5. Propagation and Correlation

5.1 Introduction

In this chapter we treat difference propagation and input-output correlation in Boolean mappings and iterated Boolean transformations. Difference propagation is specifically exploited in differential cryptanalysis (DC), invented by Eli Biham and Adi Shamir [BiSh91]. Input-output correlation is exploited in linear cryptanalysis (LC), invented by Mitsuru Matsui [Ma93]. Both DC and LC were successfully applied on the block cipher DES [Fi77]. DC was the first chosen-plaintext attack, LC the first known-plaintext attack more efficient than exhaustive key search for DES.

We start with a brief description of DES and the original DC and LC attacks using the terminology of their inventors. For a more detailed treatment of the attacks, we refer to the original publications [BiSh91,Ma94]. The only aim of our description is to indicate the aspects of the attacks that determine their expected work factor. For DC the critical aspect is the maximum *probability* for difference propagations, for LC it is the maximum *deviation* from 1/2 of the probability that linear expressions hold.

We introduce a number of algebraic tools that more adequately describe the essential mechanisms of LC and DC. This includes a number of powerful new concepts, such as the *correlation matrix* of a Boolean mapping. Using these new concepts a number of new relations and equalities are derived. These tools are further refined to describe propagation and correlation in iterated Boolean transformations.

Finally, we formulate and motivate our new structural design strategy for the round transformation of block ciphers and, more generally, the updating transformation of cryptographic finite state machines.

5.2 The Data Encryption Standard

The cipher that was the most important object of the attacks to be discussed is the Data Encryption Standard (DES) [Fi77]. Therefore, we start with a brief description of its structure.

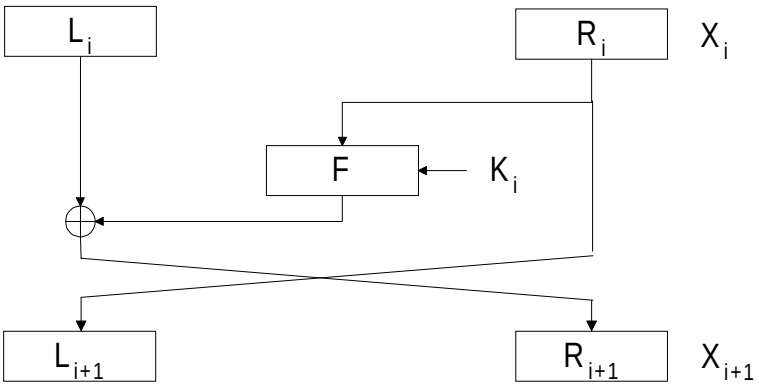

Figure 1 Computational graph of the DES round transformation.

DES is a block cipher with a block length of 64 bits and a key length of 56 bits. Its main body consists of 16 iterations of the *keyed round transformation*. The computational graph of the round transformation is depicted in Figure 1. It can be seen that the intermediate encryption value is split into a 32-bit left part L_i and a 32-bit right part R_i. The latter is the argument of the keyed F-function. The output of the F-function is added bitwise to L_i. Subsequently, left and right part are interchanged.

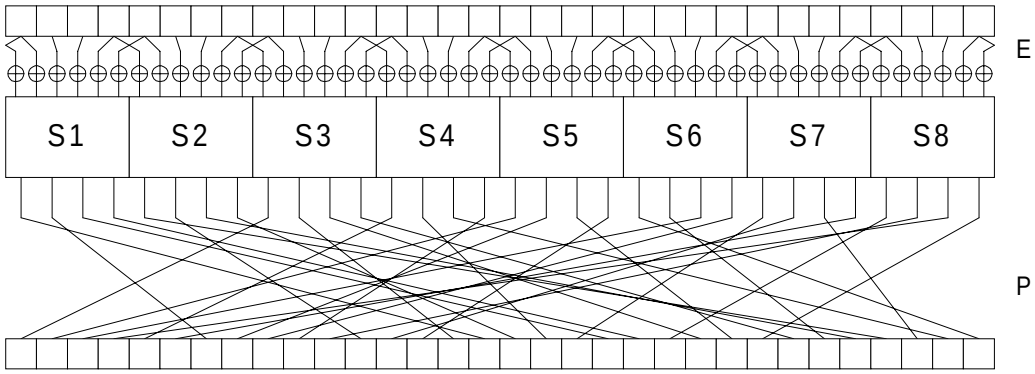

Figure 2 Computational graph of the DES F-function.

The computational graph of the F-function is depicted in Figure 2. It can be seen that it consists of the succession of four steps. In the expansion *E* the 32 input bits are expanded to 48 bits. Subsequently, a 48-bit round key is added bitwise to this 48-bit vector. The resulting 48-bit vector is mapped onto a 32-bit vector by 8 nonlinear S-boxes that each convert 6 input bits into 4 output bits. Finally, these 32 bits are transposed by the bit permutation *P*. Observe that the only nonlinear step in the F-function (and also in the round transformation) consists of the S-boxes. The 48-bit round keys are extracted from the 56-bit cipher key by means of a linear key schedule.

5.3 Differential and linear cryptanalysis

In this section we summarise differential cryptanalysis as described in [BiSh91] and linear cryptanalysis as presented in [Ma94].

5.3.1 Differential cryptanalysis

Differential cryptanalysis is a chosen-plaintext (difference) attack in which a large amount of plaintext-ciphertext pairs are used to determine the value of key bits. Statistical key information is deduced from ciphertext blocks obtained by encrypting pairs of plaintext blocks with a specific bitwise difference A' under the target key. The work factor of the attack depends critically on the largest probability $\mathrm{P}(B'|A')$ with B' a difference at some fixed intermediate stage of the cryptographic function, e.g., at the input of the last round. In a first approximation, the probabilities $\mathrm{P}(B'|A')$ for DES are assumed to be independent of the specific value of the key.

Key information is extracted from the output pairs in the following way. For each pair it is assumed that the intermediate difference is equal to B'. The absolute values of the outputs and the (assumed) intermediate difference B' impose restrictions upon a number υ of key bits of the last round key. A pair is said to *suggest* the subkey values that are compatible with these restrictions. While for some pairs many keys are suggested, no keys are found for other pairs, implying that the output values are incompatible with B' For each suggested subkey value a corresponding entry in a frequency table is incremented.

The attack is successful if the right value of the subkey is suggested significantly more often than any other value. Pairs with an intermediate difference not equal to B' are called wrong pairs. Subkey values suggested by these pairs are in general wrong. Right pairs, with an intermediate difference equal to B', do not only suggest the right subkey value but often also a number of wrong subkey values. For DES the wrong suggestions may be considered uniformly distributed among the possible key values if the value $\mathrm{P}(B'|A')$ is significantly larger than $\mathrm{P}(C'|A')$ for any $C' \neq B'$.

Under these conditions it makes sense to calculate the ratio between the number of times the right value is suggested and the average number of suggestions per entry, the so-called *signal-to-noise or S/N ratio*. If the size of the table is 2^{υ} and the average number of suggested subkeys per pair is γ, this ratio is equal to $\mathrm{P}(B'|A')2^{\upsilon}\gamma$. The S/N ratio strongly affects the number of right pairs needed to uniquely identify the right subkey value. Experimental results [BiSh91] showed that for a ratio of 1-2 about 20-40 right pairs are enough. For larger ratios only a few right pairs are needed and for ratios that are much smaller than 1 the required amount of right pairs can make a practical attack infeasible.

Large probabilities $\mathrm{P}(B'|A')$ are localized by the construction of so-called *characteristics*. An m-round characteristic constitutes an m+1-tuple of difference patterns: $(X'_0, X'_1, \ldots, X'_m)$. The probability of this characteristic is the probability that an initial difference pattern X'_0 propagates to difference patterns $X'_1, X'_2, \ldots, X'_m$ respectively after 1, 2, …m rounds. In the assumption that the propagation probability from X'_{i-1} to X'_i is independent of the propagation from X'_0 to X'_{i-1}, this probability is given by

$$\prod_{\mathrm{i}} \mathrm{P}(X'_i|X'_{i-1}), \qquad \textbf{(5.1)}$$

with $\mathrm{P}(X'_i|X'_{i-1})$ the probability that the difference pattern X'_{i-1} at the input of the round transformation gives rise to X'_i at its output. Hence, the multiple-round characteristic is decomposed into a number of single-round characteristics (X'_{i-1}, X_i) with probability $\mathrm{P}(X'_i|X'_{i-1})$.

In the construction of high-probability characteristics for DES, advantage is taken from the linearity in the round transformation. Single-round characteristics of the form $(L'_{i-1}|R'_{i-1}, L'_i|R'_i)$ with $R'_i = L'_{i-1}$ and $L'_{i-1} = R'_{i-1} = 0$ have probability 1 and are called *trivial*. The most probable nontrivial single-round characteristics have an input difference pattern that only affects a small number of the eight S-boxes.

Trivial characteristics have been exploited to construct high-probability iterative characteristics, i.e., characteristics with a periodic sequence of differences. The iterative characteristic with highest probability has period 2. Of the two involved single-round characteristics, one is trivial. In the other one there is a nonzero difference pattern at the input of three neighbouring S-boxes, that propagates to a zero difference pattern at the output of the S-boxes with probability 1/234. Hence, the resulting iterative characteristics have a probability of 1/234 per 2 rounds.

5.3.2 Linear cryptanalysis

Linear cryptanalysis is a known-plaintext attack in which a large amount of plaintext-ciphertext pairs are used to determine the value of key bits. For the 8-round variant of DES, linear cryptanalysis can also be applied in a ciphertext-only context.

A condition for applying linear cryptanalysis to a block cipher is to find ``effective'' linear expressions. Let $A[i_1, i_2, \ldots, i_a]$ be the bitwise sum of the bits of *A* with indices in a *selection set* $\{i_1, i_2, \ldots, i_a\}$, i.e.,

$$A[i_1, i_2, \ldots, i_a] = A[i_1] + A[i_2] + \cdots + A[i_a]$$

Let *P*,*C* and *K* denote respectively the plaintext, the ciphertext and the key. A linear expression is an expression of the following type:

$$P[i_1, i_2, \ldots, i_a] = C[j_1, j_2, \ldots, j_b] + K[k_1, k_2, \ldots, k_c], \qquad \textbf{(5.2)}$$

with $i_1, i_2, \ldots, i_a, i_1, i_2, \ldots, j_b$ and $k_1, k_2, \ldots, k_c$ fixed bit locations. The effectiveness of such a linear expression in linear cryptanalysis is given by $|p - 1/2|$ with p the probability that it holds. By checking the value of the left-hand side of (5.2) for a large number of plaintext-ciphertext pairs, the right-hand side can be guessed by taking the value that occurs most often. This gives a single bit of information about the key. In [Ma94] it is shown that the probability of making a wrong guess is very small if the number of plaintext-ciphertext pairs is larger than $|p - 1/2|^{-2}$.

In [Ma94] another algorithm is given that determines more than a single bit of key information using a similar linear expression. Instead of (5.2), an expression is used that contains no plaintext or ciphertext bits, but instead bits of the intermediate encryption values I_1 and I_{15} respectively after a single and all but a single round:

$$I_1[i_1, i_2, \ldots, i_a] = I_{15}[j_1, j_2, \ldots, j_b] + K[k_1, k_2, \ldots, k_c], \qquad \textbf{(5.3)}$$

By assuming values for a subset χ_k of the subkey bits of the first and last round, the bits of I_1 and I_{15} that occur in (5.3) can be calculated. These bits are correct if the values assumed for the key bits with indices in χ_k are correct. Given a large number l of plaintext-ciphertext pairs, the correct values of all bits in χ_k and the value of the right-hand side of (5.3) can be determined in the following way. For all values of the key bits with indices in χ_k, the number of plaintext-ciphertext pairs are counted for which (5.3) holds. For the correct assumption the expected value of this sum is pl or $(1-p)l$. Thanks to the nonlinear behavior of the round transformation this sum is expected to have significantly less bias for all wrongly assumed subkey values. Given a linear expression (5.3) that holds with probability p, the probability that this algorithm leads to a wrong guess is very small if the number of plaintext-ciphertext pairs is significantly (say more than a factor 8) larger than $|p-1/2|^{-2}$. In a recent improvement of this attack this factor 8 is reduced to 1 [Ma94]. Still, in both variants of the attack the value of $|p-1/2|$ is critical for the work factor of the attack.

Effective linear expressions (5.2) and (5.3) are constructed by ``chaining'' single-round linear expressions. An *m*–1-round linear expression can be turned into an m-round linear expression by appending a single-round linear expression such that all the intermediate bits cancel:

$$\begin{gathered} P[i_1,i_2,\ldots,i_a]+I_{m-1}[j_1,j_2,\ldots,j_b]=K[k_1,k_2,\ldots,k_c] \\ + \\ I_{m-1}[j_1,j_2,\ldots,j_b]+I_m[m_1,m_2,\ldots,m_a]=K[k_2,k_5,\ldots,k_d]. \\ = \\ P[i_1,i_2,\ldots,i_a]+I_m[m_1,m_2,\ldots,m_a]=K[k_1,k_3,\ldots,k_d] \end{gathered}$$

In [Ma94] it is shown that the probability that the resulting linear expression holds can be approximated by $1/2+2(p_1-1/2)(p_2-1/2)$, given that the component linear expressions hold with probability p_1 and p_2 respectively.

The DES single-round linear expressions and their probabilities can be studied by observing the dependencies in the computational graph of the round transformation. The selected round output bits completely specify a selection pattern at the output of the S-boxes. If only round output bits are selected from the left half, this involves no S-box output bits at all, resulting in linear expressions that hold with probability 1. These are of the following type:

$$I_{l-1}[j_1+32,j_2+32,\ldots,j_a+32]=I_l[j_1,j_2,\ldots,j_a].$$

This is called a *trivial* expression. Apparently, the most useful nontrivial single-round linear expressions only select bits coming from a single S-box. For a given S-box, all possible linear expressions and their probabilities can be exhaustively calculated. Together with the key application before the S-boxes, each of these linear expressions can be converted into a single-round linear expression. The most effective multiple-round linear expressions for DES are constructed by combining single-round trivial expressions with linear expressions involving output bits of only a single S-box. The resulting most effective 14-round linear expression has a probability of $1/2\pm 1.19\cdot 2^{-21}$.

5.4 Analytical and descriptive tools

In this section we present a formalism and some useful tools for the description and analysis of difference propagation and the chaining of linear expressions. We establish a relation between Boolean mappings and linear mappings over real vector spaces, allowing a much simpler treatment of linear expressions. More importantly, the proposed formalisms force us to look at the phenomena from a different angle, giving new insights. In the original descriptions of LC and DC, the propagation and chaining are described as *probabilistic* phenomena, with an emphasis on probabilities of events. In our formalism we describe the phenomena in terms of ratios and correlations, reflecting a more deterministic view.

5.4.1 The Walsh-Hadamard transform

Linear cryptanalysis can be seen as the exploitation of *correlations* between linear combinations of bits of different intermediate encryption values (or *states*). The correlation between two Boolean functions with domain $\{0,1\}^n$ an be expressed by a *correlation coefficient* that ranges between -1 and 1:

Definition 5.1: The correlation coefficient $\mathrm{C}(f,g)$ associated with a pair of Boolean functions $f(a)$ and $g(a)$ is given by

$$\mathrm{C}(f,g) = 2 \cdot \mathrm{Prob}\big(f(a) = g(a)\big) - 1 \ .$$

From this definition it follows that $\mathrm{C}(f,g) = \mathrm{C}(g,f)$. If the correlation coefficient is different from zero, the functions are said to be *correlated*.

A selection vector w is a binary vector that *selects* all components *i* of a vector for which $w_i = 1$. Analogous to the inner product of vectors in linear algebra, the linear combination of the components of a vector a selected by w can be expressed as $w^{\mathrm{t}}a$ with the t suffix denoting transposition of the vector w. A linear Boolean function $w^{\mathrm{t}}a$ is completely specified by its corresponding selection vector w.

Let $\hat{f}(a)$ be a real-valued function that is -1 for $f(a) = 1$ and +1 for $f(a) = 0$. This can be expressed by $\hat{f}(a) = (-1)^{f(a)}$. In this notation the real-valued function corresponding to a linear Boolean function $w^{\mathrm{t}}a$ becomes $(-1)^{w^{\mathrm{t}}a}$. The bitwise sum of two Boolean functions corresponds to the bitwise product of their real-valued counterparts, i.e.,

$$h(a) = f(a) + g(a) \Leftrightarrow \hat{h}(a) = \hat{f}(a)\hat{g}(a) \ .$$

We define an *inner product* for real-valued functions, not to be confused with the inner product of *vectors*, by

$$\left\langle \hat{f}, \hat{g} \right\rangle = \sum_a \hat{f}(a)\hat{g}(a) \,, \qquad \textbf{(5.4)}$$

and the corresponding *norm* by

$$\left| \hat{f} \right| = \sqrt{\left\langle \hat{f}, \hat{f} \right\rangle} \,.$$

For a Boolean function $f(a)$, the norm of $\hat{f}(a)$ is equal to the square root of its domain size, i.e., $2^{n/2}$. From Definition 5.1 it follows that

$$\mathrm{C}(f,g) = \frac{\langle \hat{f}, \hat{g} \rangle}{|\hat{f}| \cdot |\hat{g}|} \ .$$

In the space of all Boolean functions, the real-valued functions corresponding to the linear Boolean functions form an orthogonal basis with respect to the defined inner product:

$$\left\langle (-1)^{u^{\mathrm{t}}a}, (-1)^{v^{\mathrm{t}}a} \right\rangle = 2^n \delta(u+v).$$

with $\delta(w)$ the Kronecker delta function that is equal to 1 if w is the zero vector and 0 otherwise. The representation of a Boolean function with respect to this basis is called its Walsh-Hadamard transform [Go67,Pr93]. If the correlation coefficients $\mathrm{C}(f(a), w^{\mathrm{t}}a)$ are denoted by $\hat{F}(w)$, we have

$$\hat{f}(a) = \sum_w \hat{F}(w)(-1)^{w^{\mathrm{t}}a} \ , \qquad \textbf{(5.5)}$$

and dually,

$$\hat{F}(w) = \sum_a \hat{f}(a)(-1)^{w^{\mathrm{t}}a} \ , \qquad \textbf{(5.6)}$$

summarized by

$$\hat{F}(w) = \mathcal{W}(f(a)).$$

Hence, a Boolean function is completely specified by the set of correlation coefficients with all linear functions.

Taking the square of the norm of both sides of (5.5) and dividing by 2^n yields the theorem of Parseval [Pr93]:

$$1 = \sum_v \sum_w \hat{F}(v)\hat{F}(w) \left\langle (-1)^{v^{\mathrm{t}}a}, (-1)^{w^{\mathrm{t}}a} \right\rangle = \sum_w \hat{F}^2(w) \ , \qquad \textbf{(5.7)}$$

expressing a relation between the number of linear functions that are correlated with a given Boolean function and the amplitude of their correlations.

The Walsh-Hadamard transform of the sum of two Boolean functions $f(a) + g(a)$ can be derived using (5.5):

$$\begin{aligned} \hat{f}(a)\hat{g}(a) &= \sum_u \hat{F}(u)(-1)^{u^{\mathrm{t}}a} \sum_v \hat{G}(v)(-1)^{v^{\mathrm{t}}a} \\ &= \sum_u \sum_v \hat{F}(u)\hat{G}(v)(-1)^{(u+v)^{\mathrm{t}}a} \\ &= \sum_w \left(\sum_v \hat{F}(v+w)\hat{G}(v) \right) (-1)^{w^{\mathrm{t}}a} \end{aligned} \ .$$

The values of $\hat{H}(w) = \mathcal{W}(f + g)$ are therefore given by

$$\hat{H}(w) = \sum_v \hat{F}(v + w)\hat{G}(v). \tag{5.8}$$

Hence, addition modulo 2 in the Boolean domain corresponds to convolution in the transform domain. If the convolution operation is denoted by $\otimes$ this can be expressed by

$$\mathcal{W}(f + g) = \mathcal{W}(f) \otimes \mathcal{W}(g). \tag{5.9}$$

For multiplication (bitwise AND) of two Boolean functions it can easily be seen that

$$h(a) = f(a)g(a) \Leftrightarrow \hat{h}(a) = \frac{1}{2}\left(1 + \hat{f}(a) + \hat{g}(a) - \hat{f}(a)\hat{g}(a)\right).$$

Hence, we have

$$W(fg) = \frac{1}{2}\left(\delta(w) + \mathcal{W}(f) + \mathcal{W}(h) - \mathcal{W}(f + g)\right). \tag{5.10}$$

Given the convolution property it is easy to demonstrate some composition properties that are useful in the study of linear cryptanalysis.

- Complementation of a Boolean function $g(a) = f(a) + 1$ corresponds to multiplication by -1 in the transform domain: $\hat{G}(w) = -\hat{F}(w)$.
- Adding a linear function $g(a) = f(a) + u^t a$ corresponds to a dyadic shift operation in the transform domain: $\hat{G}(w) = \hat{F}(w + u)$.

The subspace of $\{0,1\}^n$ generated by the vectors w for which $\hat{F}(w) \neq 0$ is called its *support space* $\mathcal{V}_f$. The support space of the sum of two Boolean functions is a subspace of the (vector) sum of their corresponding support spaces: $\mathcal{V}_{f+g} \subseteq \mathcal{V}_f + \mathcal{V}_g$. This follows directly from the convolution property. Two Boolean functions are called *disjunct* if their support spaces are disjunct, i.e., if the intersection of their support spaces only contains the origin. A vector $v \in \mathcal{V}_{f+g}$ with f and g disjunct has a unique decomposition into a component $u \in \mathcal{V}_f$ and a component $w \in \mathcal{V}_g$. In this case the transform values of $h = f + g$ are given by

$$\hat{H}(v) = \hat{F}(u)\hat{G}(w) \text{ with } v = u + w \text{ and } u \in \mathcal{V}_f,\ w \in \mathcal{V}_g. \tag{5.11}$$

A pair of Boolean functions that depend on non-overlapping sets of input bits is a special case of disjunct functions.

5.4.2 Correlation matrices

Almost all components in encryption schemes, including S-boxes, state updating transformations and block ciphers are simply mappings from a space of n-dimensional binary vectors to a space of m-dimensional binary vectors. Often $m = n$. These mappings can be represented by their *correlation matrix*.

A mapping $h:\{0,1\}^n \to \{0,1\}^m$ can be decomposed into m *component* Boolean functions: $(h_0, h_1, ..., h_{m-1})$. Each of these component functions h_i has a Walsh-Hadamard transform $\hat{H}_i$. The vector function with components $\hat{H}_i$ is denoted by $\hat{H}$ and can be considered the Walsh-Hadamard transform of the mapping h. As in the case of Boolean functions, $\hat{H}$ completely determines the transformation h. The Walsh-Hadamard transform of any linear combination of components of h is specified by a simple extension of (5.9):

$$\mathcal{W}(u^t h) = \bigotimes_{u_i=1} \hat{H}_i . \qquad \textbf{(5.12)}$$

All correlation coefficients between linear combinations of input bits and those of output bits of the mapping h can be arranged in a $2^m \times 2^n$ *correlation matrix* C^h. The element C^h_{uw} in row u and column w is equal to $C(u^t h(a), w^t a)$. The rows of this matrix can be interpreted as

$$(-1)^{u^t h(a)} = \sum_w C^h_{uw} (-1)^{w^t a} . \qquad \textbf{(5.13)}$$

In words, the real-valued function corresponding to a linear combination of output bits can be written as a linear combination of the real-valued functions corresponding to the linear combinations of input bits.

A Boolean function $f(a)$ can be seen as a special case of a mapping and has a correlation matrix with two rows: row 0 and row 1. Row 1 contains the Walsh-Hadamard transform values of $f(a)$ and row 0 the Walsh-Hadamard transform values of the Boolean function that is equal to 0.

A matrix C^h defines a linear mapping with domain R^{2^n} and range R^{2^m}. Let $\mathcal{L}$ be a mapping from the space of binary vectors to the space of real vectors, depicting a binary vector of dimension n onto a real vector of dimension 2^n. $\mathcal{L}$ is defined by

$$\mathcal{L}:\{0,1\}^n \to \{0,1\}^m : a \mapsto \mathcal{L}(a) = \alpha = (-1)^{u^t a} .$$

Since $\mathcal{L}(a+b) = \mathcal{L}(a) \cdot \mathcal{L}(b)$, $\mathcal{L}$ is a group-homomorphism from $\langle \{0,1\}^n, + \rangle$ to $\langle \mathrm{R}^{2^n}, \cdot \rangle$ with "$\cdot$" denoting the componentwise product. From (5.13) it can easily be seen that

$$C^h \mathcal{L}(a) = \mathcal{L}(h(a)) .$$

Consider the composition of two Boolean mappings $h = h_2 \circ h_1$ or $h(a) = h_2(h_1(a))$, with h_1 mapping n-dimensional vectors to p-dimensional vectors and with h_2 mapping p-dimensional vectors to m-dimensional vectors. The correlation matrix of h is determined by the correlation matrices of the component mappings. We have

$$
\begin{aligned}
(-1)^{u^t h(a)} &= \sum_v C_{uv}^{h_2} (-1)^{v^t h_1(a)} \\
&= \sum_v C_{uv}^{h_2} \sum_w C_{vw}^{h_1} (-1)^{w^t a} \quad . \\
&= \sum_w \left(\sum_v C_{uv}^{h_2} C_{vw}^{h_1} \right) (-1)^{w^t a}
\end{aligned}
$$

Hence, we have

$$C^{h_2 \circ h_1} = C^{h_2} \times C^{h_1} \ ,$$

with $\times$ denoting the matrix product, C^{h_1} a $2^p \times 2^n$ matrix and C^{h_2} a $2^m \times 2^p$ matrix. The input-output correlations of $h = h_2 \circ h_1$ are given by

$$\mathrm{C}\left(u^t h(a), w^t a\right) = \sum_v \mathrm{C}\left(u^t h_1(a), v^t a\right) \mathrm{C}\left(v^t h_2(a), w^t a\right). \qquad \textbf{(5.14)}$$

If h is an invertible transformation in $\{0,1\}^n$, we have

$$\mathrm{C}\left(u^t h^{-1}(a), w^t a\right) = \mathrm{C}\left(u^t b, w^t h(b)\right) = \mathrm{C}\left(w^t h(b), u^t b\right) \ .$$

Using this and $C^h \times C^{\left(h^{-1}\right)} = C^{h \circ h^{-1}} = I = C^h \times \left(C^h\right)^{-1}$ we obtain

$$\left(C^h\right)^{-1} = C^{\left(h^{-1}\right)} = \left(C^h\right)^t \ ,$$

hence, C^h is an orthogonal matrix. Conversely, a Boolean mapping with an orthogonal correlation matrix is invertible.

5.4.2.1 Special mappings

In the following, the suffix h will be omitted. Consider the transformation that consists of the bitwise addition of a constant vector k: $h(a) = a + k$. Since $u^t h(a) = u^t a + u^t k$, the correlation matrix is a diagonal matrix with

$$C_{uu} = (-1)^{u^t k} \ .$$

Therefore the effect of bitwise addition of a constant vector before (or after) a mapping h on its correlation matrix is a multiplication of some columns (or rows) by –1.

Consider a linear mapping $h(a) = M\,a$ with M an $2^m \times 2^n$ binary matrix. Since $u^t h(a) = u^t M a = \left(M^t u\right)^t a$, the elements of the corresponding correlation matrix are given by

$$C_{uw} = \delta\left(M^t u + w\right).$$

If M is an invertible square matrix, the correlation matrix is a permutation matrix. The single nonzero element in row u is in column $M^t u$. The effect of applying an invertible linear transformation before (or after) a transformation h on the correlation matrix is only a permutation of its columns (or rows).

Consider a mapping from $\{0,1\}^n$ to $\{0,1\}^m$ that consists of the parallel application of l component mappings (S-boxes) from $\{0,1\}^{n_i}$ to $\{0,1\}^{m_i}$ with $\sum_i n_i = n$ and $\sum_i m_i = m$. We will call such a mapping a *juxtaposed* mapping. We have

$$a = \left(a_{(0)}, a_{(1)}, \ldots, a_{(l-1)}\right) \text{ and } b = \left(b_{(0)}, b_{(1)}, \ldots, b_{(l-1)}\right),$$

with the $a_{\{i\}}$ vectors of dimension n_i and the $b_{\{i\}}$ vectors of dimension m_i. The mapping $b = h(a)$ is defined by

$$b_{(i)} = h_{(i)}\left(a_{(i)}\right) \text{ for } 0 \le i < l.$$

With every S-box $h_{(i)}$ corresponds a $2^{n_i} \times 2^{m_i}$ correlation matrix denoted by $C^{(i)}$. Since the $h_{(i)}$ are disjunct, (5.11) can be applied and the elements of the correlation matrix of h are given by

$$C_{uw} = \prod_i C^{(i)}_{u_{(i)}w_{(i)}} \text{ with } u = \left(u_{(0)}, u_{(1)}, \ldots, u_{(l-1)}\right) \text{ and } w = \left(w_{(0)}, w_{(1)}, \ldots, w_{(l-1)}\right).$$

In words this can be expressed as: the correlation associated with input selection w and output selection u is the product of its corresponding S-box input-output correlations $C^{(i)}_{u_{(i)}w_{(i)}}$.

5.4.3 Derived properties

The concept of the correlation matrix is a valuable tool to demonstrate properties of Boolean transformations, functions and their spectrum. We will illustrate this with some examples.

Lemma 5.1: The elements of the correlation matrix of a Boolean transformation satisfy

$$C_{(u+v)x} = \sum_w C_{u(w+x)} C_{vw} \text{ for all } u, v, x \in \{0,1\}^n. \qquad \textbf{(5.15)}$$

Proof : Using the convolution property, we have

$$\mathcal{W}\left((u+v)^t h(a)\right) = \mathcal{W}\left(u^t h(a) + v^t h(a)\right) = \mathcal{W}\left(u^t h(a)\right) \otimes \mathcal{W}\left(u^t h(a)\right). \qquad \textbf{(5.16)}$$

Since the components of $\mathcal{W}\left(y^t h(a)\right)$ are given by C_{yw}, the projection of (5.16) onto the component with index *x* gives rise to (5.15).

QED

A Boolean function is *balanced* if it is 1 (0) for exactly half of the elements in the domain. Clearly, being balanced is equivalent to being uncorrelated to the Boolean function equal to 0 (or 1).

Using the properties of correlation matrices we can now give an elegant proof of the following well-known theorem.

Theorem 5.1: A Boolean transformation is invertible if and only if every linear combination of output bits is a balanced Boolean function of its input bits.

Proof:

⇒:

If h is an invertible transformation, its correlation matrix *C* is orthogonal. Since $C_{00} = 1$ and all rows and columns have norm 1, it follows that there are no other elements in row 0 or column 0 different from 0. Hence, $\mathrm{C}\left(u^{\mathrm{t}}h(a),0\right) = \delta(u)$ or $u^{\mathrm{t}}h(a)$ is balanced for all $u \neq 0$.

⇐:

The condition that all linear combinations of output bits are balanced Boolean functions of input bits corresponds to $C_{u0} = 0$ for $u \neq 0$. If this is the case, it can be shown that the correlation matrix is orthogonal. The expression $C^{\mathrm{t}} \times C = I$ is equivalent to the following set of conditions

$$\sum_{w} C_{uw} C_{vw} = \delta(u+v) \text{ for all } \{0,1\}^n . \qquad \textbf{(5.17)}$$

Now, the substitution $x = 0$ in (5.15) gives rise to

$$\sum_{w} C_{uw} C_{vw} = C_{(u+v)0} \quad .$$

Since we have $C_{u0} = 0$ for all $u \neq 0$ and $C_{00} = 1$, (5.17) holds for all possible pairs *u,v*. It follows that C is an orthogonal matrix, hence h^{-1} exists and is defined by C^{-1}.

QED

Lemma 5.2: The elements of the correlation matrix of a mapping with domain $\{0,1\}^n$ and the Walsh-Hadamard transform values of a Boolean function with domain $\{0,1\}^n$ are integer multiples of 2^{1-n}.

Proof : The sum in the right-hand side of (5.6) is always even since its value is of the form $k \cdot (1) + \left(2^n - k\right) \cdot (-1) = 2k - 2^n$. It follows that the Walsh-Hadamard coordinates must be integer multiples of 2^{1-n}.

QED

A mapping from $\{0,1\}^n$ to $\{0,1\}^m$ can be converted into a mapping from $\{0,1\}^{n-1}$ to $\{0,1\}^m$ by fixing a single component of the input. More generally, a component of the input can be set equal to a linear combination of other input components, possibly complemented. Such a restriction is of the type

$$v^{\mathrm{t}}a = \varepsilon \text{ with } \varepsilon \in \{0,1\} .$$

Assume that $v_s = 1$. The restriction can be seen as the result of a mapping $a' = h_r(a)$ from $\{0,1\}^{n-1}$ to $\{0,1\}^n$ specified by $a'_i = a_i$ for $i \neq s$ and $a'_s = v^t a + a_s + \varepsilon$. The nonzero elements of the correlation matrix of h_r are

$$C^{h_r}_{ww} = 1 \quad \text{and} \quad C^{h_r}_{(v+w)w} = (-1)\varepsilon \quad \text{for all } w \text{ with } w_s = 0.$$

It can be seen that columns indexed by w with $w_s = 0$ have exactly two nonzero entries with magnitude 1 and those with $w_s = 1$ are all-zero. Omitting the latter gives a $2^n \times 2^{n-1}$ correlation matrix C^{h_r} with only columns indexed by the vectors with $w_s = 0$.

The transformation restricted to the specified subset of inputs can be seen as the consecutive application of h_r and the transformation itself. Hence, its correlation matrix C' is given by $C \times C^{h_r}$. The elements of this matrix are

$$C'_{uw} = C_{uw} + (-1)^{\varepsilon} C_{u(w+v)}, \tag{5.18}$$

if $w_s = 0$ and equal to 0 if $w_s = 1$. The elements in C' are the Walsh-Hadamard transform values of Boolean functions of *n*–1-dimensional vectors, hence, from Lemma 5.2 they must be integer multiples of 2^{2-n}. This can be easily generalized to multiple linear restrictions on the input.

Applying (5.7) to the rows of the restricted correlation matrices gives additional laws for the Walsh-Hadamard transform values of Boolean functions. For the single restrictions of the type $v^t a = \varepsilon$ we have

$$\sum_w \left(\hat{F}(w) + \hat{F}(w+v)\right)^2 = \sum_w \left(\hat{F}(w) - \hat{F}(w+v)\right)^2 = 2.$$

Lemma 5.3: The elements of a correlation matrix corresponding to an invertible transformation of n-bit vectors are integer multiples of 2^{2-n}.

Proof : Consider an element of the correlation matrix C_{uw}. If the input of the corresponding transformation is restricted by $w^t a = 0$, the correlation of the output function $u^t h(a)$ to 0 becomes $C_{uw} + C_{u0}$. According to Lemma 5.2, this value is an integer multiple of 2^{2-n}. From Theorem 5.1 it follows that $C_{u0} = 0$ and hence that C_{uw} must be an integer multiple of 2^{2-n}.

QED

With a similar argument it can be shown that either *all* elements of the Walsh-Hadamard transform of a Boolean function are an integer multiple of 2^{2-n} or none of them is.

5.4.4 Difference propagation

Consider a couple of n-dimensional vectors a and a^* with bitwise difference $a + a^* = a'$. Let $b = h(a), b^* = h(a^*)$ and $b + b^* = b'$, hence, the difference a' propagates to the difference b' through h. This is denoted by $\left(a' \xrightarrow{h} b'\right)$ or, if h is clear from the context, simply $(a' \to b')$. In general, b' is not determined by a' but depends on the value of a (or a^*).

Definition 5.2: The prop ratio $\mathrm{R_p}$ of a difference propagation $\left(a' \xrightarrow{h} b'\right)$ is given by

$$\mathrm{R_p}\left(a' \xrightarrow{h} b'\right) = 2^{-n} \sum_{a} \delta\left(b' + h(a+a') + h(a)\right).$$

If a pair is chosen uniformly from the set of all pairs $\left(a, a^*\right)$ with $a + a^* = a'$, the probability that $h(a+a') + h(a) = b'$ is given by $\mathrm{R_p}\left(a' \xrightarrow{h} b'\right)$. In this specific experimental set-up the prop ratio corresponds to a probability. This is however not the case in general and we believe that the widespread use of the term "probability" to denote what we call "prop ratio" has given rise to considerable confusion.

The prop ratio ranges between 0 and 1. Since $h(a+a') + h(a) = h(a) + h(a+a')$, it must be an integer multiple of 2^{1-n}. The difference propagation $\left(a' \xrightarrow{h} b'\right)$ restricts the values of a to a fraction of all possible inputs. This fraction is given by $\mathrm{R_p}\left(a' \xrightarrow{h} b'\right)$. It can easily be seen that

$$\sum_{b'} \mathrm{R_p}\left(a' \xrightarrow{h} b'\right) = 1.$$

If $\mathrm{R_p}\left(a' \xrightarrow{h} b'\right) = 0$, the difference propagation $\left(a' \xrightarrow{h} b'\right)$ is called *invalid*. The input difference a' and the output difference b' are said to be *incompatible* through h.

Definition 5.3: The restriction weight of a valid difference propagation $\left(a' \xrightarrow{h} b'\right)$ is the negative of the binary logarithm of the prop ratio, i.e.,

$$\mathrm{w_r}\left(a' \xrightarrow{h} b'\right) = -\log_2 \mathrm{R_p}\left(a' \xrightarrow{h} b'\right).$$

The restriction weight can be seen as the amount of information (in bits) that is given by $\left(a' \xrightarrow{h} b'\right)$ on a, or the loss in *entropy* [Sh48] of a due to the restriction $\left(a' \xrightarrow{h} b'\right)$. The restriction weight ranges between 0 and $n-1$.

If h is linear, $b' = b + b^* = h(a) + h\left(a^*\right) = h\left(a + a^*\right) = h(a')$, i.e., a' completely determines b'. From $\mathrm{w_r}\left(a' \xrightarrow{h} b'\right) = 0$ it can be seen that this difference propagation does not restrict or gives away information on a.

5.4.4.1 Special mappings

An *affine* mapping h from $\{0,1\}^n$ to $\{0,1\}^m$ is specified by

$$b = Ma + k \ ,$$

with M a $2^m \times 2^n$ matrix and k an m-dimensional vector. The difference propagation for this mapping is determined by

$$b' = Ma' \ .$$

For a juxtaposed mapping h, it can easily be seen that

$$\mathrm{R_p}\left(a' \xrightarrow{h} b'\right) = \prod_i \mathrm{R_p}\left(a'_{(i)} \xrightarrow{h} b'_{(i)}\right) ,$$

and

$$\mathrm{w_r}\left(a' \xrightarrow{h} b'\right) = \sum_i \mathrm{w_r}\left(a'_{(i)} \xrightarrow{h} b'_{(i)}\right) ,$$

with $a' = \left(a'_{(0)}, a'_{(1)}, \ldots, a'_{(l-1)}\right)$ and $b' = \left(b'_{(0)}, b'_{(1)}, \ldots, b'_{(l-1)}\right)$.

A mapping h from $\{0,1\}^n$ to $\{0,1\}^m$ can be converted into a mapping h_s from $\{0,1\}^n$ to $\{0,1\}^{m-1}$ by discarding a single output bit a_s. The prop ratios of h_s can easily be expressed in terms of the prop ratios of *h*:

$$\mathrm{R_p}\left(a' \xrightarrow{h_s} b'\right) = \mathrm{R_p}\left(a'_{(i)} \xrightarrow{h} \omega^0\right) + \mathrm{R_p}\left(a'_{(i)} \xrightarrow{h} \omega^1\right),$$

with $b'_i = \omega_i^0 = \omega_i^1$ for $i \neq s$ and $\omega_s^0 = 0$ and $\omega_s^l = 0$.

This can be generalised to the situation in which only a number of linear combinations of the output are considered. Let θ be a linear mapping corresponding to an $m \times l$ binary matrix M. The prop ratios of $\theta \circ h$ are given by

$$\mathrm{R_p}\left(a' \xrightarrow{\theta \circ h} b'\right) = \sum_{\omega | b' = M\omega} \mathrm{R_p}\left(a' \xrightarrow{h} \omega\right).$$

5.4.4.2 Prop ratios in terms of correlation coefficients

The prop ratios of the difference propagations of Boolean functions and mappings can be expressed respectively in terms of their Walsh-Hadamard transform values and their correlation matrix elements. With a derivation similar to (5.8) it can be shown that the components of the inverse transform of the componentwise product of two spectra $\hat{c}_{fg} = \mathcal{W}^{-1}\left(\hat{F}\hat{G}\right)$ are given by

$$\hat{c}_{fg}(b) = 2^{-n} \sum_a \hat{f}(a)\hat{g}(a+b) = 2^{-n} \sum_a (-1)^{f(a)+g(a+b)} , \qquad \textbf{(5.19)}$$

$\hat{c}_{fg}(b)$ is not a Boolean function. It is generally referred to as the *cross correlation function* of f and g. Hence, the cross correlation function of two Boolean functions is the inverse Walsh-Hadamard transform of the componentwise product of their spectra. If $g = f$ it is called the autocorrelation function of f and denoted by $\hat{r}_f$. The components of the spectrum of the autocorrelation function consist of the squares of the spectrum of f, i.e.,

$$\hat{F}^2 = \mathsf{W}\left(\hat{r}_f\right).$$

This is generally referred to as the Wiener-Khintchine theorem [Pr93].

The difference propagation in a Boolean function f can be expressed easily in terms of the autocorrelation function. The prop ratio of the difference propagation $\left(a' \xrightarrow{f} 0\right)$ is given by

$$\begin{aligned} \mathrm{R_p}\left(a' \xrightarrow{f} 0\right) &= 2^{-n}\sum_a \delta\left(f(a)+f(a+a')\right) \\ &= 2^{-n}\sum_a \frac{1}{2}\left(1+\hat{f}(a)\,\hat{f}(a+a')\right) \\ &= \frac{1}{2}\left(1+\hat{r}_f(a')\right) \\ &= \frac{1}{2}\left(1+\sum_w (-1)^{w^{\mathrm{t}}a'}\hat{F}^2(w)\right) \end{aligned} .$$

The component of the autocorrelation function $\hat{r}_f(a')$ corresponds to the amount that $\mathrm{R_p}\left(a' \xrightarrow{f} 0\right)$ deviates from 1/2.

For mappings from $\{0,1\}^n$ to $\{0,1\}^m$, let the autocorrelation function of $u^{\mathrm{t}}h(a)$ be denoted by $\hat{r}_u(a')$, i.e.,

$$\hat{r}_u(a') = 2^{-n}\sum_a (-1)^{u^{\mathrm{t}}h(a)+u^{\mathrm{t}}h(a+a')} .$$

Now we can easily prove the following remarkable theorem that expresses the duality between the difference propagation and the correlation properties of a Boolean mapping.

Theorem 5.2: The table of prop ratios and the table containing the squared elements of the correlation matrix of a Boolean mapping are linked by a (scaled) Walsh-Hadamard transform. We have

$$\mathrm{R_p}\left(a' \xrightarrow{h} b'\right) = 2^{-m}\sum_{u,w} (-1)^{w^{\mathrm{t}}a'+u^{\mathrm{t}}b'} C_{uw}^2 ,$$

and dually

$$C_{uw}^2 = 2^{-n}\sum_{a',b'} (-1)^{w^{\mathrm{t}}a'+u^{\mathrm{t}}b'} \mathrm{R_p}\left(a' \xrightarrow{h} b'\right).$$

Proof: we have

$$\begin{aligned}
\mathrm{R_p}\left(a' \xrightarrow{\ h\ } b'\right) &= 2^{-n}\sum_a \delta\left(h(a)+h(a+a')+b'\right) \\
&= 2^{-n}\sum_a \prod_i \frac{1}{2}\left((-1)^{h_i(a)+h_i(a+a')+b'_i}+1\right) \\
&= 2^{-n}\sum_a 2^{-m}\sum_u (-1)^{u^t h(a)+u^t h(a+a')+u^t b'} \\
&= 2^{-m}\sum_u (-1)^{u^t b'} 2^{-n}\sum_a (-1)^{u^t h(a)+u^t h(a+a')} \\
&= 2^{-m}\sum_u (-1)^{u^t b'} \hat{r}_u(a') \\
&= 2^{-m}\sum_u (-1)^{u^t b'} \sum_w (-1)^{w^t a'} C_{uw}^2 \\
&= 2^{-m}\sum_{u,w} (-1)^{w^t a'+u^t b'} C_{uw}^2
\end{aligned}$$

QED

5.5 Application to iterated transformations

The described tools and formalisms can be applied to the propagation of differences and the calculation of correlations in iterated transformations. This includes iterated block ciphers such as DES and the repeated application of state updating transformations in synchronous stream ciphers and the round transformations in cryptographic hash functions.

The studied iterated transformations are of the form

$$\beta = \rho_m \circ \cdots \circ \rho_2 \circ \rho_1 .$$

In a block cipher the ρ_i are selected from a set of round transformations $\left\{\rho[b] | b \in \{0,1\}^{\mathrm{n_b}}\right\}$ by round keys $\kappa^{(i)}$, i.e., $\rho_i = \rho\left[\kappa^{(i)}\right]$. These round keys are derived from the cipher key κ by the key schedule. In the iterated application of the state updating transformation of a synchronous stream cipher or a hash function, the ρ_i are selected by (part of) the buffer contents. The correspondence between our formalism and the terminology of the original descriptions of LC and DC is treated in Section 5.5.3.

5.5.1 Correlation

5.5.1.1 Fixed key

In the Walsh-Hadamard transform domain, a fixed succession of round transformations corresponds to a $2^n \times 2^n$ correlation matrix that is the product of the correlation matrices corresponding to the round transformations. We have

$$C^{\beta} = C^{\rho_m} \times \quad \times C^{\rho_2} \times C^{\rho_1} . \qquad \textbf{(5.20)}$$

Linear cryptanalysis exploits the occurrence of large elements in this product matrix. An m-round *linear trail* Ξ, denoted by

$$\Xi = \left(\xi_0 \xleftarrow{\rho_1} \xi_1 \xleftarrow{\rho_2} \xi_2 \quad \xi_{m-1} \xleftarrow{\rho_m} \xi_m\right). \tag{5.21}$$

consists of the chaining of m round transformation correlations of the type $\mathrm{C}\left(\xi_i{}^{\mathrm{t}}\rho_i(a), \xi_{i-1}{}^{\mathrm{t}}a\right)$.

To this linear trail corresponds a *correlation contribution coefficient* $\mathrm{C_p}$ ranging between –1 and +1. We have

$$\mathrm{C_p}(\Xi) = \prod_i C^{\rho_i}_{\xi_i \xi_{i-1}} .$$

From this definition and (5.20) we have

$$\mathrm{C}\left(u^{\mathrm{t}}\beta(a), w^{\mathrm{t}}a\right) = \sum_{\xi_0 = w, \xi_m = u} \mathrm{C_p}(\Xi) .$$

Hence, the correlation between $u^{\mathrm{t}}\beta(a)$ and $w^{\mathrm{t}}a$ is the sum of the correlation contribution coefficients of all m-round linear trails Ξ with initial selection vector w and terminal selection vector u.

5.5.1.2 Variable key

In actual cryptanalysis the succession of round transformations is not known in advance but is governed by an unknown key or some input-dependent value. In general, the elements of the correlation matrix of ρ_i depend on the specific value of the round key $\kappa^{(i)}$.

For some block ciphers the strong round-key dependence of the correlation and propagation properties of the round transformation has been cited as a design criterion. The analysis of correlation or difference propagation would have to be repeated for every specific value of the cipher key, making linear and differential analysis infeasible. A typical problem with this approach is that the *quality* of the round transformation with respect to LC or DC strongly depends on the specific value of the round key. While the resistance against LC and DC may be very good on the average, specific classes of cipher keys can exhibit linear trails with excessive correlation contributions (or differential trails with excessive prop ratios).

These complications can be avoided by designing the round transformation in such a way that the amplitudes of the elements of its correlation matrix are independent of the specific value of the round key. As was shown in Section 5.4.2, this is the case if the round transformation consists of a fixed transformation ρ followed (or preceded) by the bitwise addition of the round key $\kappa^{(i)}$ to (part of) the state.

The correlation matrix C_ρ is determined by the fixed transformation ρ. The correlation contribution coefficient of the linear trail Ξ becomes

$$\mathrm{C_p}(\Xi) = \prod_i (-1)^{\xi^t{}_i \kappa^{(i)}} C^{\rho}_{\xi_i \xi_{i-1}} = (-1)^{\varepsilon_\Xi + \sum_i \xi^t{}_i \kappa^{(i)}} \left|\mathrm{C_p}(\Xi)\right| .$$

with $\varepsilon_\Xi = 1$ if $\prod_i C^{\rho}_{\xi_i \xi_{i-1}}$ is negative and $\varepsilon_\Xi = 0$ otherwise. $\left|\mathrm{C_p}(\Xi)\right|$ is independent of the round keys, and hence, only the sign of the correlation contribution coefficient is key-dependent. Analogous to the restriction weight for differential trails, we can define:

Definition 5.4: The *correlation weight* $\mathrm{w_c}$ of a linear trail Ξ is given by

$$\mathrm{w_c}(\Xi) = -\log_2\left|\mathrm{C_p}(\Xi)\right| .$$

The correlation weight of a linear trail is the sum of the correlation weights of its linear steps given by $-\log_2\left|C^{\rho}_{\xi_i\xi_{i-1}}\right|$.

The correlation coefficient between $u^{\mathrm{t}}\beta(a)$ and $w^{\mathrm{t}}a$ can be expressed in terms of the correlation contribution coefficients of linear trails:

$$C\left(u^{\mathrm{t}}\beta(a), w^{\mathrm{t}}a\right) = \sum_{\xi_0=w,\xi_m=u} (-1)^{\varepsilon_\Xi + \sum_i \xi^{\mathrm{t}}_i k^{(i)}} \left|\mathrm{C_p}(\Xi)\right| .$$

The amplitude of this correlation coefficient is no longer independent of the round keys since the terms are added or subtracted depending on the value of the round keys.

5.5.1.3 Correlation analysis

The analysis of a round transformation with respect to its correlation properties consists of the investigation of two aspects.

The first aspect concerns the basic entities in LC, i.e., linear trails. The round transformation can be investigated by identifying the *critical* multiple-round linear trails, i.e., with the highest correlation contribution coefficient. For block ciphers the maximum correlation contribution coefficient for linear trails that span all but a few rounds has to be investigated. An efficient round transformation combines a low work factor with critical correlation contribution coefficients that decrease rapidly when the number of rounds increases. We give a strategy for the design of this type of round transformations at the end of this chapter, called the *wide trail strategy*.

The second aspect concerns the way in which linear trails combine to multiple-round correlations. Constructive interference of many linear trails with small correlation contribution coefficients may result in large correlations. Analysis includes investigating whether the round transformation can give rise to such *clustering*. For a well-designed round transformation multiple-round correlation coefficients larger than $2^{-n/2}$ are dominated by a single linear trail.

5.5.2 Difference propagation

5.5.2.1 Fixed key

An m-round *differential trail* Ω, denoted by

$$\Omega = \left(\omega_0 \xrightarrow{\rho_1} \omega_1 \xrightarrow{\rho_2} \omega_2 \quad \omega_{m-1} \xrightarrow{\rho_m} \omega_m\right)$$

consists of the chaining of difference propagations of the type $\left(\omega_{i-1} \xrightarrow{\rho_i} \omega_i\right)$. These are called the (differential) *steps* of the trail. The prop ratio of Ω, denoted by $\mathrm{R_p}(\Omega)$ is the relative portion of values of a_0 that exhibit the specified differential trail.

A differential step $\left(\omega_{i-1} \xrightarrow{\rho_i} \omega_i\right)$ imposes restrictions on the intermediate state a_{i-1}. If the succession of round transformations is assumed to be fixed, a_{i-1} is completely determined by a_0. Consequently, the restrictions on a_{i-1} can (in principle) be converted into restrictions on a_0. Since the round transformations are invertible, the relative size of the subset of allowed a_0 values is still given by $\mathrm{R_p}\left(\omega_{i-1} \xrightarrow{\rho_i} \omega_i\right)$. The relative size of the set of values a_0 that satisfy the restrictions imposed by all the differential steps of a differential trail Ω is per definition the prop ratio of Ω.

Definition 5.5: The restriction weight of a differential trail Ω is the sum of the restriction weights of its differential steps, i.e.,

$$\mathrm{w_r}(\Omega) = \sum_i \mathrm{w_r}\left(\omega_{i-1} \xrightarrow{\rho_i} \omega_i\right).$$

Now consider a two-round differential trail. The first step imposes restrictions on a_0 and the second on a_1. Typically, these restrictions involve only a subset of the components of each of the vectors. If for every selection vector v_0 of the involved components of a_0 and every selection v_1 of the involved components of a_1 the correlation $C^{\rho_1}_{v_1 v_0} = 0$, the restrictions are said to be uncorrelated. If this is the case, imposing values upon the involved components of a_0 does not restrict the involved components of a_1 and vice versa. Hence, the two restrictions are independent and the prop ratio of the two-round differential trail is equal to the product of the prop ratios of its two differential steps. This can readily be generalized to more than two rounds.

It is generally infeasible to calculate the exact value of $\mathrm{R_p}(\Omega)$, while it is easy for the restriction weight. Under the assumption that the restrictions originating from the different steps are not (or only very weakly) correlated, the prop ratio can be approximated by

$$\mathrm{R_p}(\Omega) \approx 2^{-\mathrm{w_r}(\Omega)}. \qquad \textbf{(5.22)}$$

In practice, e.g., for DES, the approximation is very good if the restriction weight is significantly below n. If $\mathrm{w_r}(\Omega)$ is of the order n or larger, (5.22) can no longer be a valid approximation. This is due to the inevitable (albeit small) correlations between the restrictions. The prop ratio multiplied by 2^n is the number of inputs a_0 that exhibit the specified differential trail, and it must therefore be an (even) integer. Of the differential trails Ω with a restriction weight $\mathrm{w_r}(\Omega)$ above n, only a fraction $2^{n\text{-}\mathrm{w_r}(\Omega)}$ can be expected to actually occur for some a_0.

5.5.2.2 Variable key

If the round transformation consists of a fixed transformation followed by the bitwise addition of the round key, the distribution of differential steps and their restriction weight is independent of the round key. Since the restriction weight of a differential trail consists of the sum of the restriction weights of its differential steps, it is independent of the cipher key.

The reduction of the restrictions imposed upon a_{i-1} by $\left(\omega_{i-1} \xrightarrow{\rho_i} \omega_i\right)$ to restrictions on a_0 involves the round keys, and hence, the prop ratio of a differential trail is in principle not independent of the cipher key. Or alternatively, the signs of the correlations between the different restrictions depend on the round keys. Since for the proposed round transformations the approximation given by (5.22) is key independent, differential trails with restriction weight significantly below n have prop ratios that can be considered independent of the round keys. Differential trails Ω with restriction weights $\mathrm{w_r}(\Omega)$ above n will only actually occur for an expected portion $2^{n-\mathrm{w_r}(\Omega)}$ of the cipher keys.

DC exploits difference propagations $\left(\omega_0 \xrightarrow{\beta} \omega_m\right)$ with large prop ratios. Since for a given input value a_0 exactly one differential trail is followed, the prop ratio of (a', b') is given by the sum of the prop ratios of all m-round differential trails with initial difference a' and terminal difference b', i.e.,

$$\mathrm{R_p}\left(a' \xrightarrow{\beta_i} b'\right) = \sum_{\omega_0 = a', \omega_1 = b'} \mathrm{R_p}(\Omega).$$

5.5.2.3 Propagation analysis

The analysis of a round transformation with respect to its difference propagation properties consists of the investigation of three aspects.

The first aspect concerns the basic entities in DC, i.e., differential trails. The round transformation can be investigated by identifying the *critical* multiple-round differential trails, i.e., with the lowest restriction weights. For block ciphers it is relevant to check the minimum restriction weight for differential trails that span all but a few rounds of the block cipher. An efficient round transformation combines a low work factor with critical restriction weights that grow rapidly as the number of rounds increases. This type of round transformations can also be designed by the *wide trail strategy*, described at the end of this chapter.

The second aspect concerns the approximation of the prop ratio of the differential trails by the product of the prop ratio of its steps. Specifically for the critical differential trails it must be checked whether the restrictions imposed by the differential steps can indeed be considered uncorrelated.

The third aspect concerns the way in which differential trails combine to difference propagations over multiple rounds. Many differential trails with high restriction weight and equal initial and terminal difference may result in difference propagation with a large prop ratio. As in the case of LC, analysis includes the investigation whether the round transformation can give rise to such *clustering*. For a well-designed round transformation multiple-round difference propagation with prop ratios larger than 2^{1-n} are dominated by a single differential trail.

5.5.3 DES cryptanalysis revisited

In this section we match the elements of linear and differential cryptanalysis as described in Section 5.3 with those of our framework.

5.5.3.1 Linear cryptanalysis

The multiple-round linear expressions described in [Ma94] correspond to what we call linear trails. The probability p that such an expression holds corresponds to $\frac{1}{2}\left(1+\mathrm{C_p}(\Xi)\right)$, with $\mathrm{C_p}(\Xi)$ the correlation contribution coefficient of the corresponding linear trail. This implies that the considered correlation coefficient is assumed to be dominated by a single linear trail. This assumption is valid because of the large amplitude of the described correlation coefficients on the one hand and the structure of the DES round transformation on the other hand.

The correlation of the linear trail is independent of the key and consists of the product of the correlations of its steps. In general, the elements of the correlation matrix of the DES round transformation are not independent of the round keys. In the described linear trails the actual independence is caused by the fact that the steps of the described linear trail only involve bits of a single S-box.

The input-output correlations of F-function of DES can be calculated by applying the rules given in Section 5.4.2.1. The 32-bit selection vector *b* at the output of the bit permutation *P* is converted into a 32-bit selection vector *c* at the output of the S-boxes by a simple linear transformation. The 32-bit selection vector *a* at the input of the (linear) expansion *E* gives rise to a set α of 2^{ℓ} 48-bit selection vectors after the expansion, with ℓ the number of neighboring S-box pairs that are addressed by *a*.

In the assumption that the round key is all-zero, the correlation between *c* and *a* can now be calculated by simply adding the correlations corresponding to *c* and all vectors in α. Since the S-boxes form a juxtaposed mapping, these correlations can be easily calculated from the correlation matrices of the individual S-boxes. For $\ell > 0$ the calculations can be greatly simplified by recursively reusing intermediate results in computing these correlations. The total number of calculations can be reduced to less than 16ℓ multiplications and additions of S-box correlations.

The effect of a nonzero round key is the multiplication of some of these correlations by –1. Hence, if $\ell > 0$, the correlation depends on the value of 2ℓ different linear combinations of round key bits. If $\ell = 0$, α only contains a single vector and the correlation is independent of the key.

5.5.3.2 Differential cryptanalysis

The characteristics with their characteristic probability described in [BiSh91] correspond to what we call differential trails and their (approximated) prop ratio. The prop ratio of a differential trail is taken to be the prop ratio of the difference propagation from its initial difference to its terminal difference. For the DC of DES this is a valid approximation because of the large prop ratios of the considered differential trails and the structure of the DES round transformation.

For the DES round transformation the distribution of the differential steps and their restriction weights are not independent of the round keys. This dependence was already recognized in [BiSh91] where in the analysis the restriction weights of the differential steps are approximated by an average value. Lars Knudsen has shown that the two-round iterative differential with approximate prop ratio 1/234 in fact has a prop ratio of either 1/146 or 1/585 depending on the value of a linear combination of round key bits [Kn93].

Later, Martin Hellman and Susan Langford published an attack on an 8-round variant of DES that combines the mechanisms of differential and linear cryptanalysis [HeLa94]. In their attack they apply plaintext pairs with a specific difference that propagates with prop ratio 1 to a certain difference in the intermediate state after 3 rounds confined to a subset of its bits. Then a 3-round linear trail is constructed between the output of round 7 and the input of round 4. The correlation between certain linear combinations of intermediate bits in the pair is exploited to gain information about key bits. It can easily be shown that this correlation is the square of the correlation contribution coefficient of the 3-round linear trail. The number of required plaintext-ciphertext pairs can be approximated by raising this correlation contribution coefficient to the power –4 while in simple linear cryptanalysis the required number of pairs is approximately the critical correlation contribution coefficient to the power –2. This limits the usability of this attack to ciphers with poor resistance against differential and linear cryptanalysis.

5.6 The wide trail strategy

In this section we present our strategy for the design of round transformations without low-weight multiple-round linear and differential trails.

For both types of trails, the weight is given by the sum of the weights of its steps. Let the round transformation consist of three steps: an invertible nonlinear transformation γ, an invertible linear transformation θ and the round key addition.

Suppose γ is a juxtaposed transformation. As explained in Section 5.4.2.1, a correlation coefficient C_{uw}^{γ} is the product of the corresponding input-output correlation coefficients of the S-boxes. With the correlation weight of the input-output correlation of an S-box equal to minus the binary logarithm of its correlation, the correlation weight of a linear step is given by the sum of the correlation weights of the corresponding input-output correlations of the S-boxes. Similarly, the restriction weight of a differential step is the sum of the restriction weights of the corresponding difference propagations of the S-boxes.

An S-box of a specific round is said to be *active* with respect to a linear trail if its output selection vector is nonzero for that linear trail. It is said to be active with respect to a differential trail if its input difference vector is nonzero for that differential trail. Now, both for linear and differential trails it can be seen that the weight of a trail is the sum of the active S-boxes.

This suggests two possible mechanisms of eliminating low-weight trails:

- Choose S-boxes with difference propagations that have high restriction weight and with input-output correlations that have high correlation weight.
- Design the round transformation in such a way that only trails with many S-boxes occur.

The wide trail strategy emphasises the second mechanism. The round transformation must be designed in such a way that linear (or differential) steps with only few active S-boxes are followed by linear (or differential) steps with many active S-boxes. This is closely linked to the concept of *diffusion*, introduced by Shannon [Sh49] to denote the quantitative spreading of information. The only requirement for the S-boxes themselves is that their input-output correlations have a certain minimum correlation weight and that their difference propagations have a certain minimum restriction weight.

The wide trail strategy does not restrict the nonlinear step to juxtaposed transformations. It can equally well be applied to the shift-invariant transformations that are treated in the following chapter.

5.6.1 Traditional approach

The wide trail strategy contrasts highly with the approach taken by the majority of cryptographic researchers working in cipher design. This traditional approach is dominated by the structure of DES and fully concentrates on the S-boxes. This is illustrated by the small width of the linear and differential trails in DES. Its most effective differential trail contains only 3 S-boxes per 2 rounds, its most effective linear trail only 3 S-boxes per 4 rounds.

Typically, the S-boxes are (tacitly) assumed to be located in the F-function of a Feistel structure or in some academic round transformation model such as *so-called* substitution-permutation (or transposition) networks [AdTa90,Oc93]. These networks consist of the alternation of parallel S-boxes and bit permutations and were proposed in [Fe75,KaDa79]. The S-boxes are considered to be *the* active elements in the cipher and must be designed to eliminate low-weight trails. In practice this requirement is translated to "criteria" for S-boxes, such as maximum input-output correlation, maximum prop ratio and diffusion criteria. These criteria impose conflicting restrictions, and finding S-boxes that have an acceptable score with respect to all them becomes less difficult when their size grows.

This has led many researchers to the conclusion that resistance against DC and LC is best realised by adopting large S-boxes. This one-sided point of view plainly ignores the potential of high diffusion provided by a well-designed round transformation.

5.7 Conclusions

We have given a number of tools to describe and investigate the propagation of differences and the correlations in Boolean mappings and iterated transformations. An explicit design strategy has been formulated and motivated.

6. References

[AdTa90] C. Adams and S. Tavares, The Structured Design of Cryptographically Good S-Boxes, *Journal of Cryptology*, Vol. 3, No. 1, 1990, pp. 27-42.

[BiSh91] E. Biham and A. Shamir, "Differential cryptanalysis of DES-like cryptosystems," Journal of Cryptology, Vol. 4, No. 1, 1991, pp. 3-72.

[Fe75] H. Feistel, W.A. Notz, and J.L. Smith, Some cryptographic techniques for machine-to-machine data communications, Proc. IEEE, Vol. 63, No. 11, 1975, pp. 1545-1554.

[Fi77] Data Encryption Standard, Federal Information Processing Standard (FIPS), Publication 46, National Bureau of Standards, U.S. Department of Commerce, Washington D.C., January 1977.

[Go67] S.W. Golomb, Shift Register Sequences, Holden-Day Inc., San Francisco, 1967.

[HeLa94]M. Hellman and S. Langford, Differential-Linear Cryptanalysis, in Advances in Cryptology, Proc. Crypto'94, LNCS~839, Y. Desmedt, Ed., Springer-Verlag, 1994, pp. 26-39.

[KaDa79] J.B. Kam and G.I. Davida, Structured design of substitution-permutation encryption networks, IEEE Trans. on Computers, Vol. C-28, 1979, pp.747-753.

[Kn93] L.R. Knudsen, Iterative Characterisitics of DES and s^2-DES, in Advances in Cryptology, Proc. Crypto'92, LNCS 740, E.F. Brickell, Ed., Springer-Verlag, 1993, pp. 497-512.

[Ma93] M. Matsui, "Linear cryptanalysis method for DES cipher," Advances in Cryptology, Proceedings Eurocrypt'93, LNCS 765, T. Helleseth, Ed., Springer-Verlag, 1994, pp. 386-397.

[Ma94] M. Matsui, The First Experimental Cryptanalysis of the Data Encryption Standard, in Advances in Cryptology, Proc. Crypto'94, LNCS 839, Y. Desmedt, Ed., Springer-Verlag, 1994, pp.1-11.

[Oc93] L. O'Connor, On the Distribution of Characteristics in Bijective Mappings, in Advances in Cryptology, Proc. Eurocrypt '93, LNCS 765, T. Helleseth, Ed., Springer-Verlag, 1994, pp. 360-370.

[Pr93] B. Preneel, Analysis and Design of Cryptographic Hash Functions, Doct. Dissertation KULeuven, 1993.

[Sh48] C.E. Shannon, A Mathematical Theory of Communication, Bell Syst. Tech. Journal, Vol. 27, No. 3, 1948, pp. 379-423 and pp. 623-656.

[Sh49] C.E. Shannon, Communication Theory of Secrecy Systems, Bell Syst. Tech. Journal, Vol. 28, 1949, pp. 656-715.