Multiple Linear Cryptanalysis of Reduced-Round SMS4 Block Cipher

Zhiqiang Liu , Dawu Gu, Jing Zhang

Department of Computer Science and Engineering,
Shanghai Jiao Tong University, Shanghai 200240, P.R. China
ilu_zq@sjtu.edu.cn

Abstract. SMS4 is a 32-round unbalanced Feistel block cipher with its block size and key size being 128 bits. As a fundamental block cipher used in the WAPI standard, the Chinese national standard for WLAN, it has been widely implemented in Chinese WLAN industry. In this paper, we present a modified branch-and-bound algorithm which can be used for searching multiple linear characteristics for SMS4-like unbalanced Feistel block ciphers. Furthermore, we find a series of 5-round iterative linear characteristics of SMS4 when applying the modified algorithm in SMS4. Then based on each 5-round iterative linear characteristic mentioned above, an 18-round linear characteristic of SMS4 can be constructed, thus leading to a list of 18-round linear characteristics of SMS4. According to the framework of Biryukov *et al.* from Crpto 2004, a key recovery attack can be mounted on 22-round SMS4 by utilizing the above multiple linear characteristics. As a matter of fact, our result has much lower data complexity than the previously best known cryptanalytic result on 22-round SMS4, which is also the previously best known result on SMS4.

Key words: SMS4, Block cipher, Linear characteristic, Multiple linear cryptanalysis, Branch-and-bound

1 Introduction

The block cipher SMS4 [1], released by Chinese government in 2006, is an underlying block cipher used in WLAN Authentication and Privacy Infrastructure (WAPI) standard, the Chinese national standard for WLAN. Although the WAPI standard hasn't been approved as the security amendment to the ISO/IEC 8802-11 WLAN standard, it is still officially mandated for Chinese WLAN industry, thus SMS4 has been widely implemented in China.

SMS4 has an unbalanced Feistel network structure with 32 rounds, a block size of 128 bits as well as a key size of 128 bits. Up to now, several attacks have been presented on reduced-round SMS4. For instance, a differential fault analysis of SMS4 has been given in [2], an integral attack on 13-round SMS4 has been proposed in [3], an analysis of the structure of SMS4 from a viewpoint of algebra has been provided in [4], a rectangle attack on 14-round SMS4 and an impossible differential attack on 16-round SMS4 have been devised in [5], a rectangle attack on 16-round SMS4 and a differential attack on 21-round SMS4 have been presented in [6], a boomerang attack and a rectangle attack on 18-round SMS4, a linear attack and a differential attack on 22-round SMS4 have been introduced in [7], a more comprehensive analysis for the results given in [5] has been done in [8], and an improved differential attack on 22-round SMS4 has been demonstrated in [9] . In this paper, we firstly propose a modified branch-and-bound algorithm which can be used for searching multiple linear characteristics for SMS4-like unbalanced Feistel block ciphers. Moreover, a series of 5-round iterative linear characteristics of SMS4 have been obtained by applying the modified algorithm in SMS4. Then for each 5-round iterative linear characteristic mentioned above, an 18-round linear characteristic of SMS4 can be constructed, resulting in a list of 18-round linear characteristics of SMS4. Based on the framework given in [10], a key recovery attack can be mounted on 22-round SMS4 by utilizing the above multiple linear characteristics. Compared with the previously best known cryptanalytic result on 22-round SMS4, which is also the previously best known result on SMS4, our result has much lower data complexity.

The remainder of this paper is organized as follows. Section 2 introduces the notations used throughout this paper, gives a brief description of SMS4 as well as the method of multiple linear cryptanalysis. Section 3 presents our modified branch-and-bound algorithm which can be used for searching multiple linear characteristics for SMS4-like unbalanced Feistel block ciphers. Section 4 provides a series of 5-round iterative linear characteristics of SMS4 obtained by the modified algorithm. Section 5 presents our multiple linear cryptanalysis on 22-round SMS4. Finally, Section 6 summarizes the paper.

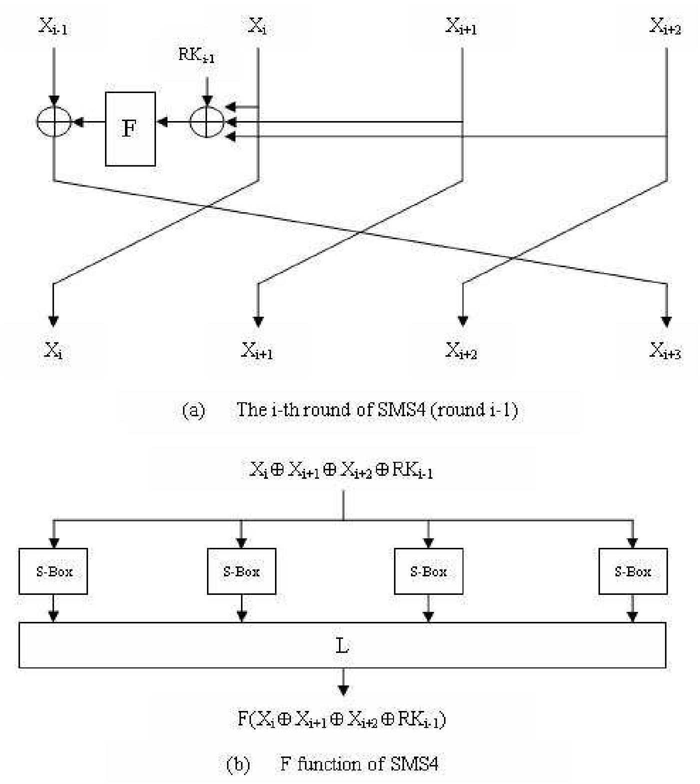
2 Preliminaries

2.1 Notations

The following notations are used throughout the paper.

- $\oplus$ denotes bitwise exclusive OR (XOR).
- $\cdot$ denotes bitwise inner product.
- $||$ denotes concatenation operation.
- $|x|$ denotes the absolute value of a real number x.
- Z_2^8 denotes the set $\{0,1\}^8$.
- Z_2^{32} denotes the set $\{0,1\}^{32}$.
- 0x denotes the hexadecimal notation.

2.2 A Brief Description of SMS4


X_{i-1}
X_i
X_{i+1}
X_{i+2}
RK_{i+1}
F
X_i
X_{i+1}
X_{i+2}
X_{i+3}
$X_i \oplus X_{i+1} \oplus X_{i+2} \oplus RK_{i-1}$
S-Box
S-Box
S-Box
S-Box
L
$F(X_i \oplus X_{i+1} \oplus X_{i+2} \oplus RK_{i-1})$

(a) The i-th round of SMS4 (round i-1)

(b) F function of SMS4

Fig. 1. Schematic Description of the SMS4 Round Function

SMS4 is an unbalanced Feistel block cipher with 32 rounds, a block size of 128 bits as well as a key size of 128 bits. Let $(P_0, P_1, P_2, P_3) \in (Z_2^{32})^4$, and $(C_0, C_1, C_2, C_3) \in (Z_2^{32})^4$ denote the plaintext P and the ciphertext C respectively. Let $(X_i, X_{i+1}, X_{i+2}, X_{i+3}) \in (Z_2^{32})^4$, and $(X_{i+1}, X_{i+2}, X_{i+3}, X_{i+4}) \in (Z_2^{32})^4$ denote the input and the output of round i respectively, where $i = 0, 1, \ldots, 31$. Note that the first round is referred to as round 0, the second round is referred to as round 1, and so on. Then the cipher can be described as follows:
(1). $(X_0, X_1, X_2, X_3) \leftarrow (P_0, P_1, P_2, P_3)$,
(2). $X_{i+4} \leftarrow X_i \bigoplus F(X_{i+1} \bigoplus X_{i+2} \bigoplus X_{i+3} \bigoplus RK_i)$, for $i = 0, 1, \ldots, 31$,
(3). $(C_0, C_1, C_2, C_3) \leftarrow R(X_{32}, X_{33}, X_{34}, X_{35}) = (X_{35}, X_{34}, X_{33}, X_{32})$,
where $RK_i \in Z_2^{32}$ denotes the round subkey for round i, the function F is composed of a non-linear confusion function S which applies a same 8×8 bijective S-Box four times in parallel and a linear diffusion function L, and R denotes a switch transformation.

Fig. 1 gives a schematic description of the SMS4 round function. Please refer to [1] for detailed information about the S-Box, the linear transformation L and the key schedule algorithm. As for the switch transformation R, we will omit it in the following cryptanalysis of SMS4 since it has no impact on our attack.

2.3 Multiple Linear Cryptanalysis

Here we just review the method of multiple linear cryptanalysis in this subsection. For more details, please refer to [10].

Linear cryptanalysis [11], proposed by Matsui in 1993, is one of the most powerful known plaintext attacks against modern block ciphers. It analyzes a block cipher E by investigating a correlation between the inputs and outputs of E and then obtains a linear approximation (also called linear characteristic and denoted as $\Gamma_P \rightarrow \Gamma_C$) of E with following type:

$$\Gamma_P \cdot P \oplus \Gamma_C \cdot C = \Gamma_K \cdot K, \tag{1}$$

where P, C and K denote plaintext, ciphertext, and secret key respectively, Γ_P, Γ_C and Γ_K stand for the mask of plaintext P, ciphertext C, and secret key respectively.

If equation (1) holds with probability $p \neq 1/2$, we call it an effective linear approximation of the block cipher E, and the linear approximation can be used to distinguish E from a random permutation since equation (1) holds with probability 1/2 for a random permutation. Let $\varepsilon = p - 1/2$ be the bias of a linear approximation on E, then the greater $|\varepsilon|$ is, the more effective the corresponding linear approximation will be.

Based on this technique, Kaliski *et al*. [12] presented the idea of generalizing linear cryptanalysis using multiple linear approximations in 1994. However, a strict constraint exists within their method as it requires to use approximations deriving the same parity bit of the secret key, which greatly restricted the number and the quality of the approximations available. As a result, an approach removing the constraint was proposed by Biryukov *et al*. in 2004 [10], which results in decreasing of data complexity compared with the original linear cryptanalysis.

Suppose that one has access to m approximations on E of the following form:

$$\Gamma_P^i \cdot P \oplus \Gamma_C^i \cdot C = \Gamma_K^i \cdot K \quad (1 \leq i \leq m). \tag{2}$$

Let ε_i, $c_i = 2\varepsilon_i$ be the bias and the imbalance of the i-th linear approximation respectively. According to [10], the multiple linear cryptanalysis requires a number of plaintext-ciphertext pairs inversely proportional to the capacity of the system (linear equations used by adversary as in equation(2)) that is defined as:

$$\overline{C}^2 = \sum_{i=1}^{m} c_i^2 = 4 \times \sum_{i=1}^{m} \varepsilon_i^2. \tag{3}$$

Therefore, one can reduce the number of necessary plaintext-ciphertext pairs to perform a successful key recovery attack by increasing the capacity when using the multiple linear cryptanalytic tool.

3 Modified Branch-and-bound Algorithm

Actually, the first step in a multiple linear cryptanalysis is to find linear approximations with biases as high as possible. In 1994, Matsui [13] proposed a branch-and-bound algorithm making it possible to effectively find the best linear approximation of DES. The algorithm works by induction: knowing the maximal bias on $(n-1)$-round DES, it manages to find the maximal bias on n-round DES as well as the corresponding input and output masks. Based on this idea, we give a modified branch-and-bound algorithm which can be used for searching multiple linear characteristics for SMS4-like unbalanced Feistel block cipher as below.

Let E be an n-round SMS4-like unbalanced Feistel block cipher. Let the bias of a linear approximation on the i-th round F-function of E be defined as:

$$(\mathit{\Gamma} I_i, \mathit{\Gamma} O_i) = \delta_i = \Pr\{\mathit{\Gamma} I_i \cdot I_i \oplus \mathit{\Gamma} O_i \cdot O_i = 0\} - 1/2, \tag{4}$$

where I_i and O_i denote the input and output of the i-th round F-function of E, and $\mathit{\Gamma} I_i$, $\mathit{\Gamma} O_i$ represent their masks respectively. We first note that if there are n linear approximations on each round F-function of E respectively (denoted as $\mathit{\Gamma} I_i \rightarrow \mathit{\Gamma} O_i, 1 \leq i \leq n$) satisfying

$$\mathit{\Gamma} O_i = \mathit{\Gamma} I_{i-1} \oplus \mathit{\Gamma} I_{i-2} \oplus \mathit{\Gamma} I_{i-3} \oplus \mathit{\Gamma} O_{i-4} \quad (5 \leq i \leq n), \tag{5}$$

the above n one-round linear approximations can be concatenated sequentially to form a linear approximation on the whole cipher E. According to the piling up lemma in [11], the total bias ε_{tot} of the n-round linear approximation is given by:

$$\varepsilon_{tot} = [\delta_1, \delta_2, \ldots, \delta_n] = 2^{n-1} \prod_{i=1}^{n} \delta_i, \tag{6}$$

and the best linear approximation on E is then defined as:

$$B_n = \max_{\substack{\mathit{\Gamma} O_i = \mathit{\Gamma} I_{i-1} \oplus \mathit{\Gamma} I_{i-2} \oplus \mathit{\Gamma} I_{i-3} \oplus \mathit{\Gamma} O_{i-4} \\ (5 \leq i \leq n)}} |[(\mathit{\Gamma} I_1, \mathit{\Gamma} O_1), (\mathit{\Gamma} I_2, \mathit{\Gamma} O_2), \ldots, (\mathit{\Gamma} I_n, \mathit{\Gamma} O_n)]|. \tag{7}$$

Moreover, let $Q_n = (q_n^1, \ldots, q_n^m)$ denote the queue sorted in the order of decreasing bias, where $q_n^i = (pattern_n^i, B_n^i)$ stores both the linear mask pattern for all the intermediate rounds and the bias of the i-th linear approximation on E. Let $\overline{B_n^m}$ be the initial estimation of B_n^m. Then the framework of our modified branch-and-bound algorithm for searching m best linear characteristics of E can be designed by the following procedures including essentially recursive calls:

```
Procedure Round-1:
    Begin the program
    For each candidate for $\mathit{\Gamma} I_1$ and $\mathit{\Gamma} O_1$, do as follows:
        • $\delta_1 \Leftarrow (\mathit{\Gamma} I_1, \mathit{\Gamma} O_1)$.
        • If $|[\delta_1, B_{n-1}]| \geq \overline{B_n^m}$, then call Procedure Round-2.
    End the program

Procedure Round-2:
    For each candidate for $\mathit{\Gamma} I_2$ and $\mathit{\Gamma} O_2$, do as follows:
        • $\delta_2 \Leftarrow (\mathit{\Gamma} I_2, \mathit{\Gamma} O_2)$.
        • If $|[\delta_1, \delta_2, B_{n-2}]| \geq \overline{B_n^m}$, then call Procedure Round-3.
    Return to the upper procedure.

Procedure Round-3:
    For each candidate for $\mathit{\Gamma} I_3$ and $\mathit{\Gamma} O_3$, do as follows:
        • $\delta_3 \Leftarrow (\mathit{\Gamma} I_3, \mathit{\Gamma} O_3)$.
        • If $|[\delta_1, \delta_2, \delta_3, B_{n-3}]| \geq \overline{B_n^m}$, then call Procedure Round-4.
    Return to the upper procedure.

Procedure Round-4:
    For each candidate for $\Gamma I_4$ and $\Gamma O_4$, do as follows:
        • $\delta_4 \Leftarrow (\Gamma I_4, \Gamma O_4)$.
        • If $|[\delta_1, \delta_2, \delta_3, \delta_4, B_{n-4}]| \geq \overline{B_n^m}$, then call Procedure Round-5.
    Return to the upper procedure.

Procedure Round-i ($5 \leq i \leq n-1$):
    For each candidate for $\Gamma I_i$, do as follows:
        • $\Gamma O_i \Leftarrow \Gamma I_{i-1} \oplus \Gamma I_{i-2} \oplus \Gamma I_{i-3} \oplus \Gamma O_{i-4}$.
        • $\delta_i \Leftarrow (\Gamma I_i, \Gamma O_i)$.
        • If $|[\delta_1, \delta_2, \ldots, \delta_i, B_{n-i}]| \geq \overline{B_n^m}$, then call Procedure Round-(i+1).
    Return to the upper procedure.

Procedure Round-n:
    For each candidate for $\Gamma I_n$, do as follows:
        • $\Gamma O_n \Leftarrow \Gamma I_{n-1} \oplus \Gamma I_{n-2} \oplus \Gamma I_{n-3} \oplus \Gamma O_{n-4}$.
        • $\delta_n \Leftarrow (\Gamma I_n, \Gamma O_n)$.
        • If $|[\delta_1, \delta_2, \ldots, \delta_n]| \geq \overline{B_n^m}$, then insert $[\delta_1, \delta_2, \ldots, \delta_n]$ and corresponding linear mask pattern into $Q_n$,
          and do $\overline{B_n^m} \Leftarrow \min_{1 \leq j \leq m} \{B_n^j\}$.
    Return to the upper procedure.
```

4 5-Round Iterative Linear Characteristics of SMS4

Let E be an n-round SMS4-like unbalanced Feistel block cipher. As mentioned above, if there are n linear approximations $\Gamma I_i \rightarrow \Gamma O_i$ $(1 \leq i \leq n)$ on each round F-function of E respectively satisfying

$$\Gamma O_i = \Gamma I_{i-1} \oplus \Gamma I_{i-2} \oplus \Gamma I_{i-3} \oplus \Gamma O_{i-4} \quad (5 \leq i \leq n), \tag{8}$$

these n one-round linear approximations can be concatenated sequentially to formulate a linear approximation on the whole cipher E. Thus the linear approximation on E can be explicitly expressed as

$$\begin{aligned}
&\Gamma O_1 \cdot P_0 \oplus \Gamma I_1 \cdot P_1 \oplus \Gamma O_2 \cdot P_1 \oplus \Gamma I_1 \cdot P_2 \oplus \Gamma I_2 \cdot P_2 \oplus \Gamma O_3 \cdot P_2 \oplus \\
&\Gamma I_1 \cdot P_3 \oplus \Gamma I_2 \cdot P_3 \oplus \Gamma I_3 \cdot P_3 \oplus \Gamma O_4 \cdot P_3 \oplus \Gamma O_n \cdot C_3 \oplus \Gamma I_n \cdot C_2 \oplus \Gamma O_{n-1} \cdot C_2 \oplus \\
&\Gamma I_n \cdot C_1 \oplus \Gamma I_{n-1} \cdot C_1 \oplus \Gamma O_{n-2} \cdot C_1 \oplus \Gamma I_n \cdot C_0 \oplus \Gamma I_{n-1} \cdot C_0 \oplus \Gamma I_{n-2} \cdot C_0 \oplus \Gamma O_{n-3} \cdot C_0 \\
= &\Gamma I_1 \cdot RK_0 \oplus \Gamma I_2 \cdot RK_1 \oplus \ldots \oplus \Gamma I_n \cdot RK_{n-1}.
\end{aligned} \tag{9}$$

Specifically, let n be 5 in equation (9). Let ΓI_1, ΓO_1, ΓI_2, ΓO_2, ΓI_3 and ΓO_3 be 0, and let ΓI_4, ΓO_4, ΓI_5, ΓO_5 be any mask $\Gamma_m \in Z_2^{32}$ such that the bias of the one-round linear characteristic $\Gamma_m \rightarrow \Gamma_m$ (denoted as (Γ_m, Γ_m)) is not equal to 0. Then we can obtain a series of 5-round iterative linear characteristics of SMS4 as follows:

$$\Gamma_m \cdot X_3 \oplus \Gamma_m \cdot X_8 = \Gamma_m \cdot RK_3 \oplus \Gamma_m \cdot RK_4. \tag{10}$$

The bias of the above 5-round iterative linear characteristic is $2 \times (\Gamma_m, \Gamma_m)^2$. Let $\Gamma_P' = (0, 0, 0, \Gamma_m) \in (Z_2^{32})^4$ and $\Gamma_C' = (0, 0, 0, \Gamma_m) \in (Z_2^{32})^4$ be the input mask and the output mask of the 5-round SMS4 respectively, then the above 5-round iterative linear characteristic of SMS4 can also be denoted as $\Gamma_P' \rightarrow \Gamma_C'$.

Furthermore, regarding the one-round linear characteristic $\Gamma_m \rightarrow \Gamma_m$ on F-function, suppose that the input mask Γ_m (denoted as $(m_1, m_2, m_3, m_4) \in (Z_2^8)^4$) goes to Γ_m' (denoted as $(m_1', m_2', m_3', m_4') \in (Z_2^8)^4$) through the non-linear layer S with biases α_1, α_2, α_3 and α_4 for each S-Box respectively(where α_i corresponds to the bias of the linear characteristic $m_i \rightarrow m_i'$ on SMS4 S-Box, $1 \leq i \leq 4$), and Γ_m' goes to Γ_m again through the linear layer L, then the bias (Γ_m, Γ_m) can be calculated as $2^3\alpha_1\alpha_2\alpha_3\alpha_4$. Accordingly, in order to use our modified branch-and-bound algorithm to search multiple 5-round iterative linear characteristics with bias as high as possible, we investigate the linear distribution table of the SMS4 S-Box and then find that the most effective linear characteristics of the SMS4 S-Box have the biases $\pm 2^{-4}$ and the second most effective ones have the biases $\pm 14/256$ (approximately $\pm 2^{-4.19}$). After that, we have gained 32 best 5-round iterative linear characteristics of SMS4 by applying our modified branch-and-bound algorithm. Table 1 illustrates the search results in detail.

Table 1. Γ_m/Γ'_m for the Above 5-Round Iterative Linear Characteristics of SMS4 with Best Biases

Bias of the 5-round linear characteristic	$\lvert(\Gamma_m, \Gamma_m)\rvert$	Γ_m	Γ'_m
$2^{-19.38}$	$2^{-10.19}$	(0x11, 0xff, 0xba, 0x00)	(0x84, 0xbe, 0x2f, 0x00)
$2^{-19.38}$	$2^{-10.19}$	(0x78, 0x52, 0xb3, 0x00)	(0x58, 0x2b, 0x15, 0x00)
$2^{-19.38}$	$2^{-10.19}$	(0x79, 0x05, 0xe1, 0x00)	(0x5a, 0xfb, 0xc6, 0x00)
$2^{-19.38}$	$2^{-10.19}$	(0xa1, 0xb4, 0x33, 0x00)	(0xf1, 0x02, 0x7a, 0x00)
$2^{-19.38}$	$2^{-10.19}$	(0xed, 0xca, 0x7c, 0x00)	(0x83, 0xff, 0xaa, 0x00)
$2^{-19.38}$	$2^{-10.19}$	(0xfa, 0x70, 0x99, 0x00)	(0xd2, 0x0b, 0x1d, 0x00)
$2^{-19.38}$	$2^{-10.19}$	(0x05, 0xe1, 0x00, 0x79)	(0xfb, 0xc6, 0x00, 0x5a)
$2^{-19.38}$	$2^{-10.19}$	(0x52, 0xb3, 0x00, 0x78)	(0x2b, 0x15, 0x00, 0x58)
$2^{-19.38}$	$2^{-10.19}$	(0x70, 0x99, 0x00, 0xfa)	(0x0b, 0x1d, 0x00, 0xd2)
$2^{-19.38}$	$2^{-10.19}$	(0xb4, 0x33, 0x00, 0xa1)	(0x02, 0x7a, 0x00, 0xf1)
$2^{-19.38}$	$2^{-10.19}$	(0xca, 0x7c, 0x00, 0xed)	(0xff, 0xaa, 0x00, 0x83)
$2^{-19.38}$	$2^{-10.19}$	(0xff, 0xba, 0x00, 0x11)	(0xbe, 0x2f, 0x00, 0x84)
$2^{-19.38}$	$2^{-10.19}$	(0x33, 0x00, 0xa1, 0xb4)	(0x7a, 0x00, 0xf1, 0x02)
$2^{-19.38}$	$2^{-10.19}$	(0x7c, 0x00, 0xed, 0xca)	(0xaa, 0x00, 0x83, 0xff)
$2^{-19.38}$	$2^{-10.19}$	(0x99, 0x00, 0xfa, 0x70)	(0x1d, 0x00, 0xd2, 0x0b)
$2^{-19.38}$	$2^{-10.19}$	(0xb3, 0x00, 0x78, 0x52)	(0x15, 0x00, 0x58, 0x2b)
$2^{-19.38}$	$2^{-10.19}$	(0xba, 0x00, 0x11, 0xff)	(0x2f, 0x00, 0x84, 0xbe)
$2^{-19.38}$	$2^{-10.19}$	(0xe1, 0x00, 0x79, 0x05)	(0xc6, 0x00, 0x5a, 0xfb)
$2^{-19.38}$	$2^{-10.19}$	(0x00, 0xfa, 0x70, 0x99)	(0x00, 0xd2, 0x0b, 0x1d)
$2^{-19.38}$	$2^{-10.19}$	(0x00, 0xed, 0xca, 0x7c)	(0x00, 0x83, 0xff, 0xaa)
$2^{-19.38}$	$2^{-10.19}$	(0x00, 0xa1, 0xb4, 0x33)	(0x00, 0xf1, 0x02, 0x7a)
$2^{-19.38}$	$2^{-10.19}$	(0x00, 0x79, 0x05, 0xe1)	(0x00, 0x5a, 0xfb, 0xc6)
$2^{-19.38}$	$2^{-10.19}$	(0x00, 0x78, 0x52, 0xb3)	(0x00, 0x58, 0x2b, 0x15)
$2^{-19.38}$	$2^{-10.19}$	(0x00, 0x11, 0xff, 0xba)	(0x00, 0x84, 0xbe, 0x2f)
$2^{-19.76}$	$2^{-10.38}$	(0x1d, 0xde, 0xab, 0x00)	(0xae, 0xc5, 0x71, 0x00)
$2^{-19.76}$	$2^{-10.38}$	(0x38, 0xa5, 0x45, 0x00)	(0x82, 0x87, 0x33, 0x00)
$2^{-19.76}$	$2^{-10.38}$	(0x0f, 0x8c, 0x00, 0xe0)	(0xb8, 0x54, 0x00, 0x34)
$2^{-19.76}$	$2^{-10.38}$	(0x19, 0xd8, 0x00, 0x70)	(0xdb, 0xb4, 0x00, 0x03)
$2^{-19.76}$	$2^{-10.38}$	(0x19, 0x00, 0x99, 0x7e)	(0xe6, 0x00, 0x07, 0x5e)
$2^{-19.76}$	$2^{-10.38}$	(0x28, 0x00, 0x7a, 0xc3)	(0x4c, 0x00, 0x6d, 0x45)
$2^{-19.76}$	$2^{-10.38}$	(0x00, 0xe0, 0x0f, 0x8c)	(0x00, 0x34, 0xb8, 0x54)
$2^{-19.76}$	$2^{-10.38}$	(0x00, 0xd7, 0x3c, 0x60)	(0x00, 0xc6, 0xa6, 0x82)

5 Multiple Linear Cryptanalysis on 22-Round SMS4

Based on each 5-round iterative linear characteristics mentioned above, an 18-round linear characteristic can be constructed as follows:

1. $(0,0,0,\Gamma_m) \rightarrow (0,0,0,\Gamma_m)$ for the 5-round SMS4 from the i-th round to the $(i+4)$-th round with bias $2 \times (\Gamma_m, \Gamma_m)^2$.
2. $(0,0,0,\Gamma_m) \rightarrow (0,0,0,\Gamma_m)$ for the 5-round SMS4 from the $(i+5)$-th round to the $(i+9)$-th round with

bias $2 \times (\Gamma_m, \Gamma_m)^2$.

3. $(0, 0, 0, \Gamma_m) \rightarrow (0, 0, 0, \Gamma_m)$ for the 5-round SMS4 from the $(i + 10)$-th round to the $(i + 14)$-th round with bias $2 \times (\Gamma_m, \Gamma_m)^2$.
4. $(0, 0, 0, \Gamma_m) \rightarrow (\Gamma_m, 0, 0, 0)$ for the 3-round SMS4 from the $(i + 15)$-th round to the $(i + 17)$-th round with bias $1/2$.

Then we obtain a linear characteristic $(0, 0, 0, \Gamma_m) \rightarrow (\Gamma_m, 0, 0, 0)$ for the 18-round SMS4 from the i-th round to the $(i + 17)$-th round, and the total bias of the 18-round linear characteristic is $2^5 \times (\Gamma_m, \Gamma_m)^6$.

In our attack, the linear characteristics for the 18-round SMS4 from the second round to the 18th round will be used, and eight 5-round iterative linear characteristics of SMS4 from Table 1 with Γ_m = (0x11, 0xff, 0xba, 0x00), (0x78, 0x52, 0xb3, 0x00), (0x79, 0x05, 0xe1, 0x00), (0xa1, 0xb4, 0x33, 0x00), (0xed, 0xca, 0x7c, 0x00), (0xfa, 0x70, 0x99, 0x00), (0x1d, 0xde, 0xab, 0x00), (0x38, 0xa5, 0x45, 0x00) will be chosen to derive eight 18-round linear characteristics respectively which can be expressed as below:

$$\begin{aligned}&(\text{0x11, 0xff, 0xba, 0x00}) \cdot (X_4 \oplus X_{19})\\ =\ &(\text{0x11, 0xff, 0xba, 0x00}) \cdot (RK_4 \oplus RK_5 \oplus RK_9 \oplus RK_{10} \oplus RK_{14} \oplus RK_{15}),\end{aligned} \tag{11}$$

$$\begin{aligned}&(\text{0x78, 0x52, 0xb3, 0x00}) \cdot (X_4 \oplus X_{19})\\ =\ &(\text{0x78, 0x52, 0xb3, 0x00}) \cdot (RK_4 \oplus RK_5 \oplus RK_9 \oplus RK_{10} \oplus RK_{14} \oplus RK_{15}),\end{aligned} \tag{12}$$

$$\begin{aligned}&(\text{0x79, 0x05, 0xe1, 0x00}) \cdot (X_4 \oplus X_{19})\\ =\ &(\text{0x79, 0x05, 0xe1, 0x00}) \cdot (RK_4 \oplus RK_5 \oplus RK_9 \oplus RK_{10} \oplus RK_{14} \oplus RK_{15}),\end{aligned} \tag{13}$$

$$\begin{aligned}&(\text{0xa1, 0xb4, 0x33, 0x00}) \cdot (X_4 \oplus X_{19})\\ =\ &(\text{0xa1, 0xb4, 0x33, 0x00}) \cdot (RK_4 \oplus RK_5 \oplus RK_9 \oplus RK_{10} \oplus RK_{14} \oplus RK_{15}),\end{aligned} \tag{14}$$

$$\begin{aligned}&(\text{0xed, 0xca, 0x7c, 0x00}) \cdot (X_4 \oplus X_{19})\\ =\ &(\text{0xed, 0xca, 0x7c, 0x00}) \cdot (RK_4 \oplus RK_5 \oplus RK_9 \oplus RK_{10} \oplus RK_{14} \oplus RK_{15}),\end{aligned} \tag{15}$$

$$\begin{aligned}&(\text{0xfa, 0x70, 0x99, 0x00}) \cdot (X_4 \oplus X_{19})\\ =\ &(\text{0xfa, 0x70, 0x99, 0x00}) \cdot (RK_4 \oplus RK_5 \oplus RK_9 \oplus RK_{10} \oplus RK_{14} \oplus RK_{15}),\end{aligned} \tag{16}$$

$$\begin{aligned}&(\text{0x1d, 0xde, 0xab, 0x00}) \cdot (X_4 \oplus X_{19})\\ =\ &(\text{0x1d, 0xde, 0xab, 0x00}) \cdot (RK_4 \oplus RK_5 \oplus RK_9 \oplus RK_{10} \oplus RK_{14} \oplus RK_{15}),\end{aligned} \tag{17}$$

$$\begin{aligned}&(\text{0x38, 0xa5, 0x45, 0x00}) \cdot (X_4 \oplus X_{19})\\ =\ &(\text{0x38, 0xa5, 0x45, 0x00}) \cdot (RK_4 \oplus RK_5 \oplus RK_9 \oplus RK_{10} \oplus RK_{14} \oplus RK_{15}),\end{aligned} \tag{18}$$

where equations (11), (12), (13), (14), (15) and (16) hold with bias $2^{-56.14}$, and equations (17), (18) hold with bias $2^{-57.28}$.

5.1 Attack Procedure

Following the framework given in [10], a key recovery attack can be mounted on the 22-round SMS4 from the first round to the 22nd round by applying the above eight 18-round linear characteristics. Based on the *Attack Algorithm MK 2* in [10], we perform partial encryptions of the first round and partial decryptions of the 20th, 21st and 22nd rounds by guessing the partial bits of RK_0, RK_{19}, RK_{20} and RK_{21} involved in the linear characteristics, and then determine the probability that the guessed subkey bits are correct by exploiting the linear characteristics over the remaining 18 rounds. For the linear system consisting of the above eight linear equations, the capacity of the system is about $2^{-107.38}$ according to equation (3). Thus the necessary number of known plaintext-ciphertext pairs to perform a successful key recovery attack on the 22-round SMS4 is about $\mathcal{O}(2^{107.38})$.

Before applying the *Attack Algorithm MK 2*, we extend the above eight equations to the expressions of a plaintext $P = (P_0, P_1, P_2, P_3)$, its corresponding 22-round ciphertext $C = (C_0, C_1, C_2, C_3)$ and subkeys of the first, 20th, 21st and 22nd rounds $(RK_0, RK_{19}, RK_{20}, RK_{21})$ similarly as in [7]. The extended expressions can be described as follows:

$$\begin{aligned}
&\Gamma_m^i \cdot (P_0 \oplus C_1) \oplus \Gamma_m^i \cdot F(P_1 \oplus P_2 \oplus P_3 \oplus RK_0) \oplus \\
&\Gamma_m^i \cdot F(C_0 \oplus C_2 \oplus C_3 \oplus RK_{19} \oplus F(C_0 \oplus C_1 \oplus C_2 \oplus RK_{21}) \oplus \\
&\qquad F(C_0 \oplus C_1 \oplus C_3 \oplus RK_{20} \oplus F(C_0 \oplus C_1 \oplus C_2 \oplus RK_{21}))) \\
= &\Gamma_m^i \cdot (RK_4 \oplus RK_5 \oplus RK_9 \oplus RK_{10} \oplus RK_{14} \oplus RK_{15}), \;\; 1 \le i \le 8,
\end{aligned} \tag{19}$$

where Γ_m^i corresponds to the Γ_m included in the above eight linear equations. Moreover, from table 1 we observe that Γ_m^i goes to $(\Gamma_m^i)'$ through the inverse of the linear layer L, which only influences 3 active S-Boxes in the non-linear layer S. Thus for the left side of equation (19), 3 active S-Boxes are involved for the first and 20th round respectively, and 4 active S-Boxes are impacted for the 21st and 22nd round respectively. If we rewrite equation (19) in a more compact form as below:

$$\begin{aligned}
&\Gamma_m^i \cdot (P_0 \oplus C_1) \oplus \Gamma_m^i \cdot f(RK, P, C) \\
= &\Gamma_m^i \cdot (RK_4 \oplus RK_5 \oplus RK_9 \oplus RK_{10} \oplus RK_{14} \oplus RK_{15}), \;\; 1 \le i \le 8,
\end{aligned} \tag{20}$$

the value of $f(RK, P, C)$ depends on 24 bits from $(P_1 \oplus P_2 \oplus P_3)$ and RK_0 respectively, 24 bits from $(C_0 \oplus C_2 \oplus C_3)$ and RK_{19} respectively, 32 bits from $(C_0 \oplus C_1 \oplus C_3)$ and RK_{20} respectively, and 32 bits from $(C_0 \oplus C_1 \oplus C_2)$ and RK_{21} respectively according to the above analysis. Let $\eta = \eta_1 \parallel \eta_2 \parallel \eta_3 \parallel \eta_4$ represent the impacted 112 subkey bits, where η_1, η_2, η_3 and η_4 denote the impacted 24 bits of RK_0, the impacted 24 bits of RK_{19}, RK_{20} and RK_{21} respectively. Let $\theta = \theta_1 \parallel \theta_2 \parallel \theta_3 \parallel \theta_4$ represent the impacted 112 plaintext-ciphertext bits, where θ_1, θ_2, θ_3 and θ_4 denote the impacted 24 bits of $(P_1 \oplus P_2 \oplus P_3)$, the impacted 24 bits of $(C_0 \oplus C_2 \oplus C_3)$, $(C_0 \oplus C_1 \oplus C_3)$, and $(C_0 \oplus C_1 \oplus C_2)$ respectively. Let $\xi = \xi_1 \parallel \xi_2 \parallel \xi_3 \parallel \xi_4$ represent the 112 bits of $\eta \oplus \theta$, where $\xi_1 = \eta_1 \oplus \theta_1$, $\xi_2 = \eta_2 \oplus \theta_2$, $\xi_3 = \eta_3 \oplus \theta_3$ and $\xi_4 = \eta_4 \oplus \theta_4$. Then we have

$$\begin{aligned}
&\Gamma_m^i \cdot f(RK, P, C) \\
= &\Gamma_m^i \cdot (F(\xi_1) \oplus F(\xi_2 \oplus F(\xi_4) \oplus F(\xi_3 \oplus F(\xi_4))) \;\; 1 \le i \le 8.
\end{aligned} \tag{21}$$

According to the success probability formula given in [14], the success probability of linear cryptanalysis depends not only on the amount of plaintext-ciphertext pairs, but also on the number of guessed subkey bits. Consequently, we need to prepare 2^{112} (that is, $2^{4.62} \times 2^{107.38}$) plaintext-ciphertext pairs in our multiple linear attack so as to achieve a high success probability of 88% approximately. Following gives the detailed description of our multiple linear cryptanalysis on 22-round SMS4.

Pre-computation phase

Initialize eight vectors Z^i $(1 \le i \le 8)$, each consisting of 2^{112} elements which correspond to all possible values of ξ. Then for each value of ξ, compute the parity of $\Gamma_m^i \cdot f(RK, P, C)$ according to equation (21) (i.e., the parity is calculated by partial encryptions of the first round and partial decryptions of the 20th, 21st and 22nd rounds). Keep the value $+1$ in the relative element of Z^i if the parity is 0, and -1 otherwise. Thus eight $2^{112} \times 2^{112}$ matrices M^i $(1 \le i \le 8)$ can be derived from the above eight vectors Z^i respectively, with $M^i[\eta][\theta] = Z^i[\xi]$, where $\xi = \eta \oplus \theta$.

Distillation phase

- Initialize eight vectors T^i $(1 \le i \le 8)$, each composed of 2^{112} counters which correspond to all possible values of θ. Then for each plaintext-ciphertext pair, compute the parity of $\Gamma_m^i \cdot (P_0 \oplus C_1)$. Increase the the relevant counter in T^i by 1 if the parity is 0, and decrease by 1 otherwise.
- Let $\hat{\mathbf{c}} = (\hat{c}_{i,\eta})_{1 \le i \le 8,\ 0 \le \eta \le 2^{112}-1}$ denote a vector consisting of 8×2^{112} elements, where $\hat{c}_{i,\eta}$ represents the estimated imbalance for the i-th linear characteristic and the subkey candidate η. Then for a given i, $1 \le i \le 8$, compute the estimated imbalance $\hat{c}_{i,\eta}$ for each possible subkey candidate η by the matrix-vector product $M^i T^i$ ($\hat{c}_{i,\eta}$ is equal to the corresponding entry of $M^i T^i$ divided by 2^{112}, the

number of plaintext-ciphertext pairs).

Analysis phase

- Compute $\|\hat{\mathbf{c}}\|^2 = \sum_{i=1}^{8}\sum_{\eta=0}^{2^{112}-1}\hat{c}_{i,\eta}^2$, and for each subkey candidate η, calculate $\|\hat{\mathbf{c}}_\eta\|^2 = \sum_{i=1}^{8}\hat{c}_{i,\eta}^2$.
- Let $\mathbf{k} = (k_i)_{1\le i\le 8}$ denote a vector composed of eight elements, where k_i represents the parity of $\Gamma_m^i \cdot (RK_4 \oplus RK_5 \oplus RK_9 \oplus RK_{10} \oplus RK_{14} \oplus RK_{15})$. For a given $\mathbf{k}$ and a given subkey candidate η, a vector $\mathbf{c}_{\mathbf{k},\eta}$ of theoretical imbalances with 8×2^{112} elements is then constructed as follows:

$$\mathbf{c}_{\mathbf{k},\eta} = (0,\ldots,0,(-1)^{k_1}c_1,\ldots,(-1)^{k_8}c_8,0,\ldots,0), \tag{22}$$

  where c_i $(1 \le i \le 8)$ corresponds to the imbalance of the i-th linear characteristic used in our attack, and the location of the subvector $((-1)^{k_1}c_1,\ldots,(-1)^{k_8}c_8)$ depends on the value of η.
- For each possible value of $\mathbf{k} \in Z_2^8$ and each possible subkey candidate η, the Euclidean distance between the vector of estimated imbalances and the vector of theoretical imbalances is measured by the following equation:

$$\begin{aligned}\|\hat{\mathbf{c}} - \mathbf{c}_{\mathbf{k},\eta}\|^2 &= \sum_{i=1}^{8}(\hat{c}_{i,\eta} - (-1)^{k_i}c_i)^2 + \sum_{\eta'\ne\eta}\sum_{i=1}^{8}\hat{c}_{i,\eta'}^2 \\ &= \sum_{i=1}^{8}(\hat{c}_{i,\eta} - (-1)^{k_i}c_i)^2 + (\|\hat{\mathbf{c}}\|^2 - \|\hat{\mathbf{c}}_\eta\|^2).\end{aligned} \tag{23}$$

- Take the value of $\mathbf{k}$ and the subkey candidate η as the correct key information if the corresponding Euclidean distance $\|\hat{\mathbf{c}} - \mathbf{c}_{\mathbf{k},\eta}\|^2$ is minimal.

The data complexity of the attack is 2^{112} known plaintext-ciphertext pairs. The time complexity of the attack is dominated mainly by the eight matrix-vector products M^iT^i ($1 \le i \le 8$) in the distillation phase, and the calculation of the Euclidean distance between the vector of estimated imbalances and the vector of theoretical imbalances for each possible value of $\mathbf{k}$ and each possible subkey candidate η in the analysis phase. Regarding each matrix-vector product M^iT^i, the time complexity is about $3 \times 112 \times 2^{112} \approx 2^{120.39}$ arithmetic operations by applying the technique given in [15], thus leading to a time complexity of $2^{123.39}$ arithmetic operations for all of the eight matrix-vector products. As for the calculations of all above possible Euclidean distances, the time complexity is about $2^8 \times 2^{112} \times 8 = 2^{123}$ arithmetic operations. Consequently, the total time complexity of our attack is approximately $2^{123.39} + 2^{123} \approx 2^{124.21}$ arithmetic operations. Furthermore, the memory complexity of the attack is primarily owing to keeping the eight vectors T^i $(1 \le i \le 8)$ in the distillation phase. In order to keep a vector T^i of 2^{112} elements, 2^{112}-bit memory is required, equivalently 2^{109}-byte memory, which results in the total memory complexity of our attack being $8 \times 2^{109} = 2^{112}$ bytes approximately.

6 Conclusion

In this paper, firstly we propose a modified branch-and-bound algorithm which allows searching multiple linear characteristics for SMS4-like unbalanced Feistel block ciphers. Then a series of 5-round iterative linear characteristics of SMS4 are obtained by applying the modified algorithm in SMS4. Furthermore, for each 5-round iterative linear characteristic mentioned above, an 18-round linear characteristic of SMS4 can be derived, resulting in a list of 18-round linear characteristics of SMS4. After that, a key recovery multiple linear attack is presented on 22-round SMS4 by exploiting the above multiple linear characteristics. As far as we know, our result has much lower data complexity than the previously best known cryptanalytic result on 22-round SMS4, which is also the previously best known result on SMS4. The complexities of our attack as well as the previously known attacks on SMS4 are summarized in Table 2. However, it should be noted that neither the previously known attacks nor our attack can endanger the full 32-round SMS4. We hope our result can be helpful in evaluating the security of SMS4 against multiple linear cryptanalysis. As a scope of further research, assessing SMS4 against combined attacks such as differential-linear cryptanalysis should be done and such work is in progress.

Table 2. Summary of Attacks on SMS4 with Reduced Number of Rounds

Number of Rounds	Type of Attack	Complexity		
		Data	Time	Memory
13	Integral [3]	2^{16} CP	2^{114} Enc	2^{20} B
14	Rectangle [5]	$2^{121.82}$ CP [1]	$2^{116.66}$ Enc [1]	$2^{125.82}$ B
14	Rectangle [8]	$2^{107.89}$ CP	$2^{107.89}$ MA	$2^{111.89}$ B
16	Impossible differential [5]	2^{105} CP [1]	2^{107} Enc [1]	2^{109} B
16	Impossible differential [8]	$2^{117.06}$ CP	$2^{132.06}$ MA	$2^{121.06}$ B
16	Rectangle [6]	2^{125} CP	2^{116} Enc	2^{125} B
18	Boomerang [7]	2^{120} ACPC	$2^{116.83}$ Enc	2^{123} B
18	Rectangle [7]	2^{124} CP	$2^{112.83}$ Enc	2^{128} B
21	Differential [6]	2^{118} CP	$2^{126.6}$ Enc	2^{123} B
22	Differential [7]	2^{118} CP	$2^{125.71}$ Enc	2^{123} B
22	Linear [7]	$2^{119.18}$ KP [2]	$2^{109.86}$ Enc + $2^{120.39}$ AO	2^{109} B
22	Differential [9]	2^{117} CP	$2^{112.3}$ Enc	2^{122} B
22	Multiple linear (this paper)	2^{112} KP	$2^{124.21}$ AO	2^{112} B

KP - Known plaintexts, CP - Chosen plaintexts, ACPC - Adaptive chosen plaintexts and ciphertexts
Enc - Encryptions, MA - Memory accesses, AO - Arithmetic operations, B - Bytes.

References

1. Office of State Commercial Cryptography Administration, P.R. China: The SMS4 block cipher (in Chinese). Arichive available at `http://www.oscca.gov.cn/UpFile/200621016423197990.pdf`.
2. Zhang, L., Wu, W.: Differential fault analysis on SMS4 (in Chinese). Chinese Journal of Computers **29**(9) (2006) 1596–1602
3. 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. Volume 4586 of Lecture Notes in Computer Science., Springer (2007) 158–170
4. Ji, W., Hu, L.: New description of SMS4 by an embedding over GF(2^8). In Srinathan, K., Rangan, C.P., Yung, M., eds.: INDOCRYPT. Volume 4859 of Lecture Notes in Computer Science., Springer (2007) 238–251
5. 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. Volume 4861 of Lecture Notes in Computer Science., Springer (2007) 306–318
6. Zhang, L., Zhang, W., Wu, W.: Cryptanalysis of reduced-round SMS4 block cipher. In Mu, Y., Susilo, W., Seberry, J., eds.: ACISP. Volume 5107 of Lecture Notes in Computer Science., Springer (2008) 216–229
7. Taehyun Kim, Jongsung Kim, S.H., Sung, J.: Linear and differential cryptanalysis of reduced SMS4 block cipher. Cryptology ePrint Archive, Report 2008/281 (2008) `http://eprint.iacr.org/`.
8. 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. Volume 5308 of Lecture Notes in Computer Science., Springer (2008) 141–156
9. 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. Volume 5451 of Lecture Notes in Computer Science., Springer (2009) 324–335

[1] As noted in [8], these figures are underestimated.

[2] The success rate table (i.e., Table 3 presented in [11]) of M. Matsui's Algorithm 2 for some special case, from which the original data complexity of 2^{117} known plaintexts is derived in [7], is not applicable for the linear attack on 22-round SMS4 indeed. Moreover, a general success probability formula has been proposed for linear cryptanalysis in [14]. Thus following this formula, the data complexity of the linear attack in [7] should be $2^{119.18}$ known plaintexts in order to achieve the same success probability as in our attack.

10. Biryukov, A., Cannière, C.D., Quisquater, M.: On multiple linear approximations. In Franklin, M.K., ed.: CRYPTO. Volume 3152 of Lecture Notes in Computer Science., Springer (2004) 1–22
11. Matsui, M.: Linear cryptanalysis method for DES cipher. In Helleseth, T., ed.: EUROCRYPT. Volume 765 of Lecture Notes in Computer Science., Springer (1993) 386–397
12. Kaliski, B.S., Robshaw, M.J.B.: Linear cryptanalysis using multiple approximations. In Desmedt, Y., ed.: CRYPTO. Volume 839 of Lecture Notes in Computer Science., Springer (1994) 26–39
13. Matsui, M.: On correlation between the order of S-boxes and the strength of DES. In Santis, A.D., ed.: EUROCRYPT. Volume 950 of Lecture Notes in Computer Science., Springer (1994) 366–375
14. Selçuk, A.A.: On probability of success in linear and differential cryptanalysis. Journal of Cryptology **21**(1) (2008) 131–147
15. Collard, B., Standaert, F.X., Quisquater, J.J.: Improving the time complexity of Matsui's linear cryptanalysis. In Nam, K.H., Rhee, G., eds.: ICISC. Volume 4817 of Lecture Notes in Computer Science., Springer (2007) 77–88