Linearity within the SMS4 Block Cipher

Muhammad Reza Z'aba, Leonie Simpson, Ed Dawson, and Kenneth Wong

Information Security Institute, Queensland University of Technology, GPO Box 2434, Brisbane, Queensland 4001, Australia
m.zaba@isi.qut.edu.au, {lr.simpson,e.dawson,kk.wong}@qut.edu.au

Abstract. We present several new observations on the SMS4 block cipher, and discuss their cryptographic significance. The crucial observation is the existence of fixed points and also of simple linear relationships between the bits of the input and output words for each component of the round functions for some input words. This implies that the non-linear function T of SMS4 does not appear random and that the linear transformation provides poor diffusion. Furthermore, the branch number of the linear transformation in the key scheduling algorithm is shown to be less than optimal. The main security implication of these observations is that the round function is not always non-linear. Due to this linearity, it is possible to reduce the number of effective rounds of SMS4 by four. We also investigate the susceptibility of SMS4 to further cryptanalysis. Finally, we demonstrate a successful differential attack on a slightly modified variant of SMS4. These findings raise serious questions on the security provided by SMS4.

Key words: SMS4, block cipher, round function, fixed point, encryption, key scheduling algorithm, linearity, cryptanalysis

1 Introduction

SMS4 [14,7] is a 32-round block cipher with 128-bit input block and 128-bit master key. It is used in the Chinese Wireless LAN Wired Authentication and Privacy Infrastructure (WAPI). Using the terminology of Schneier and Kelsey [16], the cipher employs a homogeneous, complete, source-heavy unbalanced Feistel network structure. The encryption and the key scheduling algorithms are nearly identical. The only difference between the structures of these two algorithms is the linear transformation used in each round function.

Since SMS4 was made public in January 2006, the cipher has endured extensive cryptanalysis. Reduced-round versions of the cipher have been cryptanalyzed using integral [12], rectangle [13,17,19,10], impossible differential [13,17], boomerang [10], differential [10,19] and linear [10,8] attacks. The best attack so far is a differential attack on 22 rounds by Zhang et al. [18]. In the same paper, they observe that the number of rotations and XOR operations used in the linear transformation of the SMS4 block cipher is the minimum required to

achieve an optimal branch number. They also show that the linear transformation is bijective and present the distribution of input and output patterns of this transformation to assist in differential attacks.

In this paper, we present further observations on both the encryption and the key scheduling algorithms of the SMS4 block cipher. The crucial observation is the existence of fixed points and also of simple linear relationships between the bits of the input and output words for each component in the round functions. In particular, we show that the non-linear function T has 11 fixed points. Note that the expected number of fixed points for a random permutation is one [9, Chap. 6]. Therefore, the function T does not behave like a random permutation. We also identified a set of input words for which the round functions of both the encryption and the key scheduling algorithms produce the same output words. Furthermore, we show that the branch number of the linear transformation in the key scheduling algorithm is four, which is less than optimal.

One of the implications of these observations is that the first four round functions of SMS4 are not always non-linear. Under this condition, the number of effective rounds is reduced by four: from 32 to 28. We briefly explore the susceptibility of SMS4 against algebraic and advanced variants of the slide attacks. Finally, we demonstrate that if the linear transformation in the key scheduling algorithm was used in the encryption algorithm, then this variant of SMS4, reduced to 27 rounds, is vulnerable to a differential attack. In contrast, the best differential attack on the original SMS4 is on 22 rounds [18], which is also the best existing attack so far. These observations might potentially be useful in attacking SMS4 itself.

This paper is organized as follows. Section 2 describes the specification of the SMS4 block cipher. The observations on the components in the round functions of both the encryption and the key scheduling algorithms are analyzed in Section 3. Section 4 discusses the cryptographic significance of these observations. Section 5 presents a differential attack on a slightly modified variant of SMS4. A summary of our observations and conclusions are given in Section 6.

2 Specification of SMS4

SMS4 [14,7] is a block cipher that accepts a 128-bit plaintext block P, and a 128-bit master key K. The master key is used as input to the key scheduling algorithm to produce a set of thirty-two 32-bit round subkeys. The plaintext block and the round subkeys are used as input to the encryption algorithm to produce the ciphertext block C. The encryption algorithm consists of 32 applications of the round function.

2.1 Round Function of the Encryption Algorithm

Let $P = (X_0, X_1, X_2, X_3)$ denote the 128-bit plaintext block formed from the concatenation of four 32-bit words X_i. Let K_i denote the 32-bit i-th round subkey derived from the 128-bit master key K. The derivation of these subkeys

is explained in Section 2.2. Let $T = L \circ S$ denote the function composed of the non-linear transformation S and the linear transformation L. Both S and L are described in detail later. The i-th round function of the encryption algorithm can be described as follows:

$$X_{i+4} = X_i \oplus T(X_{i+1} \oplus X_{i+2} \oplus X_{i+3} \oplus K_i),\ i = 0, 1, \ldots, 31$$

and is depicted in Figure 1. The ciphertext consists of the concatenation of the four 32-bit words $C = (X_{35}, X_{34}, X_{33}, X_{32})$, which is obtained in the reverse order from the output of the final round function to facilitate decryption.

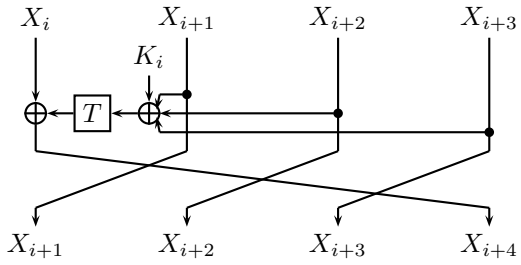

Fig. 1. Round Function of SMS4 in Round i

Decryption is the same as encryption with the only difference being the order in which the subkeys are used; this is in the reverse order as follows:

$$X_i = X_{i+4} \oplus T(X_{i+3} \oplus X_{i+2} \oplus X_{i+1} \oplus K_i),\ i = 31, 30, \ldots, 0.$$

The function T is the composition of the two transformations S and L, where S is applied first, followed by L. These transformations operate on 32-bit words. Let $X_i = (X_{i,0}, X_{i,1}, X_{i,2}, X_{i,3})$ denote a 32-bit word formed from the concatenation of four 8-bit words $X_{i,j}$. The application of the non-linear transformation S to X_i consists of the application of a single 8×8 S-box s to $X_{i,j}$ as follows:

$$S(X_i) = (s(X_{i,0}), s(X_{i,1}), s(X_{i,2}), s(X_{i,3})).$$

Let $X_i \lll k$ denote the rotation of X_i by k bits to the left. The linear transformation L is defined as:

$$L(X_i) = X_i \oplus (X_i \lll 2) \oplus (X_i \lll 10) \oplus (X_i \lll 18) \oplus (X_i \lll 24).$$

2.2 Round Function of the Key Scheduling Algorithm

In the initialization phase of the key scheduling algorithm, a 128-bit constant FK is XORed with the 128-bit master key K to produce the initial inputs for the

key scheduling algorithm. Let $K = (MK_0, MK_1, MK_2, MK_3)$ denote the master key formed from the concatenation of four 32-bit words MK_i. Similarly, let $FK = (FK_0, FK_1, FK_2, FK_3)$ denote the constant as the concatenation of four 32-bit words FK_i, where $FK_0 = \texttt{A3B1BAC6}$, $FK_1 = \texttt{56AA3350}$, $FK_2 = \texttt{677D9197}$ and $FK_3 = \texttt{B27022DC}$ (in hexadecimal). Then, the initial input words to the key scheduling algorithm are $K_{i-4} = MK_i \oplus FK_i$ for $i = 0, 1, 2, 3$. Note that this initialization phase has no cryptographic significance because the operation is linear and the constants are known.

Let $T' = L' \circ S$ denote the function composed of the non-linear transformation S and the linear transformation L' (L' is described later). Note that this transformation L' is the only difference between the round functions of the encryption and the key scheduling algorithms. Let K_i and CK_i denote the i-th round 32-bit subkey and constant, respectively. The i-th round function of the key scheduling algorithm can be described as follows:

$$K_i = K_{i-4} \oplus T'(K_{i-3} \oplus K_{i-2} \oplus K_{i-1} \oplus CK_i),\ i = 0, 1, \ldots, 31.$$

The round constants $CK_i = (CK_{i,0}, CK_{i,1}, CK_{i,2}, CK_{i,3})$, which are composed of the concatenation of four 8-bit words $CK_{i,j}$, are defined as

$$CK_{i,j} = (28i + 7j) \bmod 256,\ i = 0, 1, \ldots, 31 \text{ and } j = 0, 1, 2, 3.$$

The linear transformation L' is defined as:

$$L'(X) = X \oplus (X \lll 13) \oplus (X \lll 23).$$

3 Observations on Components in the Round Functions

This section presents several new observations on each component in the round functions of both the encryption and the key scheduling algorithms of SMS4.

3.1 Simple Linear Relationships between Input and Output Words

We observe the existence of a simple linear relationship between the bits of some input and output words of each component in the encryption and the key scheduling algorithms. For a component F, there exist a set of output words of F which are equivalent to a simple rotation of the input word. That is, for some 32-bit words X_i,

$$F(X_i) = X_i \lll j \tag{1}$$

for some particular rotation values of $j \in \{0, 1, \ldots, 31\}$. A fixed point is a special case of this relationship when $j = 0$. For example, consider the linear transformation L, i.e. $F(X_i) = L(X_i)$ and the input word $X_i = \texttt{02020202}$. The output word is $F(\texttt{02020202}) = \texttt{08080808}$, so Equation 1 is valid for $j = 2, 10, 18, 26$.

In the remainder of this section, N_F denotes the total number of distinct values X_i that satisfy the relationship described in Equation 1 for a particular component F. The set containing these input words X_i is denoted by Θ_F.

Additionally, $N_{F,j}$ denotes the number of individual values that satisfy this relationship for a specific rotation value j. Note that the sum $\sum_{j=0}^{31} N_{F,j}$ may be higher than N_F because some input words satisfy this relationship for multiple values of j. For instance, in the previous example, i.e. $F(\texttt{02020202}) = \texttt{08080808}$, the input word `02020202` is counted four times.

Non-linear Transformation S Recall that the non-linear transformation S consists of the application of a single 8×8 S-box s, applied four times in parallel. By reverse engineering, Liu et al. [12] managed to deduce how the S-box for SMS4 is constructed. They found that the S-box s uses an inversion in the finite field, which is similar to that of the AES. Note that the design of the S-box for the AES explicitly avoids fixed points [6]. However, we identified one fixed point in s. The fixed point is the 8-bit value `AB` (in hexadecimal). Thus, the non-linear transformation S also has a fixed point, which is the hexadecimal value `ABABABAB`.

In addition to this fixed point, there also exist other input words X_i that satisfy the relationship $S(X_i) = X_i \lll j$ for some $j > 0$. For these particular input words, the transformation S is basically linear. There are $N_S = 39$ (including the fixed point) distinct input words X_i that have a relationship of this form. Let Θ_S denote the set containing the exact values of these X_i, which are given in Table 3 in the Appendix. The number, $N_{S,j}$, of values that satisfy this relationship for S, for each rotation value j is given in Table 1.

Linear Transformation L We identified four fixed points ($j = 0$) and 1020 other ($j > 0$) input words X_i that satisfy the relationship $L(X_i) = X_i \lll j$, i.e. $N_L = 1024$. Let Θ_L denote the set containing the exact values of these X_i. For these input words, the linear transformation L provides poor diffusion because the input bits of these words are not well scattered by L when producing the output words. The number, $N_{L,j}$, of values that satisfy this relationship for L, for each rotation value j is given in Table 1.

Function T As a non-linear cryptographic component, the function T of SMS4 should behave like a random permutation. The probability that a given permutation of n elements has c fixed points is given by [15, Chap. 3]

$$p_{n,c} = \frac{1}{n!} \cdot \binom{n}{c} \cdot (n-c)! \cdot \sum_{k=0}^{n-c} \frac{(-1)^k}{k!} \approx \frac{1}{c!e}.$$

For both $c = 0$ and $c = 1$, as n tends to infinity, the probabilities $p_{n,0}$ and $p_{n,1}$ approach $e^{-1} = 0.3679$. Therefore, the number of permutations having at least 2 fixed points is approximately $1 - 2(0.3679) = 0.2642$. Note that the expected number of fixed points for a random permutation is one [9, Chap. 6].

By exhaustive search, we found 11 fixed points in the function T of SMS4, i.e. values X_i such that $T(X_i) = X_i$ (for $j = 0$). The fixed points are `0B0B0B0B`,

Table 1. Number of output words which are equivalent to the rotation of the input word by j bits to the left ($0 \leq j \leq 31$), for each component function

j	$N_{S,j}$	$N_{L,j}$	$N_{T,j}$	$N_{L',j}$	$N_{T',j}$	j	$N_{S,j}$	$N_{L,j}$	$N_{T,j}$	$N_{L',j}$	$N_{T',j}$
0	1	4	11	4	0	16	9	4	3	4	0
1	16	2	4	2	6	17	4	2	4	2	2
2	7	1024	7	8	8	18	3	1024	3	8	4
3	0	2	2	2	4	19	0	2	2	2	4
4	1	4	1	4	5	20	1	4	1	4	5
5	3	2	0	2	3	21	3	2	0	2	3
6	1	16	1	8	6	22	1	16	1	8	2
7	0	2	1	2	3	23	0	2	1	2	7
8	3	4	1	4	2	24	3	4	1	4	2
9	2	2	4	2	2	25	2	2	0	2	2
10	1	256	1	8	4	26	1	256	9	8	4
11	0	2	2	2	4	27	0	2	2	2	12
12	1	4	1	4	3	28	1	4	1	4	3
13	1	2	2	2	1	29	1	2	2	2	1
14	1	16	5	8	6	30	1	16	1	8	6
15	0	2	7	2	1	31	0	2	11	2	1

`3E973E97`, `3AE2C6AD`, `62D367B9`, `973E973E`, `E2C6AD3A`, `D367B962`, `C6AD3AE2`, `67B962D3`, `AD3AE2C6` and `B962D367`. For a random permutation, the probability of having 11 fixed points is approximately $p_{n,11} = 1/(11! \cdot e) \approx 9.216\text{E} - 9$, which is quite low. Interestingly, if the S-box of SMS4 is replaced by the S-box of the AES, there are no fixed points in the resulting function T.

Similarly, there exist input words X_i that satisfy the relationship $T(X_i) = X_i \lll j$ for $j > 0$. In total, there are $N_T = 59$ distinct input words X_i (including the fixed points) that satisfy this relationship. Let Θ_T denote the set containing the exact values of these X_i, which are given in Table 4 in the Appendix. The number $N_{T,j}$ for each value of j is given in Table 1.

Recall that the function T is composed of S and L, i.e. $T = L \circ S$. The 39 input words contained in the set Θ_S do not all appear in the set Θ_T. However, there are seven input words that appear in the intersection of these two sets, $\Theta_S \cap \Theta_T$. These input words are `0A0A0A0A`, `0B0B0B0B`, `21212121`, `26262626`, `ABABABAB`, `E7E7E7E7` and `FAFAFAFA`.

Linear Transformation L' We found, by exhaustive search, that there are no fixed points for L'. However, we found $N'_L = 8$ distinct input words X_i that satisfy the relationship $L'(X_i) = X_i \lll j$ for some $j > 0$. Let $\Theta_{L'}$ denote the set containing the exact values of these X_i. As a linear transformation, the diffusion provided by L' is poor for these input words. Note that the size of the set $\Theta_{L'}$ is smaller than the size of Θ_L, despite the fact that L' has fewer rotations than L. The number $N_{L',j}$ of values for each rotation value j is given in Table 1.

Function T' Unlike the function T, the function T' has no fixed points. However, there still exist some input words X_i that satisfy the relationship $T'(X_i) = X_i \lll j$ for some $j > 0$. In total, there are $N_{T'} = 59$ distinct input words X_i that satisfy this relationship. Let $\Theta_{T'}$ denote the set containing the exact values of these X_i, which are given in Table 5 in the Appendix. The number $N_{T',j}$ for each value of j is given in Table 1.

Recall that the function T' is composed of S and L', i.e. $T' = L' \circ S$. The number $N_{T'}$ of input words in the set $\Theta_{T'}$ is about 7 times more than the same number for $\Theta_{L'}$, and 20 more than Θ_S. Unlike the function T, the input words contained in the set Θ_S do not appear at all in the set $\Theta_{T'}$, i.e. $\Theta_S \cap \Theta_{T'} = \emptyset$. However, there exist a set of input words for which the functions T and T' produce the same output words. This relationship is discussed in the following section.

3.2 Relationship between T and T'

As noted in Section 2, the encryption and the key scheduling algorithms are nearly identical, differing only in the linear transformation. We identified eight input words for which the transformation L and L' produce the same output words, i.e. $L(Y_i) = L'(Y_i)$. These input words Y_i are `00000000`, `33333333`, `55555555`, `66666666`, `99999999`, `AAAAAAAA`, `CCCCCCCC` and `FFFFFFFF`.

Recall that the non-linear transformation S is the same in both the functions T and T'. If there exist some input words Y_i such that $L(Y_i) = L'(Y_i)$, then there exist words $X_i = S^{-1}(Y_i)$ such that $T(X_i) = L(S(X_i)) = L'(S(X_i)) = T'(X_i)$. The eight input words X_i are `71717171`, `28282828`, `97979797`, `A5A5A5A5`, `1F1F1F1F`, `18181818`, `04040404` and `B9B9B9B9`.

3.3 On the Branch Number of L'

A commonly used measure of diffusion for Substitution-Permutation-Network (SPN) block ciphers is the notion of the branch number [6]. For an SPN cipher, this number denotes the minimum number of active S-boxes for any two consecutive rounds. However, in the context of a generic Feistel cipher such as SMS4, this is not always true. Therefore, the branch number of a linear transformation L, denoted $\mathcal{B}(L)$, can be defined as the minimum number of non-zero subword differences for any input and output pair of L. If the input word to L is partitioned into m sub-words, then the optimal branch number for L is $\mathcal{B}(L) = m+1$ [6].

The branch number is calculated as follows. Let $X_i = (X_{i,0}, X_{i,1}, \ldots, X_{i,m-1})$ denote a mb-bit word formed from the concatenation of m b-bit words. Let $\Gamma_{X_i} = \Gamma_{X_{i,0}}\Gamma_{X_{i,1}} \ldots \Gamma_{X_{i,m-1}}$ denote a binary vector of length m where $\Gamma_{X_{i,j}} = 1$ if $X_{i,j}$ is nonzero and $\Gamma_{X_{i,j}} = 0$ otherwise. Let $wt(\Gamma_{X_i})$ denote the Hamming weight (i.e. the number of non-zero bits) of Γ_{X_i} . The branch number of L, denoted $\mathcal{B}(L)$, is defined as

$$\mathcal{B}(L) = \min\{wt(\Gamma_{X_i}) + wt(\Gamma_{Y_i}) : X_i \neq 0 \text{ and } Y_i = L(X_i)\}.$$

Table 2. The input-output pattern distribution of L'

Γ_{X_i}	Γ_{Y_i}															
	0	1	2	4	8	3	5	6	9	A	C	7	B	D	E	F
0	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
1	.	.	.	.	.	.	.	.	.	.	.	1	3	31	.	220
2	.	.	.	.	.	.	.	.	.	.	.	3	31	.	1	220
4	.	.	.	.	.	.	.	.	.	.	.	31	.	1	3	220
8	.	.	.	.	.	.	.	.	.	.	.	.	1	3	15	236
3	.	.	.	.	.	7	1	1	3	1	.	242	210	220	252	n_{22}
5	.	.	.	.	.	1	3	1	1	.	1	218	250	218	250	n_{21}
6	.	.	.	.	.	3	1	7	.	1	1	210	220	252	242	n_{22}
9	.	.	.	.	.	1	1	.	7	1	3	380	370	338	236	n_0
A	.	.	.	.	.	1	.	.	1	3	1	251	218	250	235	n_{19}
C	.	.	.	.	.	.	1	1	1	1	7	222	252	242	228	n_{20}
7	.	1	3	1	1	240	248	242	249	249	251	n_{20}	n_{18}	n_{16}	n_{11}	n_{23}
B	.	1	.	.	1	245	252	254	242	249	250	n_1	n_5	n_4	n_{12}	n_{29}
D	.	.	.	1	1	253	249	252	245	252	242	n_3	n_2	n_5	n_{13}	n_{30}
E	.	.	1	3	.	250	250	243	253	249	243	n_{17}	n_{14}	n_9	n_{15}	n_{24}
F	.	253	251	250	252	n_8	n_6	n_8	n_7	n_6	n_{10}	n_{28}	n_{27}	n_{26}	n_{25}	n_{31}

For SMS4, the input word to both L and L' is partitioned into $m = 4$ subwords. Therefore, the optimal branch number for both L and L' is 5. Zhang et al. [18] showed that the branch number of L is indeed optimal, and noted that the number of rotations and XOR operations used in L are the minimum needed to reach this optimal branch number. However, they did not investigate the branch number for L'. We determine the branch number for L' using a computer program and by observing the input-output pattern distribution table defined as follows.

Let both Γ_{X_i} and Γ_{Y_i} denote binary vectors of length $m = 4$. Furthermore, let $W[\Gamma_{X_i}][\Gamma_{Y_i}]$ denote the Γ_{X_i}-th row and Γ_{Y_i}-th column entry for the input-output pattern distribution table. The entries for this table are computed as follows. Initialize the counter W to all-zero. For every input $X_i = 0, 1, \ldots, 2^{32}-1$, calculate the output $Y_i = L'(X_i)$ and increment the counter $W[\Gamma_{X_i}][\Gamma_{Y_i}]$. The resulting table for L' is given by Table 2 where the entry '.' denotes zero, for simplicity. Due to size constraints, some values are denoted by n_i given as follows.

$n_0 = 63688$, $n_7 = 64023$, $n_{14} = 64049$, $n_{21} = 64082$, $n_{28} = 16323877$,
$n_1 = 63894$, $n_8 = 64024$, $n_{15} = 64050$, $n_{22} = 64088$, $n_{29} = 16324086$,
$n_2 = 63895$, $n_9 = 64025$, $n_{16} = 64051$, $n_{23} = 16323681$, $n_{30} = 16324087$,
$n_3 = 63919$, $n_{10} = 64026$, $n_{17} = 64057$, $n_{24} = 16323702$, $n_{31} = 4229286763$.
$n_4 = 63930$, $n_{11} = 64027$, $n_{18} = 64061$, $n_{25} = 16323764$,
$n_5 = 63939$, $n_{12} = 64032$, $n_{19} = 64065$, $n_{26} = 16323875$,
$n_6 = 64019$, $n_{13} = 64040$, $n_{20} = 64070$, $n_{27} = 16323876$,

The branch number of L' can be determined by first searching in Table 2 for a non-zero entry $W[\Gamma_{X_i}][\Gamma_{Y_i}]$ with $\Gamma_{X_i} \neq 0$ for which the sum of the Hamming

weight for Γ_{X_i} and Γ_{Y_i} is the lowest among other entries. Then, the branch number is calculated as $\mathcal{B}(L') = wt(\Gamma_{X_i}) + wt(\Gamma_{Y_i})$. An example of such an entry is $W[1][7]$ and thus, the branch number of L' is $\mathcal{B}(L') = wt(1) + wt(7) = 1+3 = 4$, which is not optimal.

The input-output pattern distribution table also gives information regarding possible and impossible subword difference paths propagated by L'. This is useful for differential-type attacks. The sub-optimal branch number for L' is an indication of a potential weakness. This is exploited in Section 5 in a differential attack on a slightly modified variant of SMS4.

4 Cryptographic Significance

This section discusses the cryptographic significance of the observations made in Section 3.

4.1 Implications for the Key Scheduling Algorithm

The length of the master key for SMS4 is 128 bits, hence there are 2^{128} possible values of the master key. The key scheduling algorithm produces 32 subkeys, each of 32 bits, thus the sequence of subkeys forms a $32 \times 32 = 1024$-bit binary sequence. Clearly, there are extremely many sequences of subkeys that are impossible.

Note that the function T', which is a 32-bit to 32-bit map, is bijective (using the theorem provided by Zhang et al. [18]). In every round, the value of a single 32-bit word is updated using the output of T', a function which takes the other three 32-bit words as input. After four rounds, all 128 bits of the master key are completely updated by the round functions. Therefore, we can reasonably conjecture that all possible values of the first four subkeys are equally likely to occur (statistically independent), whereas the values for the remaining 28 subkeys are determined entirely by these four subkeys. This conjecture allows us to make the following claim.

We know from Section 3.1 that there are 59 distinct words X_i contained in the set $\Theta_{T'}$. Recall that the value of the master key after the initialization phase is partitioned into four 32-bit words $(K_{-4}, K_{-3}, K_{-2}, K_{-1})$ and the i-th round constant is denoted by CK_i. If the input words to the first four consecutive functions T' of the key scheduling algorithm are in the set $\Theta_{T'}$, then the first four subkeys consist of merely linear combinations of the master key[1]. This event is illustrated as follows. If $(K_{-3} \oplus K_{-2} \oplus K_{-1} \oplus CK_0) \in \Theta_{T'}$, then

$$K_0 = K_{-4} \oplus [(K_{-3} \oplus K_{-2} \oplus K_{-1} \oplus CK_0) \lll j_0].$$

[1] Note that the initialization phase does not have any cryptographic significance. Therefore, if we know the value of the resulting key after this phase, then we also know the value of the master key.

Similarly, if $(K_{-2} \oplus K_{-1} \oplus K_0 \oplus CK_1) \in \Theta_{T\prime}$, then

$$K_1 = K_{-3} \oplus [(K_{-2} \oplus K_{-1} \oplus K_{-4} \oplus [(K_{-3} \oplus K_{-2} \oplus K_{-1} \oplus CK_0) \lll j_0] \oplus CK_1) \lll j_1].$$

Furthermore, if $(K_{-1} \oplus K_0 \oplus K_1 \oplus CK_2) \in \Theta_{T'}$, then

$$K_2 = K_{-2} \oplus [(K_{-1} \oplus K_{-4} \oplus [(K_{-3} \oplus K_{-2} \oplus K_{-1} \oplus CK_0) \lll j_0] \oplus K_{-3} \oplus [(K_{-2} \oplus K_{-1} \oplus K_{-4} \oplus [(K_{-3} \oplus K_{-2} \oplus K_{-1} \oplus CK_0) \lll j_0] \oplus CK_1) \lll j_1] \oplus CK_2) \lll j_2].$$

Finally, if $(K_0 \oplus K_1 \oplus K_2 \oplus CK_3) \in \Theta_{T'}$, then

$$\begin{aligned} K_3 = {} & K_{-1} \oplus [(K_{-4} \oplus [(K_{-3} \oplus K_{-2} \oplus K_{-1} \oplus CK_0) \lll j_0] \oplus \\ & K_{-3} \oplus [(K_{-2} \oplus K_{-1} \oplus K_{-4} \oplus [(K_{-3} \oplus K_{-2} \oplus K_{-1} \oplus CK_0) \lll j_0] \oplus \\ & CK_1) \lll j_1] \oplus \\ & K_{-2} \oplus [(K_{-1} \oplus K_{-4} \oplus [(K_{-3} \oplus K_{-2} \oplus K_{-1} \oplus CK_0) \lll j_0] \oplus \\ & K_{-3} \oplus [(K_{-2} \oplus K_{-1} \oplus K_{-4} \oplus [(K_{-3} \oplus K_{-2} \oplus K_{-1} \oplus CK_0) \lll j_0] \oplus \\ & CK_1) \lll j_1] \oplus CK_2) \lll j_2] \oplus CK_3) \lll j_3]. \end{aligned}$$

The above linear equations are valid for specific values of $j_i \in \{0, 1, \ldots, 31\}$. This event occurs with probability $(59/2^{32})^4 \approx 2^{-104.5}$ and thus, there are approximately $2^{23.5}$ values of the master key which cause such an event to happen.

4.2 Implications for the Encryption Algorithm

As noted in Section 3.1, there are 59 distinct words X_i contained in the set Θ_T. If the input words to the first four consecutive functions T of the encryption algorithm are in the set Θ_T, then the output block after four rounds consist of merely linear combinations of the plaintext block and subkeys. In general, this event is similar to that described in Section 4.1. Let us demonstrate the specific case in which only fixed points occur in the first four consecutive rounds. Let $\hat{\Theta}_T$ denote a subset of Θ_T containing the 11 fixed points for T (Refer to Section 3.1). This event is shown as follows for the plaintext block $P = (X_0, X_1, X_2, X_3)$ and subkeys K_0, K_1, K_2 and K_3. If $(X_1 \oplus X_2 \oplus X_3 \oplus K_0) \in \hat{\Theta}_T$ then

$$X_4 = X_0 \oplus X_1 \oplus X_2 \oplus X_3 \oplus K_0.$$

Similarly, if $(X_2 \oplus X_3 \oplus X_4 \oplus K_1) \in \hat{\Theta}_T$, then

$$X_5 = X_0 \oplus K_0 \oplus K_1. \tag{2}$$

Furthermore, if $(X_3 \oplus X_4 \oplus X_5 \oplus K_2) \in \hat{\Theta}_T$, then

$$X_6 = X_1 \oplus K_1 \oplus K_2. \tag{3}$$

Finally, if $(X_4 \oplus X_5 \oplus X_6 \oplus K_3) \in \hat{\Theta}_T$, then

$$X_7 = X_2 \oplus K_2 \oplus K_3. \tag{4}$$

Clearly, for the specific case of fixed points, the linear relationships above are much simpler than the general case because some words X_i and subkeys K_i cancel. This specific event occurs with probability $(11/2^{32})^4 \approx 2^{-114.2}$ and thus, there are approximately $2^{13.8}$ values of the plaintext block that cause such an event to happen for the full SMS4. In the general case, there are $2^{23.5}$ values of the plaintext block that cause the four-round linearity to happen.

4.3 Further Implications for Both the Key Scheduling and the Encryption Algorithms

The points discussed in Sections 4.1 and 4.2 have further security implications for SMS4. In the (admittedly rare) event that both the key scheduling and the encryption algorithms behave linearly for the first four rounds, the output block after four rounds of SMS4 is composed of merely linear combinations of the plaintext block and subkeys. The subkeys, in turn, are composed of linear combinations of the master key. Theoretically, if both of these events occur at the same time, then the number of effective rounds for SMS4 is reduced by four, from 32 to 28.

The above discussions only consider the case for which the linearity occurs in the key scheduling and the encryption algorithms in the first four consecutive rounds. Note that it may be possible for the linearity to occur in any four of the 32 rounds of SMS4. Furthermore, for certain particular combinations of plaintext block and master key, the linearity might possibly exist in more than four rounds. In this case, the number of effective rounds for SMS4 can be further reduced.

4.4 Susceptibility to Algebraic Attack

The algebraic attack [5] introduced by Courtois and Pieprzyk consists of building a system of binary equations that link the plaintext block, subkeys and ciphertext block. The binary equations describing an S-box that uses a finite field inversion, such as the AES and SMS4, are quadratic whereas the remaining equations are linear. The system is then solved to obtain the key bits. One of the obstacles in solving the system of equations for ciphers such as the AES and SMS4 is the existence of quadratic equations. The claimed advantage of this attack is that it only needs very few number of plaintext and ciphertext pairs.

As discussed in Sections 4.1, 4.2 and 4.3, there exist a few exceptional cases in which the non-linear functions T and T' are linear in the first four rounds of SMS4. Under these conditions, the binary equations describing the first four rounds are also entirely linear. Therefore, there is no need to describe the S-boxes in these rounds as systems of quadratic equations [12]. Since the occurrence of this event is statistical in nature, we may need more plaintext and ciphertext pairs compared to a conventional algebraic attack. However, the removal of some quadratic equations might help in reducing the complexity of solving the equation system.

4.5 Susceptibility to Advanced Variants of the Slide Attack

The slide attack was introduced by Biryukov and Wagner [3,4]. Given two different plaintexts, the attack permits the *sliding* of the two encryptions by a certain number of rounds. This is due to the similarity that exists between the structure of the two encryptions. The attack also allows the sliding of encryption with decryption [4].

We have shown in Section 3.2 that there are eight input words for which the functions T and T' produce the same output words. This similarity might provide an avenue for advanced variants of the slide attack. However, it is an open problem to determine whether it is useful to slide the encryption algorithm with the key scheduling algorithm if both algorithms are nearly identical, as is the case for SMS4.

4.6 Subkeys and Related-Keys

As discussed in Section 4.1, we conjecture that all possible 32-bit subkey values of the first four rounds of SMS4 are equally likely to occur. This allows us to explore the relationship between subkeys in these rounds and subkeys in the subsequent rounds. One possible relationship is described as follows. If the first four 32-bit round subkeys are identical (that is $K_i = \hat{K}$ for $i = 0, 1, 2, 3$ where $\hat{K}$ denotes an arbitrary 32-bit value), then a total of 2^{32} (out of 2^{128}) master keys have the following forms: $K_{-1} = \hat{K} \oplus T'(\hat{K} \oplus CK_3)$, $K_{-2} = \hat{K} \oplus T'(K_{-1} \oplus CK_2)$, $K_{-3} = \hat{K} \oplus T'(K_{-2} \oplus K_{-1} \oplus \hat{K} \oplus CK_1)$ and $K_{-4} = \hat{K} \oplus T'(K_{-3} \oplus K_{-2} \oplus K_{-1} \oplus CK_0)$. If this event and the event discussed in Section 4.2 occur at the same time, then the subkeys that exist in Equations 2, 3 and 4 will cancel and the subkeys in the first four rounds will have no effect on the intermediate words X_5, X_6 and X_7.

Similarly, if the subkeys in the first four rounds are identical, then the subkeys in rounds four (K_4) and five (K_5) have the following form:

$$K_4 = \hat{K} \oplus T'(\hat{K} \oplus \hat{K} \oplus \hat{K} \oplus CK_4) = \hat{K} \oplus T'(\hat{K} \oplus CK_4)$$

Suppose that $K_4 = \hat{K}$, which implies that $K_4 = \hat{K} = \hat{K} \oplus T'(\hat{K} \oplus CK_4)$ and $T'(\hat{K} \oplus CK_4) = 0$. Since CK_4 is a known fixed round constant, only one value of $\hat{K}$ can satisfy this equation, that is $\hat{K} = CK_4 \oplus \texttt{71717171} = \texttt{1060FF4}$. Therefore, for all 2^{32} master keys that have the form $K_i = \hat{K}$ for $i \in \{0, 1, \ldots, 4\}$, only one master key satisfies the relationship $K_3 = K_4$. The remaining $2^{32} - 1$ master keys have the relationship $K_3 \neq K_4$. Stated differently, if we are given a sequence of subkeys containing five identical words $K_i = \hat{K}$ for $i \in \{0, 1, \ldots, 4\}$, and $K_i \neq$ `1060FF4`, then we know that the subkeys are not the first four subkeys derived from the SMS4 key scheduling algorithm. These kinds of relationships can be further investigated beyond the first four rounds by taking into consideration the relationship between the round constants. The algorithm to derive these constants is already given in Section 2.2. In a key recovery attack, if the attacker knows the relationship of the words in the master key beforehand, then guesses that are impossible can be skipped. This reduces the key space that the attacker needs to guess.

The previous single-key discussion may be extended to the related-key model. Related-key attacks [1,11] allow the attacker to choose the relationship between two different master keys but not the actual value of the keys. The relationship is chosen such that the round subkeys of the first master key are related in some way to the round subkeys of the second master key. Then, several (known or chosen) plaintexts are encrypted using these related master keys to obtain the corresponding ciphertexts. The ciphertexts are then used to recover both master keys. This is an area for further investigation.

5 A Differential Attack on Modified SMS4

This section presents a differential attack [2] on a modified variant of SMS4, created by replacing the linear transformation L in the encryption algorithm with L'. This basically means that we are attacking the key scheduling algorithm, if it was used for encryption. We demonstrate that a differential attack is possible on a 27-round version of this variant.

5.1 23-Round Characteristic

We use a 5-round self-iterating differential characteristic based on previous differential attacks on SMS4 [10,18,19]. The characteristics used in these attacks have six active S-boxes: three in the fourth round and three in the fifth. Based on the entries of the input-output pattern distribution of L' given in Table 2, we know that there exist a number of differential paths where only two S-boxes are active in one round. An example of such a path is the entry $W[3][3]$.

Let $\alpha = (\alpha_0, \alpha_1, \alpha_2, \alpha_3)$ denote a 32-bit difference formed from the concatenation of four 8-bit differences α_i. The 5-round self-iterating characteristic satisfies $0 \xrightarrow{T} 0$ in the first, second and third rounds; and $\alpha \xrightarrow{T} \alpha$ in the fourth and fifth rounds. This characteristic is given as follows: $(\alpha, \alpha, \alpha, 0) \rightarrow (\alpha, \alpha, 0, \alpha) \rightarrow (\alpha, 0, \alpha, \alpha) \rightarrow (0, \alpha, \alpha, \alpha) \rightarrow (\alpha, \alpha, \alpha, \alpha) \rightarrow (\alpha, \alpha, \alpha, 0)$.

By exhaustive search, we found six values of α that satisfy the above 5-round self-iterating characteristic such that only two bytes of α are nonzero (i.e. two bytes are active). The values are `0000900C`, `00C900C9`, `00900C00`, `0C000090`, `900C0000` and `C900C900`. The probability that $\alpha \xrightarrow{T} \alpha$ for each of these values is 2^{-14}. The probability for the 5-round self-iterating characteristics is therefore $(2^{-14})^2 = 2^{-28}$. This characteristic can be concatenated four and a half times to produce a 23-round differential characteristic with total probability $(2^{-28})^4 = 2^{-112}$ given below.

$$(\alpha, \alpha, \alpha, 0) \xrightarrow{5 \text{ Rounds}} (\alpha, \alpha, \alpha, 0) \xrightarrow{5 \text{ Rounds}} (\alpha, \alpha, \alpha, 0) \xrightarrow{5 \text{ Rounds}}$$
$$(\alpha, \alpha, \alpha, 0) \xrightarrow{5 \text{ Rounds}} (\alpha, \alpha, \alpha, 0) \xrightarrow{3 \text{ Rounds}} (0, \alpha, \alpha, \alpha)$$

In comparison, the best 5-round differential characteristic on the original SMS4 has probability 2^{-38} and can only be concatenated up to three and a half times (to construct a 18-round differential characteristic) with total probability 2^{-114} [18].

5.2 27-Round Key Recovery Attack

The previous 23-round differential characteristic can be used in a 27-round key recovery attack on the modified variant of SMS4. Since the attack is heavily based on previous differential attacks [10,18,19], we only briefly describe the attack.

Choose $\alpha = (\texttt{00,00,90,0C})$ and let Λ be the set of all output differences of T' where only 2 S-boxes are active. For each S-box, there is only 127 possible output differences. Therefore, the set contains $127 \cdot 2 \approx 2^8$ possible values.

Let P and P^* denote a plaintext pair and let C and C^* denote the corresponding ciphertext pair after 27 rounds, where $P = (X_0, X_1, X_2, X_3)$, $P^* = (X_0^*, X_1^*, X_2^*, X_3^*)$, $C = (X_{27}, X_{28}, X_{29}, X_{30})$ and $C^* = (X_{27}^*, X_{28}^*, X_{29}^*, X_{30}^*)$. The attack proceeds as follows.

1. Generate $m \cdot (2^{16})^3 = m \cdot 2^{48}$ plaintext blocks where bytes 2, 3, 6, 7, 10 and 11 are set to all possible values whereas the remaining bytes are fixed. These propose $m \cdot 2^{48}/2 = m \cdot 2^{47}$ plaintext pairs (P, P^*) having the difference $(\alpha, \alpha, \alpha, 0)$.
2. Encrypt the plaintexts using 27 rounds of the modified SMS4.
3. Filter the ciphertexts so that we only choose $(X_{27} \oplus X_{27}^*) \in \Lambda$. This filtering causes about $m \cdot 2^{47} \cdot 2^{-8} = m \cdot 2^{39}$ pairs to remain.
4. Let $\gamma_{i,j} = s(X_{i,j} \oplus X_{i+1,j} \oplus X_{i+2,j} \oplus K_{i-1,j}) \oplus s(X_{i,j}^* \oplus X_{i+1,j}^* \oplus X_{i+2,j}^* \oplus K_{i-1,j})$ and $\delta_{i,j} = L'(X_{i+3,j} \oplus X_{i+3,j}^* \oplus \alpha_j)$.
5. For each round $i = 27, 26, 25$, do the following
   (a) For each byte $j = 0, 1, 2, 3$, do the following
      i. For each byte guess $K_{i-1,j} = \texttt{0}, \texttt{1}, \ldots, \texttt{FF}$, do the following
         A. Calculate $\gamma_{i,j}$ and $\delta_{i,j}$.
         B. If $\gamma_{i,j} = \delta_{i,j}$, then store $K_{i-1,j}$ as a possible correct candidate key byte.
      ii. After all values have been guessed for this byte, wrong pairs are expected to be discarded by a factor of 2^{-8}.
6. After Step (5), we have guessed 12 bytes of key material and about $m \cdot 2^{39} \cdot (2^{-8})^{12} = m \cdot 2^{-57}$ pairs are expected to remain.
7. For round $i = 24$, do the following
   (a) For each byte guess $K_{23,0} = \texttt{0}, \texttt{1}, \ldots, \texttt{FF}$, calculate $\gamma_{24,0}$ and $\delta_{24,0}$. If $\gamma_{24,0} = \delta_{24,0}$, then store $K_{23,0}$ as a possible correct candidate key byte.
   (b) After all values have been guessed for this byte, wrong pairs are expected to be discarded by a factor of 2^{-8}.
8. After Step (7), about $m \cdot 2^{-57} \cdot (2^{-8}) = m \cdot 2^{-65}$ pairs are expected to remain. If $m = 2^{68}$, then for a wrong key guess, the expected number of remaining ciphertext pairs is approximately $2^{68} \cdot 2^{-65} = 2^3 = 8$. However, for a right key guess, the expected number of remaining ciphertext pairs is approximately $2^{68} \cdot 2^{48} \cdot 2^{-112} = 2^4 = 16$.
9. If the guesses for $K_{23,0}$, K_{24}, K_{25} and K_{26} suggest more than 16 remaining ciphertext pairs, then the guesses are candidates for correct subkeys.

The data complexity of this 27-round attack is $2^{68} \cdot 2^{48} = 2^{116}$ chosen plaintexts. The time complexity of the attack is dominated by Steps (5) and (7). At the beginning of Step (5), there are about $2^{68} \cdot 2^{39}$ pairs of texts. We guess 12 bytes of key material and for each guess, wrong pairs are discarded by a factor of 2^{-8}. At the beginning of Step (7), there are roughly $2^{68} \cdot 2^{-57}$ pairs of texts and we only guess one byte of key material. Adding these two complexities together, we obtained the time complexity of approximately $(\sum_{k=0}^{11} 2^8 \cdot 2^{68} \cdot 2^{39} \cdot 2^{-8k}) + 2^8 \cdot 2^{68} \cdot 2^{-57} \approx 2^{115}$ encryptions. In contrast, the best existing cryptanalysis on the original SMS4 is a differential attack on 22 rounds with a data complexity of 2^{117} chosen plaintexts and time complexity of $2^{112.3}$ 22-round encryptions [18].

5.3 Comments on the Security of SMS4

As mentioned at the beginning of Section 5, the attack described above is the same as attacking the key scheduling algorithm, as if it was used for encryption. We use the original components of the SMS4 and did not modify the function of these components. The key scheduling algorithm might therefore be exploited in related-key differential attacks.

In the light of our discussion in Section 4.3, there is a small possibility that the first four rounds of SMS4 is deprived of non-linearity. Under these conditions, the number of effective rounds for SMS4 is theoretically reduced by four, from 32 to 28. In this section, we have demonstrated an attack against 27 rounds of a slightly modified variant of SMS4. This is only one round short of the effective 28 rounds. Note that the four-round linearity event discussed in Section 4.3 refers to the event in which the function T was used in the encryption, instead of T', as is the case here. However, if T' was used in the encryption, the probability of this event to occur for T', in the general case, is the same as if T was used in the encryption. This is because the number of input words in the set Θ_T is the same as the set $\Theta_{T'}$.

Recall that the best attack on the original SMS4 is on 22-rounds [18], which is six rounds short of the effective 28 rounds. However, note that the security margin is reduced from 32 to 28 rounds only if the linearity in the first four rounds can be detected and utilized in an attack. A method to detect this remains an open problem.

6 Summary and Conclusion

This paper presents several new observations on both the encryption and the key scheduling algorithms of the SMS4 block cipher. We have shown the existence of fixed points and of simple linear relationships between the bits of the input and output words for each component of the round functions for some input words. Furthermore, we show that the branch number of the linear transformation in the key scheduling algorithm is less than optimal.

The major security implication of these observations is that the round function is not always non-linear. Due to this linearity, for some combinations of plaintext block and master key, the number of effective rounds of SMS4 is theoretically reduced by four, from 32 to 28. We also briefly explored the susceptibility of SMS4 against algebraic and advanced variants of the slide attacks.

Finally, we demonstrated that if the linear transformation L of the encryption algorithm is replaced with the linear transformation L' of the key scheduling algorithm, then this variant of SMS4 is weaker than the original SMS4 with regard to differential cryptanalysis. We show this by attacking four more rounds than the best existing differential attack on SMS4. This is possible due to the sub-optimal branch number of L'. This property of L' might be an indication of further weakness that can be exploited in an attack. We strongly believe that this variant is also weaker than SMS4 against other differential-type attacks.

Given the number of expected fixed points, it is unlikely that the components in the round functions are generated randomly, that is, they were selected specifically. However, the criteria for selecting the components are not known. The findings made in this paper raise serious questions on the security provided by SMS4, and might provide clues on the existence of a flaw in the design of the cipher.

References

1. Biham, E.: New Types of Cryptanalytic Attacks Using Related Keys. In: Helleseth, T. (ed.) EUROCRYPT '93. LNCS, vol. 765, pp. 398–409. Springer, Heidelberg (1994)
2. Biham, E., Shamir, A.: Differential Cryptanalysis of the Data Encryption Standard. Springer-Verlag (1993)
3. Biryukov, A., Wagner, D.: Slide Attacks. In: Knudsen, L. (ed.) FSE '99. LNCS, vol. 1636, pp. 245–259. Springer, Heidelberg (1999)
4. Biryukov, A., Wagner, D.: Advanced Slide Attacks. In: Preneel, B. (ed.) EUROCRYPT 2000. LNCS, vol. 1807, pp. 589–606. Springer, Heidelberg (2000)
5. Courtois, N.T., Pieprzyk, J.: Cryptanalysis of Block Ciphers with Overdefined Systems of Equations. In: Zheng, Y. (ed.) ASIACRYPT 2002. LNCS, vol. 2501, pp. 267–287. Springer, Heidelberg (2002)
6. Daemen, J., Rijmen, V.: The Design of Rijndael, AES – The Advanced Encryption Standard. Springer-Verlag, (2002)
7. Diffie, W., Ledin, G.: SMS4 Encryption Algorithm for Wireless Networks. In: Cryptology ePrint Archive, Report 2008/329 (2008)
8. Etrog, J., Robshaw, M.J.B.: Improved Cryptanalysis of Reduced-Round SMS4. In: Avanzi, R., Keliher, L., Sica, F. (eds.) SAC 2008. LNCS, vol. 5381, pp. 51–65. Springer, Heidelberg (2009)
9. Grinstead, C.M., Snell, J.L.: Introduction to Probability. 2nd Revised Ed. American Mathematical Society, (1997)
10. Kim, T., Kim, J., Hong, S., Sung, J.: Linear and Differential Cryptanalysis of Reduced SMS4 Block Cipher. In: Cryptology ePrint Archive, Report 2008/281 (2008)
11. Knudsen, L.: Cryptanalysis of LOKI91. In: Seberry, J., Zheng, Y. (eds.) ASIACRYPT '92. LNCS, vol. 718, pp. 22–35. Springer, Heidelberg (1993)

12. 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)
13. 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)
14. Office of State Commercial Cryptography Administration, P.R. China: The SMS4 Block Cipher (in Chinese). (2006), http://www.oscca.gov.cn/UpFile/200621016423197990.pdf
15. Riordan, J.: An Introduction to Combinatorial Analysis. Princeton University Press, (1980)
16. Schneier, B., Kelsey, J.: Unbalanced Feistel Networks and Block Cipher Design. In: Gollmann, D. (ed.) FSE 1996. LNCS, vol. 1039, pp. 121–144. Springer, Heidelberg (1996)
17. 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)
18. Zhang, W., Wu, W., Feng, D., Su, B.: Some New Observations on the SMS4 Block Cipher in the Chinese WAPI Standard. In: Bao, F., Li, H., Wang, G. (eds.) ISPEC 2009. LNCS, vol. 5451, pp. 324–335. Springer, Heidelberg (2009)
19. Zhang, L., Zhang, W., Wu, W.: Cryptanalysis of Reduced-Round SMS4 Block Cipher. In: Mu, X., Susilo, W., Seberry, J. (eds.) ACISP 2008. LNCS, vol. 5107, pp. 216–229. Springer, Heidelberg (2008)

A Appendix

Table 3. Values of X_i (in the set Θ_S) and j such that $S(X_i) = X_i \lll j$.

X_i	j	X_i	j	X_i	j
0A0A0A0A	1, 9, 17, 25	21210A0A	1	ABB4ABDE	16
0A0A0A21	1	21210A21	1	ABDEABB4	16
0A0A210A	1	2121210A	1	B4ABDEAB	16
0A0A2121	1	21212121	1, 9, 17, 25	B4B4DEDE	16
0A210A0A	1	245C245C	2, 18	B4DEB4DE	8, 24
0A210A21	1, 17	245C2626	2	B4DEDEB4	16
0A21210A	1	26245C26	2	D056D056	5, 21
0A212121	1	2626245C	2	DEABB4AB	16
0B0B0B0B	6, 14, 22, 30	26262626	2, 10, 18, 26	DEB4B4DE	16
210A0A0A	1	56D056D0	5, 21	DEB4DEB4	8, 24
210A0A21	1	5C245C24	2, 18	DEDEB4B4	16
210A210A	1, 17	5C262624	2	E7E7E7E7	4, 12, 20, 28
210A2121	1	ABABABAB	0, 8, 16, 24	FAFAFAFA	5, 13, 21, 29

Table 4. Values of X_i (in the set Θ_T) and j such that $T(X_i) = X_i \lll j$.

X_i	j	X_i	j	X_i	j
02740274	2, 18	4F13E4B4	2	BB06C4A3	26
039A039A	1, 17	58434DF7	26	BE6CBE6C	15, 31
06C4A3BB	26	5CDE9B16	14	C4A3BB06	26
0A0A0A0A	3, 11, 19, 27	62D367B9	0	C6AD3AE2	0
0B0B0B0B	0, 8, 16, 24	67B962D3	0	C7E7C7E7	13, 29
1079D3A1	31	6CBE6CBE	15, 31	D367B962	0
13E4B44F	2	74027402	2, 18	D3A11079	31
165CDE9B	14	79D3A110	31	DE9B165C	14
16AF4D4B	15	973E973E	0, 16	E0E1F7E3	9
1A2A1A2A	1, 17	9A039A03	1, 17	E1F7E3E0	9
21212121	3, 11, 19, 27	9B165CDE	14	E2C6AD3A	0
22E59CB6	31	9CB622E5	31	E3E0E1F7	9
26262626	4, 12, 20, 28	A11079D3	31	E4B44F13	2
2A1A2A1A	1, 17	A3BB06C4	26	E59CB622	31
3AE2C6AD	0	ABABABAB	2, 10, 18, 26	E7C7E7C7	13, 29
3E973E97	0, 16	AD3AE2C6	0	E7E7E7E7	6, 14, 22, 30
434DF758	26	AF4D4B16	15	F758434D	26
4B16AF4D	15	B44F13E4	2	F7E3E0E1	9
4D4B16AF	15	B622E59C	31	FAFAFAFA	7, 15, 23, 31
4DF75843	26	B962D367	0		

Table 5. Values of X_i (in the set $\Theta_{T'}$) and j such that $T'(X_i) = X_i \lll j$.

X_i	j	X_i	j	X_i	j
02020202	4, 12, 20, 28	5228B69C	6	A66BA66B	10, 26
06C206C2	1, 17	52505250	4, 20	AAA027D5	23
087B087B	4, 20	5522DB49	27	B0B0B0B0	3, 11, 19, 27
10B78569	2	58F758F7	8, 24	B69C5228	6
12121212	6, 14, 22, 30	5A5A5A5A	2, 10, 18, 26	B7856910	2
12161216	7, 23	61F161F1	14, 30	B8B8B8B8	2, 10, 18, 26
16121612	7, 23	64C164C1	9, 25	BAC74FDD	27
1B341B34	5, 21	6910B785	2	C164C164	9, 25
1D411D41	2, 18	6BA66BA6	10, 26	C206C206	1, 17
22DB4955	27	74747474	3, 11, 19, 27	C74FDDBA	27
25A498A2	1	7B087B08	4, 20	CBA1CBA1	14, 30
27D5AAA0	23	856910B7	2	D5AAA027	23
28B69C52	6	94949494	6, 14, 22, 30	D69AD69A	12, 28
32323232	3, 11, 19, 27	98A225A4	1	DB495522	27
341B341B	5, 21	9AD69AD6	12, 28	DDBAC74F	27
411D411D	2, 18	9C5228B6	6	DFDFDFDF	3, 11, 19, 27
495522DB	27	A027D5AA	23	E1E1E1E1	7, 15, 23, 31
4F4F4F4F	5, 13, 21, 29	A1CBA1CB	14, 30	F161F161	14, 30
4FDDBAC7	27	A225A498	1	F758F758	8, 24
50525052	4, 20	A498A225	1		