A Strategy for Constructing Fast Round Functions with Practical Security Against Differential and Linear Cryptanalysis

Masayuki Kanda[1], Youichi Takashima[2], Tsutomu Matsumoto[3], Kazumaro Aoki[1], and Kazuo Ohta[1]

[1] NTT Information and Communication Laboratories
[2] NTT Human Interface Laboratories
1-1 Hikarino-oka Yokosuka-shi Kanagawa 239-0847 Japan
{kanda, ohta}@sucaba.isl.ntt.co.jp, maro@isl.ntt.co.jp
yoh@mistral.hil.ntt.co.jp
[3] Yokohama National University
79-5 Tokiwadai Hodogaya-ku Yokohama-shi Kanagawa 240-8501 Japan
tsutomu@mlab.dnj.ynu.ac.jp

Abstract. In this paper, we study a strategy for constructing fast and practically secure round functions that yield sufficiently small values of the maximum differential and linear probabilities p, q. We consider mn-bit round functions with 2-round SPN structure for Feistel ciphers.
In this strategy, we regard a linear transformation layer as an $n \times n$ matrix P over $\{0,1\}$. We describe the relationship between the matrix representation and the actual construction of the linear transformation layer. We propose a search algorithm for constructing the optimal linear transformation layer by using the matrix representation in order to minimize probabilities p, q as much possible. Furthermore, by this algorithm, we determine the optimal linear transformation layer that provides $p \leq p_s^5$, $q \leq q_s^5$ in the case of $n = 8$, where p_s, q_s denote the maximum differential and linear probabilities of s-box.

1 Introduction

1.1 Background

Differential cryptanalysis (DC) [6] proposed by Biham and Shamir and linear cryptanalysis (LC) [15] proposed by Matsui are the most powerful approaches to attacking most symmetric block ciphers. Accordingly, the designer should evaluate the security of any new proposed symmetric cipher against DC/LC and prove that it is sufficient invulnerable against them. It is known that there are four measures in order to evaluate the security of a cipher against DC/LC.

Precise measure The maximum average of differential and linear probabilities. They are also called differential probability [14] and approximate linear hull [18].

S. Tavares and H. Meijer (Eds.): SAC'98, LNCS 1556, pp. 264–279, 1999.

Lai et al. [14] and Nyberg [18] stated that the precise evaluation of the security of cipher against DC/LC should be done using this measure. Generally speaking, however, since it is infeasible to evaluate these probabilities, this measure is not practical.

Theoretical measure The upper bounds of the maximum average of differential and linear probabilities. This measure was applied to evaluate the security of MISTY [17] and the cipher $\mathcal{KN}$ [20].

These probabilities are evaluated as $2p^2,\ 2q^2$ for an R-round Feistel cipher ($R \geq 4$), respectively [20], where $p,\ q$ is the maximum differential and linear probabilities of the round function. Furthermore, for an R-round Feistel cipher with bijective round function ($R \geq 3$), they are evaluated as $p^2,\ q^2$, respectively [3]. Nyberg and Knudsen [20] stated that Feistel ciphers evaluated with this measure are **provably secure** against DC/LC, which means that they are theoretically invulnerable to DC/LC, if these probabilities are sufficiently small. However, when designers intend to construct provably secure ciphers, they have to construct round functions yielding extremely small values of the probabilities $p,\ q$. This is a very strong constraint on the design of round functions.

Heuristic measure The maximum differential and linear characteristic probabilities. This measure was applied to estimate the security of DES [15], FEAL [21,2] and so on.

Biham and Shamir [6] claimed that the larger differential characteristic probability is, the higher is the success rate of DC, since they exploited a single path between plaintexts and ciphertexts which holds with significant differential characteristic probability. Matsui [15] also claimed the same for LC. However, these probabilities only give the lower bounds of the maximum average of differential and linear probabilities for some ciphers, since there are multiple paths between the same plaintexts and ciphertexts in practice [14,18]. Furthermore, it takes much time to estimate them, since this involves the use of path searching algorithms, e.g., [16,21,2].

Practical measure The upper bounds of the maximum differential and linear characteristic probabilities.

As shown (3) and (4) in Sect. 2.2 below, given the maximum differential and linear probabilities of round function $p,\ q$, these probabilities are evaluated to be decreasing functions with the number of rounds of a Feistel cipher. Knudsen [11] noted that Feistel ciphers evaluated with this measure are **practically secure** against DC/LC, i.e., that they were believed to be invulnerable against DC/LC if these probabilities are less than a secure criterion, e.g., 2^{-64} for 64-bit ciphers and 2^{-128} for 128-bit ciphers. This implies that designers can construct practically secure Feistel ciphers that consist of good round functions while also considering their invulnerability to other known attacks and implementation efficiency with a moderate number of rounds.

Heys and Tavares studied the upper bounds of the maximum differential and linear characteristic probabilities of traditional SPN [9]. They focused on diffusion layers constructed by one bit operations. However, we consider

that the SPN structures with diffusion layers constructed around word operations, e.g., 8-bit length, should be studied, since bit operations are hard to implement in software. Of interest, Rijmen et al. [22] introduced the branch number $\mathcal{B}$ for the SPN cipher, which is a lower bound of the number of active s-boxes in two consecutive rounds of a non-trivial linear trail or a non-trivial differential characteristic. The branch number $\mathcal{B}$ is a very similar in concept to these probabilities, since the security of an SPN cipher against DC/LC is evaluated by piling up the branch numbers every two rounds.

Of course, new ciphers have to be invulnerable against other known attacks; e.g., higher order differential attack [13,12,10] and interpolation attack [10]. Unfortunately, all the above measures evaluate the security against only DC/LC. That is, these evaluations do not prove that the cipher is also secure against other known attacks. For example, the cipher $\mathcal{KN}$ with 6-round, which was evaluated by the theoretical measure, was broken by higher order differential attack [10], even though the maximum average of differential and linear probabilities is less than 2^{-60}, which means that the cipher is provably secure against DC/LC [20].

Moreover, it is known empirically that, for good round functions, increasing the number of rounds will make the cipher more secure against most known attacks. Therefore, we consider that the number of rounds is more important in order to construct a secure cipher than the values of the maximum differential and linear probabilities of round function p, q. Thus, we believe that the practical measure is more useful that the theoretical measure evaluating the security of a cipher against DC/LC.

1.2 Design

Later we discuss the round functions of Feistel ciphers that are practically secure against DC/LC. Our design strategy is as follows:

(a) To enable the maximum differential and linear probabilities of round function to be evaluated, and to prove our cipher to be sufficiently invulnerable against DC/LC.
(b) To realize a fast round function.
(c) To enable it to be efficiently implemented on multiple platforms, e.g., 8-bit, 16-bit, 32-bit, and 64-bit processors.

As mentioned above, we believe that the number of rounds is more important in constructing secure ciphers than the values of the maximum differential and linear probabilities of round function p, q. Unfortunately, though increasing the number of rounds makes the cipher more secure, it also slows the encryption speed. This means that it is very difficult to construct a fast cipher that has a sufficient number of rounds, even if the round function is fast. Thus, we believe that our cipher should have a reasonable number of rounds and suitable differential and linear probabilities of round function p, q.

Furthermore, to satisfy strategy (c), we intend to use s-boxes as non-linear transformations, since the approach of realizing s-boxes by table look-up has

no limitation with respect to the selection of function, software instructions of arithmetic operations depend on architectures, etc.

Thus, in this paper, we consider an mn-bit round function with m-bit s-boxes[1] to construct a fast and practically secure Feistel cipher. It is known that, for Feistel ciphers, major current round functions with s-boxes are based on (A) *1-round SPN structure*, or (B) *Recursive structure*. Examples of former are DES [8], CAST [1] and LOKI91 [5]; an example of the latter is MISTY [17].

Next, we focus on the total number of s-boxes in a cipher, which means how many s-boxes the cipher contains, not how many different s-boxes there are in the cipher. Roughly speaking, the encryption time is proportional to the total number of s-boxes under the assumptions that the run time except s-box process is negligible and that parallel processing is left out of consideration. That is, in order to construct a fast cipher, it is desirable that the total number of s-boxes in it be as small as possible. On the other hand, in order to construct a practically secure cipher, we must ensure that there are as many number of active s-boxes in it as the designer specified.

(A) 1-round SPN Structure This structure has one non-linear transformation layer with n parallel m-bit s-boxes and one linear transformation (permutation) layer.

Since the minimum number of active s-boxes in this structure is 1, the maximum differential and linear probabilities of round function $p,\ q$ are equal to the maximum differential and linear probabilities of s-box $p_s,\ q_s$; i.e., $p = p_s,\ q = q_s$. It is known that $p_s,\ q_s$ are equal to or larger than 2^{-m+1} when m is odd or 2^{-m+2} when m is even [7]. This means that too many rounds are required in order to construct a practically secure Feistel cipher; e.g., 33 rounds are required for a 128-bit Feistel cipher with a 64-bit bijective round function using 8-bit s-boxes whose the maximum differential and linear probabilities are $p_s = q_s = 2^{-6}$.

By the way, Nyberg [19] investigated the upper bounds of the maximum differential and linear probabilities of a generalized Feistel cipher with 1-round SPN structure, where "generalized" means "generalized Feistel structure," not "generalized linear layer." That is, unlike our research, the structure of the round function is not discussed. Furthermore, it only showed that the upper bounds of the maximum differential and linear probabilities of a cipher are of order $p_s^{2n},\ q_s^{2n}$ in the case of $n \leq 4$.

(B) Recursive Structure The recursive structure consists of nested functions with different input bit lengths.

For example, consider the 3-round recursive structure. In this case, since the number of s-boxes triples with each recursion, it is difficult to construct fast round functions because the number of s-boxes is too large. Furthermore, increasing the number of rounds rapidly increases the total number of s-boxes. Generally speaking, though the maximum differential and linear probabilities of round function $p,\ q$ are extremely small, the number of rounds must be at least 8 in order to provide invulnerability against other known attacks.

[1] In this paper, we do not consider expand permutations in round functions.

We consider that too many s-boxes would be required in a practically secure cipher using 1-round SPN structure, since the maximum differential and linear probabilities of round function p, q are not sufficiently small. On the other hand, we also consider that it is difficult to increase the number of rounds of a cipher that use the recursive structure, since the number of s-boxes in the round function is too large. Accordingly, we need a new round function structure that makes the probabilities p, q much smaller than those achieved with the 1-round SPN structure and that requires a lot fewer s-boxes than the recursive structure.

1.3 2-round SPN Structure Approach

Our strategy is based on using mn-bit round functions consisting of three-layers (see Fig. 1): "1st non-linear transformation layer with n parallel m-bit s-boxes, linear transformation layer, and 2nd non-linear transformation layer with n parallel m-bit s-boxes" in this order. We call this the **2-round SPN structure** hereafter. Each layer has the following feature.

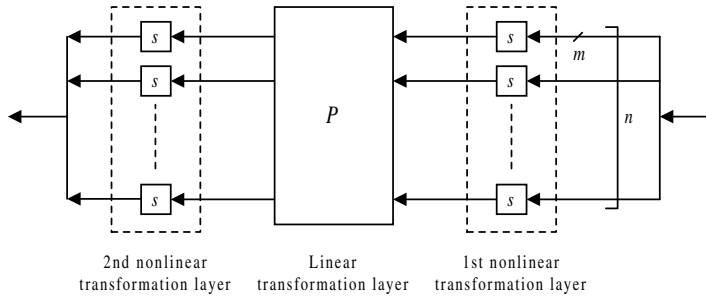

Fig. 1. 2-round SPN structural round function

1st non-linear transformation layer n parallel m-bit s-boxes are set in DES-like manner. In the following, we call this "1st non-linear layer."

2nd non-linear transformation layer n parallel m-bit s-boxes are set similar to the 1st non-linear layer. With the addition of this layer, the maximum differential and linear probabilities of round function become much smaller. In the following, we call this "2nd non-linear layer."

Linear transformation layer This makes the maximum differential and linear probabilities as small as possible for all non-zero input differences and all non-zero output mask values of round function[2]. Here, we consider that this is constructed only with bitwise XORs. That is, inputs are transformed

[2] Vaudenay proposed to use multipermutation as diffusion (permutation) [23]. Multipermutation is a good cryptographic primitive, but, it is very hard to satisfy the multipermutation requirements, and to make multipermutation with large bit size.

linearly to outputs per m-bits, especially per byte in the case of $m = 8$. In the following, we call this "linear layer."

By the way, all round subkeys are bitwise XORed with data before each s-box in order to form a Markov cipher [14]. Accordingly, we take no account of key addition layer here.

1.4 Result

For the 2-round SPN structure, an interesting issue is how to construct the optimal linear layer, since the two non-linear layers are already set. To construct the optimal linear layer, we represent a linear layer as an $n \times n$ matrix P over $\{0, 1\}$, and describe how to determine the matrix elements in order to minimize the differential and linear probabilities of round function $p,\ q$ as much as possible.

In this paper, we propose a new search algorithm based on matrix representation for constructing the optimal linear layer. Furthermore, with this algorithm, we determine the optimal linear transformation layer that provides probabilities $p \leq p_s^5,\ q \leq q_s^5$ where $n = 8$, where $p_s,\ q_s$ are the maximum differential and linear probabilities of the m-bit s-boxes in the non-linear layers.

As shown in (3) and (4) in Sect. 2.2 below, the number of rounds must be dependent on the maximum differential and linear probabilities of round function $p,\ q$ to construct a practically secure cipher. We show that the round function with the 2-round SPN structure requires one-fourth as many rounds as the 1-round SPN structure to achieve the same differential and linear probabilities, while each round has double the number of s-boxes. Hence, the round function using the 2-round SPN structure is twice as efficient as that using the 1-round SPN structure.

1.5 Organization

This paper is organized as follows. In Sect. 2, we introduce some preliminary definitions. In Sect. 3, we describe the relationship between the matrix representation and the actual construction of the linear transformation layer. We propose a new search algorithm for constructing the optimal linear layer. In Sect. 4, we show the optimal linear layer for the case of $n = 8$. Finally, we conclude with a summary in Sect. 5.

2 Preliminaries

2.1 Notations

- $\#\{S\}$: number of elements in set S.
- $\Delta x, \Gamma x$: difference of x, mask value of x.
- $a \bullet b$: parity of bitwise product a and b.
- $P = [t_{ij}],\ t_{ij} \in \{0, 1\},\ 0 \leq i, j < n$: matrix representation of a linear layer
- $\boldsymbol{z} = {}^T[z_0, \ldots, z_{n-1}],\ z_i \in GF(2)^m,\ 0 \leq i < n$: an input to a linear layer
- $\boldsymbol{z'} = {}^T[z'_0, \ldots, z'_{n-1}] = P\boldsymbol{z},\ z'_i \in GF(2)^m,\ 0 \leq i < n$: an output from a linear layer

2.2 Definitions

In this paper, we consider Feistel ciphers with mn-bit round function. We assume that a round key, which is used within one round, consists of independent and uniformly random bits and is bitwise XORed with data. Furthermore, we assume that an input also consists of independent and uniformly random bits and that all m-bit s-boxes work independently. Accordingly, we neglect the effect of the round key.

We use the following definitions in this paper.

Definition 1. *For given Δx, Γx and Δy, Γy, the differential and linear probabilities of an s-box are defined as:*

$$DP^s(\Delta x \to \Delta y) = \frac{\#\{x \in GF(2)^m | s(x) \oplus s(x \oplus \Delta x) = \Delta y\}}{2^m}$$

$$LP^s(\Gamma y \to \Gamma x) = \left(2 \times \frac{\#\{x \in GF(2)^m | x \bullet \Gamma x = s(x) \bullet \Gamma y\}}{2^m} - 1\right)^2$$

Definition 2. *The maximum differential and linear probabilities of an s-box, p_s, q_s, are defined as:*

$$p_s = \max_{\Delta x \neq 0, \Delta y} DP^s(\Delta x \to \Delta y)$$

$$q_s = \max_{\Gamma x, \Gamma y \neq 0} LP^s(\Gamma y \to \Gamma x)$$

Definition 3. *For given $\Delta \boldsymbol{x}$, $\Gamma \boldsymbol{x}$ and $\Delta \boldsymbol{y}$, $\Gamma \boldsymbol{y}$, differential and linear probabilities of a round function with the 2-round SPN structure are defined as:*

$$p(\Delta \boldsymbol{x} \to \Delta \boldsymbol{y}) = \max_{\Delta \boldsymbol{z}} \prod_{i=0}^{n-1} DP^s(\Delta x_i \to \Delta z_i) p(\Delta \boldsymbol{z} \to \Delta z_i') DP^s(\Delta z_i' \to \Delta y_i)$$

$$q(\Gamma \boldsymbol{y} \to \Gamma \boldsymbol{x}) = \max_{\Gamma \boldsymbol{z'}} \prod_{i=0}^{n-1} LP^s(\Gamma y_i \to \Gamma z_i') p(\Gamma \boldsymbol{z'} \to \Gamma z_i) LP^s(\Gamma z_i \to \Gamma x_i)$$

where $\Delta \boldsymbol{x}$ denotes $(\Delta x_0, \ldots, \Delta x_{n-1})$. Also, $\Delta \boldsymbol{y}$, $\Gamma \boldsymbol{x}$, and $\Gamma \boldsymbol{y}$ are denoted in the same way as $\Delta \boldsymbol{x}$. $\Delta \boldsymbol{z}$, $\Gamma \boldsymbol{z}$ denote the input difference and mask value of the linear layer, and yield $(\Delta z_0, \ldots, \Delta z_{n-1})$, $(\Gamma z_0, \ldots, \Gamma z_{n-1})$, respectively. Similarly, $\Delta \boldsymbol{z'}$, $\Gamma \boldsymbol{z'}$ denote the output difference and mask value of the linear layer, respectively. Here, $\Delta \boldsymbol{z}$ transforms for each i into $\Delta z_i'$ with the linear layer, $\{GF(2)^m\}^n \to GF(2)^m$; i.e., for each i, $\exists \Delta z_i'$ s.t. $p(\Delta \boldsymbol{z} \to \Delta z_i') = 1$. Similarly, $\exists \Gamma z_i$ s.t. $p(\Gamma \boldsymbol{z'} \to \Gamma z_i) = 1$.

Definition 4. *The maximum differential and linear probabilities of round function p, q are defined as:*

$$p = \max_{\Delta \boldsymbol{x} \neq 0, \Delta \boldsymbol{y}} p(\Delta \boldsymbol{x} \to \Delta \boldsymbol{y})$$

$$q = \max_{\Gamma \boldsymbol{x}, \Gamma \boldsymbol{y} \neq 0} q(\Gamma \boldsymbol{y} \to \Gamma \boldsymbol{x})$$

Definition 5. Active s-box *is defined as an s-box given a non-zero input difference or a non-zero output mask value.*
Note: When an s-box is bijective, the s-box given a non-zero output difference or a non-zero input mask value is also an active s-box.

Definition 6. *Assume that all s-boxes are bijective. The minimum number of active s-boxes n_d, n_l in a round function with the 2-round SPN structure for DC and LC are defined as:*

$$\begin{aligned} n_d &= \min_{\Delta \boldsymbol{z} \neq 0} [w_H(\Delta \boldsymbol{z}) + w_H(\Delta \boldsymbol{z'})] \\ n_l &= \min_{\Gamma \boldsymbol{z'} \neq 0} [w_H(\Gamma \boldsymbol{z}) + w_H(\Gamma \boldsymbol{z'})] \end{aligned} \tag{1}$$

where $w_H(\boldsymbol{z})$ denotes the Hamming weight of $\boldsymbol{z}$, which means the number of non-zero subblocks from $\mathrm{GF}(2)^m$ of $\boldsymbol{z}$; i.e. $w_H(\boldsymbol{z}) = \#\{0 \leq i < n | z_i \neq 0\}$.

Theorem 1. *Let n_d, n_l denote the minimum number of active s-boxes in a round function for DC and LC, respectively. Then, the probabilities p, q hold for*

$$p \leq p_s^{n_d}, \quad q \leq q_s^{n_l} \tag{2}$$

Definition 7. *For an R-round Feistel cipher, assume that $\boldsymbol{x_i}$ $(0 \leq i \leq R+1)$, which is an input to the i-th round function, is an independent random variable. The maximum differential and linear characteristic probabilities $DCP_{max}^{(R)}$, $LCP_{max}^{(R)}$ are defined as:*

$$DCP_{max}^{(R)} = \max_{(\Delta \boldsymbol{x_0}, \Delta \boldsymbol{x_1}) \neq 0, (\Delta \boldsymbol{x_R}, \Delta \boldsymbol{x_{R+1}})} \prod_{i=1}^{R} p(\Delta \boldsymbol{x_i} \rightarrow \Delta \boldsymbol{x_{i-1}} \oplus \Delta \boldsymbol{x_{i+1}})$$

$$LCP_{max}^{(R)} = \max_{(\Gamma \boldsymbol{x_1}, \Gamma \boldsymbol{x_0}), (\Gamma \boldsymbol{x_R}, \Gamma \boldsymbol{x_{R+1}}) \neq 0} \prod_{i=1}^{R} q(\Gamma \boldsymbol{x_i} \rightarrow \Gamma \boldsymbol{x_{i-1}} \oplus \Gamma \boldsymbol{x_{i+1}})$$

Theorem 2. *For a Feistel cipher with $R = 2r$, $2r+1$ rounds, the upper bounds of the maximum differential and linear characteristic probabilities are estimated as follows [11]:*

$$DCP_{max}^{(R)} \leq p^r, \quad LCP_{max}^{(R)} \leq q^r \tag{3}$$

Theorem 3. *For a Feistel cipher with $R = 3r$, $3r+1$, $3r+2$ rounds and a bijective round function, the upper bounds of the maximum differential and linear characteristic probabilities are estimated as follows:*

$$\begin{aligned} &\text{In case of } R = 3r, 3r+1: \quad DCP_{max}^{(R)} \leq p^{2r}, \quad LCP_{max}^{(R)} \leq q^{2r} \\ &\text{In case of } R = 3r+2: \quad DCP_{max}^{(R)} \leq p^{2r+1}, LCP_{max}^{(R)} \leq q^{2r+1} \end{aligned} \tag{4}$$

Theorem 4. *Concatenation Rules [4,16]*
When "$X \dashv (Y, Z)$" denotes that X branches into Y and Z, i.e., $X = Y = Z$, the following relations hold.

$X = Y \oplus Z \Rightarrow \Delta X = \Delta Y \oplus \Delta Z,\ \Gamma X = \Gamma Y = \Gamma Z$ *(XOR operation)*
$X \dashv (Y, Z) \Rightarrow \Delta X = \Delta Y = \Delta Z,\ \Gamma X = \Gamma Y \oplus \Gamma Z$ *(BRANCH operation)*

3 Relationship between Matrix Representation and Structure

We represent the linear layer as an $n \times n$ matrix P over $\{0, 1\}$. This means that inputs are transformed linearly to outputs per m-bits, per byte in the case of $m = 8$, and the linear layer can be constructed with just bitwise XORs. For example,

$$z'_i = \bigoplus_{j=0}^{n-1} t_{ij} z_j = \bigoplus_{t_{ij}=1} \{z_j\}$$

where t_{ij} denotes the element of matrix P in row i and column j.

We assume that the maximum differential and linear probabilities of the s-box is $p_s,\ q_s$. From (2) in Theorem 1, it turns out that evaluating the upper bounds of the maximum differential and linear probabilities of round function $p,\ q$ is equivalent to counting the minimum number of active s-boxes $n_d,\ n_l$. On the other hand, the optimal linear layer leads to the optimal round function; i.e., the upper bounds of the probabilities $p_s^{n_d},\ q_s^{n_l}$ are minimum. Accordingly, constructing the optimal linear layer is equivalent to determining the matrix elements $P = [t_{ij}]$ yielding the maximum value of number $n_d,\ n_l$, because $p_s \leq 1,\ q_s \leq 1$. Here, we note that $n_d,\ n_l$ is obviously equal to or less than $n + 1$, i.e.,$n_d, n_l \leq n + 1$, from (1) in Definition 6, because $w_H(\boldsymbol{z}) \leq n$.

For example, we consider the 4×4 matrix P_E such as

$$\begin{bmatrix} z'_0 \\ z'_1 \\ z'_2 \\ z'_3 \end{bmatrix} = \begin{bmatrix} 0\,1\,1\,1 \\ 1\,0\,1\,1 \\ 1\,1\,1\,0 \\ 1\,1\,1\,1 \end{bmatrix} \begin{bmatrix} z_0 \\ z_1 \\ z_2 \\ z_3 \end{bmatrix}$$

3.1 Relationship between Matrix Representation and Features of Round Function

The round function using the linear layer represented as matrix P_E has the following features. Here, we assume that all s-boxes are bijective.

- We obtain $z'_0 = 0 \cdot z_0 \oplus 1 \cdot z_1 \oplus 1 \cdot z_2 \oplus 1 \cdot z_3 = z_1 \oplus z_2 \oplus z_3$. Similarly, $z'_1,\ z'_2,\ z'_3$ can be represented by XORs among $z_0,\ z_1,\ z_2,\ z_3$. Furthermore, the differential characteristic can be represented in the same way.
- The matrix P_E represents which inputs of the linear layer are combined into each output of the linear layer. Each row corresponds to each output of the linear layer, and each column corresponds to each input of the linear layer, see Fig. 2.
- Invulnerability against DC and LC can be evaluated as the minimum number of active s-boxes $n_d,\ n_l$ in the round function, respectively. $n_d,\ n_l$ can be obtained with the search algorithm described in Sect. 3.2.
- The Hamming weight of the column vector denotes the avalanche effect. A large value of the Hamming weight means that the round function has a good avalanche effect.

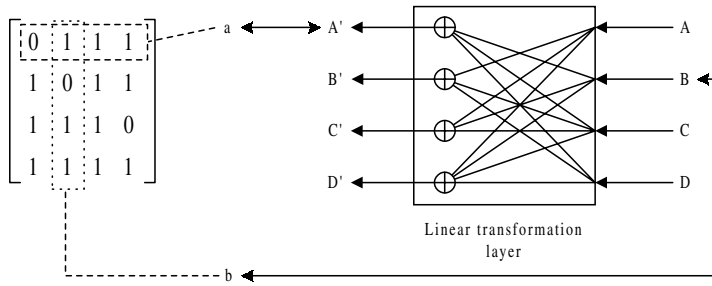

Fig. 2. Relationship between matrix representation and linear layer

As mentioned above, the preliminary features of the round function, i.e., invulnerability against DC/LC and the avalanche effect, are obtained by just matrix elements regardless of construction of linear layer.

3.2 Determination of Matrix Elements

Generally speaking, given matrix P, there are many constructions of the linear layer that correspond to matrix P. This is because matrix P denotes only the relationship between inputs and outputs of the linear layer, not the construction of the linear layer. That is, when several linear layers can be represented by the same matrix P, the round functions have same the features regardless of the constructions of the linear layers.

Accordingly, at first, we determine matrix elements in order to provide good invulnerability against DC/LC and good avalanche effect, and then realize the optimal linear layer.

Consideration of Differential Characteristic Based on the nature of differential characteristic, the matrix elements of $n \times n$ matrix P are determined by the following algorithm.

Search Algorithm

Step1 Define *security threshold* T $\quad (2 \leq T \leq n)$.
Step2 Prepare a set of column vectors C whose Hamming weights are equal to or larger than $T - 1$.
Step3 Select a subset of n column vectors P_c from set C. Repeat the following steps3-1 and step3-2 until all subsets have been checked.
Step3-1 Compute the minimum number of active s-boxes n_d for subset P_c. This is represented as $n_d(P_c)$.
Step3-2 If $n_d(P_c) \geq T$, then accept the matrix consisting of subset P_c as a candidate matrix.
Step4 Output matrix P and $n_d(P)$ that yields the maximum value of n_d among the candidate matrices.

The security threshold T is used in order to restrict candidate subsets in Step 2 and Step 3-2. If candidate matrices are found by the above algorithm, then it can be proven that the minimum number of s-boxes is equal to or larger than T. That is, $p \leq p_s^T$.

Consideration of Linear Expression Similarly, linear expression represents which output mask values of the linear layer are combined to yield each input mask value of the linear layer. The linear expression may be obtained with the concatenation rules in Theorem 4, when the construction of a linear layer is given. If each input mask value of the linear layer is represented by XORs among output mask values of the linear layer, the result is an other transformation matrix similar to that yielded by differential characteristic.

Here, based on our study of the relationship between the matrix for differential characteristic and the matrix for linear expression, we make the following two conjectures.

Conjecture 1. Given an $n \times n$ matrix P over $\{0, 1\}$ for a linear layer. The matrix for differential characteristic, i.e., relationship between input and output differences, is represented as the same matrix P, while the matrix for linear expression, i.e., relationship between output and input mask values, is represented as the transposed matrix ${}^TP$. That is, $\Delta z' = P\Delta z$, $\Gamma z = {}^TP\Gamma z'$.

Conjecture 2. The minimum number of active s-boxes n_d for differential characteristic represented as matrix P is equal to the minimum number of active s-boxes n_l for linear expression represented as the transposed matrix ${}^TP$.

Because of Conjecture 2, the minimum number of s-boxes n_l is also equal to or larger than T, when candidate matrices are found by the above algorithm. That is, $q \leq q_s^T$. For example, matrix P_E holds the following relationship. It is proven that $n_d = 3$, $n_l = 3$ for matrices P_E and ${}^TP_E$, respectively.

$$\begin{bmatrix} 0&1&1&1 \\ 1&0&1&1 \\ 1&1&1&0 \\ 1&1&1&1 \end{bmatrix} \iff \begin{bmatrix} 0&1&1&1 \\ 1&0&1&1 \\ 1&1&1&1 \\ 1&1&0&1 \end{bmatrix}$$

Matrix for differential characteristic P_E Matrix for linear expression ${}^TP_E$

3.3 Determining the Construction of Linear Layer

In this section, we describe how to determine the construction of a linear layer given matrix P.

Determination algorithm

Step 1 Choose two rows, and one row (row a) is XORed to the other row (row b) (called "primitive operation" hereafter).

Step 2 Transform the matrix P into the unit matrix I by the primitive operation, and count the number of the operations in order to find the transformation order yielding the minimum number of operations.

Step 3 Line A, which corresponds to row a, is XORed to line B, corresponds to row b, in reverse order of the transformation order found in Step 2.

For example, consider the matrix P_E. First, row 1 is XORed to row 2, and then row 4 is XORed to row 1. Repeat until matrix P_E is transformed into the unit matrix I. All states of transformation are described as follows.

$$\begin{bmatrix} 0\,1\,1\,1 \\ 1\,0\,1\,1 \\ 1\,1\,1\,0 \\ 1\,1\,1\,1 \end{bmatrix} \xrightarrow[(1 \to 2)]{} \begin{bmatrix} 0\,1\,1\,1 \\ 1\,1\,0\,0 \\ 1\,1\,1\,0 \\ 1\,1\,1\,1 \end{bmatrix} \xrightarrow[(4 \to 1)]{} \begin{bmatrix} 1\,0\,0\,0 \\ 1\,1\,0\,0 \\ 1\,1\,1\,0 \\ 1\,1\,1\,1 \end{bmatrix} \xrightarrow[(3 \to 4)]{} \begin{bmatrix} 1\,0\,0\,0 \\ 1\,1\,0\,0 \\ 1\,1\,1\,0 \\ 0\,0\,0\,1 \end{bmatrix}$$

$$\xrightarrow[(2 \to 3)]{} \begin{bmatrix} 1\,0\,0\,0 \\ 1\,1\,0\,0 \\ 0\,0\,1\,0 \\ 0\,0\,0\,1 \end{bmatrix} \xrightarrow[(1 \to 2)]{} \begin{bmatrix} 1\,0\,0\,0 \\ 0\,1\,0\,0 \\ 0\,0\,1\,0 \\ 0\,0\,0\,1 \end{bmatrix}$$

For the above transformation, the number of operations is 5. That is, the linear layer can be constructed with 5 XORs.

In reverse order of the above transformation order, line 1 is XORed to line 2, and then line 2 is XORed to line 3, and so on. Finally, we construct the linear layer described in Fig. 3.

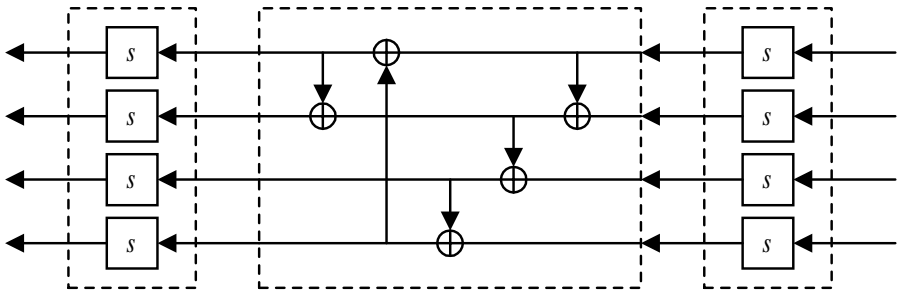

Fig. 3. Example of linear layer represented as the matrix P_E

4 Round Function with $n = 8$

In this section, we consider a round function with $n = 8$. Here, we assume that the linear layer and s-boxes are bijective, because this enables us to evaluate the invulnerability against DC/LC more precisely in the case of the bijective round function, described as (4) in Theorem 3 (Also, shown in [3]).

4.1 Determination for Matrix Elements

We determine an 8×8 matrix P ($n = 8$) yielding the maximum value of n_d using the search algorithm of Sect. 3.2. At this time, we set the security threshold $T = 8, 7, \ldots$ in this order until we found candidate matrices. Furthermore, the following conditions were added within the algorithm.

Step3 Select subset of n column vectors P_c from the set C, and compute $\mathrm{rank}(P_c)$. If $\mathrm{rank}(P_c) \neq 8$, then reject subset P_c.
Repeat the following steps until all subsets have been checked.

Step3-1 Compute the minimum number of active s-boxes n_d for P_c as follows.

- $n_d = \min\{n_{di} | 0 \leq i \leq 9\}$
- For any two columns (Columns a, b, $0 \leq i < 8$)
  $n_{d0} = 2 + \min\limits_{(a,b)} \#\{(t_{ia}, t_{ib}) | t_{ia} \oplus t_{ib} \neq 0\}$
- For any three columns (Columns a, b, c, $0 \leq i < 8$)
  $n_{d1} = 3 + \min\limits_{(a,b,c)} \#\{(t_{ia}, t_{ib}, t_{ic}) | t_{ia} \oplus t_{ib} \oplus t_{ic} \neq 0\}$
  $n_{d2} = 3 + \min\limits_{(a,b,c)} \#\{(t_{ia}, t_{ib}, t_{ic}) | \text{Exception of } (0,0,0), (1,1,1)\}$
- For any four columns (Columns a, b, c, d, $0 \leq i < 8$)
  $n_{d3} = 4 + \min\limits_{(a,b,c,d)} \#\{(t_{ia}, t_{ib}, t_{ic}, t_{id}) | t_{ia} \oplus t_{ib} \oplus t_{ic} \oplus t_{id} \neq 0\}$
  $n_{d4} = 4 + \min\limits_{(a,b,c,d)} \#\left\{(t_{ia}, t_{ib}, t_{ic}, t_{id}) \middle| \begin{array}{l}\text{Exception of } (0,0,0,0),\\ (1,1,0,0), (0,1,1,1), (1,0,1,1)\end{array}\right\}$
  $n_{d5} = 4 + \min\limits_{(a,b,c,d)} \#\left\{(t_{ia}, t_{ib}, t_{ic}, t_{id}) \middle| \begin{array}{l}\text{Exception of } (0,0,0,0),\\ (1,0,1,0), (0,1,1,1), (1,1,0,1)\end{array}\right\}$
  $n_{d6} = 4 + \min\limits_{(a,b,c,d)} \#\left\{(t_{ia}, t_{ib}, t_{ic}, t_{id}) \middle| \begin{array}{l}\text{Exception of } (0,0,0,0),\\ (1,0,0,1), (0,1,1,1), (1,1,1,0)\end{array}\right\}$
  $n_{d7} = 4 + \min\limits_{(a,b,c,d)} \#\left\{(t_{ia}, t_{ib}, t_{ic}, t_{id}) \middle| \begin{array}{l}\text{Exception of } (0,0,0,0),\\ (0,1,1,0), (1,0,1,1), (1,1,0,1)\end{array}\right\}$
  $n_{d8} = 4 + \min\limits_{(a,b,c,d)} \#\left\{(t_{ia}, t_{ib}, t_{ic}, t_{id}) \middle| \begin{array}{l}\text{Exception of } (0,0,0,0),\\ (0,1,0,1), (1,0,1,1), (1,1,1,0)\end{array}\right\}$
  $n_{d9} = 4 + \min\limits_{(a,b,c,d)} \#\left\{(t_{ia}, t_{ib}, t_{ic}, t_{id}) \middle| \begin{array}{l}\text{Exception of } (0,0,0,0),\\ (0,0,1,1), (1,1,0,1), (1,1,1,0)\end{array}\right\}$

Intuitively, the equations from n_{d0} to n_{d9} denote the minimum number of total active s-boxes when the number of active s-boxes in the 1st non-linear layer is given.

For example, consider equation n_{d0}.

When there are two active s-boxes in the 1st non-linear layer, two input differences of the linear layer can be represented as $\Delta z_a \neq 0$, $\Delta z_b \neq 0$. Each output difference of the linear layer shows $[\Delta z'_i] = [t_{ia} \cdot \Delta z_a \oplus t_{ib} \cdot \Delta z_b]$ ($0 \leq i < 8$). Here, when we assume $\Delta z_a = \Delta z_b$ as the relationship among two input differences, it shows $[\Delta z'_i] = [(t_{ia} \oplus t_{ib}) \cdot \Delta z_a]$ ($0 \leq i < 8$). By the way, an active s-box in the 2nd non-linear layer means an s-box in the 2nd non-linear layer whose input difference is non-zero, i.e., $\Delta z'_i \neq 0$. Accordingly, the minimum number of

active s-boxes in the 2nd non-linear layer is $\min_{(a,b)} \#\{\Delta z_i' | \Delta z_i' \neq 0,\ 0 \leq i < 8\} = \min_{(a,b)} \#\{(t_{ia} \oplus t_{ib}) | t_{ia} \oplus t_{ib} \neq 0,\ 0 \leq i < 8\}$. This yields equation n_{d0}.

Similarly, equations n_{d1} and n_{d2} are obtained with the relationships among three input differences, and the equations from n_{d3} and n_{d9} are obtained using the relationships among four input differences of the linear layer.

Using the above search algorithm, we find that there is no matrix with $n_d \geq 6$, and that there are 10080 candidate matrices with $n_d = 5$. Accordingly, the invulnerability of the round function with one of the candidate matrices against DC is evaluated as $p \leq p_s^5$. Conversely, its invulnerability against LC is evaluated as $q \leq q_s^5$ because of Conjecture 2 in Sect. 3.2.

Furthermore, for all 10080 candidate matrices, the total Hamming weight is equal to 44; each matrix consists of 4 column (row) vectors with the Hamming weight of six and 4 other column (row) vectors with the Hamming weight of five.

4.2 Construction of Linear Layer

Next, from the above 10080 candidate matrices, we construct the linear layer by the determination algorithm in Sect. 3.3. Since, however, the computation complexity of determining the construction is very large, there are $(8 \times 7)^{16} \approx 2^{93}$ candidates when the linear layer consists of 16 XORs, it is impossible to determine the construction exhaustively. Accordingly, we consider a linear layer that consists of four blocks with 8 m-bit inputs and 4 m-bit outputs, see Fig. 4(A). Each block consists of 4 XORs, and all inputs pass only one XOR, described in Fig. 4(B). The linear layer consists of 16 XORs, and the computation complexity is much lower with about $(4 \times 3 \times 2 \times 1)^4 \approx 2^{18}$ candidates.

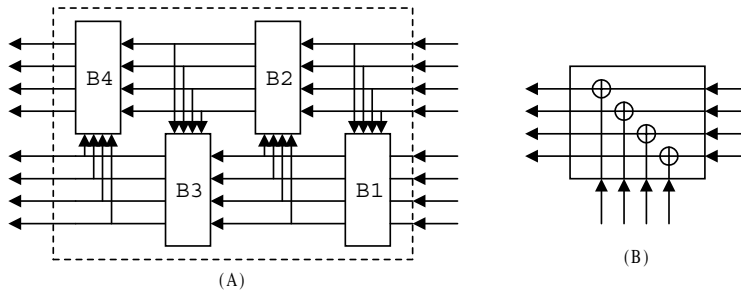

Fig. 4. Candidate construction of linear layer

This restricted search determines that 57 linear layers can be constructed from among the 10080 candidate matrices.

One of these matrices is shown below, and the linear layer with matrix P is shown in Fig. 5. Furthermore, it can be proven that, for this linear layer, the matrix for linear expression is represented as the matrix ${}^TP$, which is the matrix P transposed using the concatenation rules.

$$P = \begin{bmatrix} 0\,1\,1\,1\,1\,1\,1\,0 \\ 1\,0\,1\,1\,0\,1\,1\,1 \\ 1\,1\,0\,1\,1\,0\,1\,1 \\ 1\,1\,1\,0\,1\,1\,0\,1 \\ 1\,1\,0\,1\,1\,1\,0\,0 \\ 1\,1\,1\,0\,0\,1\,1\,0 \\ 0\,1\,1\,1\,0\,0\,1\,1 \\ 1\,0\,1\,1\,1\,0\,0\,1 \end{bmatrix} \iff {}^{T}P = \begin{bmatrix} 0\,1\,1\,1\,1\,1\,0\,1 \\ 1\,0\,1\,1\,1\,1\,1\,0 \\ 1\,1\,0\,1\,0\,1\,1\,1 \\ 1\,1\,1\,0\,1\,0\,1\,1 \\ 1\,0\,1\,1\,1\,0\,0\,1 \\ 1\,1\,0\,1\,1\,1\,0\,0 \\ 1\,1\,1\,0\,0\,1\,1\,0 \\ 0\,1\,1\,1\,0\,0\,1\,1 \end{bmatrix}$$

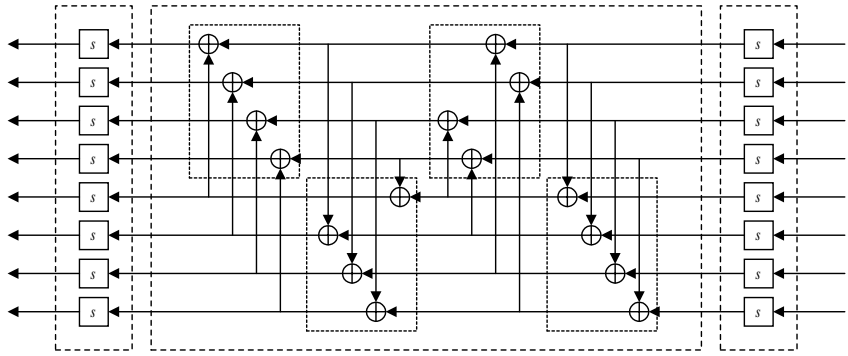

Fig. 5. Example of linear layer in the case of $n = 8$

5 Conclusion

This paper studied a strategy for constructing mn-bit round functions with the 2-round SPN structure that yield sufficiently small values of the maximum differential and linear probabilities p, q. This strategy regards a linear transformation layer as an $n \times n$ matrix P over $\{0,1\}$. We described the relationship between the matrix representation and the actual construction of the linear transformation layer. We proposed a search algorithm for constructing the optimal linear transformation layer by using the matrix representation.

Furthermore, with this algorithm, we determined the optimal linear transformation layer yielding probabilities $p \leq p_s^5,\ q \leq q_s^5$ in the case of $n = 8$.

References

1. C. M. Adams, "Simple and Effective Key Scheduling for Symmetric Ciphers," *Workshop on Selected Areas in Cryptology SAC'94*, 1994.
2. K.Aoki, K. Kobayashi, S. Moriai, "Best Differential Characteristic Search of FEAL," *Fourth International Workshop on Fast Software Encryption (FSE4)*, LNCS **1267**, 1997.

3. K. Aoki, 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*, Vol. E80-A, No. 1, pp. 2-8, 1997.
4. E. Biham, "On Matsui's Linear Cryptanalysis," *Advances in Cryptology - EUROCRYPT'94*, LNCS **950**, 1995.
5. L. Brown, M. Kwan, J. Pieprzyk, J. Seberry, "Improving Resistance to Differential Cryptanalysis and the Redesign of LOKI," *Advances in Cryptology - ASIACRYPT'91*, LNCS **739**, 1993.
6. E. Biham, A. Shamir, "Differential Cryptanalysis of DES-like Cryptosystems," *Journal of Cryptology*, Vol. 4 No. 1, pp. 3-72, 1991. (The extended abstract appeared at *CRYPTO'90*)
7. F. Chabaud, S. Vaudenay, "Links Between Differential and Linear Cryptanalysis," *Advances in Cryptology - EUROCRYPT'94*, LNCS **950**, 1995.
8. *Data Encryption Standard*, FIPS-PUB-46, 1977.
9. H. M. Heys, S. E. Tavares, "Substitution-Permutation Networks Resistant to Differential and Linear Cryptanalysis," *Journal of Cryptology*, Vol. 9 No. 1, pp. 1-19, 1996.
10. T. Jakobsen, L. R. Knudsen, "The Interpolation Attack on Block Ciphers," *Fourth International Workshop on Fast Software Encryption (FSE4)*, LNCS **1267**, 1997.
11. L. R. Knudsen, "Practically Secure Feistel Ciphers," *Cambridge Security Workshop on Fast Software Encryption (FSE1)*, LNCS **809**, 1994.
12. L. R. Knudsen, "Truncated and Higher Order Differentials," *Second International Workshop on Fast Software Encryption (FSE2)*, LNCS **1008**, 1995.
13. X. Lai, "Higher order derivatives and differential cryptanalysis," In *Proc. of Symposium on Communication, Coding, and Cryptography," in honor of James L. Massey on the occasion of his 60'th birthday*, Feb. 10-13, 1994, Monte-Verita, Ascona, Switzerland, 1994.
14. X. Lai, J. L. Massey, S. Murphy, "Markov Ciphers and Differential Cryptanalysis," *Advances in Cryptology - EUROCRYPT'91*, LNCS **547**, 1991.
15. M. Matsui, "Linear Cryptanalysis Method for DES Cipher," *Advances in Cryptology - EUROCRYPT'93*, LNCS **765**, 1994.
16. M. Matsui, "On Correlation Between the Order of S-boxes and the Strength of DES," *Advances in Cryptology - EUROCRYPT'94*, LNCS **950**, 1995.
17. M. Matsui, "New Block Encryption Algorithm MISTY," *Fourth International Workshop on Fast Software Encryption (FSE4)*, LNCS **1267**, 1997.
18. K. Nyberg, "Linear Approximation of Block Ciphers," *Advances in Cryptology - EUROCRYPT'94*, LNCS **950**, 1995.
19. K. Nyberg, "Generalized Feistel Networks," *Advances in Cryptology - ASIACRYPT'96*, LNCS **1163**, 1996.
20. K. Nyberg, L. R. Knudsen, "Provable Security Against a Differential Attack," *Journal of Cryptology*, Vol. 8 No. 1, pp. 27-37, 1995. (The extended abstract appeared at *CRYPTO'92*)
21. K. Ohta, S. Moriai, K. Aoki, "Improving the Search Algorithm for the Best Linear Expression.," *Advances in Cryptology - CRYPTO'95*, LNCS **963**, 1995.
22. V. Rijmen, J. Daemen, B. Preneel, A. Bosselaers, E. DcWin, "The Cipher SHARK," *Third International Workshop on Fast Software Encryption (FSE3)*, LNCS **1039**, 1996.
23. S. Vaudenay, "On the Need for Multipermutations: Cryptanalysis of MD4 and SAFER," *Second International Workshop on Fast Software Encryption (FSE2)*, LNCS **1008**, 1995.