On Correlation Between the Order of S-boxes and the Strength of DES

Mitsuru Matsui

Computer & Information Systems Laboratory
Mitsubishi Electric Corporation
5-1-1, Ofuna, Kamakura, Kanagawa, 247, Japan
matsui@mmt.isl.melco.co.jp

Abstract. This paper introduces a practical algorithm for deriving the best differential characteristic and the best linear expression of DES. Its principle is based on a duality between differential cryptanalysis and linear cryptanalysis, and applicable to various block ciphers. Then using this program, we observe how the order of S-boxes affects the strength of DES. We show that the order of the S-boxes is well-arranged against differential cryptanalysis, though it is not the best choice. On the other hand, our experimental results indicate that it is a very weak choice in regard to linear cryptanalysis. In other words, DES can be strengthened by just rearranging the order of the S-boxes.

1 Introduction

Differential cryptanalysis [1] and linear cryptanalysis [2] are known as most effective attacks applicable to various block ciphers. They proved for the first time that DES is breakable by a chosen-plaintext attack and a known-plaintext attack faster than an exhaustive key search, respectively.

The principle of linear cryptanalysis is similar to that of differential cryptanalysis in several aspects, as has also been pointed out by Biham [3]. These methods both analyze S-boxes statistically, then extend the local property of the S-boxes to the entire cipher structure through F-functions, and finally reach simple probabilistic relations among plaintexts, ciphertexts and the fixed secret key. The procedure for deriving the key is achieved by counting up pre-defined counters which essentially correspond to key candidates.

In this paper we begin by directing our attention to this similarity from the viewpoint of duality between differential cryptanalysis and linear cryptanalysis. The former traces the flow of differential values, which are defined as an XORed value of two series of texts, whereas the latter follows that of masking values, where the parity of the masked bits plays an essential role. We will easily see that "XOR branch" and "three-forked branch" are mutually dual operations in regard to differential values and masking values.

The next purpose of this paper is to show, on the basis of this duality, an algorithm for searching for the best differential characteristic and the best linear expression in practical time. Our program has completely determined the best

differential characteristic of DES for the first time; this was an open problem, while Knudsen [4] estimated, under a limited situation, the characteristics found by Biham and Shamir to be best.

Then using this search program, we observe how the order of S-boxes affects the strength of DES from the viewpoints of the best characteristic probability and the best linear approximate probability. We have calculated, under some assumptions, these probabilities for all possible permutations of the S-boxes, and as a result, reached new interesting properties of DES.

Biham and Shamir pointed out that changing the order of S-boxes can weaken DES in regard to differential cryptanalysis, and illustrated an example of the weaker permutations [1]. Our results prove that the order of the S-boxes is well-arranged though no permutation can resist differential cryptanalysis.

On the other hand, as for linear cryptanalysis, we face the opposite situation; the order of the S-boxes is a very weak choice. Our experimental results indicate that once we change the order of the S-boxes, the modified DES can be strengthened in almost cases. We have determined the best permutation of the S-boxes, which is now immune against linear cryptanalysis. We also show that there exist permutations that are stronger than the original DES in regard to differential cryptanalysis and linear cryptanalysis as well.

2 Notations and Preliminaries

Figure 1 illustrates the data randomization part and the F-function of DES, whose notations are used throughout this paper. We will discuss differential cryptanalysis and linear cryptanalysis in parallel, and for this purpose, it is convenient to define the term "the best n-round probability B_n" depending on the context as follows:

In the case of differential cryptanalysis:

$$(\Delta X_i, \Delta Y_i) \stackrel{\text{def}}{=} Prob\{\ F_i(X_i \oplus \Delta X_i, K_i) = F_i(X_i, K_i) \oplus \Delta Y_i\ \},$$

$$[p_1, p_2, ..., p_t] \stackrel{\text{def}}{=} \prod_{i=1}^{t} p_i,$$

$$B_n \stackrel{\text{def}}{=} \max_{\Delta X_i = \Delta X_{i-2} \oplus \Delta Y_{i-1} (3 \le i \le n)} [(\Delta X_1, \Delta Y_1), (\Delta X_2, \Delta Y_2), ..., (\Delta X_n, \Delta Y_n)].$$

In the case of linear cryptanalysis:

$$(\Gamma Y_i, \Gamma X_i) \stackrel{\text{def}}{=} |Prob\{parity(X_i \bullet \Gamma X_i) = parity(F_i(X_i, K_i) \bullet \Gamma Y_i)\} - 1/2|,$$

$$[p_1, p_2, ..., p_t] \stackrel{\text{def}}{=} 2^{t-1} \prod_{i=1}^{t} p_i,$$

$$B_n \stackrel{\text{def}}{=} \max_{\Gamma Y_i = \Gamma Y_{i-2} \oplus \Gamma X_{i-1}\ (3 \le i \le n)} [(\Gamma Y_1, \Gamma X_1), (\Gamma Y_2, \Gamma X_2), ..., (\Gamma Y_n, \Gamma X_n)],$$

where X_i and K_i are randomly given, and the symbol $\bullet$ represents a bitwise AND operation. ΔX_i and ΓX_i are called the differential value and the masking value of X_i, respectively.

Note: In general, our definition of the best probability does not necessarily indicate the following; it is unknown whether this difference is negligible or not in the case of DES:

$$B_n = \max_{(\Delta P, \Delta C) \neq (0,0)} Prob\{\ Cipher(P \oplus \Delta P, K) = Cipher(P, K) \oplus \Delta C\ \}, \quad or$$

$$B_n = \max_{(\Gamma P, \Gamma C) \neq (0,0)} |Prob\{parity(P \bullet \Gamma P) = parity(Cipher(P, K) \bullet \Gamma C)\} - 1/2|.$$

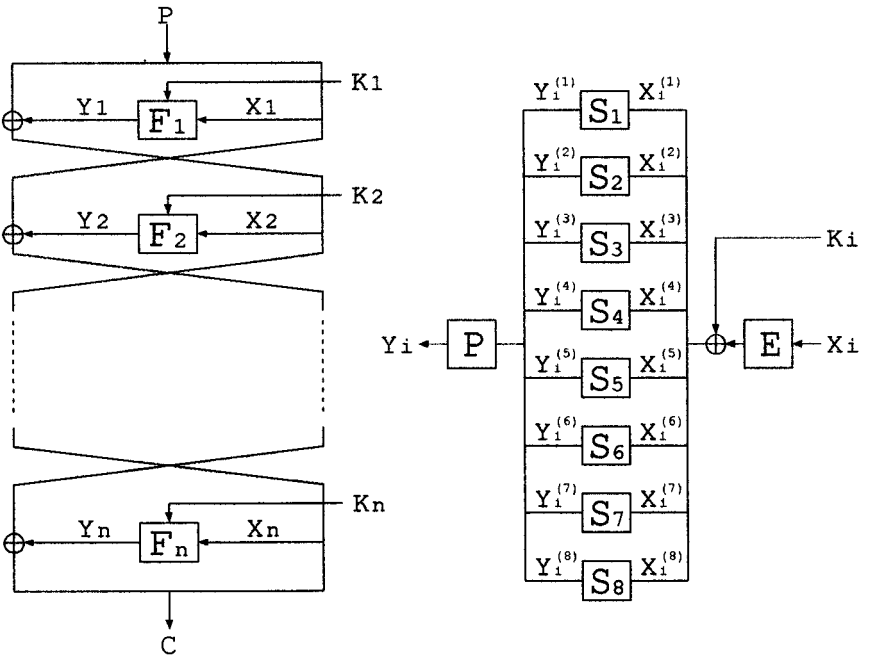

Fig. 1. The data randomization part and F-function of DES.

3 Duality Between Differential Cryptanalysis and Linear Cryptanalysis

This chapter briefly discusses correlation between differential cryptanalysis and linear cryptanalysis from the viewpoint of duality. A similar discussion is also seen in [3]. Differential cryptanalysis traces the spread of differential values to

establish the characteristic probability of the entire cipher structure. Figure 2 illustrates the flow of differential values in one round. The right input differential value ΔX_i is supplied into the F-function and also outputted as the right output differential value. We now assume that ΔX_i changes into ΔY_i through the F-function with probability p_i. Then ΔY_i is XORed with the left input differential value ΔX_{i-1} and finally outputted as the left output differential value $\Delta X_{i-1} \oplus \Delta Y_i$. This flow must be consistent throughout the entire cipher structure. In other words, equation $\Delta X_i = \Delta X_{i-2} \oplus \Delta Y_{i-1}$ $(3 \leq i \leq n)$ must hold. The total probability is represented as $\prod_{i=1}^{n} p_i$.

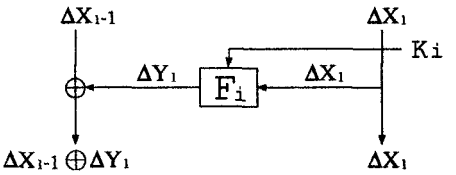

Fig. 2. The spread of differential values.

In the case of linear cryptanalysis, our description is completely reversed. Figure 3 explains how masking values flow through one round. This time we can consider that the left input masking value ΓY_i is supplied into the output side of the F-function and also outputted as the left output masking value. In the F-function, we interpret ΓY_i to be changed into ΓX_i with probability p_i. Then ΓX_i is XORed with the right input masking value ΓY_{i-1}, and finally outputted as the right output masking value $\Gamma X_i \oplus \Gamma Y_{i-1}$. The consistency of the flow requires that equation $\Gamma Y_i = \Gamma Y_{i-2} \oplus \Gamma X_{i-1}$ $(3 \leq i \leq n)$ holds, and the effectiveness of linearity is represented as $2^{n-1} \prod_{i=1}^{n} |p_i - 1/2|$ by piling-up lemma [2].

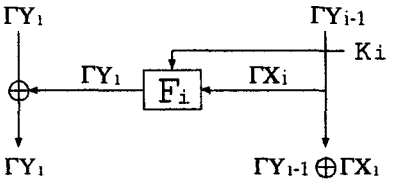

Fig. 3. The spread of masking values.

It follows that "XOR branch" after the F-function and "three-forked branch" before the F-function are mutually dual operations in regard to differential values and masked values. We may hence say that differential cryptanalysis goes downstream through an F-function, whereas linear cryptanalysis goes upstream through it. This fact is essential to the search algorithm in the following chapter.

4 The Search for the Best Probability

In this chapter we present a practical algorithm for deriving the best probability of DES in terms of differential cryptanalysis and linear cryptanalysis. This technique is also available to various block ciphers that have S-box-like tables. Since the search for the best differential characteristic is essentially the same as that for the best linear expression, we focus on differential cryptanalysis in the following. To apply the results of this chapter to linear cryptanalysis, just replace ΔX_i and ΔY_i to ΓY_i and ΓX_i, respectively.

Basically, our program works by induction on the number of rounds n. In other words, it derives the best n-round probability B_n from knowledge of the best i-round probability B_i $(1 \leq i \leq n-1)$. Since it is generally easy to calculate the best probability up to three rounds, we can usually start with $n = 4$. The program also requires an "initial value" for B_n, which is represented as $\overline{B_n}$, though it works correctly for any $\overline{B_i}$ as long as $\overline{B_i} \leq B_i$ $(1 \leq i \leq n-1)$; we will later discuss how to determine $\overline{B_n}$ for faster search. The framework of our algorithm is now established by the following procedures including essentially recursive calls:

Procedure Round-1:
Begin the program.
For each candidate for ΔX_1, do the following:
- Let $p_1 = max_{\Delta Y}(\Delta X_1, \Delta Y)$.
- If $[p_1, B_{n-1}] \geq \overline{B_n}$, then call *Procedure Round-2*.

Exit the program.

Procedure Round-2:
For each candidate for ΔX_2 and ΔY_2, do the following:
- Let $p_2 = (\Delta X_2, \Delta Y_2)$.
- If $[p_1, p_2, B_{n-2}] \geq \overline{B_n}$, then call *Procedure Round-3*.

Return to the upper procedure.

Procedure Round-i $(3 \leq i \leq n-1)$:
For each candidate for ΔY_i, do the following:
- Let $\Delta X_i = \Delta X_{i-2} \oplus \Delta Y_{i-1}$ and $p_i = (\Delta X_i, \Delta Y_i)$.
- If $[p_1, p_2, ..., p_i, B_{n-i}] \geq \overline{B_n}$, then call *Procedure Round-(i+1)*.

Return to the upper procedure.

Procedure Round-n:
Let $\Delta X_n = \Delta X_{n-2} \oplus \Delta Y_{n-1}$ and $p_n = max_{\Delta Y}(\Delta X_n, \Delta Y)$.
If $[p_1, p_2, ..., p_n] \geq \overline{B_n}$, then $\overline{B_n} = [p_1, p_2, ..., p_n]$.
Return to the upper procedure.

This program rewrites the initial value $\overline{B_n}$ while running, if it finds a better candidate for the best probability (*Procedure Round-n*). When it completes the search, $\overline{B_n}$ is equal to the best n-round probability B_n. We can easily see that

the initial value $\overline{B_n}$ is also effectively used for finding unnecessary branches and breaking them as soon as possible. Hence if we start with too small $\overline{B_n}$, it may take much time to complete the search, though the program works correctly for any initial value $\overline{B_n}$ as long as $\overline{B_n} \leq B_n$. In general, we should first derive a conditional best n-round probability instead of B_n by restricting the form of differential values or masking values, and thereby initialize $\overline{B_n}$ for faster search.

Next, we discuss the detailed inner structure of *Procedure Round-i*. Since it includes a big loop on ΔX_i and/or ΔY_i, it is not practical to try all 2^{32} or 2^{64} candidates one by one. In the following, we show an explicit implementation of *Procedure Round-2* that realizes a practical search using another recursive calls. Other procedures can be also carried out in a similar way:

Procedure Round-2: (detailed)
Let $a_0 = 0$.
Call *Procedure Round-2-1*.
Return to the upper procedure.

Procedure Round-2-j $(1 \leq j \leq 8)$:
For each candidate for a_j $(a_{j-1} < a_j \leq 8)$, $\Delta X_2^{(a_j)}$ and $\Delta Y_2^{(a_j)}$, do the following:
- Let $p_2^{(j)} = (\Delta X_2^{(a_j)}, \Delta Y_2^{(a_j)})$ and $p_2 = [p_2^{(1)}, p_2^{(2)}, ..., p_2^{(j)}]$.
- If $[p_1, p_2, B_{n-2}] \geq \overline{B_n}$ and $j \neq 8$, then call *Procedure Round-2-(j+1)*.
- Call *Procedure Round-3*.

Return to the upper procedure.

We should try $\Delta X_i^{(a_j)}$ and $\Delta Y_i^{(a_j)}$ in the order of magnitude of $(\Delta X_i^{(a_j)}, \Delta Y_i^{(a_j)})$ for fixed j so that we can avoid unnecessary calculations for $\Delta X_i^{(a_j)}$ and $\Delta Y_i^{(a_j)}$.

Our program has completely determined the best characteristic probability of DES, which was partially studied by Knudsen [4]. It took about 100 minutes on one HP9735 (PA-RISC/99MHz) computer to complete the search. As a result, we have found that DES reduced to seven or more rounds achieves the actual best probability by piling up 2-round iterative characteristics. Moreover, the best 5-round probability is better than that found by Biham and Shamir [1]. Table 1 summarizes the best n-round probability B_n $(4 \leq n \leq 16)$ of DES, where equation $B_n = B_{n-2} / 234$ $(9 \leq n \leq 16)$ holds.

Round	4	5	6	7	8	9
Probability	1.31×2^{-10}	1.72×2^{-14}	1.03×2^{-20}	1.31×2^{-24}	1.43×2^{-31}	1.43×2^{-32}
10	11	12	13	14	15	16
1.57×2^{-39}	1.57×2^{-40}	1.71×2^{-47}	1.71×2^{-48}	1.87×2^{-55}	1.87×2^{-56}	1.02×2^{-62}

Table 1: The best characteristic probability of DES.

We can also derive the best linear approximate probability of DES in the same manner. Our program completed the search in one minute on the same computer; the results can be seen in [2].

5 The Order of S-boxes and the Strength of DES

In this chapter, through various experimental results using our search program, we observe how the order of S-boxes affects the strength of DES from the viewpoints of differential cryptanalysis and linear cryptanalysis. Since it is time-consuming to make the complete search for all possible 8!=40320 permutations of the S-boxes, we begin by discussing conditional best probability for faster search, which will lead to general correlation between the order of S-boxes and the strength of DES.

5.1 Differential Cryptanalysis

First, we treat 2-round iterative characteristics, which have been effectively used for attacking the full 16-round DES by a chosen-plaintext attack faster than an exhaustive key search. Table 2 shows the distribution of the best 2-round iterative characteristic probability for 40320 possible permutations of the S-boxes of DES. Since the best 2-round iterative characteristic probability of the original DES is $1/234 = 1.09 \times 2^{-8}$, table 2 indicates that at most 256 permutations may be stronger against differential cryptanalysis.

We have confirmed, using the search program, that some of these 256 permutations actually achieve the best 16-round probability by piling up the 2-round iterative characteristic. For example, the modified DES with the order of the S-boxes "**27643158**" attains the actual best 16-round probability 1.50×2^{-64}, whereas the original DES has the probability 1.02×2^{-62}. However, even these 256 permutations cannot protect differential cryptanalysis, because their best 13-round probability is $(1.00 \times 2^{-8})^6 = 1.00 \times 2^{-48}$, while the original DES has the probability $(1.09 \times 2^{-8})^6 = 1.71 \times 2^{-48}$.

Probability	1.00×2^{-8}	1.09×2^{-8}	1.13×2^{-8}	1.25×2^{-8}	1.31×2^{-8}	1.50×2^{-8}
Frequency	256	832	832	7488	1152	5568
Probability	1.53×2^{-8}	1.75×2^{-8}	1.00×2^{-7}	1.09×2^{-7}	1.53×2^{-7}	1.75×2^{-7}
Frequency	3456	8256	2880	7680	960	960

Table 2: The distribution of the best 2-round iterative characteristic probability.

Next, we have calculated the conditional best 16-round probability based on 2-round iterative characteristics for 40320 possible permutations of the S-boxes; to be concrete, we have located three consecutive active S-boxes in the 2nd, 4th, ... and 14th rounds, and the locally best characteristic in the final round, respectively. Figure 4 shows the resultant distribution, where the arrow denotes the (actual) best 16-round probability of the original DES; namely, 1.02×2^{-62}. The number of permutations that have this probability is 32, and they are distributed throughout 2.38% – 2.46% from the best of 40320 permutations, which shows that the order of the S-boxes of DES is well-arranged against differential cryptanalysis. We do not know whether 2-round iterative characteristics establish the actual best 16-round probability for any permutation.

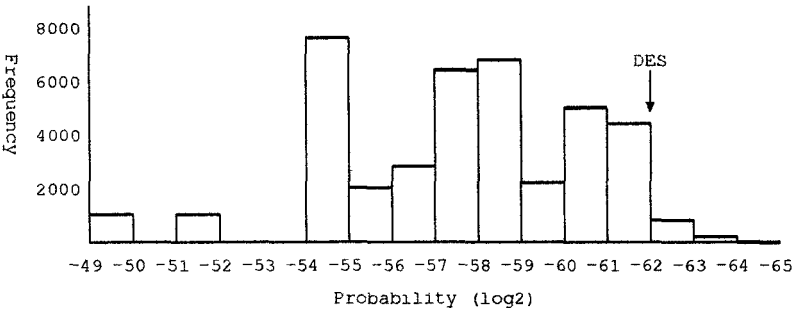

Fig. 4. The distribution of 16-round probability by 2-round iterative characteristics.

5.2 Linear Cryptanalysis

We here observe how the order of S-boxes affects the strength of DES with regard to linear cryptanalysis. Since the complete search for all 40320 permutations is time-consuming again, we begin with a partial search for the best 16-round probability.

First, we restrict ourselves to the case where at most one S-box is approximated in each round. This approximation is referred to as "Type I". The original DES achieves the actual best 16-round probability 1.49×2^{-24} by Type I approximation as described in [2]. This conditional search is easily executed for all 40320 permutations by eliminating the line including "*Procedure Round-2-(j+1)*" in "*Procedure Round-2-j*". In this case, our program works fast enough for arbitrarily small initial values $\overline{B_n}$. Figure 5 shows the resultant distribution, where the arrow denotes the original DES; namely, 1.49×2^{-24}. The number of permutations that have this probability is 2880, and this time they are distributed throughout 8.9% – 16.1% from the worst of 40320 permutations.

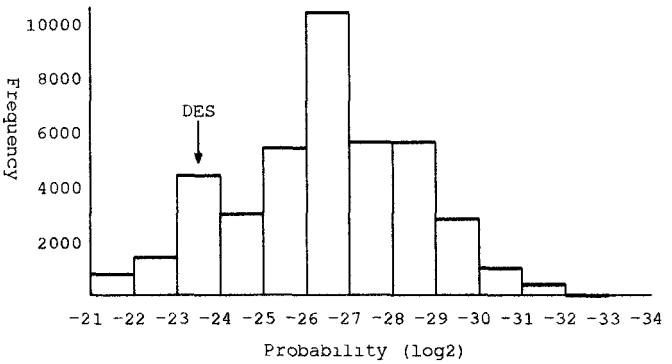

Fig. 5. The distribution of 16-round probability by Type I approximation.

However, there exist permutations that do not establish the actual best 16-round probability by Type I approximation. To explain this, we now introduce a linear approximation of F-function which is similar to 2-round iterative characteristics of differential cryptanalysis. Consider, for example, the following two linear approximations of F-function (see [2] for the notations):

$$X[3,4] \oplus F(X,K)[0,10,20,25] = K[6,7],$$
$$X[3,4] \oplus F(X,K)[5,11,17] = K[4,5].$$

These equations are derived from $NS_7(3,15)$ and $NS_8(48,13)$, and hold with probability 40/64 and 20/64, respectively. Then we have the following equation that holds with probability $1/2 + 2(40/64 - 1/2)(20/64 - 1/2) = 0.453$ by canceling the common term X:

$$F(X,K)[0,5,10,11,20,25,27] = K[4,5,6,7]$$

The left side of this equation does not contain any input information on the F-function. In other words, if input data X is random, we can guess one key bit from only output information without any input information. We can also obtain linear approximate expressions of arbitrary round DES by piling up this equation in every other round. This approximation is referred to as "Type II". We have calculated the conditional best 16-round probability by Type II approximation for all 40320 permutations. Figure 6 illustrates the resultant distribution.

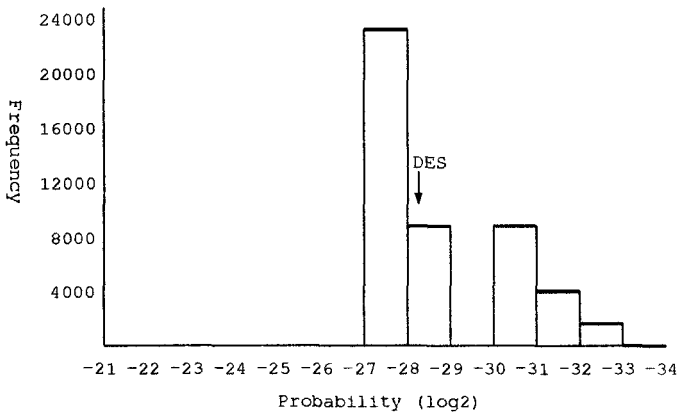

Fig. 6. The distribution of 16-round probability by Type II approximation.

Figure 7 summarizes the distribution of the better of the 16-round probabilities by Type I and Type II approximations. The probability of the original DES is again distributed throughout 8.9% – 16.1% (2880 permutations) from the worst of 40320 permutations, which suggests that the order of the S-boxes of the original DES is a very weak choice in regard to linear cryptanalysis.

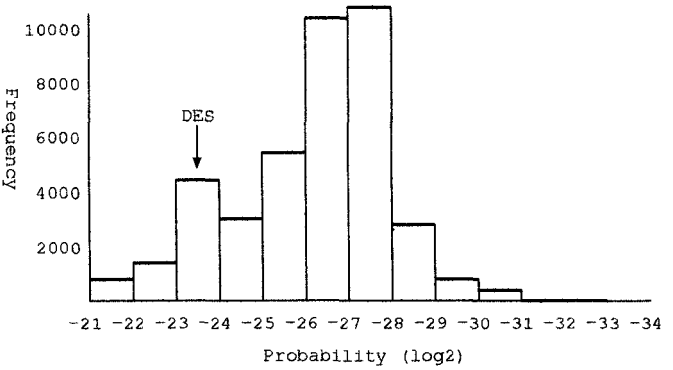

Fig. 7. The distribution of the best known 16-round probability.

Note: Recently, the author has found a new type of 16-round approximation which is not Type-I nor Type-II but attains the actual best 16-round probability. As far as the author knows, the rate of such permutations is small (2% - 3%) and moreover their best 16-round probability is at most 1.61×2^{-26}; therefore the distribution of the best 16-round probability of the original DES (8.9% - 16.1%) does not seem to be affected. More detailed data are under calculation.

We have determined the order of the S-boxes of DES which achieves the best "the actual best 16-round probability" of all 40320 permutations. It has the order of the S-boxes "**56412738**" and attains the actual best 16-round probability 1.60×2^{-33}, whereas the original DES has the probability 1.49×2^{-24}. This modified DES is now immune against linear cryptanalysis, though not good in regard to differential cryptanalysis since the 13-round probability is 1.61×2^{-45}.

There also exist permutations that are better than the original DES in regard to both differential cryptanalysis and linear cryptanalysis. One of such permutations is the order of the S-boxes "**24673158**", whose actual best 16-round characteristic probability is 1.75×2^{-63}, which is achieved by the best 2-round iterative characteristic. Its actual best 16-round linear approximate probability is 1.48×2^{-31}, which is best of 256 permutations that have the best 2-round iterative probability 1.00×2^{-8}. This choice again protects linear cryptanalysis.

References

1. Biham, E., Shamir, A.: Differential Cryptanalysis of the Data Encryption Standard. Springer-Verlag (1993)
2. Matsui, M.: Linear Cryptanalysis Method of DES Cipher. Advances in Cryptology — Eurocrypt'93, Lecture Notes in Computer Science **765** (1993) 386–397
3. Biham, B.: On Matsui's Linear Cryptanalysis. Pre-proceedings of Eurocrypt'94 (1994) 349–361
4. Knudsen, L.R.: Iterative Characteristics of DES and s^2-DES. Advances in Cryptology — Crypto'92, Lecture Notes in Computer Science **740** (1992)