On Maximum Non-averaged Differential Probability

Kazumaro Aoki

NTT Laboratories
1-1 Hikarinooka, Yokosuka-shi, Kanagawa-ken, 239-0847 Japan
maro@isl.ntt.co.jp

Abstract. Maximum average of differential probability is one of the security measures used to evaluate block ciphers such as the MISTY cipher. Here *average* means the average for all keys. Thus, there are keys which yield larger maximum differential probability even if the maximum average of differential probability is sufficiently small.
This paper presents the cases in which the maximum differential probability is larger than the maximum average of differential probability for some keys, and we try to determine the maximum differential probability considering the key effect.
Keywords: Differential cryptanalysis, linear cryptanalysis, differential, maximum average of differential probability, linear hull, maximum average of linear probability, maximum non-averaged differential probability, maximum non-averaged linear probability, DES-like cipher

1 Introduction

The security of symmetric key block ciphers against differential cryptanalysis [2] can be evaluated using several security measures. The maximum average of differential probability [5] is one such measure. We can regard that differential cryptanalysis fails for a block cipher if the maximum average of differential probability for the block cipher is sufficiently small. Thus, designers of a block cipher should guarantee that the maximum average of differential probability is sufficiently small. Some block ciphers were shown to have maximum averages of differential probability that were sufficiently small by Knudsen et al. [8,7].

It is important to note that *average* of the maximum average of differential probability means the average over all keys. That is, even if the maximum average of differential probability is sufficiently small, the block cipher may be insecure for some keys. Canteaut evaluated the maximum differential probability for all keys not just the *average* of all keys for some types of DES-like ciphers [3]. However, the proof of her main theorem was flawed.

This paper points out a flaw in the proof of [3], and extends the theorems that have correct proof to linear cryptanalysis [6]. Our conclusion is that inequalities similar to [1] hold if the number of rounds is less than 3. Moreover, we report experimental results. The results show that more rigorous inequalities may be proved on non-averaged differential probability for a specific F-function.

S. Tavares and H. Meijer (Eds.): SAC'98, LNCS 1556, pp. 118–130, 1999.

2 Preliminaries

We define the following.

1. $\mathrm{Prob}[A|B] = 0$ if B is an empty set.
2. Operations not specified here are as given in $\mathrm{GF}(2)^i$.
3. Suffixes L and R are the left and right half of the variable letter regarded as bit string, respectively.

We define the following notations.

$\mathrm{dom}\, f$	the domain of function f
$a \bullet b$	Even parity of bitwise logical AND operation of bit strings a and b
$\Delta_f(a,b)$	$\{x \mid f(x) + f(x+a) = b\}$
$\delta_f(a,b)$	$^{\#}\Delta_f(a,b)$
$\delta_f^{\max}$	$\max_{a \neq 0, b} \delta_f(a,b)$
ΔX	$X + X^*$
$\Lambda_f(a,b)$	$\{x \mid x \bullet a = f(x) \bullet b\}$
$\lambda_f(a,b)$	$2^{\#}\Lambda_f(a,b) - {}^{\#}\,\mathrm{dom}\, f$
$\lambda_f^{\max}$	$\max_{a,b \neq 0} \lambda_f(a,b)^2$
(a,b)	Concatenation of bit strings a and b
$\exp_b(e)$	b^e
$S(X)$	$\{f : X \to X \mid f$: bijective$\}$
$C(f)$	equivalence class of f
$T/\sim$	quotient set of set T with the equivalence relation $\sim$, i.e. $\{C(f) \mid f \in T\}$
$\mathcal{P}(T)$	power set of T

We define the precedence of operations as the following.

$$\Delta \succ \bullet \succ +$$

We define the cipher $E_{k_1,k_2,\ldots,k_r}$, the analysis target, as follows.

1. $L(i), R(i)$ (n-bit)
2. $Y(0)$: plaintext ($2n$-bit)
3. $(L(i), R(i)) = Y(i)$
4. $$\begin{cases} Z(i) = F_{k_{i+1}}(R(i)) \\ L(i+1) = R(i) \\ R(i+1) = L(i) + Z(i) \end{cases} \quad \text{for } 0 \leq i < r$$

We define F-function as $F_{k_{i+1}}(R(i)) = f(R(i) + k_{i+1})$, and f as bijective.

Moreover, we define

$$\delta_E^{\max} = \max_{k_1,k_2,\ldots,k_r} \delta^{\max}_{E_{k_1,k_2,\ldots,k_r}}$$
$$\lambda_E^{\max} = \max_{k_1,k_2,\ldots,k_r} \lambda^{\max}_{E_{k_1,k_2,\ldots,k_r}}$$

for the block cipher E.

The following lemma is useful for evaluating linear probability.

Lemma 1. *Following equation holds for n-input Boolean function* $f : \mathrm{GF}(2)^n \to \mathrm{GF}(2)$.

$$\frac{1}{2^n} \sum_{x \in \mathrm{GF}(2)^n} \exp_{-1}(f(x)) = 2 \operatorname*{Prob}_x[f(x) = 0] - 1$$

3 Previous Results

3.1 Averaged Case

Nyberg and Knudsen showed a bound of the maximum average of differential probability for $r \geq 3$ [8], and Matsui pointed out that a similar inequality holds in linear cryptanalysis using duality [7]; later Aoki and Ohta showed the strict bounds for bijective F-Function [1].[1]

Lemma 2 ([8,7,1]).

$$\max_{\alpha \neq 0, \beta} \operatorname*{Average}_{k_1, k_2, \ldots, k_r} \delta_{E_{k_1, k_2, \ldots, k_r}}(\alpha, \beta) \leq (\delta_f^{\max})^2 \quad \textit{if } r \geq 3$$

$$\max_{\alpha, \beta \neq 0} \operatorname*{Average}_{k_1, k_2, \ldots, k_r} \lambda_{E_{k_1, k_2, \ldots, k_r}}(\alpha, \beta) \leq \lambda_f^{\max} \quad \textit{if } r \geq 3$$

3.2 Non-Averaged Case

Canteaut showed some results of differential probability as dependent on keys [3].

Lemma 3 (2-round differential probability).

$$\operatorname*{Prob}_{Y(0), Y^*(0)}[\Delta Y(2) = \beta | \Delta Y(0) = \alpha] = \frac{\delta_f(\alpha_R, \alpha_L + \beta_L)\delta_f(\beta_L, \alpha_R + \beta_R)}{2^{2n}}$$

Lemma 4 (3-round differential probability with trivial 1-st round).

$$\operatorname*{Prob}_{Y(0), Y^*(0)}[\Delta Y(3) = \beta | \Delta Y(0) = (\alpha_L, 0)] = \frac{\delta_f(\alpha_L, \beta_L)\delta_f(\beta_L, \alpha_L + \beta_R)}{2^{2n}}$$

Lemma 5 (3-round differential probability with trivial 2-nd round).

$$\operatorname*{Prob}_{Y(0), Y^*(0)}[\Delta Y(3) = (\alpha_R, \beta_R) | \Delta Y(0) = \alpha] = \frac{\delta_f(\alpha_R, \alpha_L)\delta_f(\alpha_R, \beta_R)}{2^{2n}}$$

[1] Kaneko et al. showed similar inequalities for general F-function [4].

4 Extension to Linear Cryptanalysis

We can prove the lemmas in Sect. 3.2 for the case of linear probability using Lemma 1. We show the lemmas here and prove them in Appendices.

Lemma 6 (2-round linear probability).

$$2\Prob_{Y(0)}[Y(0) \bullet \alpha = Y(2) \bullet \beta] - 1$$
$$= \exp_{-1}(k_1 \bullet (\alpha_R + \beta_R) + k_2 \bullet (\alpha_L + \beta_L)) \times \frac{\lambda_f(\alpha_R + \beta_R, \alpha_L)\lambda_f(\alpha_L + \beta_L, \beta_R)}{2^{2n}}$$

Lemma 7 (3-round linear probability with trivial 3-rd round).

$$2\Prob_{Y(0)}[Y(0) \bullet \alpha = Y(3) \bullet (\beta_L, 0)] - 1$$
$$= \exp_{-1}(k_1 \bullet (\alpha_R + \beta_L) + k_2 \bullet \alpha_L) \times \frac{\lambda_f(\alpha_R + \beta_L, \alpha_L)\lambda_f(\alpha_L, \beta_L)}{2^{2n}}$$

Lemma 8 (3-round linear probability with trivial 2-nd round).

$$2\Prob_{Y(0)}[Y(0) \bullet \alpha = Y(3) \bullet (\beta_L, \alpha_L)] - 1$$
$$= \exp_{-1}(k_1 \bullet \alpha_R + k_3 \bullet \beta_L) \times \frac{\lambda_f(\alpha_R, \alpha_L)\lambda_f(\beta_L, \alpha_L)}{2^{2n}}$$

5 Computer Evaluations

5.1 Extension Trial to General Rounds

Canteaut evaluated DES-like ciphers with general rounds similar to Sect. 3.2 [3]. Unfortunately, the results contain errors. First, she described

$$P[\Delta Y(3) = (\beta_L, \beta_R) | \Delta Y(0) = (\alpha_L, \alpha_R), K_1 = k_1, K_2 = k_2, K_3 = k_3]$$
$$= \sum_d P[\Delta Z(2) = \beta_R + d | \Delta R(2) = \beta_L, \Delta Y(0) = (\alpha_L, \alpha_R),$$
$$K_1 = k_1, K_2 = k_2, K_3 = k_3]$$
$$\times \frac{\delta_f(\alpha_R, d + \alpha_L)\delta_f(d, \beta_L + \alpha_R)}{2^{2n}}$$

in [3, Theorem 1]. The conditional part of the probability formula of the right side of this equality misses $\Delta Y(2) = (d, \beta_L)$. So, this equality does not hold.[2] In addition, we believe that the induction part of the proof that she did not describe has the same error.

[2] Her probability formulas of [3] does not contain information on random variables, so her proofs are hard to understand. We did not understand the correctness of the proof of Propositions 2 and 3 in [3]. So, the proof of Propositions 2 and 3 may be flawed, however, we confirmed that the statements of these propositions are correct.

5.2 Preliminaries

We tried to prove [3, Theorem 1], but computer based approaches seemed feasible.

We used computers to evaluate the differential probabilities for all f in the case that n is small and for randomly generated f in the case that n is not small.

The evaluations consider all round keys and all bijective functions for f-function. Thus the evaluations are enormously complex.

Following lemmas are effective for decreasing the complexity of computer evaluations. Proofs are in Appendices.

Lemma 9. *We can obtain the same value which is an element of a set T of a measure $h : \mathcal{P}(\mathrm{GF}(2)^{2n} \times \mathrm{GF}(2)^{2n}) \to T$ by adjusting other keys for r-round cipher $Y(r) = E(Y(0))$ even if we change one of any even round key and one of any odd round key in an arbitrary manner, where the measure $h(\{(Y(0), Y(r))|Y(r) = E(Y(0))\})$ satisfies the following equation.*

$$\begin{aligned} &\forall \mu, \nu \ [h(\{(Y(0) + \mu, Y(r) + \nu)|Y(r) = E(Y(0))\}) \\ &= h(\{(Y(0), Y(r))|Y(r) = E(Y(0))\})] \end{aligned} \tag{1}$$

Corollary 1. *$\delta_E^{\max}$ is independent of (k_1, k_2).*

Corollary 2. *$\lambda_E^{\max}$ is independent of (k_1, k_2).*

These corollaries suggest that to evaluate an r-round cipher, using all round keys is not necessary; considering only $r - 2$ round keys is sufficient.

Moreover, since the evaluations consider all keys, it is sufficient to evaluate f or g if $\forall x[f(x) = g(x)+k]$ holds. We introduce the following equivalence relation for achieving this purpose.

Lemma 10.

$$f \sim g \overset{\text{def}}{\Leftrightarrow} \exists k \in \mathrm{GF}(2)^n, \forall x \in \mathrm{GF}(2)^n [f(x) = g(x) + k]$$

is an equivalence relation over $S(\mathrm{GF}(2)^n)$.

It is trivial using this equivalence relation to show that it is sufficient for considering a complete set of the representatives of $S(\mathrm{GF}(2)^n)/\sim$.

Lemma 11. *For any x_0, y_0,*

$$\{f \in S(\mathrm{GF}(2)^n)|f(x_0) = y_0\}$$

is a complete set of representatives in $S(\mathrm{GF}(2)^n)/\sim$.

Using this lemma, it is sufficient for us to consider only the elements of, for example, $f(0) = 0$ in $S(\mathrm{GF}(2)^n)$.

5.3 Experimental Results

We calculated the maximum differential probability of F-function $\delta_f^{\max}$ and the maximum differential probability of cipher $\delta_E^{\max}$ for at least 3-round ciphers using the lemmas of the previous sections.

We calculated all f for the case that number of bits of F-function is equal to 3,[3] and randomly generated f in the case that number of bits of F-function is greater than 3. We show the results in Tables 1 and 2.[4] *Ratio* here means the ratio of the number of F-functions which derives pairs $(\delta_f^{\max}, \delta_E^{\max})$ to all bijective functions (or all of randomly generated bijective functions in the case of Table 2).

In the tables, $*$ denotes the items that do not satisfy the evaluation inequality of Lemma 2, the maximum average of differential probability is replaced with maximum differential probability, $\frac{\delta_E^{\max}}{2^{2n}} \leq (\frac{\delta_f^{\max}}{2^n})^2$. These tables show that the inequality is 4.5 times looser for some F-functions. However, these tables also show that the maximum differential probability is smaller than $(\frac{\delta_f^{\max}}{2^n})^2$ for some F-functions.

We obtained the following interesting examples.

Example 1. The following example shows that the statement of [3, Theorem 1] does not hold.

$$\begin{cases} n = 4,\ r = 3 \\ f = \begin{pmatrix} 0\,1\,2\,3\,4\,5 & 6 & 7\,8 & 9\,10\,11\,12\,13\,14\,15 \\ 0\,1\,2\,3\,4\,8\,15\,11\,7\,12 & 6\,13 & 5\,10\,14 & 9 \end{pmatrix} \end{cases} \Rightarrow \begin{cases} \frac{\delta_f^{\max}}{2^4} = \frac{1}{4} \\ \frac{\delta_E^{\max}}{2^8} = 1.5 \times (\frac{\delta_f^{\max}}{2^4})^2 \end{cases}$$

Example 2. The following example shows that the maximum differential probability of cipher is less than the square of maximum differential probability of F-function.

$$\begin{cases} n = 3,\ r = 5 \\ f = \begin{pmatrix} 0\,1\,2\,3\,4\,5\,6\,7 \\ 0\,1\,3\,7\,6\,2\,5\,4 \end{pmatrix} \end{cases} \Rightarrow \begin{cases} \frac{\delta_f^{\max}}{2^3} = 1 \\ \frac{\delta_E^{\max}}{2^6} = \frac{1}{4} \times (\frac{\delta_f^{\max}}{2^3})^2 \end{cases}$$

Example 3. The following example shows that the maximum differential probability of cipher is 4.5 times greater than the square of maximum differential probability of F-function.

$$\begin{cases} n = 3,\ r = 5 \\ f = \begin{pmatrix} 0\,1\,2\,3\,4\,5\,6\,7 \\ 0\,1\,2\,6\,7\,5\,4\,3 \end{pmatrix} \end{cases} \Rightarrow \begin{cases} \frac{\delta_f^{\max}}{2^3} = \frac{1}{4} \\ \frac{\delta_E^{\max}}{2^6} = 4.5 \times (\frac{\delta_f^{\max}}{2^3})^2 \end{cases}$$

[3] We omit less than 3-bit F-function since its cases are trivial.
[4] We calculated cases of $n > 5$, however, we could not find an interesting case.

Table 1. Differential probability of an F-function and differential probability of a cipher ($n = 3$)

n	r	$\delta_f^{\max}$	$\delta_E^{\max}$	Ratio
3	3	2	4	4/15
		4	16	7/15
		8	64	4/15
	4	2	*8	12/45
		4	8	6/45
			16	14/45
			*32	1/45
		8	32	9/45
			64	3/45
	5	2	*10	2/45
			*12	4/45
			*18	6/45
		4	8	6/45
			16	12/45
			*32	2/45
			*48	1/45
		8	16	6/45
			32	3/45
			64	3/45
	6	2	*10	2/45
			*12	4/45
			*18	6/45
		4	12	6/45
			*18	12/45
			*32	2/45
			*48	1/45
		8	16	6/45
			32	3/45
			64	3/45
	7	2	*12	2/45
			*16	4/45
			*18	6/45
		4	12	6/45
			*18	12/45
			*32	2/45
			*48	1/45
		8	16	6/45
			32	3/45
			64	3/45

Table 2. Differential probability of an F-function and differential probability of a cipher ($n > 3$)

n	r	$\delta_f^{\max}$	$\delta_E^{\max}$	Ratio(%)
4	3	4	16	1.4
			*18	1.0
			*20	0.1
			*24	0.0
		6	36	49.2
		8	64	37.8
		10	100	8.7
		12	144	1.7
		16	256	0.1
5	3	4	*18	0.0
			*20	0.0
		6	36	21.0
		8	64	62.6
		10	100	14.7
		12	144	1.5
		14	196	0.1
		16	256	0.0
		18	324	0.0

There exists 16 differentials which have differentially weak keys in 1/32 key space, and all weak keys for each differential are different. Thus, a key which is in one half of the key space of the cipher is differentially weak.

6 Conclusion

This paper has extended the evaluation of maximum differential probability and maximum linear probability with keys for DES-like ciphers in the case that the F-function is bijective. As a result, strict evaluations we derived for 2-round and some 3-round ciphers. These results are the same as those gained by evaluating the maximum average of differential probability and the maximum average of linear probability, parameters were used as security measures against differential cryptanalysis and linear cryptanalysis, respectively.

Moreover, we have evaluated the maximum differential probability with keys over 3-round using computers. As a result, it is proved that there are cases in which the maximum differential probability is 4.5 times greater than the maximum average of differential probability.

There are three open problems.

1. obtaining general case evaluation
2. characterizing the F-functions whose maximum differential probability with keys is small
3. constructing design procedures of key scheduling which does not produce weak keys against differential cryptanalysis

References

1. K. Aoki and K. Ohta. Strict Evaluation of the Maximum Average of Differential Probability and the Maximum Average of Linear Probability. *IEICE Transactions Fundamentals of Electronics, Communications and Computer Sciences (Japan)*, Vol. E80-A, No. 1, pp. 2–8, 1997. (A preliminary version written in Japanese was presented at SCIS96-4A).
2. E. Biham and A. Shamir. Differential Cryptanalysis of DES-like Cryptosystems. *Journal of Cryptology*, Vol. 4, No. 1, pp. 3–72, 1991. (The extended abstract was presented at CRYPTO'90).
3. A. Canteaut. Differential cryptanalysis of Feistel ciphers and differentially δ-uniform mappings. In *Workshop on Selected Areas in Cryptography (SAC'97)*, pp. 172–184, Ottawa, Ontario, Canada, 1997. School of Computer Science, Carleton University, Entrust Technologies, and the Interac Association.
4. Y. Kaneko, S. Moriai, and K. Ohta. On Strict Estimation Method of Provable Security against Differential and Linear Cryptanalysis. In Y. Han, T. Okamoto, and S. Qing, editors, *Information and Communications Security — First International Conference ICICS'97*, Volume 1334 of *Lecture Notes in Computer Science*, pp. 258–268. Springer-Verlag, Berlin, Heidelberg, New York, 1997.
5. X. Lai, J. L. Massey, and S. Murphy. Markov Ciphers and Differential Cryptanalysis. In D. W. Davies, editor, *Advances in Cryptology — EUROCRYPT'91*, Volume 547 of *Lecture Notes in Computer Science*, pp. 17–38. Springer-Verlag, Berlin, Heidelberg, New York, 1991.

6. M. Matsui. Linear Cryptanalysis Method for DES Cipher. In T. Helleseth, editor, *Advances in Cryptology — EUROCRYPT'93*, Volume 765 of *Lecture Notes in Computer Science*, pp. 386–397. Springer-Verlag, Berlin, Heidelberg, New York, 1994. (A preliminary version written in Japanese was presented at SCIS93-3C).
7. M. Matsui. New Structure of Block Ciphers with Provable Security against Differential and Linear Cryptanalysis. In D. Gollmann, editor, *Fast Software Encryption, Third International Workshop, Cambridge, UK, February 1996, Proceedings*, Volume 1039 of *Lecture Notes in Computer Science*, pp. 205–218. Springer-Verlag, Berlin, Heidelberg, New York, 1996. (Japanese version was presented at SCIS96-4C).
8. K. Nyberg and L. R. Knudsen. Provable Security Against a Differential Attack. *Journal of Cryptology*, Vol. 8, No. 1, pp. 27–37, 1995. (A preliminary version was presented at CRYPTO'92 rump session).

A Proof of Lemma 6

We prove this lemma using Fig. 1.

$$
\begin{aligned}
&2 \Prob_{Y(0)}[Y(0) \bullet \alpha = Y(2) \bullet \beta] - 1 && (2)\\
&= \frac{1}{2^{2n}} \sum_{L(0),R(0)} \exp_{-1}(L(0) \bullet \alpha_L + R(0) \bullet \alpha_R + (L(0) + f(R(0) + k_1)) \bullet \beta_L \\
&\quad + (R(0) + f(L(0) + f(R(0) + k_1) + k_2)) \bullet \beta_R) && (3)\\
&= \frac{1}{2^{2n}} \sum_{L(0),R(0)} \exp_{-1}(R(0) \bullet (\alpha_R + \beta_R) + f(R(0) + k_1) \bullet \beta_L \\
&\quad + L(0) \bullet (\alpha_L + \beta_L) + f(L(0) + f(R(0) + k_1) + k_2) \bullet \beta_R) && (4)\\
&= \frac{1}{2^{2n}} \sum_{L(0),R(0)} \exp_{-1}(R(0) \bullet (\alpha_R + \beta_R) + f(R(0) + k_1) \bullet \beta_L \\
&\quad + (f(R(0) + k_1) + k_2) \bullet (\alpha_L + \beta_L) \\
&\quad + (L(0) + f(R(0) + k_1) + k_2) \bullet (\alpha_L + \beta_L) \\
&\quad + f(L(0) + f(R(0) + k_1) + k_2) \bullet \beta_R) && (5)\\
&= \frac{1}{2^{2n}} \sum_{R(0)} \exp_{-1}(R(0) \bullet (\alpha_R + \beta_R) + f(R(0) + k_1) \bullet \alpha_L + k_2 \bullet (\alpha_L + \beta_L)) && (6)\\
&\quad \times \sum_{L(0)} \exp_{-1}((L(0) + f(R(0) + k_1) + k_2) \bullet (\alpha_L + \beta_L) \\
&\quad + f(L(0) + f(R(0) + k_1) + k_2) \bullet \beta_R) \\
&= \frac{1}{2^{2n}} \sum_{R(0)} \exp_{-1}((R(0) + k_1) \bullet (\alpha_R + \beta_R) + f(R(0) + k_1) \bullet \alpha_L \\
&\quad + k_1 \bullet (\alpha_R + \beta_R) + k_2 \bullet (\alpha_L + \beta_L)) \\
&\quad \times \lambda_f(\alpha_L + \beta_L, \beta_R) && (7)
\end{aligned}
$$

$$= \exp_{-1}(k_1 \bullet (\alpha_R + \beta_R) + k_2 \bullet (\alpha_L + \beta_L)) \times \frac{\lambda_f(\alpha_R + \beta_R, \alpha_L)\lambda_f(\alpha_L + \beta_L, \beta_R)}{2^{2n}} \tag{8}$$

B Proof of Lemma 7

We prove this lemma using Fig. 2.

$$2 \Prob_{Y(0)}[Y(0) \bullet \alpha = Y(3) \bullet (\beta_L, 0)] - 1 \tag{9}$$

$$= \frac{1}{2^{2n}} \sum_{L(0),R(0)} \exp_{-1}(L(0) \bullet \alpha_L + R(0) \bullet \alpha_R + (R(0) + f(L(0) + f(R(0) + k_1) + k_2)) \bullet \beta_L) \tag{10}$$

$$= \frac{1}{2^{2n}} \sum_{L(0),R(0)} \exp_{-1}(L(0) \bullet \alpha_L + R(0) \bullet (\alpha_R + \beta_L) + f(L(0) + f(R(0) + k_1) + k_2) \bullet \beta_L) \tag{11}$$

$$= \frac{1}{2^{2n}} \sum_{R(0)} \exp_{-1}(R(0) \bullet (\alpha_R + \beta_L) + (f(R(0) + k_1) + k_2) \bullet \alpha_L) \times \sum_{L(0)} \exp_{-1}((L(0) + f(R(0) + k_1) + k_2) \bullet \alpha_L + f(L(0) + f(R(0) + k_1) + k_2) \bullet \beta_L) \tag{12}$$

$$= \frac{1}{2^{2n}} \sum_{R(0)} \exp_{-1}((R(0) + k_1) \bullet (\alpha_R + \beta_L) + f(R(0) + k_1) \bullet \alpha_L + k_1 \bullet (\alpha_R + \beta_L) + k_2 \bullet \alpha_L) \times \lambda_f(\alpha_L, \beta_L) \tag{13}$$

$$= \exp_{-1}(k_1 \bullet (\alpha_R + \beta_L) + k_2 \bullet \alpha_L) \times \frac{\lambda_f(\alpha_R + \beta_L, \alpha_L)\lambda_f(\alpha_L, \beta_L)}{2^{2n}} \tag{14}$$

C Proof of Lemma 8

We prove this lemma using Fig. 3.

$$2 \Prob_{Y(0)}[Y(0) \bullet \alpha = Y(3) \bullet (\beta_L, \alpha_L)] - 1 \tag{15}$$

$$= \frac{1}{2^{2n}} \sum_{L(0),R(0)} \exp_{-1}(L(0) \bullet \alpha_L + R(0) \bullet \alpha_R + (R(0) + f(L(0) + f(R(0) + k_1) + k_2)) \bullet \beta_L + (L(0) + f(R(0) + k_1) + f(R(0) + f(L(0) + f(R(0) + k_1) + k_2) + k_3)) \bullet \alpha_L) \tag{16}$$

$$= \frac{1}{2^{2n}} \sum_{L(0),R(0)} \exp_{-1}(R(0) \bullet (\alpha_R + \beta_L) + f(L(0) + f(R(0) + k_1) + k_2) \bullet \beta_L$$

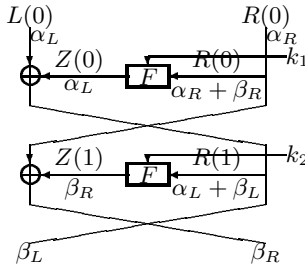

Fig. 1. 2-round linear probability

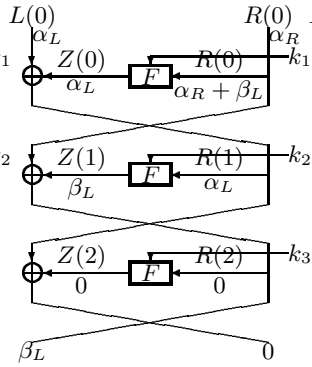

Fig. 2. 3-round linear probability with trivial 3-rd round

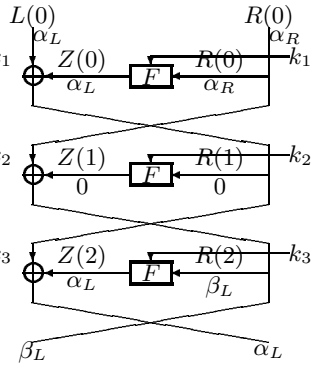

Fig. 3. 3-round linear probability with trivial 2-nd round

$$
\begin{aligned}
&+ f(R(0)+k_1) \bullet \alpha_L \\
&+ f(R(0)+f(L(0)+f(R(0)+k_1)+k_2)+k_3) \bullet \alpha_L) \qquad (17)\\
&= \frac{1}{2^{2n}} \sum_{R(0)} \exp_{-1}(R(0) \bullet (\alpha_R+\beta_L) + f(R(0)+k_1) \bullet \alpha_L + (R(0)+k_3) \bullet \beta_L) \\
&\quad \times \sum_{L(0)} \exp_{-1}((R(0)+f(L(0)+f(R(0)+k_1)+k_2)+k_3) \bullet \beta_L \\
&\quad + f(R(0)+f(L(0)+f(R(0)+k_1)+k_2)+k_3) \bullet \alpha_L) \qquad (18)\\
&= \frac{1}{2^{2n}} \sum_{R(0)} \exp_{-1}((R(0)+k_1) \bullet \alpha_R + f(R(0)+k_1) \bullet \alpha_L + k_1 \bullet \alpha_R + k_3 \bullet \beta_L) \\
&\quad \times \lambda_f(\beta_L, \alpha_L) \quad (\text{since } f\text{: bijective}) \qquad (19)\\
&= \exp_{-1}(k_1 \bullet \alpha_R + k_3 \bullet \beta_L) \times \frac{\lambda_f(\alpha_R, \alpha_L)\lambda_f(\beta_L, \alpha_L)}{2^{2n}} \qquad (20)
\end{aligned}
$$

D Proof of Lemma 9

We define transformed key $K^* = (k_1^*, k_2^*, \ldots, k_r^*)$ corresponding to original key $K = (k_1, k_2, \ldots, k_r)$.

We assume changing $(k_a, k_b) \mapsto (k_a^*, k_b^*)$ (a: odd, b: even). If we define

$$
k_i^* = \begin{cases} k_i + \Delta k_a & i\text{: odd} \\ k_i + \Delta k_b & i\text{: even} \end{cases}, \quad \mu = (\Delta k_b, \Delta k_a), \quad \text{and } \nu = \begin{cases} (\Delta k_a, \Delta k_b) & r\text{: odd} \\ (\Delta k_b, \Delta k_a) & r\text{: even} \end{cases},
$$

$E_{K^*}(Y(0)+\mu) = E_K(Y(0)) + \nu$ holds.

E Proof of Corollary 1

It is sufficient to prove that differential probability satisfies (1).

$$\forall \mu, \nu \underset{Y(0),Y^*(0)}{\mathrm{Prob}}[E(Y(0)+\mu)+\nu+E(Y^*(0)+\mu)+\nu=\beta|\Delta Y(0)=\alpha] \quad (21)$$

$$= \underset{Y(0),Y^*(0)}{\mathrm{Prob}}[E(Y(0)+\mu)+E(Y^*(0)+\mu)=\beta$$
$$|(Y(0)+\mu)+(Y^*(0)+\mu)=\alpha] \quad (22)$$

$$= \underset{Y(0)+\mu,Y^*(0)+\mu}{\mathrm{Prob}}[E(Y(0))+E(Y^*(0))=\beta|Y(0)+Y^*(0)=\alpha] \quad (23)$$

$$= \underset{Y(0),Y^*(0)}{\mathrm{Prob}}[E(Y(0))+E(Y^*(0))=\beta|\Delta Y(0)=\alpha] \quad (24)$$

F Proof of Corollary 2

It is sufficient to prove that linear probability satisfies (1).

$$\forall \mu, \nu\ (2\underset{Y(0)}{\mathrm{Prob}}[(Y(0)+\mu)\bullet\alpha=(E(Y(0))+\nu)\bullet\beta]-1)^2 \quad (25)$$

$$= (2\underset{Y(0)}{\mathrm{Prob}}[Y(0)\bullet\alpha=E(Y(0))\bullet\beta+(\mu\bullet\alpha+\nu\bullet\beta)]-1)^2 \quad (26)$$

$$= (2\underset{Y(0)}{\mathrm{Prob}}[Y(0)\bullet\alpha=E(Y(0))\bullet\beta]-1)^2 \quad (27)$$

G Proof of Lemma 10

Completeness $\forall k \in \mathrm{GF}(2)^n[x+k \in S(\mathrm{GF}(2)^n)]$ holds, and composite function of bijective functions is bijective.

Reflexive law If $k = 0 \in \mathrm{GF}(2)^n$, $\forall f \in \mathrm{GF}(2)^n[f(x) = f(x) + k]$ holds. So, $f \sim f$.

Symmetric law If $f \sim g$, $\exists k \in \mathrm{GF}(2)^n[f(x) = g(x) + k]$ holds. Thus, $g(x) = f(x) + k$ holds, i.e. $g \sim f$.

Transitive law If $f \sim g$ and $g \sim h$ holds, since $\exists k \in \mathrm{GF}(2)^n[f(x) = g(x) + k]$ and $\exists l \in \mathrm{GF}(2)^n[g(x) = h(x) + l]$ holds, then $f(x) = g(x) + k = (h(x) + l) + k = h(x) + (l + k)$ holds. So, since $l + k \in \mathrm{GF}(2)^n$ holds, $f \sim h$ holds.

H Proof of Lemma 11

We define $R = \{f \in S(\mathrm{GF}(2)^n)|f(x_0) = y_0\}$. We prove that R is a set of representatives. Let $f, g \in R$ and assume $C(f) = C(g)$. In this case, $f \sim g$ holds, and given the definition of R, $f(x_0) = g(x_0) = y_0$ holds. Thus, $f = g$ holds since $f(x) = g(x) + 0$.

We prove $C(f) = \{f(x) + k|k \in \mathrm{GF}(2)^n\}$. $g \in \{f(x) + k|k \in \mathrm{GF}(2)^n\} \Leftrightarrow \exists k \in \mathrm{GF}(2)^n[g(x) = f(x) + k] \Leftrightarrow g \sim f \Leftrightarrow g \in C(f)$.

$S(\mathrm{GF}(2^n)) \supseteq \bigcup_{f \in R} C(f)$ holds. On the other hand, since ${}^{\#}S(\mathrm{GF}(2^n)) = 2^n!$ holds and ${}^{\#}(\bigcup_{f \in R} C(f)) = {}^{\#}R \times {}^{\#}C(f) = (2^n - 1)! \times 2^n = 2^n!$ holds, so $S(\mathrm{GF}(2^n)) = \bigcup_{f \in R} C(f)$ holds. That is, R is a complete set of representatives in $S(\mathrm{GF}(2)^n)/\sim$.