Linear Cryptanalysis Using Multiple Approximations-Revisited

Jun Choi[1], Deukjo Hong[1], Seokhie Hong[1], Sangjin Lee[1], and Jongin Lim[1]

Center for Information Security Technologies(CIST),
Korea University,
{jun, hongdj, hsh, sangjin, jilim }@cist.korea.ac.kr[1]

Abstract. We present a technique which uses multiple linear approximations in the linear cryptanalysis of a block cipher and allows for a reduction in the amount of data required for a successful attack. Although the method using many linear approximations was already suggested by B. Kaliski and M. Robshaw in 1994, this paper describes an revisited version utilizing a lot of approximations. In this paper, we present a theoretical and experimental complexity analysis of our attack. The experimental results suggest a complexity. The complexity is bounded by $2^{18.7}$ known texts on 8-round DES with a success rate of 95%, $2^{42.6}$ known texts on 16-round DES with a success rate of 86%, and $2^{40.6}$ chosen texts on 16-round DES with a success rate of 86%. We believe that the results in this paper contain the most efficient attack on the DES full round reported so far in the open literature.

Keywords : Block Cipher, Linear Cryptanalysis, Multiple Linear Cryptanalysis, A Chosen-Plaintext Linear Attack on DES, Multiple Linear Approximations

1 Introduction

In general, an attack for a block cipher is based on properties which hold for at most r rounds where the cipher has r rounds, because the attacker does not know intermediate values of block ciphers and valuable observations are only plaintexts and ciphertexts. Differential and linear cryptanalysis which are most popular attacks for block ciphers utilizing such $r-$round property as differential characteristics and linear approximations, respectively. When Differential Cryptanalysis(DC)[2] was introduced firstly by E. Biham and A. Shamir, it required only one differential characteristic with relatively high probability. L. R. Knudsen proposed the variant of differential cryptanalysis which turned out to be efficient for block ciphers to have not any differential characteristic with high probability but many differential characteristics associated with the same target subkey bits. This attack is called Truncated Differential Cryptanalysis[8]. It takes the advantage that the sum of the probabilities of the characteristics is greater than each of the probabilities of the characteristics.

Linear cryptanalysis has similar history to differential cryptanalysis. M. Matsui introduced Linear Cryptanalysis(LC)[3] for DES firstly. His attack used only one linear approximation with relatively high probability.

The development of linear cryptanalysis means that the number of required known plaintexts decreases for the fixed success rate. A method of the developed examples of linear cryptanalysis is a method[6] using multiple linear approximations by B. Kaliski and M. Robshaw. We call this method as"Multiple Linear Cryptanalysis(MLC)" for convenience. However, their method requires a strong condition that all the linear approximations use the same subkeybits. Although they suggested a technique to use good linear approximations which potentially involve different subkey bits, additional computations were required to utilize the technique. Therefore the algorithm of MLC may be inefficient if additional computations are very large.

However an efficient method using multiple linear approximations without additional computations exists if the algorithm of MLC is revisited, and so we can find the key for a block cipher by gathering many linear approximations with the number of required plaintexts smaller than that of Linear Cryptanalysis using one linear approximation.

This paper is composed by six sections. We summarizes Linear Cryptanalysis introduced by M. Matsui in section 2 and Multiple Linear Cryptanalysis introduced by B. Kaliski and M. Robshaw in section 3. In section 4, We will introduce a new method using multiple linear approximations and the efficiency of it. The experimental results for attack on 8, 16 round DES will be exhibited in section 5. Finally, we consider conclusions in section 6.

2 Notations and Linear Cryptanalysis(LC)

Most of notations used in this paper are according to [6, 3]'s notations except a few notations defined by us. The right most bit of each symbol is referred as the 0-th (lowest) bit.

- $P(= P_H||P_L)$: Plaintexts 64-bit(|| means the concatenation of bits)
- $C(= C_H||C_L)$: Ciphertexts 64-bit
- P_H, P_L : The upper and the lower 32-bit data of P, respectively.
- C_H, C_L : The upper and the lower 32-bit data of C, respectively.
- X_i : The 32-bit intermediate value in the i-th round
- K_i : The 48-bit subkey in the i-th round
- $F_i(X_i, K_i)$: The i-th round F-function
- $A[i]$: The i-th bit of A
- $A[i, j, \cdots, k] : A[i] \oplus A[j] \oplus \cdots \oplus A[k]$
- χ_P : Plaintext bits which a linear approximation includes.
- χ_C : Ciphertext bits which a linear approximation includes.
- χ_K : Key bits which a linear approximation includes.
- χ_{F_r} : Output bits of F-function which a linear approximation includes.
- $P[\chi_P] : P[i_1] \oplus \ldots \oplus P[i_a]$ when χ_P are the $i_1, \ldots, i_a$-th bits of a plaintext

- $C[\chi_C] : C[j_1] \oplus \ldots \oplus C[j_b]$ when χ_C are the $j_1, \ldots, j_b$-th bits of a ciphertext
- $K[\chi_K] : K[k_1] \oplus \ldots \oplus K[k_c]$ when χ_K are the $k_1, \ldots, k_c$-th bits of a key
- $F_r(C_L, K_r)[\chi_{F_r}] : F_r(C_L, K_r)[l_1] \oplus \ldots \oplus F_r(C_L, K_r)[l_d]$ when χ_{F_r} are the $l_1, \ldots, l_d$-th bits of the output of F-function.

In LC, we find the following linear approximation (1) which holds with the probability $p \neq 1/2$ for randomly given plaintext P and the corresponding ciphertext C.

$$P[\chi_P] \oplus C[\chi_C] = K[\chi_K] \tag{1}$$

If equation (1) holds with probability $p \neq 1/2$ and the guess of $K[\chi_K]$ by the maximum likelyhood method is right, then we can determine one bit of the key $K[\chi_K]$.

- Algorithm 1

Step 1. Let T be the number of plaintext/ciphertext pairs such that the left side of equation (1) is equal to zero and let N be the total number of pairs.

Step 2. If $T > N/2$, then guess $K[\chi_K] = 0$(when $p > 1/2$), or 1(when $p < 1/2$), else guess $K[\chi_K] = 1$(when $p > 1/2$) , or 0(when $p < 1/2$).

The success rate of Algorithm 1 clearly increases when N or $|p - 1/2|$ does. We now refer to the most effective linear expression (i.e. $|p - 1/2|$ is maximal) as the best expression and the probability p as the best probability.

As you see the above method, Algorithm 1 gets only one bit key information. If Algorithm 2[3] is used in LC, we can obtain more key bit information.

For a practical attack of r-round DES cipher, we use the linear approximation (2) which holds with the best probability of $(r-1)$-round DES cipher.

$$P[\chi_P] \oplus C[\chi_C] \oplus F_r(C_L, K_r)[\chi_{F_r}] = K[\chi_K] \tag{2}$$

If one substitutes an incorrect candidate for K_r in equation (2), the effectiveness of this equation clearly decreases. Therefore, the following maximum likelyhood method can be applied to deduce K_r and $K[\chi_K]$.

- Algorithm 2

Step 1. For each candidate $K_r^{(i)}(i = 1, 2, \ldots)$ of K_r, let T_i be the number of plaintexts such that the left side of equation (2) is equal to zero.

Step 2. Let T_{max} be the maximum value and T_{min} be the minimum value all T_i's.

- If $|T_{max} - N/2| > |T_{min} - N/2|$, then adopt the key candidate corresponding to T_{max} and guess $K[\chi_K] = 0$(when $p > 1/2$), or 1(when $p < 1/2$).
- If $|T_{max} - N/2| < |T_{min} - N/2|$, then adopt the key candidate corresponding to T_{min} and guess $K[\chi_K] = 1$(when $p > 1/2$), or 0(when $p < 1/2$).

By analyzing the form of the round function it is possible to identify which bits of the subkey K_r and which bits of the input C_L effect the value of $F_r(C_L, K_r)[\chi_{F_r}]$. Such subkey bits which are termed effective key bits and such bits of the input C_L which are termed effective text bits.

Sufficient data is then taken to ensure that the correct guess can be distinguished from among the incorrect guesses thereby identifying the correct guess for the effective key bits.

We sometimes use linear approximations based on the $(r-2)$-round expression[4] which reduces the number of required plaintexts and increases the success rate of our attack.

3 Multiple Linear Cryptanalysis(MLC)

Suppose we have n linear approximations which involve the same χ_K but differ in the plaintext and ciphertext bits that they use.

Suppose the i-th linear approximation for $n \geq 1$ has the following form:

$$P[\chi_P^i] \oplus C[\chi_C^i] = K[\chi_K] \tag{3}$$

For the sake of analysis, we will suppose, without loss of generality, that each bias ε_i is positive.

- Algorithm 1M

Step 1 For $1 \leq n$ let T_i be the number of plaintext/ciphertext pairs such that the left side of equation (3) is equal to zero. Let N denote the total number of plaintexts.

Step 2 Calculate $U = \sum_{i=1}^{n} a_i T_i$ for some set of weights $a_1, a_2, \ldots, a_n$ where $\sum_{i=1}^{n} a_i = 1$

Step 3 If $U > N/2$ the guess $K[\chi_K] = 0$, else guess $K[\chi_K] = 1$.

For Algorithm 1M, the assumption that all the linear approximations use the same χ_K is requested. Therefore Algorithm 1M can't be used for linear approximations which involve different subkey bits. At this point B. Kaliski and M. Robshaw [6] presented an extension to Algorithm 1M. We call the extension as Algorithm 1MG for convenience. Algorithm 1MG is following:

Suppose we have n linear approximations and the i-th linear approximation has the form

$$P[\chi_P^i] \oplus C[\chi_C^i] = K[\chi_K^i]. \tag{4}$$

To consider the approximations together, we must first guess, for each $i, j (2 \leq i, j \leq n, i \neq j)$, whether $K[\chi_K^i]$ and $K[\chi_K^j]$ are equal or not. There are at most 2^{n-1} guesses and for each guess we can obtain n linear approximations of the form

$$P[\chi_P^i] \oplus C[\chi_C^i] \oplus \Delta^i = K[\chi_K^i] \tag{5}$$

where $\Delta^1 = 0$ and Δ^i for $i > 1$ depends on the guess.

Note that in practice one would not repeat Algorithm 1M for each guess; instead one would consider up 2^{n-1} ways of combining the statistics T_i for $1 \leq i \leq n$.

B. Kaliski and M. Robshaw[6] constructed Algorithm 2M from Algorithm 1M like the method that Matsui organized Algorithm 2 from Algorithm 1. To begin with, we can get the i-th linear equation (6) among the n linear approximations from the extension of equation (3).

$$P[\chi_P^i] \oplus C[\chi_C^i] \oplus F_1(P_L, K_1)[\chi_{F_1}^i] \oplus F_r(C_L, K_r)[\chi_{F_r}^i] = K[\chi_K] \tag{6}$$

We will again suppose, that each bias ε_i is positive.

- Algorithm 2M

Step 1 Let $K_1^{(g)}$ $(g = 1, 2, \ldots)$ and $K_r^{(h)}$ $(h = 1, 2, \ldots)$ be possible candidates for K_1 and K_r respectively. Then for each key pair $(K_1^{(g)}, K_r^{(h)})$ and each linear approximation i let $T_{g,h}^i$ be the number of plaintexts such that the left side of equation (6) is equal to zero when K_1 is replaced by $K_1^{(g)}$ and

K_r by $K_r^{(h)}$.
Let N be the total number of plaintexts.

Step 2 Let $a_i = \varepsilon_i / \sum_{i=1}^{n} \varepsilon_i$. Calculate $U_{g,h} = \sum_{i=1}^{n} a_i T_{g,h}^i$ for each g, h.

Step 3 Let U_{max} be the maximum value and U_{min} be the minimum value of all $U_{g,h}$'s.
- If $|U_{max} - N/2| > |U_{min} - N/2|$, adopt the key candidate corresponding to U_{max} and guess $K[\chi_K] = 0$.
- If $|U_{max} - N/2| < |U_{min} - N/2|$, adopt the key candidate corresponding to U_{min} and guess $K[\chi_K] = 1$.

For Algorithm 2M, the assumption that all the linear approximations use the same χ_K is requested. Therefore Algorithm 2M can't be used for linear approximations which involve different subkey bits. At this point B. Kaliski and M. Robshaw [7] presented an extension to Algorithm 1M. We call the extension as Algorithm 2MG for convenience. Algorithm 2MG is following:

Suppose we have n linear approximations and the i-th linear approximation has the form (7)

$$P[\chi_P^i] \oplus C[\chi_C^i] \oplus F_1(P_L, K_1)[\chi_{F_1}^i] \oplus F_r(C_L, K_r)[\chi_{F_r}^i] = K[\chi_K^i]. \tag{7}$$

We will also suppose, without loss of generality, that the bias of each linear approximation is positive.

- Algorithm 2MG

Step 1 Let $K_1^{(g)}$ $(g = 1, 2, \ldots)$ and $K_r^{(h)}$ $(h = 1, 2, \ldots)$ be possible candidates for K_1 and K_r respectively. Then for each key pair $(K_1^{(g)}, K_r^{(h)})$ and each linear approximation i let $T_{g,h}^i$ be the number of plaintexts such that the left side of equation (7) is equal to zero when K_1 is replaced by $K_1^{(g)}$ and K_r by $K_r^{(h)}$.
Let N be the total number of plaintexts.

Step 2 Let $a_i = \varepsilon_i / \sum_{i=1}^{n} \varepsilon_i$. Define the n-tuple[1] $C = (c_1, \cdots, c_n)$. Calculate for each g, h and each C,

$$U_{g,h}[C] = \sum_{i=1, c_i=0}^{n} a_i T_{g,h}^i + \sum_{i=1, c_i=1}^{n} a_i (N - T_{g,h}^i)$$

.

Step 3 Let U_{max} be the maximum value and U_{min} be the minimum value of all $U_{g,h}[C]$'s. Adopt the key candidate corresponding to U_{max} and guess $K[\chi_K^i] = c_i$ for $1 \le i \le n$

[1] This tuple represents the possible values for the bits of key information in each approximation.

When using n linear approximations to different bits of key information, one performs linear cryptanalysis n times in Algorithm 2MG, but combine the results up to 2^{n-1} times, using each possible relation between the key information in the n approximations.

4 Multiple Linear Cryptanalysis-Revisited(RMLC)

The conditions to apply Algorithm 1M or 2M are that all linear approximations have the same χ_K and the same effective key bits. However these conditions restrict the use of Algorithm 1M or 2M. The suggested and well-designed block ciphers in recent years are not likely to have such linear approximations with relatively great bias.

Furthermore, Algorithm 1MG or 2MG demand too many additional computations to attack in proportion to the number of the linear approximations used. Therefore we consider that Algorithm 1MG or 2MG associated with linear approximations which have different χ_K's were inefficient.

These problems generated by using multiple linear approximations can be solved if $U_{g,h}(=\sum_{i=1}^{n} a_i T_{g,h}^i)$ in Algorithm 2M is replaced by weighted sum of $(T_{g,h}^i - N/2)^2$. We abandon guessing the right sides of each linear approximation and focus our efforts on finding first or last effective round key bits. Note that the linear approximations for practical attacks must have the same effective key bits for efficiency of key recovery algorithm.

In the following subsections, we will introduce a new proposal Algorithm RM and the comparison of LC,MLC, and RMLC.

To help the theoretical analysis of Algorithm RM, we define some notations used in this section.

- **Notations**
  - N : The number of required plaintexts
  - α_w, λ_w : The parameters of Gamma function for wrong keys
  - α_r, λ_r : The parameters of Gamma function for the right key
  - n : The number of linear approximations
  - $p_i, \varepsilon_i (1 \le i \le n)$: The probability and the bias for each linear approximation
  - $q_i : 1 - p_i$
  - w_i : The weight mentioned in Algorithm RM

4.1 The introduction and the analysis Algorithm RM

Our attack works according to the following algorithm.

- Algorithm RM

Step 1 Let N be the total number of plaintexts and let r be the number of total round for a block cipher
$K_1^{(g)}$ $(g = 1, 2, \ldots)$ and $K_r^{(h)}$ $(h = 1, 2, \ldots)$ be possible candidates for K_1 and K_r respectively. Then for each key pair $(K_1^{(g)}, K_r^{(h)})$ and each linear approximation i let $T_{g,h}^i$ be the number of plaintexts such that the left side of equation (7) is equal to zero when K_1 is replaced by $K_1^{(g)}$ and K_r by $K_r^{(h)}$.

Step 2 Let $k = \frac{\prod_{i=1}^{n}(Np_iq_i)^2}{\sum_{i=1}^{n}(\prod_{j=1,j\neq i}^{n}(Np_jq_j)^2)}$ and $w_i = k/(Np_iq_i)^2$.
Calculate $U_{g,h} = \sum_{i=1}^{n} w_i(T_{g,h}^i - N/2)^2$ for each g, h.

Step 3 Let U_{max} be the maximum value of all $U_{g,h}$'s.
The g, h corresponding to U_{max} are the candidates of K_1, K_r.

We may modify Algorithm RM to find only either first or last round key for some case. The following equation (8) is used in this case(Generally, n is either 1 or r for r round cipher).

$$P[\chi_P^i] \oplus C[\chi_C^i] \oplus F_n(P_L, K_n)[\chi_{F_n}^i] = K[\chi_K^i] \tag{8}$$

4.2 The Complexity of Algorithm RM

To determine the complexity of Algorithm RM, $\sum_{i=1}^{n} \varepsilon_i^2$ is considered in RMLC where each bias of n linear approximations is $\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n$. We could confirm that $c \cdot 1/\sum_{i=1}^{n} \varepsilon_i^2$($c$:constant) determines the number of required plaintexts in RMLC by our simulations. Also, we could know the probability density function of $\sum_{i=1}^{n} w_i(T_{g,h}^i - N/2)^2$ for Algorithm RM in RMLC where the density function depends on $c \cdot 1/\sum_{i=1}^{n} \varepsilon_i^2$. For convenience, $\sum_{i=1}^{n} \varepsilon_i^2$ is denoted by E^2. In Theorem 1, we show the relation between the success rate and the density function when the number of required plaintexts are vary.

Assumption 1 *For the right key pair (g, h) and each i-th linear approximation, $T_{g,h}^i (1 \leq i \leq n)$ are independent random variables each having an normal density with mean $\mu_i = Np_i$ and variance $\sigma_i^2 = Np_iq_i$.*

Assumption 2 *For the wrong key pair (g, h) and all linear approximations, $T_{g,h}^i (1 \leq i \leq n)$ are independent random variables each having an normal density with mean $\mu = N/2$ and variance $\sigma^2 = N/4$.*

These assumptions lead to the following theorem:

Theorem 1. *Let $a_i = Np_i - N/2$, $Y_i = T_{g,h}^i - Np_i$, $s = \sum_{i=1}^{n}(\sqrt{w_i}Y_i)^2$, and $b = \sum_{i=1}^{n}(\sqrt{w_i}a_i)^2$. Then the success rate of Algorithm RM is*

$$\frac{\lambda_r}{\Gamma(\alpha_r)} \int_0^{\infty} (\Pi_{K_{g,h}^{(i)} \neq K_{g,h}} \int_0^{s-2\sqrt{sb}+b} \frac{\lambda_w}{\Gamma(\alpha_w)} w^{\alpha_w - 1} e^{-\lambda_w w} dw) s^{\alpha_r - 1} e^{-\lambda_r s} ds \tag{9}$$

where Γ is the gamma function and the product is taken over all subkey candidates $K_{g,h}^{(i)}$ except right key $K_{g,h}$.

Proof. Appendix

The numerical caculation of expression (9) is as follows.

N	$2E^{-2}$	$4E^{-2}$	$8E^{-2}$	$16E^{-2}$
Success rate	43%	75%	95%	99%

Table 1. The success rate of Algorithm RM by Theorem 1

Through experiments in the next section, we observe that E^2 behaves just like the square of bias for a linear approximation in conventional LC [3].

For example the success rate of Algorithm 2 is 96.7% with $8\varepsilon^{-2}$ known plaintexts [3]. Similarly the success rate of Algorithm RM is about 94 - 97% with $8E^{-2}$ known plaintexts.

4.3 The comparison of LC, MLC, and RMLC

We are going to compare LC, MLC, and RMLC via experiments. The following equations are the linear approximations used in simulations.

$$P_H[7,18,29] \oplus P_L[15] \oplus C_H[15] \oplus C_L[7,18,29] \oplus F_4(C_L,K_4)[15] = k_1[22] \oplus k_3[22] \quad (\varepsilon : 0.03) \tag{10}$$

$$P_H[3,21] \oplus P_L[9] \oplus C_H[9] \oplus C_L[3,21] \oplus F_4(C_L,K_4)[9] = k_1[14] \oplus k_3[14] \quad (\varepsilon : 0.03) \tag{11}$$

$$P_H[7,18,24,29] \oplus C_H[7,18,24] \oplus F_1(P_L,K_1)[7,18,24,29] \oplus F_7(C_L,K_7)[7,18,24] = k_3[22] \oplus k_4[44] \oplus k_5[22] \quad (\varepsilon : 25 \times 2^{-12}) \tag{12}$$

$$P_H[7,18,24] \oplus C_H[7,18,24,29] \oplus F_1(P_L,K_1)[7,18,24] \oplus F_7(C_L,K_7)[7,18,24,29] = k_3[22] \oplus k_4[44] \oplus k_5[22] \quad (\varepsilon : 25 \times 2^{-12}) \tag{13}$$

Table 2 shows the results of Algorithm 2 applying (10) and (11) separately and Algorithm RM using (10) and (11) on 4 round DES. Here, E^2 is two times as large as the square of each bias. Clearly, we can observe that Algorithm RM requires less known plaintexts for the same success rate than Algorithm 2.

Table 3 shows the the results of Algorithm 2 applying (12) and (13) separately and Algorithm 2M and Algorithm RM using (12) and (13) on 7 round DES. ¿From this experiment we believe that Algorithm 2M and RM have the same results under the situation where Algorithm 2M is available.

	experimental results			
the number of known plaintexts	2222	4444	8888	17776
(10) Algorithm 2	52.5%	63.7%	75.1%	89%
(11) Algorithm 2	48.5%	73.9%	92.3%	97%
(10)&(11) Algorithm RM	73.7%	91.4%	97.1%	99.9%

Table 2. Complexity of the LC and the RMLC on 4-round DES. Here we found 6 key bits.

	experimental results			
the number of known plaintexts	13422	26844	53688	107376
(12) Algorithm 2	3%	2%	17%	51%
(13) Algorithm 2	1%	1%	15%	45%
(12)&(13) Algorithm 2M	4%	13%	46%	94%
(12)&(13) Algorithm RM	3%	14%	45%	94%

Table 3. Complexity of the LC, MLC and RMLC on 7-round DES. Here we found 12 key bits.

5 New results for attack on DES

5.1 Attack on 8 round DES

We found 139 7-round linear approximations with $|\varepsilon_i| > 0.00003$ available to guess effective key bits of S1-box on 8-round DES. The reason why we only consider S1-box is that $\sum_{i=1}^{139} \varepsilon_i^2 \approx 0.00001838$ is the most value among 8 S-box's $\sum \varepsilon_i^2$. This value is 5.06 times as large as ε^2 of the 7-round linear approximation with the best probability in [3].

Table 4 shows the comparison of the theoretical results and the experimental results of Algorithm RM using 139 7-round linear approximations. As you see from the experimental results in Table 4, the theoretical results are similar to experimental results, and the number of required known plaintexts used to Algorithm RM is five times as small as the number of required known plaintexts used to Algorithm 2 [3] for the same success rate if we use 139 linear approximations. For instance, the number of required known plaintexts used to Algorithm 2 is 2,206,072 for the 96.7%, but the number of required known plaintexts used to Algorithm RM is 435,200 for the same success rate.

5.2 Attack on 16 round DES

B. Kaliski and M. Robshaw found 10,006 14-round linear approximations involving one active S-box at each round with bias $|\varepsilon_i| > 10^{-8}$ and $\sum_{i=1}^{10,006} \varepsilon_i^2 \approx 1.23 \times 10^{-11}$[6]. They raised the question whether there is an efficient way to apply these linear approximations. We could also find 10,098 14-round linear

	experimental results			
plaintexts	108800	217600	435200	870400
Algorithm RM	45%	76%	97%	99%

Table 4. Complexity of the RMLC on 8-round DES. Here we found 6 key bits.

approximations involving one active S-box at each round with bias $|\varepsilon_i| > 10^{-8}$ and $\sum_{i=1}^{10,098} \varepsilon_i^2 \approx 2.21 \times 10^{-11}$. Though we found more linear approximations than 10,006, we confirmed there were only a few linear approximations available to find the 6 key bits of S1-box and the 6 key bits of S5-box on 16-round DES. Because of $\sum_{i=1}^{3} \varepsilon_i^2 = 4.04 \times 10^{-13}$ for 3 linear approximations available among many linear approximations, the number of required known plaintexts used for Algorithm RM in RMLC may be about 1.25 times as small as the number of required known plaintexts(2^{43}) used to Algorithm 2 in LC using one linear approximation with the best probability($1/2 - 1.19 \times 2^{-21}$).

There is another strategy to attack full round DES. We can adopt a chosen plaintext attack instead of a known plaintext attack. L. R. Knudsen and J. E. Mathiassen introduced "A Chosen-Plaintext Linear Attack on DES[9]" using chosen plaintexts. For the 14-round characteristic used by M. Matsui there is no active S-box in the second round. However, if one takes the first 13 rounds of this characteristic and uses these in the rounds 3 to 15 one gets a single active S-box in both the first and second rounds. Then we get the following picture.($3, 4, \cdots, 15$ mean the round number. In addition to A, B, D and $A \oplus B$ mean bit-masks. For instance $A = [7, 18, 24], B = [7, 18, 24, 29], D = [15]$ and $A \oplus B[29]$ can mean bit-masks.)

$$
\begin{aligned}
3 &: - - - \\
4 &: A \leftarrow D \\
5 &: D \leftarrow A \oplus B \\
6 &: B \leftarrow D \\
7 &: - - - \\
8 &: B \leftarrow D \\
9 &: D \leftarrow A \oplus B \\
10 &: A \leftarrow D \\
11 &: - - - \\
12 &: A \leftarrow D \\
13 &: D \leftarrow A \oplus B \\
14 &: B \leftarrow D \\
15 &: - - -
\end{aligned}
\tag{14}
$$

We try to combine our work with their method. We could find five more 13-round approximations that involve the identical effective key bits with the same as (14) used by L. R. Knudsen and J. E. Mathiassen. If we calculate the bias of each linear approximation, the values are $2^{-19.85}, 2^{-19.85}, 2^{-25.95}, 2^{-25.95}, 2^{-33.7}$, and $2^{-33.7}$. Depending on this value, we estimate the number of required chosen plaintexts for Algorithm RM in RMLC using 6 linear approximations will be about $2^{40.6}$ for 86% success rate. $2^{40.6}$ chosen plaintexts have the following property: To achieve plaintexts used in our attack we need to fix 22 bits of the right half of the all plaintexts. The fixed positions are $0, 1, 2, 3, 4, 15 \sim 31$.

We suppose that each of n linear approximations has the different bias $\varepsilon_i (1 \leq i \leq)$ in Algorithm using multiple linear approximations, whereas an linear approximation has the bias ε in Algorithm 2. Then Table 5 shows the comparison of the theoretical results for Algorithm 2, Algorithm 2M, and Algorithm RM where $E^2 = \sum_{i=1}^{n} \varepsilon_i^2 = c \cdot \varepsilon^2$.

	Plaintexts		
Success \ Attacks	Algorithm 2 [3]	Algorithm 2M [6]	Algorithm RM
48%	$2\varepsilon^{-2}$	None	$2.2E^{-2} = (2.2/c)\varepsilon^{-2}$
78%	$4\varepsilon^{-2}$	None	$4.3E^{-2} = (4.3/c)\varepsilon^{-2}$
96%	$8\varepsilon^{-2}$	None	$8E^{-2} = (8/c)\varepsilon^{-2}$
99%	$16\varepsilon^{-2}$	None	$16E^{-2} = (16/c)\varepsilon^{-2}$

Table 5. The comparison of theoretical results.

6 Conclusions

In this paper we proposed what we believe is very efficient attack on the 16-round DES. The attack use many linear approximations to carry out the attack on block ciphers. The method requires known plaintexts less than that of Algorithm 2 in LC[3] and is available with more linear approximations in more cases than Algorithm 2M in MLC[6].

We defined E^2 in order to estimate the data complexity of Algorithm RM in RMLC, which behaves just like square of bias for a linear approximation in LC.

Based on Theorem 1, we could estimate the data requirements of Algorithm RM in attacks on 8 and 16 round DES. The number of required known plaintexts for 8-round DES is $2^{18.73}$ for 95% success rate and the number of required known plaintexts for 16-round DES is $2^{42.68}$ for 86% success rate.

If we adopt a chosen plaintexts instead of known plaintexts as our strategy, we can reduce the data complexity of Algorithm RM applying 16-round DES. Therefore we showed that Algorithm RM requires $2^{40.6}$ chosen plaintexts with 86% success rate if we use chosen plaintexts to attack.

The focus of our work is not the running time of key recovery algorithm but the requirements of data because, in most cases, we would rather to carry out many off-line operations than to get a lot of data using on on-line system.

Our approach might be applicable on other block ciphers.

References

1. Paul G. Hoel, Sidney C. Port and Charles J. Stone *Introduction to Probability Theory*, University of California, Los Angeles, 1971.
2. E.Biham and A.Shamir, *Differential Cryptanalysis of the Data Encryption Standard*, Springer-Verlag, New York, 1993.
3. M.Matsui, *Linear cryptanalysis method for DES cipher*, Advanced in cryptology , Eurocrypt'93, Springer-Verlag, 1993.
4. M.Matsui, *The first Experimental cryptanalysis of DES*, Advanced in cryptology , CRYPTO'94, Springer-Verlag, 1994.
5. M.Matsui, *New Structure of Block Ciphers with Provable Security against Differential and Linear Cryptanalysis*, Fast Software Encryption'96, Springer-Verlag, 1996.
6. B.Kaliski and M.Robshaw, *Linear Cyptanalysis Using Multiple Approximations*,Advanced in cryptology, CRYPTO'94, 1994.
7. B. Kaliski and M. Robshaw, *Linear cyptanalysis using multiple approximation and FEAL*, in FAST Software Encryption, FSE'94, vol. 1008 of Lecture Notes in Computer Science, pp.249-264, Springer-Verlag, 1998
8. L.R.Knudsen, *Truncated and higher order differential*, In Fast Softwqre Encryption - Second International Workshop, volume 1008 of Lecture Notes in Computer Science, pp.196-211, Springer-Verlag, 1995.
9. L.R.Knudsen and J.E. Mathiassen, *A Chosen-Plaintext Linear Attack on DES*, FSE'2000, 2000.

A The Lemma 1 and 2 to prove Theorem 1

Note that Γ is used to define a function called the gamma function and the gamma density with parameters α and λ is denoted by $\Gamma(\alpha,\lambda)$ or $\Gamma(x;\alpha,\lambda)$. We see that

$$\Gamma(x;\alpha,\lambda)=\begin{cases}\frac{\lambda^{\alpha}}{\Gamma(\alpha)}x^{\alpha-1}e^{-\lambda x} & \text{if} \quad x>0\\ 0 & \text{if} \quad x\leq 0\end{cases} \tag{15}$$

where

$$\Gamma(\alpha)=\int_0^{\infty}x^{\alpha-1}e^{-x}dx, \qquad \alpha>0.$$

Lemma 1. [1] *Let X be a random variable having the normal density $N(0,\sigma^2)$. Then the density of the random variable $Y=X^2$ corresponds to the special case $\alpha=\frac{1}{2}$ and $\lambda=\frac{1}{2\sigma^2}$ in the equation (15).*

Lemma 2. [1] *Let $X_i(1\leq i\leq n)$ be independent random variables each having the normal density $N(0,\sigma^2)$. Note that each variance must be identical for all $X_i(1\leq i\leq n)$. Then the density of $\sum_{i=1}^{n}X_i^2$ has the gamma density $\Gamma(\frac{n}{2},\frac{1}{2\sigma^2})$.*

B The Proof of Theorem 1

We consider the statistic $\sum_{i=1}^{n} w_i(T_{g,h}^i - \frac{N}{2})^2$ in Algorithm RM. If g, h is a wrong key pair, $(T_{g,h}^i - N/2)^2$ approximates to zero because the expected value of $T_{g,h}^i$ is $N/2$. Then $\sum_{i=1}^{n} w_i(T_{g,h}^i - \frac{N}{2})^2(g, h \ : \ wrong)$ also approximates to zero regardless of weight. So, the value of weight for wrong pairs may be left out of consideration in our proof.

In the following proof, the meanings of k and a_i are identical those of Algorithm RM and Theorem 1.

Proof

We consider the statistic $\sum_{i=1}^{n} w_i(T_{g,h}^i - \frac{N}{2})^2$ in Algorithm RM. Let $Y_i = T_{g,h}^i - Np_i (1 \le i \le n)$.

$$\begin{aligned}
\sum_{i=1}^{n} w_i(T_{g,h}^i - \frac{N}{2})^2 &= \sum_{i=1}^{n} w_i(T_{g,h}^i - \frac{N}{2})^2 \\
&= w_1(T_{g,h}^1 - \frac{N}{2})^2 + \cdots + w_n(T_{g,h}^n - \frac{N}{2})^2 \\
&= w_1(Y_1 + Np_1 - \frac{N}{2})^2 + \cdots + w_n(Y_n + Np_n - \frac{N}{2})^2 \\
&= w_1(Y_1 + a_1)^2 + \cdots + w_n(Y_n + a_n)^2 \\
&= \sum_{i=1}^{n}(\sqrt{w_i}Y_i)^2 + 2(\sum_{i=1}^{n} a_i w_i Y_i) + \sum_{i=1}^{n}(\sqrt{w_i}a_i)^2
\end{aligned}$$

Let $\sum_{i=1}^{n}(\sqrt{w_i}Y_i)^2, \sum_{i=1}^{n} a_i w_i Y_i$, and $\sum_{i=1}^{n}(\sqrt{w_i}a_i)^2$ be s, t and b, respectively. Then we can get the following equation (16) by Cauchy-Schwartz Inequality.

$$sb \ge t^2 \tag{16}$$

We can estimate the low and upper bound of

$$\begin{aligned}
&\sum_{i=1}^{n}(\sqrt{w_i}Y_i)^2 + 2(\sum_{i=1}^{n} a_i w_i Y_i) + \sum_{i=1}^{n}(\sqrt{w_i}a_i)^2 \\
&= s + 2t + b
\end{aligned}$$

by equation (16).

The low and upper bound for right key are determined by the following relation.

$$\begin{aligned}
&sb \ge t^2 \\
&\quad \Rightarrow -\sqrt{sb} \le t \le \sqrt{sb} \\
&\quad \Rightarrow -2\sqrt{sb} \le 2t \le 2\sqrt{sb} \\
&\quad \Rightarrow s - 2\sqrt{sb} \le s + 2t \le s + 2\sqrt{sb} \\
&\quad \Rightarrow s - 2\sqrt{sb} + b \le s + 2t + b \le s + 2\sqrt{sb} + b
\end{aligned}$$

Therefore the key candidates $K^i_{g,h}$ for the i-th approximation will be wrong if the value of $T^i_{g,h}$ is smaller than $s - 2\sqrt{sb} + b$.

Under previous Assumption 1, $T^i_{g,h}$ $(1 \le i \le n)$ be independent random variables each having the normal density $N(Np_i, Np_iq_i)$, that is, the expected value $E(T^i_{g,h})$ of $T^i_{g,h}$ is Np_i and the variance $V(T^i_{g,h})$ of $T^i_{g,h}$ is Np_iq_i. Then $E(T^i_{g,h} - \frac{N}{2}) = Np_i - \frac{N}{2}$, and $V(T^i_{g,h} - \frac{N}{2}) = Np_iq_i$. Then $E(Y_i) = 0$, and $V(Y_i) = Np_iq_i = N(1 - \frac{1}{4}\varepsilon_i^2)$ where $Y_i = T^i_{g,h} - Np_i$. Let $w_i = k/(Np_iq_i)^2$. Since each of $\sqrt{w_i}Y_i$ for right key candidates is independent random variables each having the normal density $N(0, k)$, each of $(\sqrt{w_i}Y_i)^2$ is the gamma density with parameters $\alpha_r = \frac{1}{2}$ and $\lambda_r = \frac{1}{2k}$ by Lemma 1. Then the density of $\sum_{i=1}^n (\sqrt{w_i}Y_i)^2$ has the gamma density $\Gamma(\frac{n}{2}, \frac{1}{2k})$ by Lemma 2.

Similarly, the density of $\sum_{i=1}^n (\sqrt{w_i}Y_i)^2$ has the gamma density $\Gamma(\frac{n}{2}, \frac{1}{2}\frac{1}{(\frac{1}{\sqrt{n}})^2\frac{N}{4}})$ for $T^i_{g,h}$, $(1 \le i \le n)$ corresponding to Assumption 2 where w_i's approximate to $\frac{1}{n}$ for all i($\because$ As we discuss the weight for wrong key before we begin the proof of Theorem 1, the weight for wrong key can be approximated arbitrary values.) and $Y_i = T^i_{g,h} - \frac{N}{2}$.

If we use the values which are $\alpha_w, \lambda_w, \alpha_r$, and λ_r, for wrong keys and right key, the success rate of Algorithm RM for one right key and wrong keys is

$$\int_0^\infty \frac{\lambda_r}{\Gamma(\alpha_r)} (\Pi_{K^{(i)}_{g,h} \neq K_{g,h}} \int_0^{s-2\sqrt{sb}+b} \frac{\lambda_w}{\Gamma(\alpha_w)} w^{\alpha_w - 1} e^{-\lambda_w w} dw) s^{\alpha_r - 1} e^{-\lambda_r s} ds$$

where Γ is the gamma function and the product is taken over all subkey candidates $K^{(i)}_{g,h}$ except right key $K_{g,h}$.