Su BZ, Wu WL, Zhang WT. Security of the SMS4 block cipher against differential cryptanalysis. JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY 26(1): 130–138 Jan. 2011. DOI 10.1007/s11390-011-1116-9

Security of the SMS4 Block Cipher Against Differential Cryptanalysis

Bo-Zhan Su[1,2] (苏波展), Wen-Ling Wu[1] (吴文玲), *Senior Member*, *CCF*, and Wen-Tao Zhang[1] (张文涛)

[1] *State Key Laboratory of Information Security, Institute of Software, Chinese Academy of Sciences, Beijing 100190, China*

[2] *State Key Laboratory of Information Security, Graduate University of Chinese Academy of Sciences, Beijing 100049, China*

E-mail: {subozhan, wwl, zhangwt}@is.iscas.ac.cn

Received November 22, 2009; revised November 1, 2010.

Abstract SMS4 is a 128-bit block cipher used in the WAPI standard for wireless networks in China. In this paper, we analyze the security of the SMS4 block cipher against differential cryptanalysis. Firstly, we prove three theorems and one corollary that reflect relationships of 5- and 6-round SMS4. Next, by these relationships, we clarify the minimum number of active S-boxes in 6-, 7- and 12-round SMS4 respectively. Finally, based on the above results, we present a family of about 2^{14} differential characteristics for 19-round SMS4, which leads to an attack on 23-round SMS4 with 2^{118} chosen plaintexts and $2^{126.7}$ encryptions.

Keywords block cipher, SMS4, differential cryptanalysis

1 Introduction

SMS4 is the underlying block cipher used in the WAPI (WLAN Authentication and Privacy Infrastructure) standard for securing wireless LANs in China. SMS4 has a 128-bit block size, a 128-bit user key, and a total of 32 rounds. It employs a kind of 4-branch generalized Feistel network in which only 32 bits are modified in each round. The main part of its round function is a simple SP network consisting of a subkey XOR-addition operation, four 8-bit to 8-bit S-box parallel lookups, and a linear transformation L.

SMS4 was made public[1] in January 2006 by the Chinese government ([2] gives an English translation). Since then, SMS4 has attracted much attention due to its simplicity and Chinese standard prominence. In [3], Liu *et al.* investigated the origin of the S-box employed in the cipher and presented an integral attack on 13-round SMS4. In [4], Ji *et al.* analyzed SMS4 from the viewpoint of algebraic attacks, and estimated the complexity of solving the equation system. In [5], Lu presented a rectangle attack on 14-round SMS4 and an impossible differential attack on 16-round SMS4. In [6], Toz *et al.* made a detailed analysis of the attacks given in [5] and further improved these attacks. In [7], Zhang *et al.* presented a differential attack on 21-round SMS4 and a rectangle attack on 16-round SMS4. In [8], Etrog *et al.* presented a linear attack on 22-round SMS4. In [9], Kim *et al.* presented a linear attack and a differential attack on 22-round SMS4, as well as a boomerang attack and a rectangle attack on 18-round SMS4. In [10], Zhang *et al.* gave three observations on the design of the linear transformation of SMS4, then they presented a differential attack on 22-round SMS4, which was an improvement of the previous work due to Zhang *et al.*[7] and Kim *et al.*[9]. Among the previous cryptanalytic work on SMS4, the best two distinguishers are as follows: one is an 18-round differential characteristic with a probability of 2^{-114}; the other is an 18-round linear approximation with a bias of $2^{-56.2}$. Both of these distinguishers can be used to attack up to 22-round SMS4[8,10].

In this paper, we make a more comprehensive study of the security of SMS4 against differential cryptanalysis. Here are our main results: 1) we give a clarification of the minimum number of differential active S-boxes for 6-, 7- and 12-round SMS4 respectively; 2) based on the above result, we obtain a family of $(2^7-1)^2$ effective differential characteristics for 19-round SMS4, one differential characteristics with probability 2^{-124}, 254 differential characteristic with probability 2^{-125} each, and all the others with probability 2^{-126} each. Then, we present a differential attack on 23-round SMS4 using the newly-found 19-round distinguishers, with 2^{118} chosen plaintexts and $2^{126.7}$ encryptions. The attack uses the early abort technique introduced in [11]. For comparison, the best previous attack on SMS4 can only reach 22 rounds.

The rest of this paper is organized as follows. Section

Short Paper

This work is supported by the National Natural Science Foundation of China under Grant Nos. 60873259 and 60903212, and the Knowledge Innovation Project of the Chinese Academy of Sciences.

2 provides a description of SMS4. Subsection 3.1 presents four relationships among the input and output differences of the T functions in 5- and 6-round SMS4, which is the very important basis for our follow-up work. Subsections 3.2∼3.4 present a clarification of the minimum number of active S-boxes for 6-, 7- and 12-round SMS4, respectively. Section 4 presents a family of $(2^7-1)^2$ effective 19-round differential characteristics. Based on this family of 19-round distinguishers, Section 5 gives a differential cryptanalysis of 23-round SMS4. Finally, Section 6 summarizes this paper.

2 Description of SMS4

SMS4 is a block cipher with a 128-bit block size and a 128-bit key size. Its overall structure is a kind of unbalanced Feistel network. The encryption procedure and the decryption procedure of SMS4 are identical except that the round subkeys are used in the reverse order.

2.1 Notation

The following notations are used throughout this paper.

- Z_2^{32} denotes the set of 32-bit words, and Z_2^8 denotes the set of 8-bit bytes;
- $Sbox(\cdot)$ is the 8×8 bijective S-box used in the round function F;
- $\lll i$: left rotation by i bits;
- $(X_i, X_{i+1}, X_{i+2}, X_{i+3}) \in (Z_2^{32})^4$ denotes the input of the i-th round, and RK_i is the corresponding 32-bit subkey in the i-th round $(0 \leqslant i \leqslant 31)$;
- $\Pr_T(\alpha \to \beta)$: the probability that the output difference of the function T is β when the input difference is α (T can be omitted when the context is clear);
- We call an S-box active if its input difference is nonzero; otherwise, we call it passive;
- Let $H_w(X)$ $(X \in (Z_2^8)^4)$ denote the number of non-zero bytes of X;
- Let ΔX denote the difference of X and X^*, in this paper, $\Delta X = X \oplus X^*$.

2.2 Encryption Procedure of SMS4

Let $(X_0, X_1, X_2, X_3) \in (Z_2^{32})^4$ and $(Y_0, Y_1, Y_2, Y_3) \in (Z_2^{32})^4$ denote the 128-bit plaintext P and the 128-bit ciphertext C respectively. Let $RK_i \in Z_2^{32}$, $(i = 0, 1, 2, \ldots, 31)$ denote the round subkeys. Note that the first round is referred to Round 0.

The encryption procedure of SMS4 is as follows:

$$\begin{aligned} X_{i+4} &= F(X_i, X_{i+1}, X_{i+2}, X_{i+3}, RK_i) \\ &= X_i \oplus T(X_{i+1} \oplus X_{i+2} \oplus X_{i+3} \oplus RK_i) \end{aligned}$$

for $i = 0, 1, \ldots, 31$. In the end, the 128-bit ciphertext is generated by applying the switch transformation R to the output of Round 31:

$$\begin{aligned} (Y_0, Y_1, Y_2, Y_3) &= R(X_{32}, X_{33}, X_{34}, X_{35}) \\ &= (X_{35}, X_{34}, X_{33}, X_{32}). \end{aligned}$$

Specifically, the i-th round of SMS4 can be expressed as follows:

$$(X_i, X_{i+1}, X_{i+2}, X_{i+3}) \longrightarrow (X_{i+1}, X_{i+2}, X_{i+3}, X_{i+4})$$

where $X_{i+4} = X_i \oplus T(X_{i+1} \oplus X_{i+2} \oplus X_{i+3} \oplus RK_i)$. The transformation T is composed of a non-linear transformation S and a linear diffusion function L, namely $T(\cdot) = L(S(\cdot))$. Fig.1 depicts one round of the encryption procedure of SMS4.

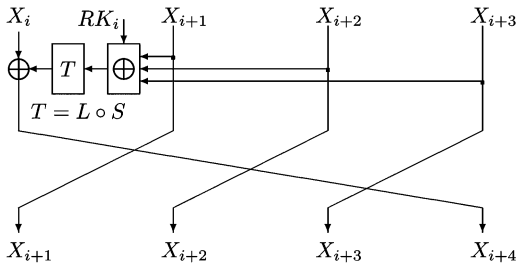


Fig.1. The i-th round of SMS4.

The non-linear transformation S applies the same 8×8 S-box four times in parallel to the 32-bit input. Let $\boldsymbol{U} = (u_0, u_1, u_2, u_3) \in (Z_2^8)^4$ denote the input of the transformation S, and $\boldsymbol{V} = (v_0, v_1, v_2, v_3) \in (Z_2^8)^4$ denote the corresponding output. Then the transformation S is defined as follows:

$$\begin{aligned} \boldsymbol{V} &= (v_0, v_1, v_2, v_3) = S(\boldsymbol{U}) \\ &= (Sbox(u_0), Sbox(u_1), Sbox(u_2), Sbox(u_3)). \end{aligned}$$

The diffusion transformation $\boldsymbol{L}$ is a simple linear function whose input is the output of the transformation S. Let $\boldsymbol{V} \in Z_2^{32}$ and $\boldsymbol{L}(\boldsymbol{V}) \in Z_2^{32}$ denote the input and output of L respectively. Then the linear function L is defined as follows.

$$\begin{aligned} \boldsymbol{L}(\boldsymbol{V}) = \boldsymbol{V} &\oplus (\boldsymbol{V} \lll 2) \oplus (\boldsymbol{V} \lll 10) \oplus \\ &(\boldsymbol{V} \lll 18) \oplus (\boldsymbol{V} \lll 24). \end{aligned}$$

We omit the key schedule algorithm of SMS4 as it is not involved in our analysis; interested readers can refer to [1] for details.

3 Minimum Number of Active S-Boxes in 6-, 7- and 12-Round SMS4

Our goal is to search for the best effective differential characteristics of SMS4 (i.e., differential characteristics

with probability as large as possible, yet not lower than 2^{-128}), so as to evaluate the security of SMS4 against differential cryptanalysis. It is known that the probability of differential characteristics can be evaluated by the minimum number of active S-boxes and the maximal differential probability of S-boxes. For the S-box of SMS4, there exist 127 possible output differences for any nonzero input difference, of which 1 output difference occurs with probability 2^{-6}, and each of the other 126 output differences occurs with probability 2^{-7}. Therefore, we believe that the best differential characteristics generally have the minimum number of active S-boxes for a fixed number of rounds in SMS4. In this section, we will clarify the minimum number of active S-boxes in some consecutive rounds.

The overall structure of SMS4 is a kind of generalized Feistel network, which produces some relationships among different rounds. In the following, we will present two relationships of 5-round SMS4, and two relationships of 6-round SMS4. We will see later that these four relationships are very helpful for clarifying the minimum number of differential active S-boxes in certain numbers of consecutive rounds. Next, using these relationships, we will study the minimum number of differential active S-boxes in 6-, 7- and 12-round SMS4 respectively.

Let "1" denote that the input difference of a T function is non-zero, and "0" denote that the difference is zero. Then, for an r-round differential characteristic, the sequence $(b_i, b_{i+1}, \ldots, b_{i+r-1})$ ($b_{i+j} = 0$ or 1 for $0 \leqslant j \leqslant (r-1)$) specifies the active or passive T functions in the r rounds. This sequence is called the differential pattern of the r-round differential characteristic.

3.1 Some Relationships of 5- and 6-Round SMS4

Let In_i and Out_i respectively denote the input and output of the T function in the i-th round, for $i = 0, 1, 2, \ldots, 31$. The round subkeys are XORed with the input of the T function, thus subkey addition has no influence on the number of active S-boxes.

Theorem 1. *For any* 5 *consecutive rounds (from i-th round to $(i+4)$-th round), the following relationship holds*:

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

Proof. According to Fig.1, we have

$$\begin{aligned} Out_{i+1} &= X_{i+1} \oplus X_{i+5}, \\ Out_{i+2} &= X_{i+2} \oplus X_{i+6}, \\ Out_{i+3} &= X_{i+3} \oplus X_{i+7}. \end{aligned}$$

On the other hand, we have

$$In_i \oplus In_{i+4} = X_{i+1} \oplus X_{i+2} \oplus X_{i+3} \oplus X_{i+5} \oplus X_{i+6} \oplus X_{i+7}.$$

Thus, we have

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

The following gives the definition of the branch number of L.

Definition 1 (Branch Number[12]). *The branch number of a function $F : Z_2^{32} \to Z_2^{32}$ is defined by*

$$\beta(F) = \min_{X \neq 0, X \in Z_2^{32}} (H_w(X) + H_w(F(X))).$$

It is easy to see that $\beta(F) \leqslant 5$. For the linear transformation L of SMS4, the branch number reaches the maximum 5, which can be easily verified by a computer experiment.

Theorem 2. *For any* 5 *consecutive rounds (from i-th round to $(i+4)$-th round), if $\Delta In_i \oplus \Delta In_{i+4} \neq 0$, then there are at least five active S-boxes in the* 5 *consecutive rounds.*

Proof. According to Theorem 1 and the linearity of L, we know that

$$\begin{aligned} \Delta In_i \oplus \Delta In_{i+4} &= \Delta Out_{i+1} \oplus \Delta Out_{i+2} \oplus \Delta Out_{i+3} \\ &= L(L^{-1}(\Delta Out_{i+1}) \oplus L^{-1}(\Delta Out_{i+2}) \oplus \\ &\quad L^{-1}(\Delta Out_{i+3})) \\ &= L(\Delta L^{-1}(Out_{i+1}) \oplus \Delta L^{-1}(Out_{i+2}) \oplus \\ &\quad \Delta L^{-1}(Out_{i+3})). \end{aligned} \tag{1}$$

Clearly $H_w(\Delta In_i) = H_w(\Delta L^{-1}(Out_i))$.

According to the formula $H_w(X) + H_w(Y) \geqslant H_w(X \oplus Y)$, we have

$$H_w(\Delta In_i) + H_w(\Delta In_{i+4}) \geqslant H_w(\Delta In_i \oplus \Delta In_{i+4}). \tag{2}$$

Similarly,

$$\begin{aligned} &H_w(\Delta In_{i+1}) + H_w(\Delta In_{i+2}) + H_w(\Delta In_{i+3}) \\ &= H_w(\Delta L^{-1}(Out_{i+1})) + H_w(\Delta L^{-1}(Out_{i+2})) + \\ &\quad H_w(\Delta L^{-1}(Out_{i+3})) \\ &\geqslant H_w(\Delta L^{-1}(Out_{i+1}) \oplus \Delta L^{-1}(Out_{i+2}) \oplus \\ &\quad \Delta L^{-1}(Out_{i+3})). \end{aligned} \tag{3}$$

The branch number of L is 5. If $\Delta In_i \oplus \Delta In_{i+4} \neq 0$, then we get the following result using the above 3 relationships:

$$\begin{aligned} &H_w(\Delta In_i) + H_w(\Delta In_{i+1}) + H_w(\Delta In_{i+2}) + \\ &\quad H_w(\Delta In_{i+3}) + H_w(\Delta In_{i+4}) \geqslant 5. \end{aligned}$$

$\square$

Theorem 3. *For any* 6 *consecutive rounds* (*from* i*-th round to* $(i+5)$*-th round*)*, if the* i*-th round and the* $(i+1)$*-th round are both differentially passive, then the following relationship holds*:

$$\Delta In_{i+4} \oplus \Delta Out_{i+4} = \Delta In_{i+5}.$$

Proof. The i-th round is differentially passive, we have

$$\Delta X_{i+1} \oplus \Delta X_{i+2} \oplus \Delta X_{i+3} = 0.$$

The $(i+1)$-th round is also differentially passive, we have

$$\begin{aligned} &\Delta X_{i+2} \oplus \Delta X_{i+3} \oplus \Delta X_{i+4} = 0, \\ &\Delta X_{i+1} = \Delta X_{i+5}. \end{aligned}$$

Then, we get $\Delta X_{i+1} = \Delta X_{i+4} = \Delta X_{i+5}$.

Since $\Delta In_{i+4} = \Delta X_{i+5} \oplus \Delta X_{i+6} \oplus \Delta X_{i+7}$, and also $\Delta In_{i+5} = \Delta X_{i+6} \oplus \Delta X_{i+7} \oplus \Delta X_{i+8}$, we can finally get

$$\begin{aligned} \Delta In_{i+4} \oplus \Delta In_{i+5} &= \Delta X_{i+5} \oplus \Delta X_{i+8} \\ &= \Delta X_{i+4} \oplus \Delta X_{i+8} \\ &= \Delta Out_{i+4}. \end{aligned}$$

□

Corollary 1. *For any* 6 *consecutive rounds* (*from* i*-th round to* $(i+5)$*-th round*)*, if the* $(i+4)$*-th round and the* $(i+5)$*-th round are both differentially passive, then the following relationship holds*:

$$\Delta In_{i+1} \oplus \Delta Out_{i+1} = \Delta In_i.$$

We omit the proof, since it is similar in nature to Theorem 3.

3.2 6-Round Differential Patterns

There are $2^6 = 64$ differential patterns for 6-round differential characteristics. Using Theorem 1, Theorem 3 and Corollary 1, we can know that 18 patterns are impossible. Details are as follows:

Applying Theorem 1 to the first 5-round segment $(0,0,0,0,1)$ of the first pattern $(0,0,0,0,1,1)$, we get that $\Delta In_4 = 0$ which leads to a contradiction. So the first pattern $(0,0,0,0,1,1)$ is an impossible pattern. Similarly, we can obtain that patterns $1 \sim 14$ are also impossible patterns which contradict with Theorem 1. Applying Theorem 3 to pattern 15 which is $(0,0,1,1,0,1)$, we can get that $\Delta In_4 \oplus \Delta Out_4 = \Delta In_5$. However, both ΔIn_4 and ΔOut_4 are zero, and ΔIn_5 does not equal zero, which is an obvious contradiction. Therefore, pattern 15 is impossible. Then applying Corollary 1 to pattern 16 which is $(1,0,1,1,0,0)$, and we obtain that $\Delta In_1 \oplus \Delta Out_1 = \Delta In_0$. However, both ΔIn_1 and ΔOut_1 equal zero, and ΔIn_0 does not equal zero, which is a contradiction. Therefore, pattern 16 is impossible.

Then, there remain $64 - 18 = 46$ possible patterns. Using Theorem 2, Theorem 3 and Corollary 1, we get the following results: there are at least 2 active S-boxes for one pattern, at least 5 active S-boxes for 12 patterns, and at least 6 active S-boxes for 33 patterns. Table 1 gives the details, and we give two patterns for illustration below.

The first example is the 6-round pattern $(0,0,0,1,1,0)$. Applying Theorem 3, we get that $\Delta In_4 \oplus \Delta Out_4 = \Delta In_5 = 0$ which means $\Delta In_4 = \Delta Out_4$. Thus, T_4 has at least 3 active S-boxes because the branch number of the linear transformation L is 5.

Table 1. The Number of Active S-Boxes in 6-Round Differential Patterns

The Number of Active S-Boxes	Patterns
At least 2 active S-boxes	$(0,0,1,1,0,0)$,
At least 5 active S-boxes	$(0,1,0,0,1,0), (1,0,1,0,1,0), (0,1,1,0,1,0), (1,0,0,1,1,0), (0,1,0,1,1,0),$ $(1,0,1,1,1,0), (0,1,1,1,1,0), (0,1,0,1,0,1), (0,1,1,0,0,1), (1,1,1,0,1,1),$ $(0,1,1,1,0,1), (1,1,0,1,1,1)$
At least 6 active S-boxes	$(0,1,1,1,0,0), (0,0,1,1,1,0), (1,1,0,1,1,0), (0,1,1,0,0,0), (1,1,1,0,0,0),$ $(0,1,0,1,0,0), (1,1,0,1,0,0), (1,1,1,1,0,0), (1,1,0,0,1,0), (0,0,1,0,1,0),$ $(1,1,1,0,1,0), (0,0,0,1,1,0), (1,1,1,1,1,0), (1,1,0,0,0,1), (1,0,1,0,0,1),$ $(1,1,1,0,0,1), (1,0,0,1,0,1), (1,1,0,1,0,1), (1,0,1,1,0,1), (1,1,1,1,0,1),$ $(1,0,0,0,1,1), (0,1,0,0,1,1), (1,1,0,0,1,1), (0,0,1,0,1,1), (1,0,1,0,1,1),$ $(0,1,1,0,1,1), (0,0,0,1,1,1), (1,0,0,1,1,1), (0,1,0,1,1,1), (0,0,1,1,1,1),$ $(1,0,1,1,1,1), (0,1,1,1,1,1), (1,1,1,1,1,1)$
Impossible patterns	$(0,0,0,0,1,1), (1,0,0,0,0,0), (0,1,0,0,0,0), (0,1,0,0,0,1), (0,0,1,0,0,0),$ $(0,0,0,1,0,1), (0,0,0,1,0,0), (1,0,0,0,0,1), (0,0,1,0,0,1), (0,0,0,0,1,0),$ $(1,0,0,0,1,0), (0,0,0,0,0,1), (1,0,0,1,0,0), (1,1,0,0,0,0), (1,0,1,0,0,0),$ $(0,0,1,1,0,1), (1,0,1,1,0,0), (0,0,0,0,0,0)$

On the other hand, we get that $\Delta Out_4 = \Delta Out_3$ by applying Theorem 1 to the last 5 consecutive rounds, hence T_3 also has at least 3 active S-boxes. Therefore, there are at least 6 active S-boxes in the pattern $(0, 0, 0, 1, 1, 0)$.

The second example is $(0, 0, 1, 0, 0, 1)$. We can deduce from Theorem 3 that $\Delta In_4 \oplus \Delta Out_4 = \Delta In_5$. However, both ΔIn_4 and ΔOut_4 are zero, and ΔIn_5 is non-zero. This is a contradiction. Therefore, the differential pattern $(0, 0, 1, 0, 0, 1)$ is impossible.

3.3 7-Round Differential Patterns

There are $2^7 = 128$ differential patterns for 7-round differential characteristics. Examining them one by one by using the three theorems and Corollary 1, we find that 47 patterns are impossible, the other 81 patterns are possible. Among the 81 possible patterns, there are at least 5 active S-boxes for 2 patterns (Table 2 lists the 2 patterns), and at least 6 active S-boxes for 79 patterns. For each of the 2 patterns which have at least 5 active S-boxes, we make experiments to search for optimal differential characteristics having the corresponding pattern. Our results show that there exist characteristics which have just 5 active S-boxes for each pattern. Here are the details: 1) for the first pattern, the number of active S-boxes in each of the 7 rounds is $3, 0, 0, 1, 1, 0, 0$ respectively, and the differential probability is 2^{-33}; 2) for the second pattern, the number of differential active S-boxes in each of the 7 rounds is $0, 0, 1, 1, 0, 0, 3$ respectively, and the differential probability is 2^{-33}.

Table 2. 7-Round Differential Patterns Which Can Have 5 Active S-Boxes

Two Patterns	Optimal Probability
$(1, 0, 0, 1, 1, 0, 0)$	2^{-33}
$(0, 0, 1, 1, 0, 0, 1)$	2^{-33}

3.4 12-Round Differential Patterns

There are $2^{12} = 4096$ differential patterns for 12-round differential characteristics. A 12-round differential characteristic can be regarded as a concatenation of two 6-round differential characteristics, hence we can easily know that there are at least 7 active S-boxes for the 12 rounds. After examining some possible 12-round differential patterns, we set our goal on the search for patterns which have at most 10 active S-boxes (actually, there are at least 10 active S-boxes in any effective 12-round differential characteristic, which will be shown below). Because of the symmetry of encryption and decryption of SMS4, we assume that there are at most 5 active S-boxes in the first 6 rounds (if we reverse the round order of a pattern, we can get another pattern which has at most 5 active S-boxes in the last 6 rounds). According to Table 1, there are 13 possible patterns for the first 6 rounds, there are 46 possible patterns for the last 6 rounds. Hence, we will only focus our attention on these $13 \times 46 = 598$ possible patterns.

As we know, there are 18 impossible patterns for 6-round differential characteristics. For a 12-round differential pattern, if it includes an impossible 6-round pattern as a subsegment, then this 12-round differential pattern is also impossible. Using this property, we can easily get that 233 patterns are impossible by a computer experiment.

There are $598 - 233 = 365$ possible patterns remaining. Examining them one by one by using the three theorems and Corollary 1, we find that all of the 365 patterns are possible. Here are the details: there are at least 10 active S-boxes for 2 patterns, at least 11 active S-boxes for 174 patterns, at least 12 active S-boxes for 105 patterns, at least 13 active S-boxes for 51 patterns, at least 14 active S-boxes for 16 patterns, at least 15 active S-boxes for 9 patterns, at least 16 active S-boxes for 6 patterns, and at least 17 active S-boxes for 2 patterns.

Take the 12-round differential pattern $(1, 0, 0, 1, 1, 0, 0, 1, 0, 1, 0, 0)$ for example. Considering the two 6-round segments: one is $(0, 0, 1, 0, 1, 0)$ (from 5th round to 10th round), the other is $(0, 1, 0, 1, 0, 0)$ (from 6th round to 11th round). Applying Theorem 3 to the first segment $(0, 0, 1, 0, 1, 0)$, we get that $\Delta In_9 = \Delta Out_9$. Applying Corollary 1 to the second segment $(0, 1, 0, 1, 0, 0)$, we can get that $\Delta In_7 \oplus \Delta Out_7 = \Delta In_6$. Because ΔIn_6 equals zero, we obtain that $\Delta In_7 = \Delta Out_7$. Thus, we can know that each of T_7 and T_9 has at least 3 active S-boxes. On the other hand, considering the two 5-round segments: one is $(1, 0, 0, 1, 0)$ (from 4th round to 8th round), the other is $(0, 0, 1, 1, 0)$ (from 1st round to 5th round), we can deduce from Theorem 1 that $\Delta Out_7 = \Delta In_4$ and $\Delta Out_4 = \Delta Out_3$. Thus, we also get that each of T_4 and T_3 has at least 3 active S-boxes. Therefore, there are at least $3 \times 4 + 1 = 13$ active S-boxes in this 12-round differential pattern.

By considering the reverse order of the 2 patterns which have at least 10 active S-boxes, we arrive at the conclusion that there are only three 12-round patterns which can possibly have 10 active S-boxes. Table 3 lists the three patterns. For each of the three patterns, we conducted experiments to search for optimal differential characteristics. And our results show that there exist some differential characteristics having just 10 active S-boxes for each pattern. Here are the details: 1) for the first pattern, the number of active S-boxes in each

of the 12 rounds is $0,0,1,1,0,0,3,3,0,0,1,1$ respectively, and the differential probability is 2^{-67}; 2) for the second pattern, the number of active S-boxes in each of the 12 rounds is $1,1,0,0,3,3,0,0,1,1,0,0$ respectively, and the differential probability is 2^{-67}; 3) for the third pattern, the number of active S-boxes in each of the 12 rounds is $0,1,1,0,0,3,3,0,0,1,1,0$ respectively, and the differential probability is 2^{-68}.

Table 3. 12-Round Differential Patterns Which Can Have 10 Active S-Boxes

Three Patterns	Optimal Probability
$(0,0,1,1,0,0,1,1,0,0,1,1)$	2^{-67}
$(1,1,0,0,1,1,0,0,1,1,0,0)$	2^{-67}
$(0,1,1,0,0,1,1,0,0,1,1,0)$	2^{-68}

4 19-Round Differential Characteristics

In this section, we will present $(2^7-1)^2 \approx 2^{14}$ 19-round differential characteristics of SMS4, which are the best effective differential characteristics of SMS4 we have found. Among them, one with probability 2^{-124}, $2\times(2^7-1)=254$ ones with probability 2^{-125}, and all the other $(2^7-1)^2-255$ ones with probability 2^{-126}. In Section 5, we will present an attack on 23-round SMS4, which is based on these newly-found 19-round differential characteristics.

We can regard a 19-round differential characteristic as a concatenation of one 12-round and one 7-round differential characteristic. There are three 12-round differential patterns, each of them has 10 active S-boxes (see Table 3). There are two 7-round differential patterns, each has 5 active S-boxes (see Table 2). By fixing the first 12 rounds as one of the 3 patterns and the second 7 rounds as one of the 2 patterns, we get 6 different patterns. Examining the 6 patterns one by one by using Theorem 3 and Corollary 1, we find that 3 patterns are impossible, and the other 3 patterns are possible, Table 4 lists the 3 possible patterns. By using the three Theorems and Corollary 1, we find that there are at least 18 active S-boxes for the first pattern, and at least 19 active S-boxes for the other two patterns.

Table 4. Some 19-Round Differential Patterns Which Have at Least 18/19 Active S-Boxes

No. Active S-Boxes	Patterns
At least 18 active S-boxes	(0,0,1,1,0,0,1,1,0,0,1,1,0,0,1,1,0,0,1)
At least 19 active S-boxes	(0,0,1,1,0,0,1,1,0,0,1,1,1,0,0,1,1,0,0) (0,1,1,0,0,1,1,0,0,1,1,0,0,0,1,1,0,0,1)

For the first pattern in Table 4, we make an experiment to search for optimal differential characteristics. We finally obtain $(2^7-1)^2 \approx 2^{14}$ 19-round differential characteristics, each one with 18 active S-boxes. Let a_1 = 0xf3f30033, a_2 = 0xf3000030, a_3 = 0x00f30003, a_4 = 0x00cf0033, a_5 = 0xf33c0000, a_6 = 0x3f0000cf, a_7 = 0xccf300fc. Define $DiffSet = \{x \in Z_2^{32} | \mathrm{Prob}_T(a_2 \to x) \neq 0\}$, there are $127^2 \approx 2^{14}$ elements in $DiffSet$. Define $\Omega = \{y \in Z_2^{32} | y = x \oplus a_3, x \in DiffSet\}$. Table 5 illustrates the family of about 2^{14} 19-round differential characteristics in detail, where $a_0 \in \Omega$, (a_0,a_1,a_1,a_2) is the input difference and (a_6,a_7,a_1,a_1) is the output difference. There exists one element which satisfies $\mathrm{Prob}_T(a_2 \to x) = 2^{-12}$. This element corresponds to a 19-round differential characteristic with probability 2^{-124}. There exist $2\times(2^7-1)=254$ elements which satisfy $\mathrm{Prob}_T(a_2 \to x) = 2^{-13}$. These elements correspond to some 19-round differential characteristics with probability 2^{-125}. All the other $(2^7-1)^2-255$ elements correspond to 19-round differential characteristics with probability 2^{-126}.

Table 5. A Family of 19-Round Differential Characteristics

$R(i)$	$\Delta X_i \Delta X_{i+1} \Delta X_{i+2} \Delta X_{i+3}$	Probability
0	(a_0,a_1,a_1,a_2)	–
1	(a_1,a_1,a_2,a_3)	$2^{-14}/2^{-13}/2^{-12}$
2	(a_1,a_2,a_3,a_1)	1
3	(a_2,a_3,a_1,a_1)	1
4	(a_3,a_1,a_1,a_4)	2^{-14}
5	(a_1,a_1,a_4,a_5)	2^{-14}
6	(a_1,a_4,a_5,a_1)	1
7	(a_4,a_5,a_1,a_1)	1
8	(a_5,a_1,a_1,a_5)	2^{-14}
9	(a_1,a_1,a_5,a_4)	2^{-14}
10	(a_1,a_5,a_4,a_1)	1
11	(a_5,a_4,a_1,a_1)	1
12	(a_4,a_1,a_1,a_3)	2^{-14}
13	(a_1,a_1,a_3,a_2)	2^{-14}
14	(a_1,a_3,a_2,a_1)	1
15	(a_3,a_2,a_1,a_1)	1
16	(a_2,a_1,a_1,a_6)	2^{-14}
17	(a_1,a_1,a_6,a_7)	2^{-14}
18	(a_1,a_6,a_7,a_1)	1
19	(a_6,a_7,a_1,a_1)	1

5 Differential Cryptanalysis of 23-Round SMS4

In this section, we will present a differential attack on 23-round SMS4 using the 19-round differential characteristics illustrated in Table 5.

We apply the 19-round differential characteristics to Rounds $0 \sim 18$. If the output difference of Round 18 is (a_6,a_7,a_1,a_1), then the input difference of the T function in Round 19 equals a_7. For the S-box of SMS4, there are 127 possible output differences for any nonzero input difference, thus the output difference of the T function in Round 19 has only about $2^{7\times 3}=2^{21}$

possible values. Let Λ denote the set of these 2^{21} possible values. Then, the output differences of Rounds $19 \sim 22$ must belong to (a_7, a_1, a_1, Λ), $(a_1, a_1, \Lambda, *)$, $(a_1, \Lambda, *, *)$, and $(\Lambda, *, *, *)$ respectively, where $*$ denotes an unknown word, as illustrated in Fig.2.

We need to choose plaintext pairs with difference (a_0, a_1, a_1, a_2), where $a_0 \in \Omega$. There are about 2^{14} possible values of a_0. Generate two pools S_1 and S_2 of 2^{32} plaintexts each:

$$S_1 = (x, const_1, const_2, const_3)$$
$$S_2 = (y, const_1 \oplus a_1, const_2 \oplus a_1, const_3 \oplus a_2)$$

where $const_1, const_2, const_3$ are 3 fixed words, x and y take all 2^{32} possible values respectively. We call the pair (S_1, S_2) a structure. One structure includes 2^{33} different plaintexts, and can produce about $2^{32} \times 2^{32} \times \frac{2^{14}}{2^{32}} = 2^{46}$ candidate plaintext pairs.

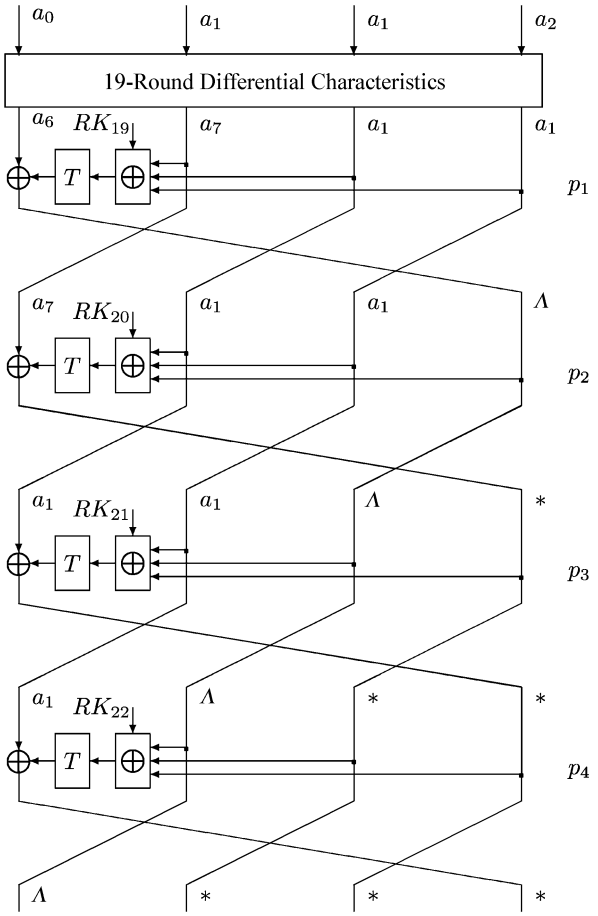


Fig.2. Differential cryptanalysis of 23-round SMS4.

5.1 Attack Procedure

Step 1. Select m structures, then we can get about $m \times 2^{46}$ plaintext pairs using $m \times 2^{33}$ plaintexts.

Step 2. For each plaintext pair (P, P^*), compute the difference of the corresponding ciphertext pair (C, C^*), and check if the first word of the ciphertext difference belongs to the set Λ. If this is not the case, discard the pair. After this test, about $m \times 2^{46} \times \frac{2^{21}}{2^{32}} = m \times 2^{35}$ pairs are expected remain.

Step 3) For every guess of the 0th byte of RK_{22}, i.e., $RK_{22,0}$, do as follows:

(a) For each remaining ciphertext pair (C, C^*), let $C = (Y_0, Y_1, Y_2, Y_3) \in (Z_2^{32})^4$, $C^* = (Y_0^*, Y_1^*, Y_2^*, Y_3^*) \in (Z_2^{32})^4$, partially decrypt to get the output difference of the 0th S-box in Round 22: $\gamma = Sbox((Y_0 \oplus Y_1 \oplus Y_2)_0 \oplus RK_{22,0}) \oplus Sbox((Y_0^* \oplus Y_1^* \oplus Y_2^*)_0 \oplus RK_{22,0})$, and compute $\delta = (L^{-1}(Y_3 \oplus Y_3^* \oplus a_1))_0$. If $RK_{22,0}$ is guessed correctly, then we must have $\gamma = \delta$ for right pairs. If not, discard the pair. After this test, about $m \times 2^{35} \times 2^{-8} = m \times 2^{27}$ ciphertext pairs are expected to remain. For each of the other 3 bytes of RK_{22}, continue making a guess, then compute γ and δ with the guessed value. If γ does not equal δ, discard the disqualified ciphertext pairs in each step. After these tests, about $m \times 2^{27} \times 2^{-3\times 8} = m \times 2^3$ ciphertext pairs are expected to remain.

(b) For each of the 4 bytes of RK_{21}, make a guess and do similarly as Step 3(a). After these tests, about $m \times 2^3 \times 2^{-4\times 8} = m \times 2^{-29}$ ciphertext pairs are expected to remain.

(c) For each of the 4 bytes of RK_{20}, make a guess and do similarly as Step 3(a). After these tests, about $m \times 2^{-29} \times 2^{-4\times 8} = m \times 2^{-61}$ ciphertext pairs are expected to remain.

(d) Guess $RK_{19,0}$, do similarly as in Step 3(a). A probability of about $\frac{1}{2}$ pairs are already discarded because of the selection of the set Λ. Hence, after this test, about $m \times 2^{-61} \times 2^{-7} = m \times 2^{-68}$ ciphertext pairs are expected to remain. For every guess of each of $RK_{19,1}$ and $RK_{19,3}$, make a guess and do similarly as Step 3(a). After these tests, about $m \times 2^{-68} \times 2^{-2\times 7} = m \times 2^{-82}$ ciphertext pairs are expected to remain. Because the input difference of the round function of Round 19 is a_7, and it consists of only three active S-boxes. Hence, we cannot get the information of $RK_{19,2}$.

Step 4. Output the 120-bit subkey guess RK_{22}, RK_{21}, RK_{20}, $RK_{19,0}$, $RK_{19,1}$ and $RK_{19,3}$ as the correct subkey, if it has maximal number of remaining pairs after Step 3(d).

5.2 Analysis of the Attack

For most of the 19-round differential characteristics in Table 5, the probability is 2^{-126}. For all the other 19-round differential characteristics in Table 5, the probability is larger than 2^{-126}. Thus, from the attacker's point of view, we can assume that the probability of each of the differential characteristics is 2^{-126} for simplicity.

It is expected that there remain about $m \cdot 2^{46} \cdot 2^{-126} = m \cdot 2^{-80}$ right pairs for the correct key. However, for the wrong subkey guesses, the expected number of remaining pairs after Step 3(d) is about $m \cdot 2^{-82}$. In the analysis that follows, we exploit the concept of "signal-to-noise ratio" introduced by Biham and Shamir in [13] to choose an appropriate value of m to make the differential attack succeed with high probability.

The signal-to-noise ratio is defined as the proportion of the probability of the right key being suggested by a right pair to the probability of a random key being suggested by a random pair with the initial difference. According to [13], the signal-to-noise ratio can be computed by the following formula:

$$S/N = \frac{2^k \times p}{\alpha \times \beta}$$

where k is the number of guessed key bits, p is the probability of the differential characteristic, α is the average number of keys suggested by a counted pair, and β is the ratio of the counted pairs to all pairs (both counted and discarded).

In the above attack, we have guessed 120 subkey bits, and we assume the probability of each of the differential characteristics is 2^{-126}.

For every test in Step 3(a), Step 3(b) and Step 3(c), there are 2^{32} possible key guesses and a counted pair needs to satisfy a 32-bit condition. For the test in Step 3(d), there are 2^{24} possible key guesses and a counted pair needs to satisfy a 21-bit condition. Hence, $\alpha = 2^3$. In Step 2, a 11-bit condition is used to discard the pairs, thus $\beta = 2^{-11}$. Therefore, the signal-to-noise ratio of the above attack is $2^{120} \times 2^{-126}/2^{-11} = 4$. We choose $m = 2^{85}$, the expectation of the remaining ciphertext pairs is about 8 for a wrong key guess, and the expectation of the remaining ciphertext pairs is about $\mu = pN = 32$ for the right key guess. According to the results in [14], the success probability can be estimated as follows:

$$P_S = \int_{-\frac{\sqrt{\mu S_N} - \Phi^{-1}(1-2^{-a})}{\sqrt{S_N+1}}}^{\infty} \phi(x)\mathrm{d}x \approx 0.9890$$

where $a = 120$ is the number of subkey bits recovered. The remaining key bits can be found by exhaustive search.

The attack requires $m \cdot 2^{32+1} = 2^{118}$ chosen plaintexts in total. The time complexity is dominated by Steps 3(a)∼3(d). In Step 3(a), 2^{120} ciphertext pairs denoted as C are treated with 2^8 subkey candidates for $RK_{22,0}$, so the time complexity is about $2^{120} \times 2^8 \times 2 \times \frac{1}{23} \times \frac{1}{4} \approx 2^{122.5}$ 23-round SMS4 encryptions. For every guess i $(0 \leqslant i \leqslant 255)$ of $RK_{22,0}$, about 2^{112} ciphertext pairs are expected to remain after this step, which are denoted as C_i ($|C_i| \approx 2^{112}$). Then all C_i are treated with 2^8 subkey candidates for $RK_{22,1}$, the time complexity of this step is about $2^{112} \times 2^8 \times 2^8 \times 2 \times \frac{1}{23} \times \frac{1}{4} \approx 2^{122.5}$. For every guess (i, j) $(0 \leqslant i, j \leqslant 255)$ of $(RK_{22,0}, RK_{22,1})$, about 2^{104} pairs are left after the step, which are denoted as $C_{i,j}$ ($|C_{i,j}| \approx 2^{104}$). Then all $C_{i,j}$ are treated with 2^8 subkey candidates for $RK_{22,2}$, the time complexity of this step is about $2^{104} \times 2^{16} \times 2^8 \times 2 \times \frac{1}{23} \times \frac{1}{4} \approx 2^{122.5}$. Similar analysis can be applied to the filtering test for every guess of $RK_{22,3}$, the time complexity of this step is about $2^{104} \times 2^{24} \times 2^8 \times 2 \times \frac{1}{23} \times \frac{1}{4} \approx 2^{122.5}$. According to the analysis above, the time complexity of Step 3(a) is about $4 \times 2^{122.5}$. Similarly, the time complexity of Step 3(b) is about $4 \times 2^{122.5}$, the time complexity of Step 3(c) is about $4 \times 2^{122.5}$. After Step 3(c), about 2^{24} pairs are left for every guess of RK_{22}, RK_{21} and RK_{20}. Therefore, the time complexity of Step 3(d) is about $(2^{24+104+1} + 2^{17+112+1} + 2^{10+120+1}) \times \frac{1}{23} \times \frac{1}{4} \approx 2^{125.3}$. Hence, the total time complexity is about $2^{122.5} \times 12 + 2^{125.3} \approx 2^{126.7}$ 23-round SMS4 encryptions.

6 Summary

In this paper, we firstly give a clarification for the minimum number of active S-boxes in 6-, 7- and 12-round SMS4 respectively. The key point is the utility of four relationships for 5- and 6-round SMS4. These relationships concern the input difference and the output difference of the T functions in different rounds, and result from the structure of SMS4 (i.e., the special 4-branch generalized Feistel network of its overall structure and the SP network of the T functions). Then, we present a family of effective 19-round distinguishers on SMS4 (the previous best distinguishers can only reach 18 rounds), which leads to a differential attack on 23-round SMS4. Table 6 summarizes our attack along with the previously known ones on reduced-round SMS4.

Table 6. Summary of our Attack and the Previously Known Attacks on SMS4

Rounds	Attack Type	Data	Time	Source
13	Integral Attack	2^{16}	2^{114}	[3]
14	Rectangle Attack	$2^{107.89}$	$2^{87.69}$	[6]
16	Rectangle Attack	2^{125}	2^{116}	[7]
16	Impossible Differential	$2^{117.06}$	$2^{95.07}$	[6]
18	Rectangle Attack	2^{124}	$2^{112.83}$	[9]
18	Boomerang Attack	2^{120}	$2^{116.83}$	[9]
21	Differential Cryptanalysis	2^{118}	$2^{126.6}$	[7]
22	Linear Cryptanalysis	$2^{118.4}$	2^{117}	[8]
22	Linear Cryptanalysis	2^{117}	$2^{109.86}$	[9]
22	Differential Cryptanalysis	2^{118}	$2^{125.71}$	[9]
22	Differential Cryptanalysis	2^{117}	$2^{112.3}$	[10]
23	Differential Cryptanalysis	2^{118}	$2^{126.7}$	This paper

References

[1] Specification of SMS4, block cipher for WLAN products – SMS4. http://www.oscca.gov.cn/UpFile/200621016423197990.pdf. (in Chinese)

[2] Diffie W, Ledin G (translators). SMS4 encryption algorithm for wireless networks. Cryptology ePrint Archive, Report 2008/329, Received Jul. 29 2008, http://eprint.iacr.org/.

[3] Liu F, Ji W, Hu L, Ding J, Lv S, Pyshkin A, Weinmann R P. Analysis of the SMS4 block cipher. In *Proc. ACISP 2007*, Townville, Australia, Jul. 2-4, 2007, pp.158-170.

[4] Ji W, Hu L. New description of SMS4 by an embedding over $GF(2^8)$. In *Proc. INDOCRYPT 2007*, Chennai, India, Dec. 9-13, 2007, pp.238-251.

[5] Lu J. Attacking reduced-round versions of the SMS4 block cipher in the Chinese WAPI standard. In *Proc. ICICS 2007*, Zhengzhou, China, Dec. 12-15, 2007, pp.306-318.

[6] Toz D, Dunkelman O. Analysis of two attacks on reduced-round versions of the SMS4. In *Proc. ICICS 2008*, Paris, France, Dec. 14-17, 2008, pp.141-156.

[7] Zhang L, Zhang W T, Wu W L. Cryptanalysis of reduced-round SMS4 block cipher. In *Proc. ACISP 2008*, Wollongong, Australia, Jul. 7-9, 2008, pp.216-229.

[8] Etrog J, Robshaw M J B. The Cryptanalysis of reduced-round SMS4. In *Proc. SAC 2008*, Fortaleza, Brazil, Mar. 16-20, 2008, pp.51-65.

[9] Kim T, Kim J, Hong S, Sun J. Linear and differential cryptanalysis of reduced SMS4 block cipher. Cryptology ePrint Archive, Report 2008/281, http://eprint.iacr.org/.

[10] Zhang W T, Wu W L, Feng D G, Su B Z. Some new observations on the SMS4 block cipher in the Chinese WAPI standard. In *Proc. ISPEC 2009*, Xi'an, China, Apr. 13-15, 2009, pp.324-335.

[11] Lu J, Kim J, Keller N, Dunkelman O. Improving the efficiency of impossible differential cryptanalysis of reduced camellia and MISTY1. In *Proc. CT-RSA 2008*, San Francisco, USA, Apr. 8-11, 2008, pp.370-386.

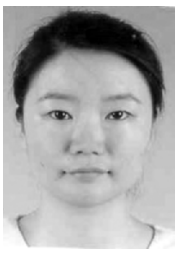
[12] Daemen J. Cipher and hash function design strategies based on linear and differential cryptanalysis [Ph.D. Dissertation]. K.U. Leuven, March 1995.

[13] Biham E, Shamir A. Differential cryptanalysis of DES-like cryptosystems. *Journal of Cryptology*, 1991, 4(1): 3-72.

[14] Selçuk A A. On probability of success in linear and differential cryptanalysis. *Journal of Cryptology*, 2008, 21(1): 131-147.

Bo-Zhan Su received his M.S. degree in maths from Northwest University in 1999. He is currently a Ph.D. candidate of Institute of Software, Chinese Academy of Sciences, and Graduate University of Chinese Academy of Sciences. His current interests include block cipher and hash function.

Wen-Ling Wu is now a professor and Ph.D. supervisor at the State Key Laboratory of Information Security, Institute of Software, Chinese Academy of Sciences. She received her B.S. degree and M.S. degree in maths from Northwest University in 1987 and 1990, respectively. She received her Ph.D. degree in cryptography from Xidian University in 1997. From 1998 to 1999 she was a postdoctoral fellow in the Institute of Software, Chinese Academy of Science. She is a senior member of China Computer Federation. Her current research interests include theory of cryptography, mode of operation, block cipher, stream cipher and hash function.



Wen-Tao Zhang is an associate professor at the State Key Laboratory of Information Security, Institute of Software, Chinese Academy of Sciences. Her main interest is block cipher.