Electronics and Communications in Japan, Part 3, Vol. 82, No. 10, 1999
Translated from Denshi Joho Tsushin Gakkai Ronbunshi, Vol. J81-A, No. 6, June 1998, pp. 988–995

On Linear Cryptanalysis of MBAL Ciphers

Kunio Kobayashi and Kazumaro Aoki

NTT Laboratories, Yokosuka, Japan 239-0847

SUMMARY

This paper studies the linear cryptanalysis of the MBAL (MultiBlock ALgorithm) Cipher. MBAL includes SXAL8 Cipher and 3-round Fm function. Moreover, the size of the input data is variable. In this paper, we propose the strategy of making linear approximations for MBAL which are independent of SXAL8 and the size of the input data. We succeed in making an approximation with bias $2^{-5.49}$ from the first to second (or second to third) rounds, where the bias is defined as $|Prob - 1/2|$. Next, we show how to reduce the number of "effective key bits" from 96 to 80 in the first (or n-th) round. Finally, we succeed in making an approximation for MBAL with bias $2^{-16.57}$, which gives 1 bit of information on the key. © 1999 Scripta Technica, Electron Comm Jpn Pt 3, 82(10): 1–8, 1999

Key words: Multiblock algorithm (MBAL) cipher; linear cryptanalysis; variable input data size; linear approximation; effective key bits.

1. Introduction

Linear cryptanalysis was proposed by Matsui in 1993 [1]. He broke the 8-round DES (Data Encryption Standard) with 2^{19} known plaintexts, and the full-round DES with 2^{43} known plaintexts [5].

This paper studies the linear cryptanalysis of the MBAL Cipher [2]. For the differential cryptanalysis of MBAL, a characteristic with probability 2^{-15} for 2-round Fm has been reported [4].

In this paper, we first focus on F_3, a part of Fm, and try making an approximation for 2-round F_3 (F_3-2R) with consideration of data reversal in section 3. Next, we try to reduce the number of effective key bits in the elimination round in section 4. Finally, in section 5 we try to make an approximation that gives 1 bit of information on the key by applying F_3-2R from the first to second Fm and approximating the third Fm.

2. Preparation

2.1. Notation

This section defines the notation. The leftmost bit or byte is the highest one in the following figures. The highest bit or byte is numbered 1, and the following bits or bytes are numbered 2, 3, ... Numbers in typewriter fonts (e.g., `45`) denote hexadecimal digits.

P:	Plaintext (its size is variable)
C:	Ciphertext (its size is variable)
KO, KI:	Subkey (8 bytes each)
$K6$, $K7$, $K8$:	Subkey (4 bytes each)
S_j^i:	The j-th sbox in the i-th Fm
$A[i]$:	The i-th bit of A, where A is any binary vector
$A \bullet B$:	Bitwise logical AND of A and B
$A \oplus B$:	Bitwise exclusive OR of A and B
$A[i_1, i_2, \ldots, i_k]$:	$A[i_1] \oplus A[i_2] \oplus \ldots \oplus A[i_k]$
ΓA:	Mask value of A
p'_x:	Bias for probability p_x ($p'_x = \|p_x - 1/2\|$)

CCC1042-0967/99/100001-08

2.2. Principle of linear cryptanalysis [1, 2]

We make a linear approximation with significant probability ($p'_x \neq 0$) such as that in Eq. (1), and we use it to derive $K[k_1, k_2, \ldots, k_c]$:

$$P[i_1, i_2, \ldots, i_a] \oplus C[j_1, j_2, \ldots, j_b] = K[k_1, k_2, \ldots, k_c] \quad (1)$$

Hereafter, we call the referenced bit positions [e.g., i_1, i_2, . . . , i_a of P in Eq. (1)] *effective bit positions*.

The *mask value* is a binary vector whose bits are defined as 1 when the bit is an effective position, and otherwise 0. The mask value is denoted by "Γ," such as ΓP or ΓC. Using these notations, Eq. (1) is rewritten as

$$P[\Gamma P] \oplus C[\Gamma C] = K[\Gamma K]$$

where $P[\Gamma P]$ is defined as the parity value of $P \bullet \Gamma P$.

If Eq. (1) holds with significant probability p, then $p \neq 1/2$.

Generally, to derive the key efficiently for n-round cryptosystems with iteration function $F(X, K)$, we approximate from the first to $(n - 1)$-th or from the second to n-th round and put the F-function in the remaining round, as in Eqs. (2) and (3). This technique is called the *one-round elimination technique*. We call the round in which we put the F-function the *elimination round*:

$$P[\Gamma P] \oplus C[\Gamma C] \oplus F(P, K^1)[\Gamma X] = K[\Gamma K] \quad (2)$$

$$P[\Gamma P'] \oplus C[\Gamma C'] \oplus F^{-1}(C, K^n)[\Gamma Y] = K[\Gamma K'] \quad (3)$$

Here K^1 or K^n are the targets to be derived (K^i is the subkey for the i-th round). ΓX in Eq. (2) is the input mask value of the linear approximation from the second to n-th round. Similarly, ΓY in Eq. (3) is the output mask value of the linear approximation from the first to $(n - 1)$-th round.

Hereafter, we assume that the linear approximation probability of each sbox is independent. We call the approximation that gives the maximum bias the *best linear approximation*, and call this maximum bias the *best bias*.

2.3. MBAL [2]

2.3.1. Overview of MBAL

MBAL is a registered ISO cipher.* MBAL consists of Fm (which will be described later) and an SXAL8 block cipher.** The size of MBAL's input and output data is variable and the size of the key is 8 bytes.

MBAL lets SXAL8 transform the highest 4 bytes and the lowest 4 bytes only. The other bytes are independent of SXAL8.

The flow chart of MBAL is shown in Fig. 1.

2.3.2. The sbox

The sbox permutes the input data (8 bits) to the output data (8 bits) by means of a fixed data table. The sbox is a part of f (described later).

2.3.3. Base algorithm f

The flow chart of f is shown in Fig. 2. P_f is divided into four single bytes. P_f is computed from t1 to t4, ta, tb, and the sboxes (t1 to t4, ta, and tb are input parameters). C_f is the output data of f, and u1 to u4, ua, and ub are output parameters.

2.3.4. Extended base algorithm Fm

Fm consists of m pieces of f which are arranged as CBC mode. The input and output data size of Fm is $4m$ bytes. The i-th f from the leftmost one in Fm is described as f^i. The input parameters of f^i are described as $t1^i$ to $t4^i$, ta^i, tb^i. Similarly, the output parameters of f^i are described as $u1^i$ to $u4^i$, ua^i, ub^i. Consequently, the connections are described as follows:

$t1^1 = Kx1$ $\quad$ $ta^1 = 0$
$t2^1 = Kx2$ $\quad$ $tb^1 = Kx4$
$t3^1 = Kx3$
$t4^1 = Kx4$

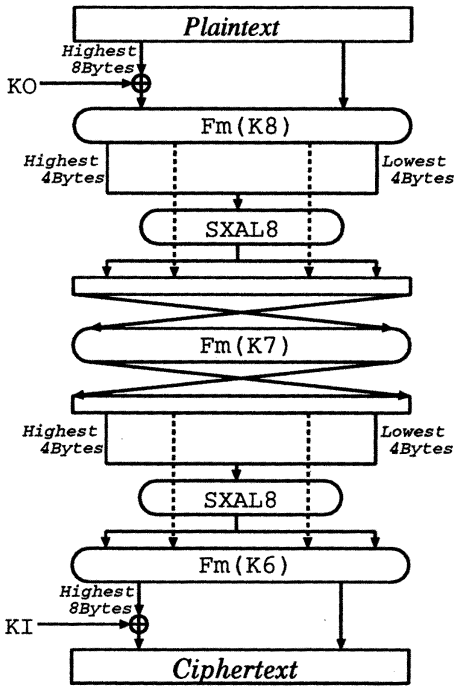


Fig. 1. Flow chart of MBAL.

* ISO/IEC9979-12.

** SXAL8 is an 8-round cryptosystem. The size of SXAL8's input and output data is 8 bytes, respectively.

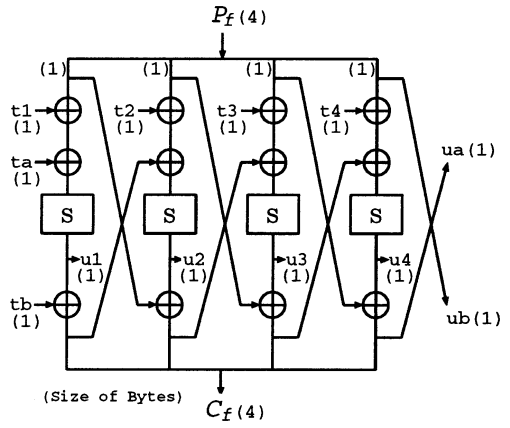


Fig. 2. Base algorithm f.

$t1^{i+1} = u2^i$ $\quad ta^{i+1} = ua^i$
$t2^{i+1} = u4^i$ $\quad tb^{i+1} = ub^i$
$t3^{i+1} = u1^i$
$t4^{i+1} = u3^i$ $\quad (1 \le i < m)$

In the above descriptions, Kx means the input subkey of Fm. Kx is one of ($K6$, $K7$, $K8$). In this paper, Kx (4 bytes) is represented as $Kx = \{Kx1, Kx2, Kx3, Kx4\}$ by dividing it into four 1-byte units.

3. Linear Approximation of MBAL

One of our goals is to make linear approximations for MBAL that are independent of SXAL8, so that our target

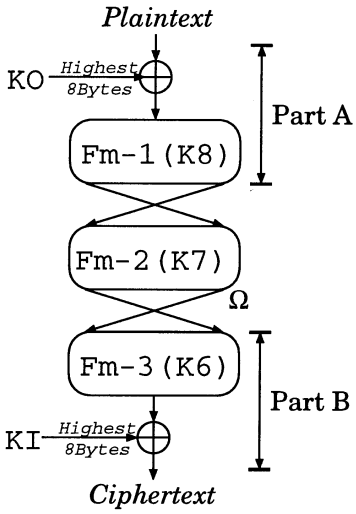


Fig. 3. MBAL without SXAL8.

Table 1. Results of approximating MBAL's sbox

Bias	Number of samples
0.500	1
0.375	4
0.281	4
0.250	64
0.188	128
0.125	1,212
0.094	376
0.063	6,016
0.031	8,836
0	48,895
Total	65,536

MBAL is equal to MBAL without SXAL8 (Fig. 3). The MBAL without SXAL8 is as follows:

Fm → data reversal → Fm → data reversal → Fm.

We attempt the linear cryptanalysis of MBAL without SXAL8 (Fig. 3). In Fig. 3, Fm-i is the i-th Fm, and Ω is the data after Fm-2 and before data reversal.

3.1. Linear approximation of sbox

A method for a linear approximation of sbox in DES was reported in Ref. 1. We make linear approximations of sbox in MBAL using the same method. Only the results are described (Table 1). In Table 1, the best bias 0.375(= 3/8) is given by the following mask value pairs,* where α is the input mask value and β is the output mask value:

(α, β) = (6B, BB), (8E, EE), (B0, D6), (E0, 63)

3.2. Linear approximation of F_3

Our next goal is to make linear approximations of MBAL that are independent of the size of MBAL's input data. We focus on F_3, which adjoins three fs in Fm. F_3 is shown in Fig. 4. In Fig. 4, we number the sboxes from the left for convenience. S_1 may possibly be the leftmost sbox in Fm. Fm is a function in which the F_3s are arranged in CBC mode. Therefore, the output parameter of a higher F_3 is the input parameter of F_3, and the output parameter of F_3 is the input parameter of a lower F_3.

In this section, we try to make linear approximations of F_3 that are independent of the input parameter and the

*The case of bias = 0.5 in Table 1 is given by only (α, β) = (00, 00), which is trivial.

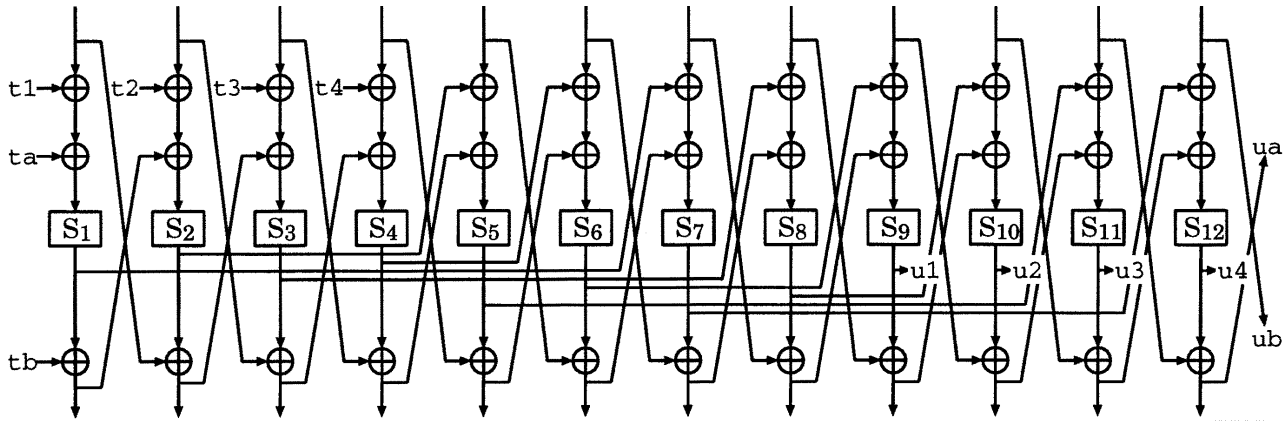


Fig. 4. F_3.

output parameter, because we want to find approximations that are independent of the size of MBAL's input data. For that reason, $\Gamma t1$ to $\Gamma t4$, Γta, $\Gamma u1$ to $\Gamma u4$, Γua, Γub must be 0. Γtb may not be 0, because we can make linear approximations without approximating sboxes in a higher f even if $\Gamma tb \neq 0$.

We must not approximate S_1 to S_4 in Fig. 4, since we make linear approximations of MBAL that are independent of the size of MBAL's input data. As a result we can approximate only S_5 to S_{12}.

Section 3 stated that our target is

Fm → data reversal → Fm → data reversal → Fm

Accordingly, we arrange F_3 vertically considering data reversal as shown in Fig. 5. We call Fig. 5 *F_3-2R*. In Fig. 5, the suffix "i" means the first round, "ii" the second round, and values in parentheses are described later.

We found four best linear approximations of F_3-2R using the above conditions. The four best linear approximations approximate six sboxes: S_5^i, S_8^i, S_{10}^i, S_6^{ii}, S_9^{ii}, and S_{11}^{ii}.

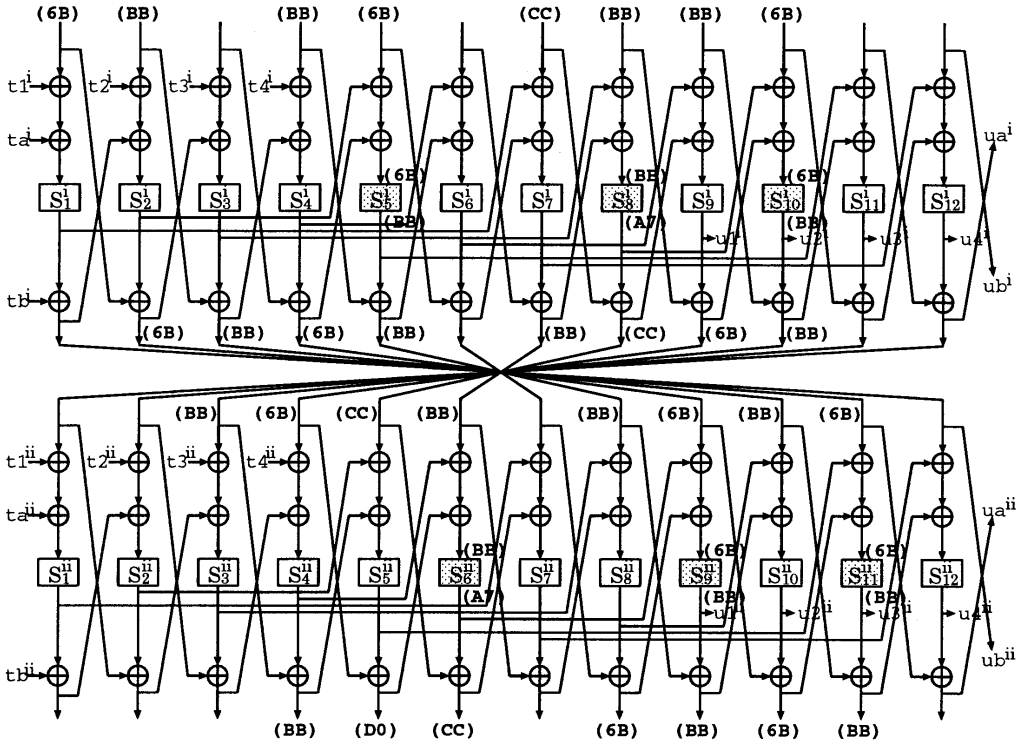


Fig. 5. "F_3-2 rounds (F_3-2R)" using pattern 4 (bias = $2^{-5.49}$).

Table 2. Mask value of sbox used by the best approximation of F_3-2R

	S^{i}_{5}	S^{i}_{8}	S^{i}_{10}	S^{ii}_{6}	S^{ii}_{9}	S^{ii}_{11}
Case 1	(8E,EE)	(EE,9F)	(8E,EE)	(EE,9F)	(8E,EE)	(8E,EE)
Case 2	(8E,EE)	(EE,F9)	(8E,EE)	(EE,F9)	(8E,EE)	(8E,EE)
Case 3	(6B,BB)	(BB,7A)	(6B,BB)	(BB,7A)	(6B,BB)	(6B,BB)
Case 4	(6B,BB)	(BB,A7)	(6B,BB)	(BB,A7)	(6B,BB)	(6B,BB)
Bias	3/8	3/16	3/8	3/16	3/8	3/8

The mask values of the sboxes are shown in Table 2. Cases 1 to 4 give the same bias value $P'_{F_3^2}$. $P'_{F_3^2}$ is

$$p'_{F_3^2} = 2^{6-1} \times \frac{3}{8} \times \frac{3}{16} \times \frac{3}{8} \times \frac{3}{16} \times \frac{3}{8} \times \frac{3}{8} = \frac{729}{32768} \approx 2^{-5.49} \quad (4)$$

The values in parentheses in Fig. 5 are the input and output mask values of F_3-2R in case 4. Similarly, the input and output mask values of F_3-2R in cases 1 to 4 are shown in Table 3. For each case of Table 3, the upper row is the input mask value and the lower row is the output mask value; blank columns mean a mask value of 0. "No." indicates the i-th byte in F_3-2R. The upper row of No. 1 is the input mask value and the lower row of No. 1 is the output mask value of the highest byte in F_3-2R. In the same way, we found a linear approximation of bias = $2^{-10.66}$ in F_2-2R (8 bytes).

4. Reducing the Number of Effective Key Bits

In this section, we attempt to reduce the number of effective key bits in the elimination round, when F_3-2R is applied to the first to $(n-1)$-th or second to n-th round. We study Parts A and B in Fig. 1 below. Both Parts A and B need 12 bytes of subkey information (Part A: KO, $K8$; Part B: KI, $K6$). This means that we need 96 bits of key information and the number of candidates is 2^{96}. These 96 bits are the targets to be attacked. However, it is difficult to overcome the enormous complexity. Moreover, this complexity is higher than the complexity of exhaustive search.*

We try reducing the size of the target keys. We can then bundle the multiple subkeys into one, reducing the bit size of the target keys from 96 to 80. Equivalent circuits of Parts A and B are shown in Figs. 6 and 7, respectively. In Figs. 6 and 7, the subkeys are displayed as 1-byte units as follows:

$$KO = \{KO1, KO2, \ldots, KO8\}$$
$$KI = \{KI1, KI2, \ldots, KI8\}$$
$$K6 = \{K61, K62, K63, K64\}$$
$$K8 = \{K81, K82, K83, K84\}$$

In Figs. 6 and 7, we can regard the subkeys encircled in dotted lines (e.g., $K61$) as the right sides of Eqs. (2) and (3).† The targets for attack are 80-bit subkeys which are not encircled in dotted lines. In addition, the bias is not changed by this reduction.

However, it is difficult to overcome the enormous complexity, and this complexity is higher than the complexity of exhaustive search. This is one problem to be solved.

5. Deriving Subkeys by Linear Approximation of Fm-3

In this section, we try to make a linear approximation of MBAL. We apply F_3-2R to Fm-1,2 and we try to make a linear approximation of Fm-3.

We apply the first round of F3-2R to the 5th to 16th bytes of Fm-1 in $4N$-byte MBAL [the second round of F_3-2R is applied to the $(4N-15)$-th to $(4N-4)$-th bytes of Fm-2]. As a result, we get a linear approximation with bias $2^{-5.49}$:

$$P[\Gamma P] \oplus \Omega[\Gamma\Omega] = KO[\Gamma KO] \quad (5)$$

Next, we try to make linear approximations of Fm-3 as follows to link to Eq. (5):

$$C[\Gamma C] \oplus \Omega[\Gamma\Omega] = KI[\Gamma KI] \oplus K6[\Gamma K6] \quad (6)$$

In this paper we make it a rule to approximate S^3_1 to S^3_{16} and not to approximate sboxes lower than S^3_{17}. We found

*The size of MBAL's original key is 64 bits.

†The subkeys within dotted lines are input or output data of SXAL8. Since we try to make linear approximations that are independent of SXAL8 in this case, we expect that the subkeys encircled in dotted lines are not present on the right side of Eqs. (2) and (3).

Table 3. Mask value of F_3-2R (from Table 2)

No.	1	2	3	4	5	6	7	8	9	10	11	12
Case 1	8E	EE		EE	8E		11	EE	EE	8E		
				EE	60	11		8E	EE	8E	EE	
Case 2	8E	EE		EE	8E		77	EE	EE	8E		
				EE	60	77		8E	EE	8E	EE	
Case 3	6B	BB		BB	6B		11	BB	BB	6B		
				BB	D0	11		6B	BB	6B	BB	
Case 4	6B	BB		BB	6B		CC	BB	BB	6B		
				BB	D0	CC		6B	BB	6B	BB	

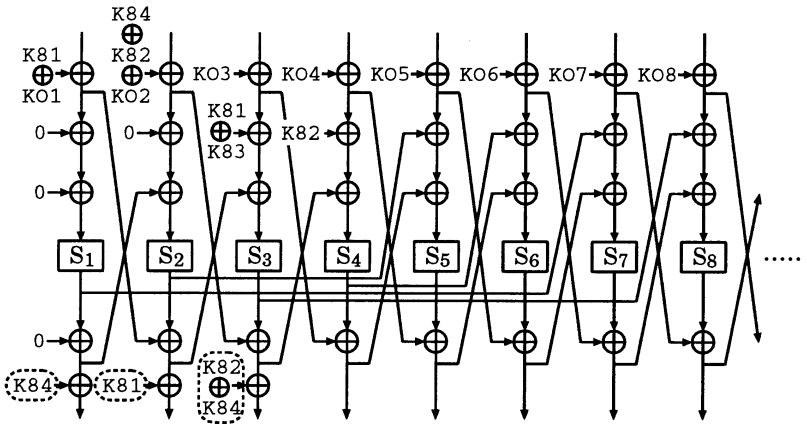


Fig. 6. Equivalent circuit of Part A (highest 8 bytes).

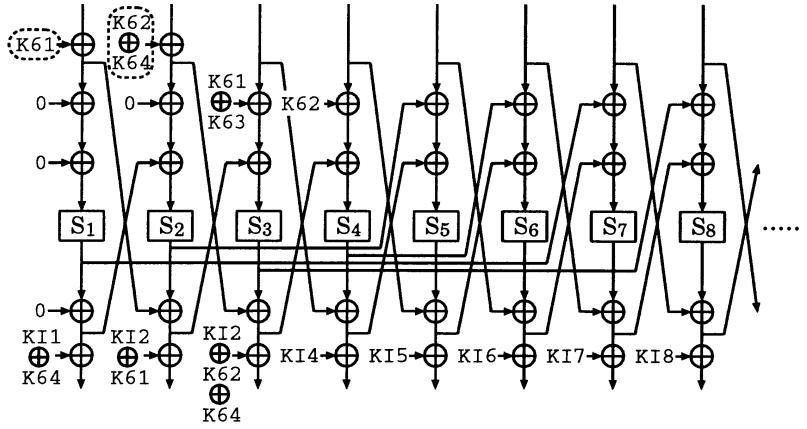


Fig. 7. Equivalent circuit of Part B (highest 8 bytes).

Table 4. Mask value of sbox in Fm-3 for F_3-2R (case 4)

Approximated sbox	S_1^3	S_4^3	S_5^3	S_6^3	S_9^3	S_{10}^3	S_{11}^3	S_{12}^3	S_{13}^3
(α, β)	(B0,D6)	(5C,CC)	(B0,D6)	(CC,DB)	(6B,BB)	(6B,BB)	(8A,6B)	(77,46)	(BB,A7)
Bias	3/8	1/4	3/8	1/16	3/8	3/8	1/8	1/8	3/16

Table 5. Effective text bit positions of Eq. (7)

	Effective test bit positions
ΓP	34,35,37,39,40,41,43,44,45,47,48,57,59,60,61,63,64,66,67,69,71,72,81,82,85,86,89,91,92,93,95,96,97,99,100 101,103,104,106,107,109,111,112
ΓC	1,2,4,6,7,9,11,12,18,20,21,22,25,27,28,33,36,41,43,44,50,51,52,54,55,56,65,66,68,73,77,79,84,85,86,89,90,91 92,93,94,96,97,99,102,103,104
ΓKO	34,35,37,39,40,41,43,44,45,47,48,57,59,60,61,63,64
ΓKI	1,2,4,6,7,9,11,12,18,20,21,22,25,27,28,33,36,41,43,44,50,51,52,54,55,56
$\Gamma K6$	1,3,4,25,29,31

an approximation [Eq. (6)] with a best bias of $2^{-12.08}$.* We found this approximation by using F_3-2R best linear approximation case 4. The mask values of sboxes in Fm-3 are shown in Table 4. Let Eq. (5) ⊕ Eq. (6); then,

$$P[\Gamma P] \oplus C[\Gamma C] = KI[\Gamma KI] \oplus K6[\Gamma K6] \oplus KO[\Gamma KO] \quad (7)$$

Equation (7) consists of only P, C, and K. It gives 1 bit of key information. The bias of Eq. (7), denoted by p'_D, is as follows:

$$p'_D = 2 \times 2^{-5.49} \times 2^{-12.08} = 2^{-16.57}$$

The effective text bit positions of Eq. (7) with bias p'_D are shown in Table 5. We do not know whether Eq. (7) is one of the best linear approximations of MBAL.

When we try getting key information using Q random known plaintexts and Eq. (7), the probability of success in deriving the keys is shown in Table 6 when we choose "Algorithm 1" of Ref. 1 for solving Eq. (7). The number of required known plaintexts is also shown in Table 6, where the block size of MBAL is the standard value of 1024 bytes (1TB = 2^{40}B).

Table 6. Probability of success in deriving the keys by using Eq.(7)

Q	Required data size (1024-byte MBAL)	Probability of success
2^{31}	2 TB	83%
2^{32}	4 TB	91%
2^{33}	8 TB	97%
2^{34}	16 TB	100%

*Similarly, we can approximate Fm-3 with bias $2^{-12.25}$ by applying F_2-2R to Fm-1,2.

6. Conclusions

This paper has proposed the following approach to the linear cryptanalysis of MBAL.

1. We find linear approximations with bias 2^{-5} of 2-round Fm. These approximations are independent of the size of input data.
2. In the elimination round, we can reduce the number of effective key bits from 96 to 80.
3. We find a linear approximation with bias 2^{-16} that consists of P, C, and K. This approximation gives 1 bit of key information.

This paper does not necessarily show that MBAL is weak. To evaluate the strength of MBAL, we must conduct more detailed studies. We will extend the size of the focused block (F_4, F_5, . . .) and reduce the number of effective key bits in the elimination rounds.

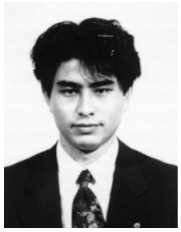
REFERENCES

1. Matsui M. Linear cryptanalysis method for DES cipher. Eurocrypt'93, Lecture Notes in Computer Science 765, Springer-Verlag, p 386–397.
2. Ito K, Kondo S, Mitsuoka Y. SXAL8/MBAL algorithm. Tech Rep IEICE 1993;IT93-66, ISEC93-68, SST93-61:19–24. (in Japanese)
3. Matsui M. On correlation between the order of S-boxes and the strength of DES. Advances in Cryptology—Eurocrypt'94, Lecture Notes in Computer Science 950, Springer-Verlag, p 366–375.
4. Noguchi K, Ashiya H, Sano Y, Kaneko T. A study on differential attack of MBAL cryptosystem. Proc 1994 Symp on Cryptography and Information Security, SCIS'94-14B. (in Japanese)
5. Matsui M. The first experimental cryptanalysis of the data encryption standard. Advances in Cryptology—

Crypto'94, Lecture Notes in Computer Science 839, Springer-Verlag, p 1–11.

6. Kobayashi K, Aoki K. On linear cryptanalysis of MBAL cipher. Trans IEICE 1998;J81-A. (in Japanese)

AUTHORS (from left to right)



Kunio Kobayashi received his B.E. and M.E. degrees from Waseda University in 1995 and 1997, respectively. He is a researcher at NTT Information and Communication Systems Laboratories. He was awarded an SCIS'97 paper prize.

Kazumaro Aoki received his B.S. and M.S. degrees from Waseda University in 1993 and 1995, respectively. He is a researcher at NTT Information and Communication Systems Laboratories. He was awarded SCIS'95 and SCIS'96 paper prizes, and a 1997 IEICE Young Engineer Award. He is a member of the International Association for Cryptologic Research.