The upper bounds on differential characteristics in block cipher SMS4

Zhang MeiLing♣, Liu JingMei, Wang XinMei

National Key Lab. of Integrated Service Networks, Xidian University. Xi'an, 710071,China

Abstract: SMS4 is a 128-bit block cipher with a 128-bit user key and 32 rounds, which is used in the Chinese National Standard for Wireless LAN WAPI. In this paper, all possible differential patterns are divided into several sections by six designed rules. In order to evaluate the security against the differential cryptanalysis of SMS4, we calculate the lower bounds on the number of active S-Boxes for all kinds of sections, based on which the lower bounds on the number of active S-Boxes in all possible differential patterns can be derived. Finally, the upper bounds on differential probabilities of reduced or and full round SMS4 are given, which can be used to estimate the upper bounds on the differential characteristic probabilities and the linear characteristic probabilities.

Keywords: SMS4, active S-Boxes, lower bounds, upper bounds, differential pattern, differential characteristic.

1 Introduction

SMS4[1], released in January 2006, is the underlying block cipher used in the WAPI(WLAN Authentication and Privacy Infrastructure) standard to protect WLAN products. SMS4 employs a 32-round unbalanced Feistel network structure on encryption algorithm and key scheduling algorithm.

So far, there have been several attacks on reduced SMS4, while differential attack is one of the most efficient cryptanalysis. In [6], by iterating a kind of 5-round differential characteristic three and a half times, an 18-round differential characteristic with a probability of 2^{-126} is derived. Then based on this 18-round differential characteristic, a differential attack is applied to 21-round SMS4 whose complexities are 2^{118} chosen plaintexts and $2^{126.6}$ encryptions. By using an early abort technique, an improved differential attack [4] on 22-round SMS4 is presented, based on the same 18-round differential characteristic, but with more efficient time complexity: $2^{125.71}$ encryptions. In [7], twelve 18-round differential characteristics, each with a higher probability of 2^{-114}, are proposed, the main idea of which is similar to that in [6]. Then a differential attack on 22-round SMS4 based on any one of the 12 differential characteristics is presented, with complexities of 2^{117} chosen plaintexts and $2^{112.3}$ encryptions.

The precise estimation of the lower bounds on the number of active S-Boxes of block ciphers has been known as one of the practical means to evaluate the strength of ciphers [5], because the lower bounds can be used to estimate the upper bounds on the differential characteristic probabilities and the linear characteristic probabilities. In this paper, we focus on the investigation of the lower bounds on the number of active S-Boxes of SMS4. Firstly, we design six new rules to divide the differential pattern into several possible sections; secondly, it is investigated that the lower bound on the number of active S-Boxes in each possible section. As any possible differential pattern consists of some sections, then the lower bound on the number of active S-Boxes in any certain differential pattern can be derived by combining the lower bounds on the number of *AS* in these sections.

The outline of this paper is as follows: in Section 2, we introduce the SMS4 algorithm. The lower bounds on the number of differential active S-Box for all possible differential patterns of

♣ E-mail: zhangmlwy@gmail.com

SMS4 are discussed in Section 3. Finally, section 4 concludes the paper.

2 Description of SMS4 algorithm

2.1 Notations

- W: W= Z_2^{32} , the set of 32-bit words.
- B: B= Z_2^8 , the set of 8-bit bytes.
- S(.) : the 8*8 bijective S-Box used in the round function F.
- L(.) : linear transformation in the round function F.
- + : bitwise exclusive-OR (XOR).
- <<<i : left rotation by i bits.
- RK_i : a 32-bit subkey in round i.
- (P_i, P_{i+1}, P_{i+2}, P_{i+3}): the input of the i-th round, where $P_i \in W$ (i=1, …33).
- V_i: the input difference of the non-linear transformation S, where $V_i \in W$, (i=1, 2, …, 32).
- Z_i: the input difference of the linear diffusion function L, also the output difference of the S. where $Z_i \in W$, (i=1, 2, …, 32).
- U_i: the output difference of L, where $U_i \in W$, (i=1, 2, …, 32).
- ZS : a section composed by consecutive 32-bit zeros.
- NZS : a section composed by consecutive 32-bit non-zeros.
- ||: concatenation.
- Nrnd： the total round number, and the Nrnd of SMS4 is 32.
- *AS*: the number of active S-Boxes.
- *AS*(SEC): the number of active S-Box in the section SEC
- Hw(x): the Hamming weight of x, i.e. the number of non-zero bytes, $x \in W$
- #Sec: the number of elements in Sec.
- dF(Δ)： dF(Δ)=S(L(x))+ S(L(x+ Δ)).

2.2 Description of SMS4

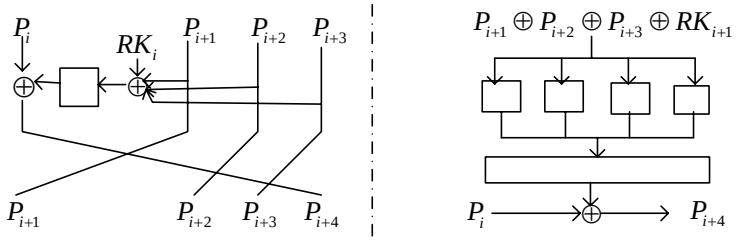

Fig.1. (1) the i-th round function of SMS4 (2) F function

SMS4 is 32-round unbalanced Feistel cipher whose block sizes and key sizes are both 128-bit. The plaintext and ciphertext are represented as four 32-bit words (P_1, P_2, P_3, P_4) $\in (W)^4$, and (C_1, C_2, C_3, C_4) $\in (W)^4$ respectively. $RK_i \in W$ is the 32-bit round key for the i-th（i=1, 2, …, 32）round. The encryption procedure of SMS4 is as follows: (see Fig.1.）

1. Input the plaintext P = (P_1, P_2, P_3, P_4),
2. For i (=1, 2, …, 32)
   - P_{i+4}=R(P_i, P_{i+1}, P_{i+2}, P_{i+3}, RK_i)= $P_i \oplus$ F($P_{i+1} \oplus P_{i+2} \oplus P_{i+3} \oplus RK_i$), （$i$=1, 2, …, 32），
3. Output the ciphertext C = (C_1, C_2, C_3, C_4)=(P_{36}, P_{35}, P_{34} , P_{33}).

Where the transformations F is composed of a non-linear transformation S and a linear diffusion

function L, namely F(.)=L(S(.)). The non-linear transformation S applies the same 8*8 S-Box four times in parallel to an 32-bit input. Then the 32-bit output of S will be the input of L. Let $X \in W$ and $Y \in W$ be the input and output of L respectively. Then the L and its inverse L^{-1} are defined as follows:

$Y=L(X)= X \oplus (X<<<2) \oplus (X<<<10) \oplus (X<<<18) \oplus (X<<<24)$,

$X = L^{-1}(Y)=Y \oplus (Y<<<2) \oplus (Y<<<4) \oplus (Y<<<8) \oplus (Y<<<12) \oplus (Y<<<14) \oplus (Y <<< 16) \oplus (Y<<<18) \oplus (Y<<<22) \oplus (Y<<<24) \oplus (Y<<<30)$.

The key scheduling algorithm of SMS4 is very similarly to encryption algorithm. The difference is that the diffusion function $L'(X)= X \oplus (X<<<13) \oplus (X<<<23)$, (see [1] for detail).

Definition 1 [2]**.** For any given Δt, $\Delta x \in B$, the differential probability of S-Box is defined as:

$$DP^s(\Delta t \to \Delta x) = \frac{\#\{t \in B \mid S(t) \oplus S(t \oplus \Delta t) = \Delta x\}}{\#B}.$$

Definition 2 [3]**.** A differential active S-Box is defined as an S-Box given a non-zero input difference.

3 the upper bounds on the differential characteristics in SMS4

Our goal is to compute the upper bounds on the differential characteristics in SMS4, i.e. the lower bounds on active S-Box in the differential characteristics, so as to evaluate the security of SMS4 against differential cryptanalysis.

Let "1" denote the nonzero input difference of a F function, and "0" denote the zero input difference. And for an r-round differential characteristic, with input difference sequence (v_i, v_{i+1},…, v_{i+r-1}). (b_i, b_{i+1},…, b_{i+r-1}) is the differential pattern of the r-round differential characteristic if the b_{i+j} satisfies:

$$b_{i+j} = \begin{cases} 1, & v_{i+j} \neq 0 \\ 0, & v_{i+j} = 0 \end{cases} \quad (j = 0,...,r-1).$$

3.1 Division of differential patterns

We divide a differential pattern to several sections by the following rules.

Rule 1: consecutive 0s comprise a ZS section, and consecutive 1s comprise a NZS section. Then the ZS and NZS appear alternatively in the differential pattern.

Rule 2: consecutive four sections, ZS||NZS||ZS||NZS, comprise a BS section (short for big section). If the number of elements in BS is equal to 4, we call it SBS (short for special BS), i.e. the pattern of SBS is (0, 1, 0, 1). And if #(ZS||NZS||ZS||NZS)>4, we call the it NBS (short for Normal BS).

Rule 3: if there are several consecutive SBS, we combine them into an *x*SBS section, i.e. SBS||SBS||…||SBS. If not, we use Rule 4.

Rule 4: if there exists SBS||NBS||SBS, we combine the three sections. If not, we use Rule 5.

Rule 5: if there is a SBS between two NBSs, we combine it with the previous NBS or the following NBS, i.e. NBS||SBS or SBS||NBS.

Rule 6: PS is a special section that is composed of parts of BS. Then all the possible PS has the following forms: (1) NZS, (2) ZS, (3) NZS||ZS, (4) ZS||NZS, (5) NZS||ZS||NZS, (6) ZS||NZS||ZS. We combine the PS with the previous section:

Example 1: pattern (0, 0, 1, 0, 1, 1, 0, 1, 0, 1, 0, 1, 0).

The pattern can be expressed as (0, 0, 1, 0, 1, 1) || (0, 1, 0, 1) || (0, 1, 0), i.e. NZS||SBS||PS.

After dividing, all kinds of patterns can be expressed by one section or combination of some

sections of (1) NBS, (2) *x*SBS, (3) SBS||NBS, (4) NBS||SBS, (5) SBS||NBS||SBS; and (6) NBS||PS, (7) *x*SBS||PS, (8) SBS||NBS||PS, (9) NBS||SBS||PS, (10) SBS||NBS||SBS||PS.

Note that the sections that contain PS section is only located at the last.

The main idea of calculating the lower bound on the number of active S-Box in the differential characteristic is: (1) calculating the lower bounds on the number of active S-Box in each section; (2) calculating the lower bounds on the number of active S-Box in all possible combination of sections.

3.2 lower bounds on *AS* in several consecutive rounds of SMS4

Theorem 1 It's impossible that the input difference of S are all zero difference for four or even more consecutive rounds. The input difference of S and the output difference of L satisfy the following relationship.

$$V_i=V_{i-4}+U_{i-1}+U_{i-2}+U_{i-3}.$$

Proof. We use reduction to absurdity to proof the first part, and deduce for the second part.

(1) Suppose that the input difference of S are all zero difference for four or even more consecutive rounds, then following equations holds.

$V_i= P_{i+1}+P_{i+2}+P_{i+3}=0,$ (1)

$V_{i+1}= P_{i+2}+P_{i+3}+P_{i+4}=0,$ (2)

$V_{i+2}= P_{i+3}+P_{i+4}+P_{i+5}=0,$ (3)

$V_{i+3}= P_{i+4}+P_{i+5}+P_{i+6}=0,$ (4)

Then $P_{i+4}= P_i$, $P_{i+5}= P_{i+1}$, and $P_{i+6}= P_{i+2}$.

Thus, $P_i=P_{i+1}=P_{i+2}=P_{i+3}=0$, which contradict non-all-zero differences.

(2)$V_i= \Delta P_{i+1}+ \Delta P_{i+2}+ \Delta P_{i+3}$

$= \Delta P_{i-3}+dF(\Delta P_{i-2}+ \Delta P_{i-1}+ \Delta P_i)+ \Delta P_{i-2}+dF(\Delta P_{i-1}+ \Delta P_i+ \Delta P_{i+1})$

$+ \Delta P_{i-1}+dF(\Delta P_i+ \Delta P_{i+1}+ \Delta P_{i+2})$

$=P_{i-3}+dF(V_{i-2})+P_{i-2}+dF(V_{i-1})+P_{i-1}+dF(V_i)= \Delta P_{i-3}+U_{i-2}+ \Delta P_{i-2}+U_{i-1}+ \Delta P_{i-1}+U_i$

$= V_{i-4}+U_{i-1}+U_{i-2}+U_{i-3}.$ Q.E.D.

Corollary 1. Two consecutive nonzero differences exist before or after three consecutive zero differences.

Corollary 2. For the differential pattern (0, 0, 1, 1), if the following input difference of next round is 0, then the two differences of the following two rounds must be both 0.

Example 2: the differential pattern (0, 0, 1, 1, 0, 1) doesn't exist, as $V_5=U_3+U_4=0$ and $V_6=U_3+U_4$, V_6 must be 0.

Corollary 3. If there exists a r-round differential characteristic (DC) for a r-round differential pattern, the reverse order of the r-round differential pattern must exist and the reverse order of DC is one of its corresponding r-round differential characteristics.

Example 3: for a 7-round differential pattern (0, 1, 1, 0, 1, 0, 0), its corresponding characteristic is (0, c_1, c_2, 0, c_3, 0, 0). Then the characteristic (0, 0, c_3, 0, c_2, c_1, 0) is one of characteristics of the pattern (0, 0, 1, 0, 1, 1, 0).

Definition 3. The differential branch number B_d of linear transformation L is defined as:

$$B_d = \min_{a,b\neq a}\{\mathrm{Hw}(a \oplus b) + \mathrm{Hw}(L(a) \oplus L(b))\}.$$

Property 1. The branch number of the linear transformation L in the round function of SMS4 is 5.

Property 2. For the S-Box of SMS4, there exist 127 possible output differences for any nonzeros input difference, of which 1 output difference occurs with probability 2^{-6}, and each of the other

126 output difference occurs with probability 2^{-7}. Then $p_s=2^{-6}$.

Property 3. For any V_1, $V_2 \in W$, $Hw(V_1+V_2) \leq Hw(V_1)+Hw(V_2)$.

Theorem 2: If $V_3=U_1+U_2 \neq 0$ or $V_3=V_1+U_2 \neq 0$, Then $Hw(V_1)+ Hw(V_2) +Hw(V_3) \geq 5$.

Proof:

For $V_3=U_1+U_2 \neq 0$, as $U_1+U_2=L(Z_1+Z_2)$, then $Hw(U_1+U_2)+Hw(Z_1+Z_2) \geq 5$.

Hence $Hw(V_1)+Hw(V_2)+Hw(V_3) = Hw(Z_1)+Hw(Z_2)+Hw(U_1+U_2) \geq Hw(Z_1+Z_2)+Hw(U_1+U_2) \geq 5$.

And for $V_3=V_1+U_2 \neq 0$, by Property 3,

$Hw(V_1)+Hw(V_2)+Hw(V_3) \geq Hw(V_1+V_3)+Hw(V_2)=Hw(U_2)+Hw(V_2) \geq 5$. Q.E.D.

Theorem 3: If $U_1=U_2 \neq 0$, $Hw(V_1)=Hw(V_2)$.

Proof:

As $U_1=U_2$, $L^{-1}(U_1)=L^{-1}(U_2) \neq 0$, i.e. $Z_1=Z_2$.

Thus, $Hw(V_1)=Hw(V_2)$ and the locations of nonzero differences of V_1 and V_2 are the same. Q.E.D.

Now we investigate the lower bound on the number of active S-Box in each section of {NBS, *x*SBS, SBS||NBS, NBS||SBS, SBS||NBS||SBS, NBS||PS, *x*SBS||PS, SBS||NBS||PS, NBS||SBS||PS, SBS||NBS||SBS||PS}. We use *AS*/N to evaluate the frequency of *AS*, where N is the number of rounds in certain section.

Note that our purpose is to calculate the lower bounds on the *AS* of differential patterns, which means that the lower frequency (*AS*/N) is, the better the result is. The higher frequency, for example *AS*/N $\geq$ 1, makes no contribution to the lower bound.

Based on the results above, the lower bound on *AS*/N of each section is derived, see Table 1。

Table 1: the lower bounds on *AS*/N of all kinds of sections.

Sections	The lower bounds on *AS*/N of the last section	The lower bounds on *AS*/N of other location section
NBS	(N-1)/N(N $\geq$ 6); (N-2)/N (N $\geq$ 7)	(N-1)/N (N $\geq$ 7)
SBS||NBS	(N-1)/N (N $\geq$ 9); (N-2)/N (N $\geq$ 10)	(N-1)/N (N $\geq$ 11)
NBS||SBS	(N-2)/N (N $\geq$ 10); (N-2)/N (N $\geq$ 10)	(N-1)/N (N $\geq$ 11) ; (N-2)/N (N $\geq$ 13)
SBS||NBS||SBS	(N-1)/N (N $\geq$ 12)	(N-2)/N (N $\geq$ 14);
*x*SBS	(N-1)/N (N $\geq$ 8)	1 (N $\geq$ 8)
NBS||PS	(N-3)/N (N $\geq$ 15); (N-2)/N (N $\geq$ 8)	-
SBS||NBS||PS	(N-3)/N (N $\geq$ 14); (N-2)/N (N $\geq$ 9)	-
NBS||SBS||PS	(N-2)/N (N $\geq$ 11)	-
SBS||NBS||SBS||PS	(N-2)/N (N $\geq$ 14)	-
*x*SBS||PS	(N-2)/N (N $\geq$ 13)	-

Now we give an example to show how to calculate the *AS*/N.

Example 4*:* for the Pattern (0, 0, 1, 0, 1, 1, 0, 1, 0, 1, 0, 1, 0).

By Theorem 1, we have

$V_5=U_3$, $V_{12}=V_8+U_{10}$, and $U_8=U_{10}=U_{12}$.

Thus $Hw(V_3)+Hw(V_5)=Hw(V_3)+Hw(U_3) \geq 5$.

By Theorem 2 and 3, $Hw(V_8)=Hw(V_{10})=Hw(V_{12})$, and $Hw(V_8)+Hw(V_{10})+Hw(V_{12}) \geq 5$.

Thus $(V_8)+Hw(V_{10})+Hw(V_{12}) \geq 6$.

Therefore, the lower bound on *AS* of the Pattern is 12 and *AS*/N $\geq$ 12/13.

Example 5: for the Pattern (0, 1, 1, 0, 0, 1).

The pattern is a NBS section. The lower bound of *AS* may be difference while it is located at the last or not.

(1) the pattern is located at the last, that is to say, there is no other sections followed.

By Theorem 1, we have U2=U3 and V6=V2+U3.

Thus, $Hw(V_2)+Hw(V_3)+Hw(V_6) \geq Hw(V_2+V_6)+Hw(V_3)=Hw(U3)+Hw(V3) \geq 5$.

Hence, $AS/N \geq 5/6$.

(2) If the location of the pattern is not at the last, then there is one zero followed at least.

Then we have V3+U6=0.

Thus, $Hw(V_2)+Hw(V_3)+Hw(V_6)=Hw(V_2)+Hw(U_6)+Hw(V_6) \geq 6$.

Hence, $AS/N \geq 1$.

According to Table 1, we can test all the combinations, and obtain the lower bound on *AS*/N of reduced or full rounds.

Corollary 4: the lower bounds on *AS*/N of reduced or full rounds:

(1) when $N \geq 14$, $AS \geq N-3$. The corresponding pattern is SBS||NBS||PS.

(2) when $N \geq 20$, $AS \geq N-4$. The corresponding pattern is NBS||SBS||NBS.

(3) when $N \geq 27$, $AS \geq N-5$. The corresponding pattern is NBS||SBS||SBS||NBS||PS.

4 Conclusion

We have studied the lower bound of *AS* or S/N of consecutive rounds of SMS4, which shows that the block cipher SMS4 is immune against differential cryptanalysis based on certain differential characteristic. The whole process is carried on analyzing the structure of the round function. It is to be noted here that the bounds in Corollary 4 may be tighter. We think that our method can also be applied to study other similar ciphers to search some differential characteristics with high probabilities.

5 Acknowledgment

We would like to thank Y. Wang, L. Li, W.G. Zhang, and W. Chen for their comments. The research presented in this paper is supported by State Key Laboratory of Information Security (Institute of Software, Chinese Academy of Sciences), National Key Laboratory of Integrated Service Networks (ISN10-11 and ISN090402), the National Natural Science Foundation of China (No. 60773002 and 60903199), and Natural Science Basic Research Plan in Shanxi Province of China (No.SJ08-ZT14).

Reference:

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

[2] E. Biham, and A. Shamir, Differential cryptanalysis of DES-like cryptosystems, Journal of Cryptology, Vol.4, No.1, pp.3-72, 1991.

[3] M.Kanda, Y.TakAshima, T.MaTsumoto, K.Aoki, and K.Ohta, “A strategy for constructing fast round function with practical security against differential and linear cryptanalysis, ” Selected Areas in Cryptography-5th Annual International Workshop, SAC’98, LNCS 1556.PP.264-279, 1999.

[4] T. Kim, J. Kim, S. Hong, J. Sun, Linear and Differential Cryptanalysis of Reduced SMS4 Block Cipher. Cryptology ePrint Archive, report 2008/281, available at http://eprint.iacr.org/.

[6][5] Taizo Shirai and Kyoji Shibutani. On Feistel Structures using a Diffusion Switching Mechanism. In M.J.B. Robshaw, editor, Proceedings of Fast Software Encryption – FSE’06, number 4047 in Lecture Notes in Computer Science, pages 41-56. Springer, 2006.

[3][6] L. Zhang, W.T. Zhang and W.L. Wu, Cryptanalysis of reduced-round SMS4 Block cipher. ACISP 2008. LNCS, vol. 5107, pp. 216-299. Springer, 2008.
[5][7] W. T. Zhang, W.L. Wu, D.G. Feng, and B.Z. Su, Some New Observations on the SMS4 Block Cipher in the Chinese WAPI Standard. ISPEC 2009. LNCS, vol. 5451, pp.324-335. Springer, 2009.