x_{4i+12+k}$. (Color figure online)

Using the similar idea, we can find some other relationships to cut off the differential modes with weight 1. We present the relevant variables as follows, and the proof is omitted.

- $(z_{4i+k}, z_{4i+4+k}, w_{4i+4+k}, z_{4i+8+k}, x_{4i+12+k}, kz_{4i+12+k})$
- $(z_{4i+k}, z_{4i+8+k}, w_{4i+4+k}, kw_{4i+4+k}, kw_{4i+8+k}, kz_{4i+k}, z_{4i+12+k}, x_{4i+20+k})$
- $(z_{4i+k}, z_{4i+8+k}, z_{4i+12+k}, w_{4i+8+k}, kw_{4i+16+k}, x_{4i+4+k}, kz_{4i+20+k})$
- $(w_{4i+4+k}, w_{4i+8+k}, w_{4i+12+k}, z_{4i+k}, z_{4i+16+k}, kw_{4i+16+k})$

The corresponding constraints can be obtained easily. Note that these relationships and constraints combine the encryption procedure and the key schedule algorithm, which are quite helpful to cut off the invalid differential modes.

Experimental results. We solve the MILP model using Gurobi, and get the lower bounds on the number of active S-boxes for SM4 in the related-key setting which is concluded in Table 1. We find that removing the constraints resulted from the relationships in encryption procedure has few effects on the lower bounds, which can accelerate the search.

Although the lower bounds given in Table 1 are not tight, the results do make some sense. We can know the security of SM4 against differential attack in related-key setting indeed becomes stronger and 19 rounds of SM4 with at least 22 active S-boxes is enough to prevent the valid differential characteristic with probability larger than 2^{-128}.

6 Conclusion

Our works provide a new insight on giving a tighter bound on the number of active S-boxes for SM4 cipher and also the unbalanced generalized Feistel structure using MILP method. By exploring the inner detail relationships, the word-oriented MILP method may play a more important role in cryptography. For example, we can try to search the actual characteristics if the lower bound given by MILP model is tight enough. Furthermore, despite of the high nonlinearity

of key schedule algorithm, our works also show that it is feasible to evaluate the security of SM4 cipher against related-key differential analysis. However, a lot of invalid differential modes still exist even we have exploited so many relationships, especially when the number of rounds is large. How to get a tighter bound on the number of active S-boxes and search the actual (related-key) differential characteristic for more consecutive rounds of SM4 is still an interesting problem.

Acknowledgments. We would like to thank anonymous referees for their helpful comments and suggestions. The research presented in this paper is supported by the National Basic Research Program of China (No. 2013CB338002) and National Natural Science Foundation of China (Nos. 61272476, 61672509 and 61232009).

References

1. Gurobi: Gurobi optimizer reference mannual. http://www.gurobi.com
2. Albrecht, M., Cid, C.: Cold boot key recovery by solving polynomial systems with noise. In: Lopez, J., Tsudik, G. (eds.) ACNS 2011. LNCS, vol. 6715, pp. 57–72. Springer, Heidelberg (2011). doi:10.1007/978-3-642-21554-4_4
3. Aoki, K., Ichikawa, T., Kanda, M., Matsui, M., Moriai, S., Nakajima, J., Tokita, T.: *Camellia*: a 128-bit block cipher suitable for multiple platforms design and analysis. In: Stinson, D.R., Tavares, S. (eds.) Selected Areas in Cryptography, SAC 2000. Lecture Notes in Computer Science, LNCS, vol. 2012, pp. 39–56. Springer, Heidelberg (2000)
4. Biryukov, A., Nikolić, I.: Automatic search for related-key differential characteristics in byte-oriented block ciphers: application to AES, Camellia, Khazad and others. In: Gilbert, H. (ed.) EUROCRYPT 2010. LNCS, vol. 6110, pp. 322–344. Springer, Heidelberg (2010). doi:10.1007/978-3-642-13190-5_17
5. Diffie, W., Ledin, G.: SMS4 encryption algorithm for wireless networks. IACR Cryptology ePrint Archive 2008:329 (2008)
6. Emami, S., Ling, S., Nikolić, I., Pieprzyk, J., Wang, H.: The resistance of PRESENT-80 against related-key differential attacks. Cryptogr. Commun. **6**(3), 171–187 (2014)
7. Etrog, J., Robshaw, M.J.B.: The cryptanalysis of reduced-round SMS4. In: Avanzi, R.M., Keliher, L., Sica, F. (eds.) SAC 2008. LNCS, vol. 5381, pp. 51–65. Springer, Heidelberg (2009). doi:10.1007/978-3-642-04159-4_4
8. Halderman, J.A., Schoen, S.D., Heninger, N., Clarkson, W., Paul, W., Calandrino, J.A., Feldman, A.J., Appelbaum, J., Felten, E.W.: Lest we remember: cold-boot attacks on encryption keys. Commun. ACM **52**(5), 91–98 (2009)
9. Kanda, M.: Practical security evaluation against differential and linear cryptanalyses for Feistel ciphers with SPN round function. In: Stinson, D.R., Tavares, S. (eds.) SAC 2000. LNCS, vol. 2012, pp. 324–338. Springer, Heidelberg (2001). doi:10.1007/3-540-44983-3_24
10. Liu, F., Ji, W., Hu, L., Ding, J., Lv, S., Pyshkin, A., Weinmann, R.-P.: Analysis of the SMS4 block cipher. In: Pieprzyk, J., Ghodosi, H., Dawson, E. (eds.) ACISP 2007. LNCS, vol. 4586, pp. 158–170. Springer, Heidelberg (2007). doi:10.1007/978-3-540-73458-1_13
11. Lu, J.: Attacking reduced-round versions of the SMS4 block cipher in the Chinese WAPI standard. In: Qing, S., Imai, H., Wang, G. (eds.) ICICS 2007. LNCS, vol. 4861, pp. 306–318. Springer, Heidelberg (2007). doi:10.1007/978-3-540-77048-0_24

12. Mouha, N., Wang, Q., Gu, D., Preneel, B.: Differential and linear cryptanalysis using mixed-integer linear programming. In: Wu, C.-K., Yung, M., Lin, D. (eds.) Inscrypt 2011. LNCS, vol. 7537, pp. 57–76. Springer, Heidelberg (2012). doi:10.1007/978-3-642-34704-7_5
13. Shibutani, K.: On the diffusion of generalized Feistel structures regarding differential and linear cryptanalysis. In: Biryukov, A., Gong, G., Stinson, D.R. (eds.) SAC 2010. LNCS, vol. 6544, pp. 211–228. Springer, Heidelberg (2011). doi:10.1007/978-3-642-19574-7_15
14. IBM software group. CPLEX. http://www-01.ibm.com
15. Su, B.-Z., Wu, W.-L., Zhang, W.-T.: Security of the SMS4 block cipher against differential cryptanalysis. J. Comput. Sci. Technol. **26**(1), 130–138 (2011)
16. Sun, S., Hu, L., Song, L., Xie, Y., Wang, P.: Automatic security evaluation of block ciphers with S-bP structures against related-key differential attacks. In: Lin, D., Xu, S., Yung, M. (eds.) Inscrypt 2013. LNCS, vol. 8567, pp. 39–51. Springer, Heidelberg (2014). doi:10.1007/978-3-319-12087-4_3
17. Sun, S., Hu, L., Wang, P., Qiao, K., Ma, X., Song, L.: Automatic security evaluation and (related-key) differential characteristic search: application to SIMON, PRESENT, LBlock, DES(L) and other bit-oriented block ciphers. In: Sarkar, P., Iwata, T. (eds.) ASIACRYPT 2014. LNCS, vol. 8873, pp. 158–178. Springer, Heidelberg (2014). doi:10.1007/978-3-662-45611-8_9
18. Toz, D., Dunkelman, O.: Analysis of two attacks on reduced-round versions of the SMS4. In: Chen, L., Ryan, M.D., Wang, G. (eds.) ICICS 2008. LNCS, vol. 5308, pp. 141–156. Springer, Heidelberg (2008). doi:10.1007/978-3-540-88625-9_10
19. Wu, S., Wang, M.: Security evaluation against differential cryptanalysis for block cipher structures. Technical report, IACR Cryptology ePrint Archive, Report 2011/551 (2011)
20. Wenling, W., Zhang, W., Lin, D.: Security on generalized Feistel scheme with SP round function. IJ Netw.Secur. **3**(3), 215–224 (2006)
21. Zhang, L., Zhang, W., Wu, W.: Cryptanalysis of reduced-round SMS4 block cipher. In: Mu, Y., Susilo, W., Seberry, J. (eds.) ACISP 2008. LNCS, vol. 5107, pp. 216–229. Springer, Heidelberg (2008). doi:10.1007/978-3-540-70500-0_16
22. Zhang, L., Zhang, W., Wu, W.: Cryptanalysis of reduced-round SMS4 block cipher. In: Mu, Y., Susilo, W., Seberry, J. (eds.) ACISP 2008. LNCS, vol. 5107, pp. 216–229. Springer, Heidelberg (2008). doi:10.1007/978-3-540-70500-0_16
23. Zhang, M., Liu, J., Wang, X.: The upper bounds on differntial characteristics in block cipher SMS4. Technical report, IACR Cryptology ePrint Archive, Report 2010/155 (2010)