MitM Fault Analysis on Word-oriented SPN Block Ciphers

Zhiqiang Liu[1,3], Ya Liu[2], Qingju Wang[1,3], Dawu Gu[1], Wei Li[4]

[1]Department of Computer Science and Engineering,
Shanghai Jiao Tong University, Shanghai 200240, P.R. China
{ilu_zq,dwgu}@sjtu.edu.cn
[2]Department of Computer Science and Engineering,
University of Shanghai for Science and Technology, Shanghai 200093, P.R. China
liuyaloccs@gmail.com
[3]ESAT/COSIC and IBBT,
Katholieke Universiteit Leuven, Leuven, Belgium
qjuwang@gmail.com
[4]School of Computer Science and Technology,
Donghua University, Shanghai 201620, P.R. China
liwei.cs.cn@gmail.com

Abstract. Meet-in-the-Middle (MitM) fault analysis is a kind of powerful cryptanalytic approach suitable for various block ciphers. When applying the method to analyze the security of block ciphers, it is very crucial to find effective MitM characteristics based on some fault models. In this paper, we investigate the security of word-oriented SPN block ciphers by means of MitM fault analysis, and observe that if the diffusion layers of the ciphers have some special properties, it is easy to derive effective MitM characteristics under the condition of single-word fault model, which can lead to efficient fault attacks on the ciphers. In order to demonstrate the effectiveness of our observation, we apply it to ARIA and AES, and obtain some effective MitM characteristics respectively, then we present efficient MitM fault attacks on the ciphers in terms of these characteristics. It is expected that our work could be helpful in evaluating the security of word-oriented SPN block ciphers against fault attack. We also hope that this work could be beneficial to the design strategy of diffusion layers of block ciphers.

Key words: MitM Fault Analysis, Word-oriented SPN Structure, Block Cipher, ARIA, AES

1 Introduction

Since the fault analysis was first introduced in 1997 by Boneh *et al* [1], it has drawn much attention from cryptanalysts all over the world and has become a very efficient cryptanalytic tool in analyzing various cryptographic devices. As a matter of fact, fault analysis is a class of implementation attacks that disturb cryptographic computations so as to recover secret keys. More precisely, an adversary can find the key of a cipher by exploiting information derived from correct and faulty ciphertexts. Currently, several techniques are known to induce faults during cryptographic computations such as triggering a spike on the power supply, a glitch on the clock, or using external methods based on laser, Focused Ion

Beam, or electromagnetic radiations [2]. Specially, a laser with certain energy and wavelength could interfere fixed parts of the memory/registers without damaging them, resulting in a single-bit error or single-byte error at some internal states accurately [3].

Up to now, much research work has been devoted to fault analysis on block ciphers and such work can be mainly diversified into three directions. The first direction is to find efficient differential fault attacks for a block cipher. So far, DES [4, 5], Triple-DES [6], AES [7–15], IDEA [16], CLEFIA [17], SMS4 [18], ARIA [19, 20], Camellia [21] and KATAN32 [22] have been analyzed by means of differential fault analysis (DFA). The second direction is to further improve the fault analysis results on a block cipher by exploring faults induced at an earlier round, reducing the number of required faults or weakening the fault injection model. For example, Derbez *et al.* [23] presented fault attacks on AES in which faults were triggered at an earlier round compared to the previous related work, Kim [20] introduced fault attacks on ARIA which used fewer faults to retrieve the 128-bit cipher key in comparison with [19]. The third direction is to extend fault analysis so as to make it more efficient. For instance, Phan *et al.* [24] gave an idea of amplifying fault attack with techniques from various cryptanalytic approaches, Derbez *et al.* [23] proposed new kinds of fault attack, i.e., MitM fault attack and impossible differential fault attack, Liu *et al.* [25] showed the feasibility of applying linear characteristics in fault analysis and presented a novel fault attack called linear fault analysis.

For a given block cipher, we denote the relation of some intermediate states of the cipher (for instance, a linear equation of some internal state bits) as a MitM characteristic of the cipher. If a MitM characteristic of a cipher could be exploited to mount an attack better than exhaustive attack on the cipher, we call it an effective MitM characteristic. MitM fault analysis is a kind of powerful fault cryptanalytic tool, which uses effective MitM characteristics to facilitate fault attacks on block ciphers. More specifically, MitM fault analysis can be described as follows:

- For a given block cipher, choose plaintexts to encrypt and get the ciphertexts.
- Encrypt the same plaintexts as above and induce faults during encryption processes so as to obtain the corresponding faulty ciphertexts.
- Try to derive some effective MitM characteristics by exploring relations among the intermediate states corresponding to the correct and faulty ciphertext pairs.
- Mount an efficient key recovery attack on the cipher based on the above MitM characteristics.

So far MitM fault analysis has been successfully applied to AES cipher [23] and KATAN32 cipher [22], and the attack results are the currently best fault cryptanalytic results on AES and KATAN32 respectively. In order to devise a good strategy for MitM fault attack on a block cipher, one needs to obtain effective MitM characteristics of the cipher given a fault model. But how can we get effective MitM characteristics of a block cipher? Are there any particular features in the diffusion layer of a cipher which could be used to construct effective MitM characteristics? As a matter of fact, some research work (e.g. [26]) has been done to explore the properties of diffusion layers of block ciphers with similar structures which make the ciphers subject to certain cryptanalytic tools such as differential cryptanalysis, linear cryptanalysis, impossible differential cryptanalysis, and so on. However, to the best of our knowledge, there has not been any known work for investigating such properties with respect to MitM fault analysis so far.

In this paper, we study the security of word-oriented SPN block ciphers by using MitM fault analysis, and find that if the diffusion layers within the ciphers have some specific properties, one can easily get effective MitM characteristics under the condition of single-word fault model, thus leading to efficient fault attacks on the ciphers. For the purpose of illustration, we apply our observation to ARIA and AES, and attain some effective MitM characteristics of the ciphers respectively, then in terms of these characteristics, we present efficient MitM fault attacks on the ciphers. For ARIA, our attack result together with the previously known fault attack results are given in Table 1. As to AES, we can derive the MitM characteristics used in [23] by applying our observation to the cipher, thus following the attack procedure given in [23], an efficient MitM fault attack could be mounted on the cipher.

Table 1. Summary of Fault Attacks on ARIA

Type of Attack	Fault Model	Fault Injection Round	Number of Faulty Ciphertexts
DFA [19]	Single-byte Fault	penultimate round	45
DFA [20]	Multi-byte Fault	penultimate round	13
MITMFA (This paper)	Single-byte Fault	antepenultimate round	44

DFA: Differential Fault Attack, MITMFA: MitM Fault Attack.

The rest of the paper is organized as follows. Section 2 introduces the notations used throughout this paper, and gives the fault model and assumption adopted in this work. Section 3 shows our observation on word-oriented SPN block ciphers, with which one can derive good strategies for MitM fault attacks on the ciphers. Section 4 and 5 apply our observation to ARIA and AES respectively, and present efficient MitM fault attacks on the ciphers. Finally, Section 6 summarizes the paper.

2 Preliminaries

The following notations are used throughout the paper.
– $\oplus$ denotes bitwise exclusive OR (XOR).
– 0x denotes the hexadecimal notation.
– $|S|$ denotes the cardinality of the set S.
– $\|$ denotes the concatenation operation.
– set $1 \setminus$ set 2 denotes the set $\{x | x \in \text{set } 1,\ x \notin \text{set } 2\}$.
– set $1 \cap$ set 2 denotes the set $\{x | x \in \text{set } 1,\ x \in \text{set } 2\}$.

2.1 Single-word Fault Model

For a given block cipher, an adversary is able to choose plaintexts to encrypt (under an unknown cipher key), and during the encryption process, he can induce single-word faults into one round and obtain the corresponding right and faulty ciphertexts. Note that the values and/or positions (within the impacted round) of the faults injected by the adversary are unknown and uniformly distributed.

3 Our Observation on Word-oriented SPN Block Ciphers

Let E be an n-round word-oriented SPN block cipher with d-bit word size and m-word block size. Let $I_i = (I_{i,0}, I_{i,1}, \ldots, I_{i,m-1}) \in GF(2^d)^m$ be the input of the i-th $(1 \leq i \leq n)$ round of E. Let $X_i = (X_{i,0}, X_{i,1}, \ldots, X_{i,m-1})$, $Y_i = (Y_{i,0}, Y_{i,1}, \ldots, Y_{i,m-1}) \in GF(2^d)^m$ be the input and output of the substitution layer of the i-th round, respectively.

The i-th round function of E is depicted in Fig. 1, where the round key addition is to XOR the i-th round key, $S_{i,j}$ $(0 \leq j \leq m-1)$ is a non-linear word permutation which operates on $X_{i,j}$, and the diffusion layer is essentially a linear transformation $P:\ GF(2^d)^m \rightarrow GF(2^d)^m$ which is performed on Y_i.

Let $\Delta I_i = (\Delta I_{i,0}, \Delta I_{i,1}, \ldots, \Delta I_{i,m-1}) \in GF(2^d)^m$ be the input difference of the i-th round of E. Let $\Delta Y_i = (\Delta Y_{i,0}, \Delta Y_{i,1}, \ldots, \Delta Y_{i,m-1}) \in GF(2^d)^m$ be the output difference of the substitution layer of the i-th round.

Note that in this paper we will only consider a kind of MitM characteristics of E, that is, **linear equations on the words of I_i (or ΔI_i)**.

We observe that if the diffusion layer within E satisfies Property 1 and 2 given below, one can easily derive some effective MitM characteristics of the cipher under the condition of single-word fault model.

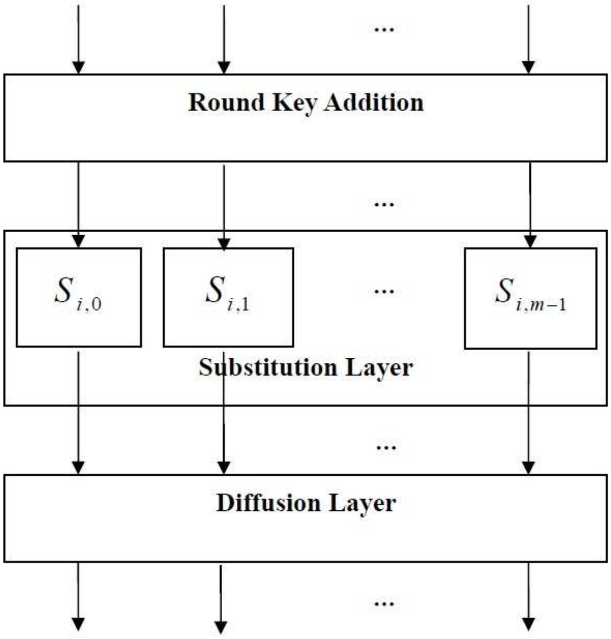


Fig. 1. The i-th round function of E

Property 1. Let $(a_0, a_1, \ldots, a_{m-1})$, $(b_0, b_1, \ldots, b_{m-1}) \in GF(2^d)^m$ be the input and output of the diffusion layer, respectively. Then there exist some pairs of subscript sets (T_1, T_1'), (T_2, T_2'), $\ldots$, (T_L, T_L') such that

$$\bigoplus_{t \in T_l} a_t = \bigoplus_{t \in T_l'} b_t, \quad 1 \le l \le L,$$

where T_l, T_l' are subsets of $\{0, 1, ..., m-1\}$, and $2 \le L < 2^m$.
Property 2. For a given i_0 $(1 \le i_0 \le n-2)$, suppose that ΔY_{i_0} satisfies

$$\Delta Y_{i_0} = (0\text{x}00, \ldots, 0\text{x}00, \Delta Y_{i_0,j_0}, 0\text{x}00, \ldots, 0\text{x}00),$$

where $0 \le j_0 \le m-1$, and $\Delta Y_{i_0,j_0}$ is an unknown nonzero value. Then for some integer r_0 $(1 \le r_0 \le n - i_0 - 1)$, there exist subscript sets $T_{l_1}, T_{\tilde{l_1}}$, $T_{l_2}, T_{\tilde{l_2}}$, $\ldots, T_{l_\Lambda}, T_{\tilde{l_\Lambda}}$ such that

$$T_{l_\lambda} \setminus T_{\tilde{l_\lambda}} \subseteq J \text{ and } T_{\tilde{l_\lambda}} \setminus T_{l_\lambda} \subseteq J, \quad 1 \le \lambda \le \Lambda,$$

where $J = \{j | \Delta Y_{i_0+r_0,j} = 0,\ 0 \le j \le m-1\}$, $l_\lambda \ne \tilde{l_\lambda}$ $(1 \le l_\lambda, \tilde{l_\lambda} \le L)$, and T_{l_λ}, $T_{\tilde{l_\lambda}}$ are the subscript sets given in Property 1.

Next we will show the reasonability of our observation in detail. Suppose that the cipher E has the above two properties. We find that the effect of injecting single-word fault before the diffusion layer in the i_0-th round of E, is somewhat like inducing a difference in Y_{i_0} (More specifically, ΔY_{i_0} is an m-word vector with one word being nonzero and the other $m-1$ words being zero). Thus according to Property 1 given above, we can obtain the following equations:

$$\bigoplus_{t \in T_{l_\lambda}} \Delta Y_{i_0+r_0,t} = \bigoplus_{t \in T'_{l_\lambda}} \Delta I_{i_0+r_0+1,t}, \quad 1 \le \lambda \le \Lambda, \tag{1}$$

$$\bigoplus_{t \in T_{\tilde{l}_\lambda}} \Delta Y_{i_0+r_0,t} = \bigoplus_{t \in T'_{\tilde{l}_\lambda}} \Delta I_{i_0+r_0+1,t}, \quad 1 \le \lambda \le \Lambda. \tag{2}$$

Moreover, following the Property 2, we have

$$\bigoplus_{t \in T_{l_\lambda}} \Delta Y_{i_0+r_0,t} = \bigoplus_{t \in T_{\tilde{l}_\lambda}} \Delta Y_{i_0+r_0,t}, \quad 1 \le \lambda \le \Lambda. \tag{3}$$

Combining the equations (1), (2) and (3), we get multiple MitM characteristics of E under the single-word fault model as shown below:

$$\bigoplus_{t \in T'_{l_\lambda}} \Delta I_{i_0+r_0+1,t} = \bigoplus_{t \in T'_{\tilde{l}_\lambda}} \Delta I_{i_0+r_0+1,t}, \quad 1 \le \lambda \le \Lambda. \tag{4}$$

Furthermore, we find that if the diffusion layer within E meets Property 3 and 4 given below, one can obtain some effective MitM characteristics of the cipher under the condition of single-word fault model.

Property 3. Let $(a_0, a_1, \ldots, a_{m-1})$, $(b_0, b_1, \ldots, b_{m-1}) \in GF(2^d)^m$ be the input and output of the diffusion layer respectively. Then there exist some pairs of subscript sets (V_1, V'_1), (V_2, V'_2), $\ldots$, (V_H, V'_H) such that each element in $\{b_v | v \in V'_h\}$ is the linear combination of all the elements in $\{a_v | v \in V_h\}$, where V_h, V'_h are subsets of $\{0, 1, ..., m-1\}$, and $1 \le h \le H$.

Property 4. For a given i_0 $(1 \le i_0 \le n-2)$, suppose that ΔY_{i_0} satisfies

$$\Delta Y_{i_0} = (\text{0x00}, \ldots, \text{0x00}, \Delta Y_{i_0,j_0}, \text{0x00}, \ldots, \text{0x00}),$$

where $0 \le j_0 \le m-1$, and $\Delta Y_{i_0,j_0}$ is an unknown nonzero value. Then for some integer r_0 $(1 \le r_0 \le n - i_0 - 1)$, there exist subscript sets V_{h_1}, V_{h_2}, $\ldots$, V_{h_Ω} such that

$$|J_{h_\omega}| = 1, \quad 1 \le \omega \le \Omega,$$

where $J_{h_\omega} = \{j | \Delta Y_{i_0+r_0,j} \ne 0, j \in V_{h_\omega}\}$, $1 \le h_\omega \le H$, and V_{h_ω} is the subscript set given in Property 3.

Suppose that the cipher E satisfies Property 3 and 4. As mentioned above, $\varDelta Y_{i_0}$ is an m-word vector with one word being nonzero and the other $m-1$ words being zero in the case of triggering single-word fault before the diffusion layer in the i_0-th round of E. Thus for any two elements in $\{\varDelta I_{i_0+r_0+1,v} | v \in V'_{h_\omega}\}$, $1 \leq \omega \leq \varOmega$, we can easily get a linear equation on the two elements (i.e., a MitM characteristic of E) according to Property 3 and 4.

With the help of the above MitM characteristics, one can mount key recovery attacks on E much better than exhaustive attack. Next, we will illustrate how to mount efficient MitM fault attacks on word-oriented SPN block ciphers by applying our observation to ARIA and AES respectively.

4 MitM Fault Attack on ARIA

4.1 A Brief Description of ARIA

ARIA [27] is a word-oriented SPN block cipher with 8-bit word size and 128-bit block size. It accepts keys of 128, 192 or 256 bits and the number of rounds is 12, 14 or 16 respectively. In the following parts, we will focus on **the ARIA cipher with 128-bit key size** unless otherwise stated.

The input and output of each round of ARIA could be treated as 16-byte vectors (also denoted as states), and each byte within the vectors could be regarded as an element in $GF(2^8)$. Each round of ARIA consists of the following three basic operations (except the last round, where the DL operation is replaced by an additional RKA operation):
Round Key Addition (RKA): XOR the 128-bit round key. All round keys are derived from the cipher key according to the key schedule.
Substitution Layer (SL): Apply 16 non-linear 8×8-bit S-boxes to the 16 bytes of the intermediate vector respectively. ARIA has two types of substitution layers, i.e., type 1 and type 2, where type 1 is used in the odd rounds and type 2 is used in the even rounds.
Diffusion Layer (DL): A linear transformation $LT : GF(2^8)^{16} \rightarrow GF(2^8)^{16}$ is performed on the intermediate 16-byte vector. Note that $LT^{-1} = LT$.

Please refer to [27] for detailed information about the substitution layer, the diffusion layer and the key schedule adopted in ARIA.

4.2 MitM Characteristics of ARIA under Single-byte Fault Model

After investigating the diffusion layer adopted in ARIA in detail, we find that the cipher has two properties given in Section 3 as shown below.

(a). Let $(a_0, a_1, \ldots, a_{15})$, $(b_0, b_1, \ldots, b_{15}) \in GF(2^8)^{16}$ be the input and output of the diffusion layer of ARIA respectively. Then there exist 6 pairs of subscript sets

$$\begin{gathered}
(T_1, T_1') = (\{0,3,12,15\}, \{0,3,12,15\}), \\
(T_2, T_2') = (\{1,3,5,7\}, \{0,2,4,6\}), \\
(T_3, T_3') = (\{2,3,10,11\}, \{0,1,8,9\}), \\
(T_4, T_4') = (\{1,2,13,14\}, \{1,2,13,14\}), \\
(T_5, T_5') = (\{0,1,8,9\}, \{2,3,10,11\}), \\
(T_6, T_6') = (\{0,2,4,6\}, \{1,3,5,7\})
\end{gathered}$$

such that

$$\bigoplus_{t \in T_l} a_t = \bigoplus_{t \in T_l'} b_t, \quad 1 \le l \le 6.$$

(b). **Case 1**: For a given i_0 $(1 \le i_0 \le 10)$, suppose that ΔY_{i_0} satisfies

$$\Delta Y_{i_0} = (\Delta Y_{i_0,0}, 0\text{x}00, \ldots, 0\text{x}00),$$

where $\Delta Y_{i_0,0}$ is an unknown nonzero value. Then for $r_0 = 1$, let $J_1 = \{j | \Delta Y_{i_0+r_0,j} = 0,\ 0 \le j \le 15\}$, there exist 2 pairs of subscript sets (T_1, T_2), (T_1, T_3) such that

$$T_1 \setminus T_2,\ \ T_2 \setminus T_1,\ \ T_1 \setminus T_3,\ \ T_3 \setminus T_1 \subseteq J_1,$$

$$T_1 \cap T_2 = T_1 \cap T_3.$$

Case 2: For a given i_0 $(1 \le i_0 \le 10)$, suppose that ΔY_{i_0} satisfies

$$\Delta Y_{i_0} = (0\text{x}00, 0\text{x}00, \Delta Y_{i_0,2}, 0\text{x}00, \ldots, 0\text{x}00),$$

where $\Delta Y_{i_0,2}$ is an unknown nonzero value. Then for $r_0 = 1$, let $J_2 = \{j | \Delta Y_{i_0+r_0,j} = 0,\ 0 \le j \le 15\}$, there exist 2 pairs of subscript sets (T_2, T_4), (T_4, T_5) such that

$$T_2 \setminus T_4,\ \ T_4 \setminus T_2,\ \ T_4 \setminus T_5,\ \ T_5 \setminus T_4 \subseteq J_2,$$

$$T_2 \cap T_4 = T_4 \cap T_5.$$

Case 3: For a given i_0 $(1 \le i_0 \le 10)$, suppose that ΔY_{i_0} satisfies

$$\Delta Y_{i_0} = (0\text{x}00, 0\text{x}00, 0\text{x}00, \Delta Y_{i_0,3}, 0\text{x}00, \ldots, 0\text{x}00),$$

where $\Delta Y_{i_0,3}$ is an unknown nonzero value. Then for $r_0 = 1$, let $J_3 = \{j | \Delta Y_{i_0+r_0,j} = 0,\ 0 \le j \le 15\}$, there exist 2 pairs of subscript sets (T_1, T_6), (T_5, T_6) such that

$$T_1 \setminus T_6,\ \ T_6 \setminus T_1,\ \ T_5 \setminus T_6,\ \ T_6 \setminus T_5 \subseteq J_3,$$

$$T_1 \cap T_6 = T_5 \cap T_6.$$

Thus we can obtain the following MitM characteristics of ARIA under the single-byte fault model:

$$\begin{aligned} & \Delta I_{i_0+2,1} \oplus \Delta I_{i_0+2,8} \oplus \Delta I_{i_0+2,9} \\ = & \Delta I_{i_0+2,2} \oplus \Delta I_{i_0+2,4} \oplus \Delta I_{i_0+2,6} \\ = & \Delta I_{i_0+2,3} \oplus \Delta I_{i_0+2,12} \oplus \Delta I_{i_0+2,15}, \end{aligned} \tag{5}$$

$$\begin{aligned} & \Delta I_{i_0+2,0} \oplus \Delta I_{i_0+2,4} \oplus \Delta I_{i_0+2,6} \\ = & \Delta I_{i_0+2,1} \oplus \Delta I_{i_0+2,13} \oplus \Delta I_{i_0+2,14} \\ = & \Delta I_{i_0+2,3} \oplus \Delta I_{i_0+2,10} \oplus \Delta I_{i_0+2,11}, \end{aligned} \tag{6}$$

$$\begin{aligned} & \Delta I_{i_0+2,0} \oplus \Delta I_{i_0+2,12} \oplus \Delta I_{i_0+2,15} \\ = & \Delta I_{i_0+2,1} \oplus \Delta I_{i_0+2,5} \oplus \Delta I_{i_0+2,7} \\ = & \Delta I_{i_0+2,2} \oplus \Delta I_{i_0+2,10} \oplus \Delta I_{i_0+2,11}. \end{aligned} \tag{7}$$

Next, we will demonstrate an effective key recovery attack on ARIA in terms of the characteristics (5), (6) and (7).

4.3 Attacking ARIA

Let $ek_i = (ek_{i,0}, ek_{i,1}, \ldots, ek_{i,15}) \in GF(2^8)^{16}$ be the round key of the i-th ($1 \leq i \leq 12$) round of ARIA. Let $ek_{13} = (ek_{13,0}, ek_{13,1}, \ldots, ek_{13,15}) \in GF(2^8)^{16}$ be the additional round key of the 12th round. We now present a key recovery attack on ARIA under the condition of single-byte fault model. Since there isn't any known fault attack on ARIA which is done by inducing faults at the round earlier than the penultimate round of the cipher so far, the general countermeasure against fault attack on ARIA could be implemented by protecting the last two rounds of the cipher if taking into account the efficiency of the implementation. However, our effective attack shows that MitM fault analysis could be a threat to the protected implementation of ARIA.

The basic idea of our attack is as follows. Firstly, collect ciphertext pairs each of which consists of a right ciphertext under ARIA and a corresponding faulty ciphertext derived by provoking single-byte fault before the diffusion layer in the antepenultimate round (i.e., the 10th round). Then we can recover all bytes of ek_{13} with the help of the MitM characteristics (5), (6) and (7) given in Section 4.1. After that, we decrypt the last round by using ek_{13} and mount an attack similarly to the above and obtain all bytes of ek_{12}. Repeat this procedure until we get the values of ek_{11} and ek_{10}. Finally, the cipher key can be derived from ek_{10}, ek_{11}, ek_{12} and ek_{13} according to the key schedule of ARIA. Following gives the detailed description of our attack in two phases.

***Phase 1. Retrieving the round key** ek_{13}*

Let $S_{i,j}^{-1}$ denote the inverse of the S-box $S_{i,j}$, $1 \le i \le 12$, $0 \le j \le 15$. Then do as below:

Step 1. Under the single-byte fault model, we assume that single-byte faults can be triggered on the first byte of the intermediate state before the diffusion layer in the 10th round, then do as follows:

(a). For a given plaintext P_r, encrypt it seven times under the unknown cipher key so as to collect six pairs of ciphertexts, each pair consisting of the right ciphertext C_r and the corresponding faulty ciphertext C_f^k $(1 \le k \le 6)$ derived by the above means.

(b). Let $C_{r,j}$, $C_{f,j}^k$ denote the $(j+1)$-th bytes of C_r and C_f^k respectively. Let Δ_j^k denote $S_{12,j}^{-1}(C_{r,j} \oplus ek_{13,j}) \oplus S_{12,j}^{-1}(C_{f,j}^k \oplus ek_{13,j})$. According to the characteristic (5), we can obtain

$$\Delta_1^k \oplus \Delta_8^k \oplus \Delta_9^k = \Delta_2^k \oplus \Delta_4^k \oplus \Delta_6^k, \quad 1 \le k \le 6, \tag{8}$$

and

$$\Delta_2^k \oplus \Delta_4^k \oplus \Delta_6^k = \Delta_3^k \oplus \Delta_{12}^k \oplus \Delta_{15}^k, \quad 1 \le k \le 6. \tag{9}$$

Fig. 2 gives a schematic description of the MitM characteristics of ARIA.

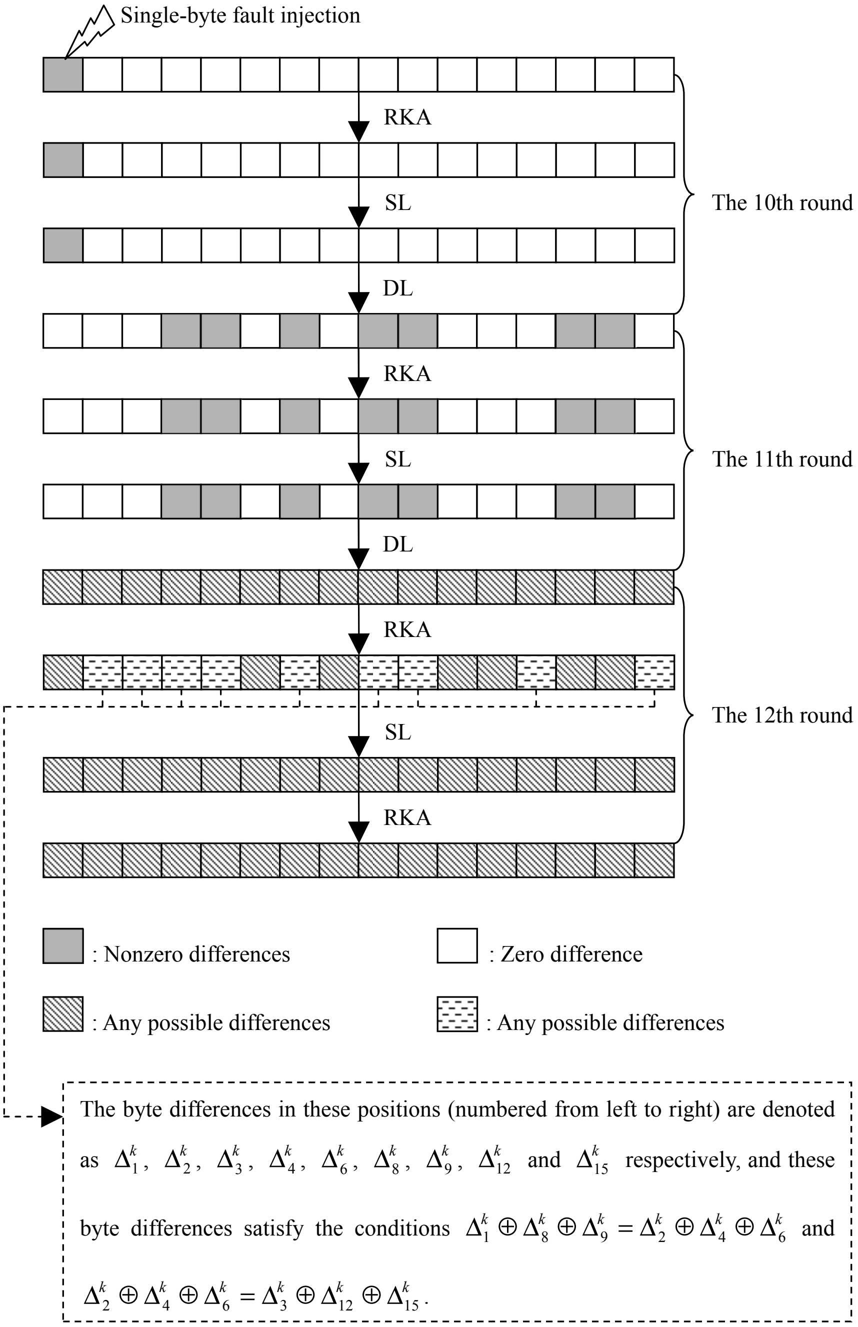
(c). Let $\Delta_{1,8,9}^k$, $\Delta_{2,4,6}^k$ denote $\Delta_1^k \oplus \Delta_8^k \oplus \Delta_9^k$ and $\Delta_2^k \oplus \Delta_4^k \oplus \Delta_6^k$ respectively. For each possible value of $\{ek_{13,1}, ek_{13,8}, ek_{13,9}\}$, $\Delta_{1,8,9}^k$'s $(1 \le k \le 6)$ are calculated for the above six ciphertext pairs, and the results are stored in a hash table with the input index coming from $\Delta_{1,8,9}^1 \| \Delta_{1,8,9}^2 \| \Delta_{1,8,9}^3 \| \Delta_{1,8,9}^4 \| \Delta_{1,8,9}^5 \| \Delta_{1,8,9}^6$ and the output being the value of $ek_{13,1} \| ek_{13,8} \| ek_{13,9}$. Then for each guess of $\{ek_{13,2}, ek_{13,4}, ek_{13,6}\}$, compute the values of $\Delta_{2,4,6}^k$'s $(1 \le k \le 6)$ for the six ciphertext pairs, and query in the above hash table by using $\Delta_{2,4,6}^1 \| \Delta_{2,4,6}^2 \| \Delta_{2,4,6}^3 \| \Delta_{2,4,6}^4 \| \Delta_{2,4,6}^5 \| \Delta_{2,4,6}^6$ as the search criteria, if a value of $ek_{13,1} \| ek_{13,8} \| ek_{13,9}$ is included in the query result, take this value together with the guessed value of $\{ek_{13,2}, ek_{13,4}, ek_{13,6}\}$ as the correct key information.

As a matter of fact, each equation in (8) holds with probability 2^{-8}, thus for each candidate of $\{ek_{13,1}, ek_{13,2}, ek_{13,4}, ek_{13,6}, ek_{13,8}, ek_{13,9}\}$, the equality

$$\Delta_{2,4,6}^1 \| \Delta_{2,4,6}^2 \| \Delta_{2,4,6}^3 \| \Delta_{2,4,6}^4 \| \Delta_{2,4,6}^5 \| \Delta_{2,4,6}^6 = \Delta_{1,8,9}^1 \| \Delta_{1,8,9}^2 \| \Delta_{1,8,9}^3 \| \Delta_{1,8,9}^4 \| \Delta_{1,8,9}^5 \| \Delta_{1,8,9}^6$$

holds with probability 2^{-48}. Since there are totally 2^{48} candidates of $\{ek_{13,1}, ek_{13,2}, ek_{13,4}, ek_{13,6}, ek_{13,8}, ek_{13,9}\}$, it is expected that only one candidate will be left as the correct key information after the above procedure, so we recover $ek_{13,1}$, $ek_{13,2}$, $ek_{13,4}$, $ek_{13,6}$, $ek_{13,8}$ and $ek_{13,9}$.

(d). For the correct value of $\{ek_{13,2}, ek_{13,4}, ek_{13,6}\}$ obtained before, derive the values of $\Delta_{2,4,6}^k$'s for the three ciphertext pairs (C_r, C_f^k) $(1 \le k \le 3)$. Let $\Delta_{3,12,15}^k$



The byte differences in these positions (numbered from left to right) are denoted as Δ_1^k, Δ_2^k, Δ_3^k, Δ_4^k, Δ_6^k, Δ_8^k, Δ_9^k, Δ_{12}^k and Δ_{15}^k respectively, and these byte differences satisfy the conditions $\Delta_1^k \oplus \Delta_8^k \oplus \Delta_9^k = \Delta_2^k \oplus \Delta_4^k \oplus \Delta_6^k$ and $\Delta_2^k \oplus \Delta_4^k \oplus \Delta_6^k = \Delta_3^k \oplus \Delta_{12}^k \oplus \Delta_{15}^k$.

Fig. 2. Schematic Description of the MitM Characteristics of ARIA

represent $\Delta_3^k \oplus \Delta_{12}^k \oplus \Delta_{15}^k$. Then for each guess of $\{ek_{13,3}, ek_{13,12}, ek_{13,15}\}$, calculate the values of $\Delta_{3,12,15}^k$'s for the three ciphertext pairs (C_r, C_f^k) $(1 \leq k \leq 3)$, and check whether the values of $\varDelta^1_{3,12,15}||\varDelta^2_{3,12,15}||\varDelta^3_{3,12,15}$ and $\varDelta^1_{2,4,6}||\varDelta^2_{2,4,6}||\varDelta^3_{2,4,6}$ are equal or not, if yes, keep the guessed value of $\{ek_{13,3}, ek_{13,12}, ek_{13,15}\}$ as the correct key information and discard it otherwise. In terms of the analysis similar to step 1(c), it is expected that only one candidate of $\{ek_{13,3}, ek_{13,12}, ek_{13,15}\}$ can pass the check and be left as the correct key information, so we retrieve $ek_{13,3}$, $ek_{13,12}$ and $ek_{13,15}$.

Step 2. Suppose that single-byte faults can be provoked on the third byte of the same intermediate state as that in step 1, then do as follows:

(a). Encrypt the plaintext P_r three times so as to collect three pairs of right and faulty ciphertexts (C_r, C^k_f), $7 \le k \le 9$.

(b). According to the characteristic (6), we can get

$$\varDelta^k_1 \oplus \varDelta^k_{13} \oplus \varDelta^k_{14} = \varDelta^k_0 \oplus \varDelta^k_4 \oplus \varDelta^k_6, \quad 7 \le k \le 9, \tag{10}$$

and

$$\varDelta^k_0 \oplus \varDelta^k_4 \oplus \varDelta^k_6 = \varDelta^k_3 \oplus \varDelta^k_{10} \oplus \varDelta^k_{11}, \quad 7 \le k \le 9. \tag{11}$$

(c). Thus with the help of the analysis similar to steps 1(c) and 1(d), we can recover the subkey bytes $ek_{13,0}$, $ek_{13,10}$, $ek_{13,11}$, $ek_{13,13}$ and $ek_{13,14}$.

Step 3. Assume that single-byte faults can be induced on the fourth byte of the same intermediate state as that in step 1, then do as follows:

(a). Do the encryptions twice for the plaintext P_r in order to obtain two pairs of correct and faulty ciphertexts (C_r, C^k_f), $10 \le k \le 11$.

(b). Based on the characteristic (7), we can acquire

$$\varDelta^k_0 \oplus \varDelta^k_{12} \oplus \varDelta^k_{15} = \varDelta^k_1 \oplus \varDelta^k_5 \oplus \varDelta^k_7, \quad 10 \le k \le 11. \tag{12}$$

(c). Then following the analysis similar to step 1(d), we can retrieve the subkey bytes $ek_{13,5}$ and $ek_{13,7}$. After that, all the bytes of ek_{13} are obtained.

Phase 2. Recovering the cipher key

Now we can decrypt the last round by using ek_{13} and mount an attack similarly to phase 1 and get all bytes of ek_{12}. Repeat this procedure until we retrieve ek_{11} and ek_{10}. After that, the cipher key can be recovered from ek_{10}, ek_{11}, ek_{12} and ek_{13} according to the key schedule of ARIA.

4.4 Theoretical Complexity of the Attack

The theoretical complexity of the above attack should be measured in three aspects, that is, data complexity, time complexity as well as memory complexity of the attack. To do so, we only need to evaluate the data, time and memory complexities required in Phase 1, from which the overall complexities of the attack could be deduced easily.

- **Data complexity of Phase 1.** It is measured in the number of faulty ciphertexts used in this phase. Since 6, 3, 2 faulty ciphertexts are needed in the steps 1, 2 and 3 respectively, the data complexity of phase 1 is 11 faulty ciphertexts.
- **Time complexity of Phase 1.** Among all the steps in phase 1, step 1(c) dominates the time complexity, thus the time complexity of phase 1 can be estimated as $2^{24} \times 6 \times \frac{3}{16\times12} \times 2 \approx 2^{21.58}$ ARIA encryptions plus $2^{24} \times \log_2 2^{24} \approx 2^{28.58}$ memory accesses. More specifically, the time complexity of step 1(c) mainly consists of two parts. One part corresponds to $2^{24} \times 6 \times 2$ partial decryptions which are used to derive $\Delta^k_{1,8,9}$ and $\Delta^k_{2,4,6}$ for all possible values of $\{ek_{13,1}, ek_{13,8}, ek_{13,9}\}$ and $\{ek_{13,2}, ek_{13,4}, ek_{13,6}\}$, $1 \le k \le 6$, and each partial decryption can be roughly approximated as $\frac{3}{16\times12}$ ARIA encryption. The other part is related to 2^{24} table lookups in a hash table with 2^{24} records, resulting in about $2^{24} \times \log_2 2^{24} \approx 2^{28.58}$ memory accesses.
- **Memory complexity of Phase 1.** Regarding the memory complexity of phase 1, it is primarily owing to storing the hash table in the step 1(c), consequently, it can be approximated as $2^{24} \times 3 \approx 2^{25.58}$ bytes.

Since we adopt the procedures similar to phase 1 to retrieve ek_{12}, ek_{11} and ek_{10} respectively, the data and time complexities of the attack could be estimated as four times those of phase 1 (Note that the time complexity of deriving the cipher key from ek_{10}, ek_{11}, ek_{12} and ek_{13} can be ignored compared with that of phase 1). Moreover, the memory complexity of the attack is just the same as that of phase 1 because the procedures for retrieving ek_{13}, ek_{12}, ek_{11} and ek_{10} could be implemented sequentially. Thus the overall data, time and memory complexities of the attack are about 44 faulty ciphertexts, $2^{23.58}$ ARIA encryptions plus $2^{30.58}$ memory accesses, $2^{25.58}$ bytes respectively.

4.5 Experiments and Results

We use a PC with Intel Pentium Dual Core E6500 processor (2.93 GHz) and 4G DDR memory to do the experiments of our key recovery attack on ARIA. The software platform of the experiments is Visual C++, and fault injections are simulated in this platform. Under this condition, we implement 500 experiments of our attack on ARIA with randomly generated cipher keys and divide all the experiments into 10 groups, each of which includes 50 experiments.

The main procedure of each experiment is as follows. Firstly, follow phase 1 to retrieve the round key ek_{13}. Generally, 11 faulty ciphertexts are sufficient for obtaining ek_{13}, but more faulty ciphertexts could be collected to derive ek_{13} if needed. Secondly, do similarly to recover the round keys ek_{12}, ek_{11} and ek_{10} respectively. Finally, the cipher key is deduced from ek_{10}, ek_{11}, ek_{12} and ek_{13} with the help of the key schedule of ARIA.

The experimental results show that our attack needs 44.39 faulty ciphertexts in average to recover the cipher key. Moreover, the average time and memory complexities of all the experiments are comparable to the theoretical ones given above. Since data complexity is the key factor for a fault attack with practical time and memory complexities, we only list the experimental results related to data requirements in our attack in Table 2. It can be seen that our experimental results match the theoretical analysis given in Section 4.4 well.

Table 2. Data Requirements in the Experiments of Our Attack on ARIA

Group No.	#AFC to recover ek_{13}	#AFC to recover ek_{12}	#AFC to recover ek_{11}	#AFC to recover ek_{10}	#AFC to recover the cipher key
1	11.06	11.08	11.14	11.14	44.42
2	11.10	11.12	11.08	11.04	44.34
3	11.16	11.08	11.06	11.08	44.38
4	11.12	11.10	11.10	11.08	44.40
5	11.08	11.14	11.04	11.06	44.32
6	11.16	11.06	11.14	11.12	44.48
7	11.14	11.04	11.08	11.10	44.36
8	11.08	11.20	11.10	11.06	44.44
9	11.06	11.08	11.06	11.10	44.30
10	11.10	11.06	11.18	11.12	44.46

#AFC represents the average number of faulty ciphertexts for a group of experiments.

5 MitM Fault Attack on AES

5.1 A Brief Introduction of AES

AES [28] is a word-oriented SPN block cipher with 8-bit word size and 128-bit block size. It accepts keys of 128, 192 or 256 bits and the number of rounds is 10, 12 or 14 respectively. Note that we will focus on **the AES cipher with 128-bit key size**.

In this paper, we treat the input and output of each round of AES as 16-byte vectors, and each byte within the vectors could be regarded as an element in $GF(2^8)$. These 16-byte vectors are also denoted as states or byte matrices of size 4×4, which can be shown as

$$\begin{pmatrix} \text{byte } 0 & \text{byte } 4 & \text{byte } 8 & \text{byte } 12 \\ \text{byte } 1 & \text{byte } 5 & \text{byte } 9 & \text{byte } 13 \\ \text{byte } 2 & \text{byte } 6 & \text{byte } 10 & \text{byte } 14 \\ \text{byte } 3 & \text{byte } 7 & \text{byte } 11 & \text{byte } 15 \end{pmatrix}.$$

Each round of AES consists of the following four basic operations (except the last round, where the MixColumns operation is replaced by an additional AddRoundKey operation):
AddRoundKey: XOR the 128-bit round key. All round keys are derived from the cipher key according to the key schedule.
SubBytes (Substitution Layer): Apply 16 non-linear 8×8-bit S-boxes to the 16 bytes of the intermediate vector respectively.
Diffusion Layer: It consists of the **ShiftRows** operation and the **MixColumns** operation, where the former cyclically shifts the i-th ($i = 0, 1, 2, 3$) row of the above byte matrix by i bytes to the left, and the latter multiplies the j-th ($j = 0, 1, 2, 3$) column of the above byte matrix with a constant 4×4 matrix over $GF(2^8)$.

Please refer to [28] for detailed information about the substitution layer, the diffusion layer and the key schedule adopted in AES.

5.2 MitM Characteristics of AES under Single-byte Fault Model

After investigating the diffusion layer adopted in AES in detail, we find that the cipher has two properties given in Section 3 as shown below.
(a). Let $(a_0, a_1, \ldots, a_{15})$, $(b_0, b_1, \ldots, b_{15}) \in GF(2^8)^{16}$ be the input and output of the diffusion layer of AES respectively. Then there exist 4 pairs of subscript sets

$$(V_1, V_1') = (\{0, 5, 10, 15\}, \{0, 1, 2, 3\}),$$

$$(V_2, V_2') = (\{3, 4, 9, 14\}, \{4, 5, 6, 7\}),$$

$$(V_3, V_3') = (\{2, 7, 8, 13\}, \{8, 9, 10, 11\}),$$

$$(V_4, V_4') = (\{1, 6, 11, 12\}, \{12, 13, 14, 15\})$$

such that each element in $\{b_v | v \in V_h'\}$ is the linear combination of all the elements in $\{a_v | v \in V_h\}$, where $1 \leq h \leq 4$.
(b). For a given i_0 ($1 \leq i_0 \leq 8$), suppose that ΔY_{i_0} satisfies

$$\Delta Y_{i_0} = (\text{0x00}, \Delta Y_{i_0,1}, \text{0x00}, \ldots, \text{0x00}),$$

where $\Delta Y_{i_0,1}$ is an unknown nonzero value. Then for $r_0 = 1$, there exist subscript sets V_1, V_2, V_3 and V_4 such that

$$|J_h| = 1, \quad 1 \leq h \leq 4,$$

where $J_h = \{j | \Delta Y_{i_0+r_0,j} \neq 0, j \in V_h\}$.

Thus for any two elements in $\{\Delta I_{i_0+2,v} | v \in V_h'\}$, we can easily obtain a linear equation on the two elements. More specially, suppose that single-byte faults can be injected on the second byte of the intermediate state before the diffusion layer in the 7th round, we can derive the following MitM characteristics of AES:

$$\Delta I_{9,1} = \Delta I_{9,0},\ \Delta I_{9,2} = 3\Delta I_{9,0},\ \text{and}\ \Delta I_{9,3} = 2\Delta I_{9,0}.$$

Then in terms of these MitM characteristics, we can mount an efficient key recovery attack on AES according to the attack procedure given in [23]. The data complexity of this attack is about 10 pairs of correct and faulty ciphertexts. The time and memory complexities of the attack are approximately 3×2^{40} AES encryptions and 2^{40} bytes, respectively. Please refer to Section 3.2 in [23] for more details about the attack procedure.

As a matter of fact, the above attack is the currently best fault cryptanalytic result on AES. This means that by using the observation given in Section 3, we could get some effective MitM characteristics of AES which can be used to mount a very efficient fault attack on the cipher.

6 Conclusion and Further Work

By exploring the security of word-oriented SPN block ciphers with the help of MitM fault analysis, we have observed that if there are some special properties in the diffusion layers of the ciphers, one can devise good strategies for MitM fault attacks on the ciphers under the condition of single-word fault model. In order to demonstrate the effectiveness of our observation, we have applied it to ARIA and AES respectively, and obtained some effective MitM characteristics by which efficient MitM fault attacks could be mounted on the ciphers.

Our results may be beneficial to fault attack on word-oriented SPN block ciphers as well as the design strategy of diffusion layers of such ciphers. Further work could be done to investigate whether our observation could be applied to Feistel ciphers with SP (or SPS) round functions.

References

1. D. Boneh, R.A. DeMillo and R.J. Lipton. On the Importance of Checking Cryptographic Protocols for Faults, EUROCRYPT 1997, LNCS 1233, 1997, pp. 37-51.
2. H. Bar-El, H. Choukri, D. Naccache, M. Tunstall and C. Whelan. The Sorcerer's Apprentice Guide to Fault Attacks, FDTC 2004 in association with DSN 2004, 2004, pp. 330-342.
3. J.M. Dutertre, A.P. Mirbaha, D. Naccache, A.L. Ribotta and A. Tria. Reproducible Single-Byte Laser Fault Injection, PASTIS 2010, 2010.
4. E. Biham and A. Shamir. Differential Fault Analysis of Secret Key Cryptosystems, CRYPTO 1997, LNCS 1294, 1997, pp. 513-525.
5. M. Rivain. Differential Fault Analysis on DES Middle Rounds, CHES 2009, LNCS 5747, 2009, pp. 457-469.

6. L. Hemme. A Differential Fault Attack against Early Rounds of (Triple-)DES, CHES 2004, LNCS 3156, 2004, pp. 254-267.
7. J. Blömer and J.-P. Seifert. Fault Based Cryptanalysis of the Advanced Encryption Standard, FC 2003, LNCS 2742, 2003, pp. 162-181.
8. C.-N. Chen and S.-M. Yen. Differential Fault Analysis on AES Key Schedule and Some Countermeasures, ACISP 2003, LNCS 2727, 2003, pp. 118-129.
9. P. Dusart, G. Letourneux and O. Vivolo. Differential Fault Analysis on AES, ACNS 2003, LNCS 2846, 2003, pp. 293-306.
10. C. Giraud. DFA on AES, International Conference Advanced Encryption Standard - AES 2004, LNCS 3373, 2004, pp. 27-41.
11. C.H. Kim and J.-J. Quisquater. New Differential Fault Analysis on AES Key Schedule: Two Faults Are Enough, CARDIS 2008, LNCS 5189, 2008, pp. 48-60.
12. G. Piret and J.-J. Quisquater. A Differential Fault Attack Technique against SPN Structures, with Application to the AES and Khazad, CHES 2003, LNCS 2779, 2003, pp. 77-88.
13. J. Takahashi, T. Fukunaga and K. Yamakoshi. DFA Mechanism on the AES Key Schedule, FDTC 2007, 2007, pp. 62-74.
14. C.H. Kim. Improved Differential Fault Analysis on AES Key Schedule, IEEE Transactions on Information Forensics and Security, 7(1), 2012, pp. 41-50.
15. S.S. Ali, D. Mukhopadhyay and M. Tunstall. Differential fault analysis of AES: towards reaching its limits, Journal of Cryptographic Engineering, 3(2), 2013, pp. 73-97.
16. C. Clavier, B. Gierlichs and I. Verbauwhede. Fault Analysis Study of IDEA, CT-RSA 2008, LNCS 4964, 2008, pp. 274-287.
17. H. Chen, W. Wu and D. Feng. Differential Fault Analysis on CLEFIA, ICICS 2007, LNCS 4861, 2007, pp. 284-295.
18. W. Li, D. Gu and Y. Wang. Differential Fault Analysis on the Contracting UFN Structure, with Application to SMS4 and MacGuffin, Journal of Systems and Software, 82(2), 2009, pp. 346-354.
19. W. Li, D. Gu and J. Li. Differential Fault Analysis on the ARIA Algorithm, Information Sciences, 10(178), 2008, pp. 3727-3737.
20. C.H. Kim. Differential Fault Analysis of ARIA in Multi-byte Fault Models, Journal of Systems and Software, 85(9), 2012, pp. 2096-2103.
21. Y. Zhou, W. Wu, N. Xu and D. Feng. Differential Fault Attack on Camellia, Chinese Journal of Electronics, 18(1), 2009, pp. 13-19.
22. W. Zhang, F. Liu, X. Liu and S. Meng. Differential fault analysis and meet-in-the-middle attack on the block cipher KATAN32, Journal of Shanghai Jiaotong University (Science), 18(2), 2013, pp. 147-152.
23. P. Derbez, P.-A. Fouque and D. Leresteux. Meet-in-the-Middle and Impossible Differential Fault Analysis on AES, CHES 2011, LNCS 6917, 2011, pp. 274-291.
24. R.C.-W. Phan and S.-M. Yen. Amplifying Side-Channel Attacks with Techniques from Block Cipher Cryptanalysis, CARDIS 2006, LNCS 3928, 2006, pp. 135-150.
25. Z. Liu, D. Gu, Y. Liu and W. Li. Linear Fault Analysis of Block Ciphers, ACNS 2012, LNCS 7341, 2012, pp. 241-256.
26. Y. Wei, P. Li, B. Sun, and C. Li. Impossible Differential Cryptanalysis on Feistel Ciphers with SP and SPS Round Functions, ACNS 2010, LNCS 6123, 2010, pp. 105-122.
27. National Security Research Institute, Korea. Specification of ARIA, Version 1.0, 2005.
28. J. Daemen and V. Rijmen. The Design of Rijndael: AES - The Advanced Encryption Standard, Springer, 2002.