Date: Thu, 15 Apr 1999 22:17:28 +0900 (JST)
From: shiho@sucaba.isl.ntt.co.jp (Shiho MORIAI)
To: AESFirstRound@nist.gov
cc: shiho@sucaba.isl.ntt.co.jp
Subject: an official comment for AESFirstRound

Dear Jim,

I'll submit the paper titled
``Security of E2 against Truncated Differential Cryptanalysis''
written by NTT Laboratories as an official comment on E2.

Please let me know if you have any problem to read it.

Best regards,

Shiho Moriai
NTT Laboratories

Security of E2 against Truncated Differential Cryptanalysis (in progress)

NTT Laboratories*

April 15, 1999

Abstract. This paper studies the security offered by E2 against truncated differential attack. At FSE'99, Matsui and Tokita presented a paper on this. They showed a possible attack on an 8-round variant of E2 without IT- and FT-Functions. To check their results and confirm that the full E2 is secure against this type of cryptanalysis, we developed a search algorithm to find all byte characteristics that lead to possible attacks on E2. As a result, we found another possible attack on an 8-round variant of E2 without IT- or FT-Function *with less complexity*. Moreover, we found that it is possible to distinguish a 7-round variant of E2 *with* IT- and FT-Functions from a random function. However, no flaw in the full E2 has been discovered by this type of cryptanalysis.

1 Introduction

Truncated differential cryptanalysis was introduced by Knudsen [2]. It deals with truncated differential, i.e. differentials where only a part of the difference can be predicted. Although the notion of truncated differential is wide, with a byte-oriented cipher it is natural to study bytewise differential characteristics as truncated differentials. Because the truncated differential can partly deal with the so called multi-path differential characteristics for a Markov cipher, the upper bound of the probabilities of *truncated differential characteristics* can be closer to that of *differentials*, the best measure of security against differential cryptanalysis [3], than that of *differential characteristics*. In other words, studying the security against truncated differential cryptanalysis provides a more accurate evaluation of the security against differential cryptanalysis.

*Contact to: Shiho Moriai (shiho@isl.ntt.co.jp)

The truncated differential cryptanalysis of reduced-round variants of E2 presented by Matsui and Tokita at FSE'99 [4] studies bytewise differential characteristics. Their analysis is based on the "byte characteristic," where the values to the difference in a byte are distinguished between non-zero and zero. Their analysis found a 7-round byte characteristic, which leads to a possible attack on an 8-round variant of E2 without IT- and FT-Functions. No flaw by the cryptanalysis above has been discovered for the full 12-round E2, i.e. E2 in the specification submitted to NIST as an AES candidate [5].

In order to check their results and confirm that the full E2 is secure against this type of cryptanalysis, we performed experiments to find all byte characteristics that lead to possible attacks on E2. As a result, we found another 7-round byte characteristic, which leads to a possible attack on an 8-round variant of E2 without IT- or FT-Function *with less complexity* than that offered by Matsui et al. Moreover, this byte characteristic is also useful in distinguishing a 7-round variant of E2 *with* IT- and FT-Functions from a random function.

This paper describes how we performed the byte characteristic search of E2. First, in Section 2, we describe the algorithm used to derive all possible byte characteristics of the round function, which Matsui et al. didn't go into details about in [4]. Second, we show a search algorithm for the byte characteristics of the whole cipher in Section 3. Section 4 describes possible scenarios of attacks on reduced-round variants of E2 and estimates their complexity. In Section 5, we introduce related works and show the future plan of this study. We also give some comments on Matsui et al.'s cryptanalysis.

2 Byte Characteristics of Round Function

This section studies the transition rules between the input and output bytewise differences of the round function of E2. Throughout this paper we follow the notations used in the specification of E2 [5] (see also Figure 1). The linear transformation in the round function (P-Function) is represented as follows.

$$
\begin{aligned}
z'_1 &= z_2 \oplus z_3 \oplus z_4 \oplus z_5 \oplus z_6 \oplus z_7 \\
z'_2 &= z_1 \oplus z_3 \oplus z_4 \oplus z_6 \oplus z_7 \oplus z_8 \\
z'_3 &= z_1 \oplus z_2 \oplus z_4 \oplus z_5 \oplus z_7 \oplus z_8 \\
z'_4 &= z_1 \oplus z_2 \oplus z_3 \oplus z_5 \oplus z_6 \oplus z_8 \\
z'_5 &= z_1 \oplus z_2 \oplus z_4 \oplus z_5 \oplus z_6 \\
z'_6 &= z_1 \oplus z_2 \oplus z_3 \oplus z_6 \oplus z_7 \\
z'_7 &= z_2 \oplus z_3 \oplus z_4 \oplus z_7 \oplus z_8 \\
z'_8 &= z_1 \oplus z_3 \oplus z_4 \oplus z_5 \oplus z_8
\end{aligned}
$$

Let $\Delta x \in \mathrm{GF}(2)^{64}$, $\Delta y \in \mathrm{GF}(2)^{64}$, and $\Delta z \in \mathrm{GF}(2)^{64}$ be the difference of the input of the round function, the difference of the output of the round

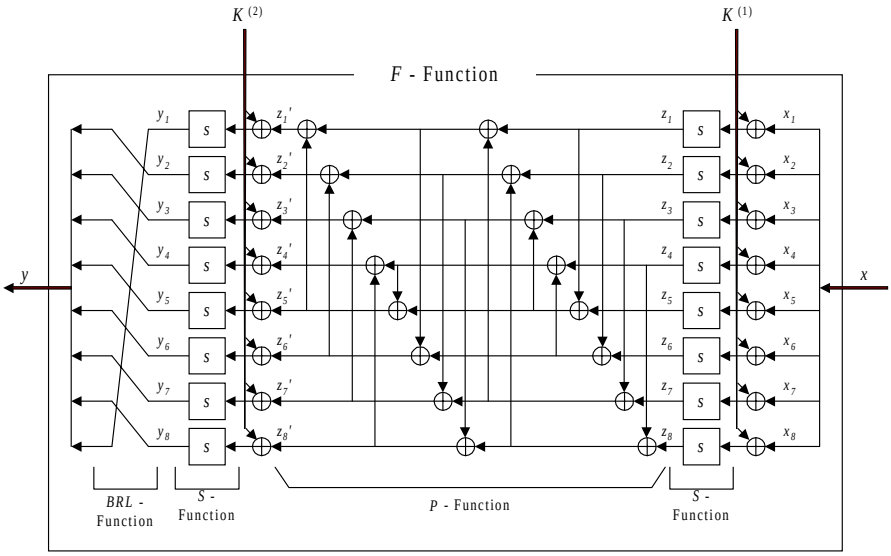

Figure 1: The round function of E2

function, and the difference of the input of P-Function, respectively.

$$\begin{aligned}
\Delta x &= (\Delta x_1, \Delta x_2, \ldots, \Delta x_8), & \Delta x_i &\in \mathrm{GF}(2)^8 \\
\Delta y &= (\Delta y_2, \ldots, \Delta y_8, \Delta y_1), & \Delta y_i &\in \mathrm{GF}(2)^8 \\
\Delta z &= (\Delta z_1, \Delta z_2, \ldots, \Delta z_8), & \Delta z_i &\in \mathrm{GF}(2)^8
\end{aligned}$$

For example, when two bytes of the input x_1 and x_5 are changed, if $\Delta z_1 = \Delta z_5$, then three bytes of the output y_2, y_6, and y_1 are changed. Otherwise (i.e., if $\Delta z_1 \neq \Delta z_5$) all bytes except y_7 are changed. Assuming that the input values $x_1, x_2, \ldots, x_8$ and the input differences Δx_1 and Δx_5 are given randomly (the other Δx_i's are fixed to 0 ($i \neq 1$ or 5)), the former happens with approximate probability 2^{-8} (though the exact value is $\frac{1}{255}$), and the latter happens with approximate probability $1 - 2^{-8}$. Following [4], we describe the transition rules above between the input and output bytewise differences as follows.

$$\begin{aligned}
(10001000) &\to (10001001) \quad p \approx 2^{-8} \\
(10001000) &\to (11111011) \quad p \approx 1 - 2^{-8}
\end{aligned}$$

We call these transition rules the byte characteristics of the round function. Formally, we define them as follows.

Definition 1 (χ-Function) *Let χ be the function $\mathrm{GF}(2)^8 \to \mathrm{GF}(2)$ defined as follows.*

$$\chi(x) = \begin{cases} 0 & \text{if } x = 0 \\ 1 & \text{if } x \neq 0 \end{cases}$$

Let $\boldsymbol{\chi}(x_1, x_2, \ldots, x_8) = (\chi(x_1), \chi(x_2), \ldots, \chi(x_8))$.

Definition 2 (Byte characteristic of round function) *Let $\delta x \in \mathrm{GF}(2)^8$ and $\delta y \in \mathrm{GF}(2)^8$ be defined as follows.*

$$\begin{aligned} \delta x &= (\delta x_1, \delta x_2, \ldots, \delta x_8), & \delta x_i &\in \mathrm{GF}(2) \\ \delta y &= (\delta y_2, \ldots, \delta y_8, \delta y_1), & \delta y_i &\in \mathrm{GF}(2) \end{aligned}$$

where

$$\begin{aligned} \delta x_i &= \chi(\Delta x_i), \\ \delta y_i &= \chi(\Delta y_i). \end{aligned}$$

The probability of the byte characteristic of the round function F is defined as follows:

$$p = \Pr_{x \in \mathrm{GF}(2)^{64}}[\, \boldsymbol{\chi}(F(x) \oplus F(x \oplus \Delta x)) = \delta y \mid \boldsymbol{\chi}(\Delta x) = \delta x \,],$$

where we assume that the output differences of s-box are uniformly distributed for any non-zero input difference. That is, p is the same for any Δx s.t. $\boldsymbol{\chi}(\Delta x) = \delta x$.

We define the triplet of δx, δy, and probability p be the byte characteristic of the round function. We represent it as follows:

$$\delta x \to \delta y \quad \text{with probability } p.$$

With the probability of the byte characteristic of the round function, the following theorem is easily proven.

Theorem 1 (Probability of byte characteristic of round function) *If $\delta x = (00000000)$, then $\delta y = (00000000)$ with probability 1. Otherwise, the non-zero probability of the byte characteristic of the round function can be approximated as follows. Note that $w_{\mathrm{H}}(\delta x)$ denotes the Hamming weight of δx.*

$$p \approx (2^{-8})^{w_{\mathrm{H}}(\delta x) - \dim_{\mathrm{GF}(2)} \langle \Delta x_1, \ldots, \Delta x_8 \rangle}$$

We need to derive all possible byte characteristics of the round function with non-zero probability to search exhaustively for the effective byte characteristics of the whole cipher. Simply listing all possible byte characteristics of the round function of E2 requires too much complexity. We used the algorithm below which requires less complexity.

Let $\{d_1, d_2, \ldots, d_8\}$ be a set of non-zero differences to $\Delta x_i \in \mathrm{GF}(2)^8$, and let $D = \#(\{\Delta x_1, \Delta x_2, \ldots, \Delta x_8\} \setminus \{0\})$. For example,

$$\text{if } \Delta x = (d_1, 0, d_1, d_2, d_3, 0, d_3, 0), \quad D = 3$$

In this algorithm, we classify byte characteristics according to D.

Algorithm 1 (Listing all possible byte characteristics of round function of E2 with non-zero probability)

1. For each D ($1 \leq D \leq 8$), list all linear relations that hold among $\{d_1, d_2, \ldots, d_D\}$. For example, when $D = 5$, all the linear relations that hold among $\{d_1, d_2, \ldots, d_5\}$ are as follows (there are 5). Note that all the linear relations where only the subscripts of d are permuted are regarded as the same one.

$$\begin{cases} \cdot \dim\langle d_1, d_2, d_3, d_4, d_5\rangle = 5 & d_1, \ldots, d_5 \text{ are linearly independent.} \\ \cdot \dim\langle d_1, d_2, d_3, d_4, d_5\rangle = 4 & \begin{cases} \cdot\ d_1 + d_2 = d_3,\ d_4 \text{ and } d_5 \text{ are linearly independent of the others.} \\ \cdot\ d_1 + d_2 + d_3 = d_4,\ d_5 \text{ is linearly independent of the others.} \\ \cdot\ d_1 + d_2 + d_3 + d_4 = d_5. \end{cases} \\ \cdot \dim\langle d_1, d_2, d_3, d_4, d_5\rangle = 3 & d_1 + d_2 = d_3 + d_4 = d_5. \end{cases}$$

2. For each linear relation above, take a set of any non-zero values for $\{d_1, d_2, \ldots, d_D\}$ which satisfy the linear relation. Then compute the output byte characteristics δy for every $\Delta x_i \in \{0, d_1, d_2, \ldots, d_D\}$. For the computation above, we use the following property: since the s-box is bijective, if the input difference of the s-box is zero, the output difference is zero, otherwise the output difference is non-zero. Thus we list all the obtained triplets "$\delta x \to \delta y$ with probability $p(> 0)$" as the byte characteristics of the round function, where probability p can be computed using Theorem 1.

3 Byte Characteristic Search of E2

This section finds all byte characteristics that lead to possible attacks on (reduced-round variants of) E2 using the byte characteristics of the round function obtained in the previous section.

Below we show a search algorithm for all "effective" byte characteristics of the 128-bit Feistel cipher with R rounds when all byte characteristics of the round function are known. In this paper, "effective" means that the byte characteristic could lead to possible attacks, in other words, the probability of

the byte characteristic is equal or higher than the probability with which the byte characteristic holds for a random permutation.

All byte characteristics of the round function should be sorted in order of the probability of byte characteristics for each input difference. This search algorithm is the depth first search rather than the breadth first search considering the required memory. The "depth" corresponds to the number of rounds of the Feistel cipher.

Algorithm 2 (Finding all effective byte characteristics of Feistel cipher with R rounds and blocksize 128 bits)

1. Let $X^{(r)} \in \mathrm{GF}(2)^8$ be the input byte characteristic of the r-th round function. Thus $(X^{(0)}, X^{(1)})$ is the byte characteristic of the plaintext. Let $\mathcal{P}$ be the probability of the byte characteristic. $\mathcal{P}$ is initialized to be 1, *i.e.*, $\mathcal{P} := 1$.

2. For each byte characteristic of the plaintext, (i.e., ${}^{\forall}X^{(0)} \in \mathrm{GF}(2)^8$ and ${}^{\forall}X^{(1)} \in \mathrm{GF}(2)^8$) call the procedure The 1st round, i.e., the procedure The r-th round for $r = 1$.

3. [The r-th round] For each $X^{(r)}$, set the output byte characteristic of the round function $Y^{(r)} \in \mathrm{GF}(2)^8$ in order of the probability of the byte characteristic.

   - Let $p_r := \Pr\{X^{(r)} \to Y^{(r)}\}$.
   - If $\mathcal{P} \times p_r < 2^{-128}$, then try another $Y^{(r)}$.
   - Call the procedure The r-th Xor.

   If $r \neq 1$, return to the procedure The $(r-1)$-st Xor,
   otherwise, exit the program.

4. [The r-th Xor] At the XOR operation of the r-th round in the Feistel cipher, $X^{(r+1)}$ is derived from $X^{(r-1)}$ and $Y^{(r)}$. Here the difference may be canceled out: $1 \oplus 1 = 0$ with probability $\frac{1}{255} (\approx 2^{-8})$, while $1 \oplus 1 = 1$ with probability $\frac{254}{255}$, assuming that the difference is independent and uniformly distributed. When the cancellation occurs in c bytes, the probability is approximately $(2^{-8})^c$. The number of all possible values of $X^{(r+1)}$ is $2^{w_{\mathrm{H}}(X^{(r-1)} \wedge Y^{(r)})}$. For each $X^{(r+1)}$, call the following procedure.

   - Let $\mathcal{P} := \mathcal{P} \times p_r \times (2^{-8})^c$.
   - If $\mathcal{P} < 2^{-128}$, then try another $X^{(r+1)}$.
   - If $\mathcal{P}$ is lower than the probability for a random function,
     i.e., if $\mathcal{P} < 2^{8 \times (w_{\mathrm{H}}(X^{(r)}) + w_{\mathrm{H}}(X^{(r+1)})) - 128}$, then try another $X^{(r+1)}$.

- If $r < R$, call the procedure THE $(r+1)$-ST ROUND, else print the byte characteristic:

$$(X^{(0)}, X^{(1)}) \to (X^{(R+1)}, X^{(R)}) \quad \text{with probability } \mathcal{P}.$$

Return to the procedure THE $(r-1)$-ST ROUND.

4 Attacks on Reduced-Round Variants of E2

The best[1] byte characteristic that leads to possible attacks on reduced-round variants of E2 is the 7-round byte characteristic shown in Figure 2. This 7-round byte characteristic holds with probability of about 2^{-104}; for a random function the probability of the byte characteristic is expected to be $(2^{-8})^{14} = 2^{-112}$, which is significantly smaller. Therefore, in a way similar to that described in [4], we can extract subkey information of the last round of an 8-round variant of E2 without FT-Function.

The number of required plaintext pairs is 2^{109}, which can be generated from 2^{94} chosen plaintext blocks ($94 = 109 - 16 + 1$). The attack on an 8-round variant of E2 without IT- and FT-Functions shown in [4] required 2^{100} chosen plaintext blocks. Moreover, we do not have to choose special plaintexts [4, Section 5.2] since the probability that correct pairs are detected is much larger than the probability that wrong pairs appear.

Moreover, this 7-round byte characteristic is useful in attacking a 7-round variant of E2 *with* IT- and FT-Functions. In IT- and FT-Functions, 32-bit multiplications with subkeys are used. Since this multiplication is modulo 2^{32} (roughly speaking, the upper 32-bit of the resultant 64-bit is discarded), this multiplication has the following trivial byte characteristic as shown in [4].

$$(1000) \to (1000) \qquad p = 1$$

Hence the 7-round byte characteristic shown in Figure 2 can skip IT- and FT-Functions with probability 1. Additionally, the positions of the bytes which have a non-zero difference are not changed by BP-Function (or BP^{-1}-Function) in IT-Function (or FT-Function). It follows that we have the following byte characteristic connecting the plaintext and ciphertext for a 7-round variant of E2.

$$(10001000\ 00000000) \to (10001000\ 00000000) \qquad p \approx 2^{-104}$$

This means that in a chosen plaintext scenario, we can distinguish the 7-round variant of E2 with IT- and FT-Functions from a random permutation. According to Matsui et al.'s theory, we then create 2^{106} plaintext pairs with the difference pattern (10001000 00000000) from 2^{91} plaintext blocks ($91 =$

[1] Here the "best" means that the ratio of the probability of the byte characteristic to the probability for a random permutation is the highest.

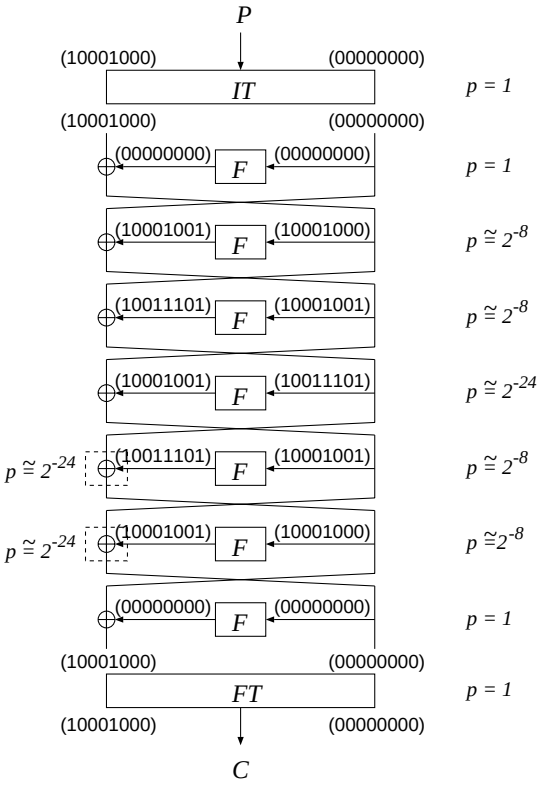

Figure 2: The best 7-round byte characteristic of E2

$106 - 16 + 1$) and encrypt them. If a ciphertext pair with the difference pattern (10001000 00000000) is found, we can regard it as the 7-round variant of E2 with IT- and FT-Functions, otherwise we regard it as a random permutation.

5 Discussion

Related works One of the results presented by Sugita et al. at the second AES conference [6] was on calculating the maximum average of differential probability of the SPN structure. By using their result, more accurate differential probability of the round function of E2 can be computed, the effects of the "multiple paths" being considered under some assumptions. The result using the strict values for the differential probabilities of the round function of E2 will be published on the E2 home page: `http://info.isl.ntt.co.jp/e2/`.

Another related work is on the method to find the best truncated differential characteristic shown by Vaudenay [7, 8]. He applied Floyd-Warshall's algorithm and performed the search effectively. Our search using the idea is ongoing.

On Matsui et al.'s cryptanalysis According to Matsui and Tokita [4, Section 6], if the byte permutation in the round function (BRL-Function) is modified, the security level of the modified version against differential cryptanalysis can be lower than the designers' estimation: the probability of the best 9-round differential characteristic is much smaller than $2^{-140.34}$. The designers' estimation of the security against differential cryptanalysis was based on the upper bound of the probabilities of *differential characteristics*, not differentials (nor truncated differentials). This is because we conjectured that the upper bound of the probabilities of differential characteristics is close to that of differentials when each round function is independent in Feistel ciphers [1]. Actually, this estimation works well for most cases including the real E2, as Matsui et al. showed. Moreover, the designers added 3 rounds (to 9 rounds) and IT- and FT-Functions so that E2 should have much more security.

6 Conclusion

We are studying the security of E2 against truncated differential cryptanalysis. In particular, in this paper we studied the bytewise differential cryptanalysis proposed by Matsui and Tokita [4]. The best attack that we found is an attack on an 8-round variant of E2 without IT- or FT-Function requiring 2^{94} chosen plaintexts. We also found that it is possible to distinguish a 7-round variant of E2 *with* IT- and FT-Functions from a random function using 2^{91} chosen plaintexts.

In spite of our severe examination, this type of cryptanalysis fails to break the full E2. We believe that this means that the full E2 offers strong security against truncated differential cryptanalysis. We will continue our study to confirm this.

References

[1] M.Kanda, Y.Takashima, and T.Matsumoto, "A round function structure consisting of few S-boxes (Part I)," (in Japanese), Technical report of IEICE, ISEC97-07, The Institute of Electronics, Information and Communication Engineers, 1997.

[2] L.R.Knudsen, "Truncated and Higher Order Differentials," Fast Software Encryption — Second International Workshop, Lecture Notes in Computer Science 1008, pp.196–211, Springer-Verlag, 1995.

[3] X.Lai, J.L.Massey, and S.Murphy, "Markov Ciphers and Differential Cryptanalysis," Advances in Cryptology — EUROCRYPT'91, Lecture Notes in Computer Science 547, pp.17–38, Springer-Verlag, 1991.

[4] M.Matsui and T.Tokita, "Cryptanalysis of a Reduced Version of the Block Cipher E2," in pre-proceedings of Fast Software Encryption'99, pp.70–79, 1999.

[5] Nippon Telegraph and Telephone Corporation, "Specification of E2 — a 128-bit Block Cipher," 1999, available at `http://info.isl.ntt.co.jp/e2/`.

[6] M.Sugita, K.Kobara, and H.Imai, "Pseudorandomness and Maximum Average of Differential Probability of Block Ciphers with SPN-Structures like E2," in proceedings of the second Advanced Encryption Standard candidate conference, pp.200–214, 1999.

[7] S.Vaudenay, "On the Security of CS-Cipher," in pre-proceedings of Fast Software Encryption'99, pp.259–274, 1999.

[8] S.Vaudenay, private communications, 1999.