Supplementary Material

Appendix A The mixture differential and the exchange attack

We give detailed descriptions of the main results in mixture differential cryptanalysis[7] and the exchange attack [10], and how they can be interpreted in our quadruple differential defined in Def. 2.

A.1 Grassi's mixture differential cryptanalysis

The property in Grassi's method [7] is revealed from the perspective of vectors and vector spaces over $\mathbb{F}_{28}^{4\times 4}$. The unit vectors of $\mathbb{F}_{28}^{4\times 4}$ is denoted by $\{e_{0,0},\cdots,e_{3,3}\}$ (e.g., $e_{i,j}$ has a single 1 in row i and column j). Then the column spaces can be denoted by the spaces spanned by the corresponding unit vectors as $C_i = \langle e_{0,i}, e_{1,i}, e_{2,i}, e_{3,i} \rangle$. For example,

$$\begin{split} \mathcal{C}_0 = & \langle e_{0,0}, e_{1,0}, e_{2,0}, e_{3,0} \rangle, \\ = & \left\{ \begin{bmatrix} x_0 & 0 & 0 & 0 \\ x_1 & 0 & 0 & 0 \\ x_2 & 0 & 0 & 0 \\ x_3 & 0 & 0 & 0 \end{bmatrix} \middle| \forall x_0, x_1, x_2, x_3 \in \mathbb{F}_{2^8} \right\}, \\ \equiv & \begin{bmatrix} x_0 & 0 & 0 & 0 \\ x_1 & 0 & 0 & 0 \\ x_2 & 0 & 0 & 0 \\ x_3 & 0 & 0 & 0 \end{bmatrix}. \end{split}$$

The diagonal spaces and the mixed spaces are defined as $\mathcal{D}_i = \mathsf{SR}^{-1}(\mathcal{C}_i)$ and $\mathcal{M}_i = \mathsf{MC}(\mathsf{SR}(\mathcal{C}_i))$. And for $J \subseteq \{0,1,2,3\}$, $\mathcal{M}_J = \oplus_{j \in J} \mathcal{M}_j$. For an element t that belongs to a coset of a subspace $\mathcal{X} \equiv \langle x_0, x_1, \cdots, x_{n-1} \rangle$, i.e., $t \in \mathcal{X} \oplus a$ for arbitraty a, say that t is generated by the generating variables $(t^0, t^1, \cdots, t^{n-1})$ (denoted by $t \equiv (t^0, t^1, \cdots, t^{n-1})$) if and only if $t = a \oplus \bigoplus_{i=0}^{n-1} t^i \cdot x_i$. The 4-round mixture differential property is given as follows [7, Thm3,Thm4]:

Theorem 1 ([7]) Given two plaintexts p^1, p^2 , generate two other plaintexts \tilde{p}^1, \tilde{p}^2 in the following two cases:

1. Given the subspace $C_0 \cap D_{0,3} \equiv \langle e_{0,0}, e_{1,0} \rangle \subseteq C_0$, two plaintexts $p^1, p^2 \in (C_0 \cap D_{0,3}) \oplus a$ where $p^1 \equiv (z^1, w^1), p^2 \equiv (z^2, w^2)$. Let $\tilde{p}^1, \tilde{p}^2 \in C_0 \oplus a$ be two other plaintexts generated by

$$\tilde{p}^1 \equiv (z^1, w^1, x, y), \tilde{p}^2 \equiv (z^2, w^2, x, y) \text{ or } \tilde{p}^1 \equiv (z^2, w^1, x, y), \tilde{p}^2 \equiv (z^1, w^2, x, y)$$

where x and y can take any possible value in \mathbb{F}_{28} ;

- 2. Given the subspace $C_0 \equiv \langle e_{0,0}, e_{1,0}, e_{2,0}, e_{3,0} \rangle$, two plaintexts $p^1, p^2 \in C_0 \oplus a$ where $p^1 \equiv (x^1, y^1, z^1, w^1), p^2 \equiv (x^2, y^2, z^2, w^2)$. Let $\tilde{p}^1, \tilde{p}^2 \in C_0 \oplus a$ be two other plaintexts generated by
 - $\begin{array}{lll} 1.(x^2,y^1,z^1,w^1) \ \ and \ (x^1,y^2,z^2,w^2); & 2.(x^1,y^2,z^1,w^1) \ \ and \ (x^2,y^1,z^2,w^2); \\ 3.(x^1,y^1,z^2,w^1) \ \ and \ (x^2,y^2,z^1,w^2); & 4.(x^1,y^1,z^1,w^2) \ \ and \ (x^2,y^2,z^2,w^1); \\ 5.(x^2,y^2,z^1,w^1) \ \ and \ (x^1,y^1,z^2,w^2); & 6.(x^2,y^1,z^2,w^1) \ \ and \ (x^1,y^2,z^1,w^2); \\ 7.(x^2,y^1,z^1,w^2) \ \ and \ (x^1,y^2,z^2,w^1). \end{array}$

The following event

$$R^4(p^1) \oplus R^4(p^2) \in \mathcal{M}_J$$
 if and only if $R^4(\tilde{p}^1) \oplus R^4(\tilde{p}^2) \in \mathcal{M}_J$ (A1)

holds with prob. 1 for 4-round AES, independently of the secret key, of the details of the Sbox and of the MixColumns matrix (except for the branch number equal to 5).

If concentrating on any byte position of the above-constructed plaintext quadruples, we will find that each byte quadruple contains two identical unordered pairs (including the case of four identical values). We use the following formal definition

Definition 1 (Mixture bytes) A byte quadruple is called the mixture bytes or a mixture byte quadruple if it contains two identical unordered pairs (including the case of four identical values).

In the mixture differential cryptanalysis, the first plaintext pair is chosen from a coset of a subspace. Determining how to construct the subspace and the way to prepare the second pair is the basic contribution in [7]. The property revealed in Eq. (A1) can only be used along with a condition that $R^4(p^1) \oplus R^4(p^2) \in \mathcal{M}_J$ where the index set J contains less than 4 elements. The probability of this condition being satisfied is estimated the same for both AES and a random permutation. So the distinguishing property on the 4-round ciphertexts is substantive that the differences in some columns of the two pairs after L^{-1} operation are both zero with a higher probability for AES than for a random permutation.

A.2 The exchange attack

The mixtures constructed in Th. 1 are limited to the case that only one column is active. In Bardeh *et al.*'s exchange attack [10], the activity in plaintexts is extended to diagonals. Moreover, the property is more explicit and is extended to five rounds. They use $\nu(s)$ to denote the binary indicator vector which is 1 in position *i* if the *i*-th column of $L^{-1}(s)$ is non-zero and 0 otherwise. Then the main property in the exchange attack is given as follows:

Theorem 2 ([10]) Let $\alpha, \beta \in \mathbb{F}_{28}^{4 \times 4}$ denote two plaintexts equal in |K| diagonals indicated by $K \subsetneq \{0, 1, 2, 3\}$ and assume $0 < wt(\nu(R^5(\alpha) \oplus R^5(\beta))) < 4$. For a non-trivial choice of $I \subsetneq \{0, 1, 2, 3\} \setminus K$, $(\tilde{\alpha}, \tilde{\beta})$ is a pair produced by exchanging the diagonals indicated by I of (α, β) . Then the relation

$$\nu(R^5(\alpha) \oplus R^5(\beta)) = \nu(R^5(\tilde{\alpha}) \oplus R^5(\tilde{\beta})) \tag{A2}$$

holds with probability

$$P_5(|I|,|K|) = \sum_{d=1}^{3} {4 \choose d} (2^{-8})^{4(|I|+d)-|K|d-2|I|d}.$$

Note that the assumption in Th. 2 that $0 < wt(\nu(R^5(\alpha) \oplus R^5(\beta))) < 4$ is satisfied with the same probability for both AES and a random permutation, so $P_5(|I|,|K|)$ gives the advantage for 5-round AES to satisfy Eq. (A2). It is estimated that $P_5(2,1) > 2^{-28.2}, P_5(1,2) > 2^{-38}$.

A commonality of the plaintext quadruples in the exchange attack with Grassi's method is that each byte quadruple also consists of two identical unordered pairs. The plaintext quadruples constructed in the second case in Th. 1 are special cases of exchanging type quadruples where $K=\varnothing$ and I takes all possible combinations. The difference between the two methods lies in that the plaintext quadruples constructed in the first case in Th. 1 cannot be constructed by merely exchanging bytes of a given pair. And the active positions in the exchange attack are no longer restricted to a column as in Grassi's method. So the mixture differential method and exchange attack are very related but neither can cover the other. Our work will unite both methods by taking plaintext quadruples consisting of all kinds of mixture bytes into account and predicting the distinguishing properties for as many rounds as possible.

A.3 Illustration in our form

Now we can express Grassi's mixture differential distinguisher and Bardeh $et\ al.$'s exchange attack in a unified way as a quadruple differential distinguisher. Grassi's mixture differential distinguisher in Th. 1 is that with plaintext quadruple pattern of

$$(\mathbf{Q}_{0,0}, \mathbf{Q}_{1,0}, \mathbf{Q}_{2,0}, \mathbf{Q}_{3,0}) \in \{ (\mathbf{c}, \mathbf{c}, \mathbf{s}, \mathbf{s}), (\mathbf{c}, \mathbf{x}, \mathbf{s}, \mathbf{s}) \}, (\mathbf{Q}_{0,j}, \mathbf{Q}_{1,j}, \mathbf{Q}_{2,j}, \mathbf{Q}_{3,j}) = (-, -, -, -), j = 1, 2, 3,$$
(A3)

$$(\mathbf{Q}_{0,0}, \mathbf{Q}_{1,0}, \mathbf{Q}_{2,0}, \mathbf{Q}_{3,0}) \in \{(\mathbf{x}, \mathbf{c}, \mathbf{c}, \mathbf{c}), (\mathbf{c}, \mathbf{x}, \mathbf{c}, \mathbf{c}), (\mathbf{c}, \mathbf{c}, \mathbf{x}, \mathbf{c}), (\mathbf{c}, \mathbf{c}, \mathbf{c}, \mathbf{c}, \mathbf{c}, \mathbf{c}), (\mathbf{c}, \mathbf{c}, \mathbf{c}, \mathbf{c}, \mathbf{c}), (\mathbf{c}, \mathbf{c}, \mathbf{c}, \mathbf{c}, \mathbf{c}, \mathbf{c}), (\mathbf{c}, \mathbf{c}, \mathbf{c}, \mathbf{c}, \mathbf{c}, \mathbf{c}, \mathbf{c}), (\mathbf{c}, \mathbf{c}, \mathbf{c}), (\mathbf{c}, \mathbf{c}, \mathbf{c$$

$$(\mathbf{Q}_{0,j}, \mathbf{Q}_{1,j}, \mathbf{Q}_{2,j}, \mathbf{Q}_{3,j}) = (-, -, -, -), j = 1, 2, 3,$$

after 4-round encryption the ciphertext quadruples $C^i = R^4(P^i) (i = 0, 1, 2, 3)$ satisfy $(L^{-1}(C^0), L^{-1}(C^1), L^{-1}(C^2), L^{-1}(C^3)) \triangleright \mathbf{Q}'$ with probability $2^{-32 \cdot |J|}$ where index set $J \subset \{0, 1, 2, 3\}, 1 \leq |J| \leq 3$, and

$$(\mathbf{Q}'_{0,j}, \mathbf{Q}'_{1,j}, \mathbf{Q}'_{2,j}, \mathbf{Q}'_{3,j}) = (\mathbf{s}, \mathbf{s}, \mathbf{s}, \mathbf{s}), j \in J$$
 (A5)

(both pairs collide on the *J*-th columns), while for a random permutation the probability is $2^{-64 \cdot |J|}$.

The exchange attack distinguisher in Th. 2 is that for non-empty index subsets $K \subset \{0,1,2,3\}, I \subset \{0,1,2,3\} \setminus K, |K| + |I| \leq 3$ with plaintext quadruple patterns being

$$(\mathbf{Q}_{0,k}, \mathbf{Q}_{1,(k+1) \mod 4}, \mathbf{Q}_{2,(k+2) \mod 4}, \mathbf{P}_{3,(k+3) \mod 4})$$

$$=(-, -, -, -), k \in K,$$

$$(\mathbf{Q}_{0,j}, \mathbf{Q}_{1,(j+1) \mod 4}, \mathbf{Q}_{2,(j+2) \mod 4}, \mathbf{Q}_{3,(j+3) \mod 4})$$

$$=(\mathbf{c}, \mathbf{c}, \mathbf{c}, \mathbf{c}), j \in \{0, 1, 2, 3\} \setminus (K \cup I),$$

$$(\mathbf{Q}_{0,i}, \mathbf{Q}_{1,(i+1) \mod 4}, \mathbf{Q}_{2,(i+2) \mod 4}, \mathbf{Q}_{3,(i+3) \mod 4})$$

$$=(\mathbf{x}, \mathbf{x}, \mathbf{x}, \mathbf{x}), i \in I,$$
(A6)

after 5-round encryption the ciphertext quadruples $C^i = R^5(P^i), (i = 0, 1, 2, 3)$ satisfy $(L^{-1}(C^0), L^{-1}(C^1), L^{-1}(C^2), L^{-1}(C^3)) \triangleright \mathbf{Q}'$ with probability $2^{-32 \cdot |J|} P_5(|I|, |K|)$ where index set $J \subset \{0, 1, 2, 3\}, 1 \leq |J| \leq 3$,

$$(\mathbf{Q'}_{0,j},\mathbf{Q'}_{1,j},\mathbf{Q'}_{2,j},\mathbf{Q'}_{3,j}) = (\mathbf{s},\mathbf{s},\mathbf{s},\mathbf{s}), j \in J$$

(both pairs collide on the *J*-th columns) and $P_5(1,2)>2^{-28.2}, P_5(2,1)>2^{-38}$ while for random permutation the probability is $2^{-64\cdot|J|}$.

Appendix B Inequality templates used in MILP model

We give all feasible point sets and the corresponding inequalities used in the modeling of AES. The inequalities concerning i variables consist of vectors of length (i+1). Each vector represents one inequality. Formally, vector $(a_0, a_1, \dots, a_{i-1}, b)$ represents inequality $a_0x_0 + a_1x_1 + \dots + a_{i-1}x_{i-1} + b \ge 0$.

Inequalities

```
(0,0,0,1,-1,1,0), (1,1,0,-1,0,0,0), (0,0,0,1,1,-1,0), (-1,-1,1,-1,1,1,1),\\ (1,-1,0,1,0,0,0), (-1,1,0,1,0,0,0), (1,1,1,-1,-1,-1,1), (1,0,-1,0,1,0,0),\\ (0,1,-1,0,0,1,0), (0,-1,1,0,0,1,0), (-1,0,1,0,1,0,1,0,0)
```

Table B2: Inequalities describing feasible patterns for $[deA_{\mathsf{Ind}}^{in}, e_{\mathsf{Ind}}^{out_k}, Ah^k, Al^k]$.

Inequalities

```
 \begin{array}{c} (1,2,3,3,1,2,-2,-1,-3,-1,-2,-1,-5,-3,0), (8,-14,-14,4,4,-10,-6,6,6,-1,-1,7,16,6,14), (-4,0,4,-2,2,-2,4,-1,-3,2,-2,3,6,4,0), (-10,-14,4,-14,4,8,7,6,-1,6,-1,-6,16,6,14), (-14,8,-14,4,-10,4,6,-6,6,-1,7,-1,16,6,14), (4,-6,4,8,4,-8,-2,4,-6,-5,-1,5,0,4,2), (-2,-4,2,0,-2,4,2,4,-2,-1,3,-3,6,4,0), (2,4,-2,-2,4,6,-1,-1,0,0,-1,-3,-6,-4,0), (4,2,-2,-4,0,-2,-3,-2,3,4,-1,2,6,4,0), (4,4,4,4,4,4,1,-1,2,2,3,-4,-4,-14,-10,0), (-2,4,4,2,4,-2,-2,-3,1,-2,0,1,-6,-2,0), (4,-2,4,4,-2,2,1,-2,3,0,1,-2,-6,-2,0), (0,0,0,-2,-2,-2,2,1,1,-2,1,2,2,4,2,2), (-4,-3,-2,-4,-3,-2,-1,3,3,2,3,-1,6,2,10), (-2,-2,-2,-2,-2,-2,-2,2,2,2,2,2,1,-1,1,3,2,7) \end{array}
```

Appendix C The exchange attack procedure

The pseudo-code for 6-round distinguishing attack from the exchange attack [10]. The 5-round attack pretty much follows the same line.

```
Algorithm 1: Pseudo-code for 6-round distinguisher.
```

```
Input: m^{\frac{1}{4}} values for each active byte. D=m^t where t is the number of
            active diagonals
   Result: 1 if AES, -1 otherwise
 1 L^{-1} \leftarrow SR^{-1} \circ MC^{-1}
    /* 3 diagonals as an example
 2 Chose m random values A = \{a_0, a_1, \cdots, a_{m-1}\} \subset \mathbb{F}_{28}^4
 3 Chose m random values B = \{b_0, b_1, \dots, b_{m-1}\} \subset \mathbb{F}_{2^8}^4
 4 Chose m random values C = \{c_0, c_1, \dots, c_{m-1}\} \subset \mathbb{F}_2^{\overline{4}}
 5 C \leftarrow \{\}
 6 T_0, T_1, T_2, T_3 = \{\} // empty hash tables containing unordered sets
        (e.g., unordered multisets)
    /* Encrypt m^t plaintexts
 7 for i from 0 to m-1 do
        for j from 0 to m-1 do
            for k from 0 to m-1 do
 9
                l \leftarrow i \cdot m^2 + j \cdot m + k
10
                p^l \leftarrow (a_i, b_i, c_k, z_3) // a_i is the fist diagonal value and b_i
11
                     is the second diagonal value and so forth.
                 c^l \leftarrow R^6(p^l)
12
                 /* T_r[z] contains indices (i,j,k) for ciphertext
                    e^{i\cdot m^2+j\cdot m+k} with value z in the collision columns of
                     L^{-1}(c^{i\cdot m^2+j\cdot m+k}). There are four cases.
                 for r from 0 to 3 do
13
                     z \leftarrow |L^{-1}(c^l)_r| // |L^{-1}(c^l)_r| is the integer value of
14
                         the collision column in the r-th case
                   T_r[z] \leftarrow T_r[z] \cup \{(i,j,k)\}
15
                C \leftarrow C \cup \{c^l\}
16
    /* Search for double collisions
17 for each c^i in C do
        /* coeffs(i) returns coefficients a,b,c s.t. a\cdot m^2 + b\cdot m + c = i
18
        i_1, j_1, k_1 \leftarrow coeffs(i)
        for j from 0 to 3 do
19
            for i_2, j_2, k_2 \in T_j[|c_i^i|] do
20
                /* G_{i,j} is the set of ciphertexts corresponding to new
                     pairs generated according to input patterns.
                 S \leftarrow G_{(i_1,j_1,k_1),(i_2,j_2,k_2)} // |G| is the number of input
21
                     patterns in Tab. 5
                 for each pair (a,b) \in S do
22
                     if |L^{-1}(R^6(a)) \oplus L^{-1}(R^6(b))| = 0 then
23
                         /* Two pairs forming double collision found
24
                         return 1
25 return -1
```

Points

Inequalities

Table B4: Inequalities describing feasible patterns for $[lab^{in}, dop^{in}, e^{out_k}_{lnd}]$.

Inequalities

Table B5: Feasible patterns and inequalities describing $[lab^{in}, dop^{in}, Ah^k, Al^k, minus]$.

Points

 $\begin{matrix} (0,0,1,1,1),\ (0,0,1,0,1),\ (0,0,0,1,0),\ (0,0,0,0,0),\ (0,1,0,0,0),\ (0,1,0,1,0),\ (0,1,1,1,0),\ (1,0,0,0,0),\ (1,0,0,1,0),\ (1,0,1,0,0),\ (1,0,1,1,0),\ (1,1,1,1,0),\ (1,1,1,1,0)\end{matrix}$

Inequalities

(0, -1, 0, 0, -1, 1), (0, 0, 1, 0, -1, 0), (1, 1, -1, 0, 1, 0), (-1, 0, 0, 0, -1, 1)

Table B6: Feasible patterns and inequalities describing $[e_{\mathsf{Ind}}, Bh, Bl]$.

Points

 $\begin{matrix} (1,1,1,1,1,1,0,0),\ (0,0,0,0,0,0,1,1),\ (1,0,1,1,0,1,1,0),\ (1,1,0,0,1,1,1,0),\ (0,1,1,1,1,0,1,0),\ (1,1,1,0,0,0,1,0),\ (1,0,0,1,1,0,1,0),\ (1,1,1,1,0,0,1),\ (1,1,1,1,0,1,0,1),\ (1,1,1,1,0,1,0,1),\ (0,1,0,1,0,1,1,0),\ (1,1,0,1,1,1,0,1),\ (1,0,1,1,1,1,0,1),\ (0,1,1,1,1,1,0,1),\ (0,0,1,0,1,1,1,0)\end{matrix}$

Inequalities

 $\begin{array}{c} (1,1,1,1,1,1,1,5,1,-6), (-2,-1,-1,-2,-2,-1,-8,-6,14), (1,0,1,1,0,1,2,2,-4), (2,0,-1,1,1,3,2,-5), (-1,1,2,1,2,0,3,2,-5), (0,-1,-1,0,0,-1,-3,-2,5), (-1,0,-1,0,-1,0,-1,0,-1,0,3), (-1,-1,0,-1,0,0,-1,0,3), (2,0,2,1,-1,1,3,2,-5) \end{array}$