On linear hulls in one round of DES

Tomer Ashur and Raluca Posteuca

imec-COSIC, KU Leuven, Leuven, Belgium
[tomer.ashur, raluca.posteuca]@esat.kuleuven.be

Abstract. At Indocrypt 2016, Ashur et al. showed that linear hulls are sometimes formed in a single round of a cipher (exemplifying on Simon ciphers) and showed that the success rate of an attack may be influenced by the quality of the estimation of one-round correlations. This paper improves the understanding regarding one-round linear hulls and trails, being dedicated to the study of one-round linear hulls of the DES cipher, more exactly of its f-function. It shows that, in the case of DES, the existence of one-round hulls is related to the number of active Sboxes and its correlation depends on a fixed set of key bits. All the ideas presented in this paper are followed by examples and are verified experimentally.

1 Introduction

Together with differential cryptanalysis, linear cryptanalysis is one of the most powerful techniques used in the security evaluation of a block cipher. It was introduced in early 1990 by Mitsuru Matsui in [1], who applied the technique to the DES cipher. The technique became intensively studied, its formalism being extended in [2], [3] and [4]. The attack has been generalised in many subsequent works such as [6], [8].

In general, the idea behind linear cryptanalysis is to find a linear approximation between a set of plaintext bits and ciphertext bits that holds with probability different from 0.5. Estimating the quality of a linear approximation, usually measured by its bias or correlation, is one of the most important problems in linear cryptanalysis, being directly related to the success rate of a linear attack.

The idea introduced by Matsui was to construct a linear approximation for n rounds of a block cipher by concatenating n one-round linear approximations. The estimation of the correlation of a linear trail is computed using the Pilling-Up Lemma, by multiplying the correlations of each one-round linear approximation.

1.1 Related work

During the last three decades, many papers tried to find ways to obtain a better estimation of the correlation of a linear approximation.

In [4] was first observed that, in some cases, there are more than one linear trail involving the same plaintext and ciphertext bits. The set of all such linear

trails, with a fixed set of input and output bits, is called a linear hull. The correlation of a hull is the sum of the linear trails correlations, so the correlation of one trail may be different than the correlation of the hull containing it. In practice, when using an attack based on linear cryptanalysis, the correlation of a linear hull is used, so the success rate of the attack is closely related to the quality of the hull correlation.

In [7] the authors show that the linear hull effect may sometimes appear at a micro-level, inside a single round of a cipher. They also show that overlooking this phenomenon may lead to wrong estimations of the linear correlation, offering examples and experiments on round-reduced versions of ciphers of the Simon family.

1.2 Our contribution

In this paper we extend the study regarding one-round linear hulls introduced in [7]. In the first part of the paper we recall some definitions and terminology of linear cryptanalysis. In the second part we present an analysis, similar to the one described in [7], but applied to the DES cipher. We prove the existence of the linear hull effect on the f-function. We also prove the connection between the number of active Sboxes, the existence and the number of linear trails in one-round linear hulls. Finally we present a simple and illustrative example of a hull containing 4 linear trails, and the manner in which its correlation is computed.

2 Notations

In this section we recall some terminology regarding linear cryptanalysis and we introduce the notations that will be used in this paper.

2.1 Masks and approximations

Let a be a hexadecimal value smaller than 2^n and let $a^t x = \sum_{i=0}^{n-1} a_i x_i$, where a_i and x_i represent the i^{th} bit of a and x, respectively. We will call a the mask of x. Given the fact that applying a mask to a number represents, in essence, a selection of bits of x, in this paper we will also use the description of a mask as a set of positions:

$$\bar{a} = \{i_1, i_2, ..., i_v\} \Leftrightarrow \begin{cases} a_j = 1, \forall j \in \{i_1, i_2, ..., i_u\} \\ a_j = 0, \forall j \notin \{i_1, i_2, ..., i_u\} \end{cases}$$

Let $R_k(x) = y$ denote the round function of a cipher, where k denotes the key. The explicit function will always be clear from context. A linear approximation for R_k is the tuple (iM, oM, kM), where iM represents the input mask, oM the output mask, and kM the key mask. Let p be the probability that the equation $iM^t x \oplus oM^t y \oplus kM^t k = 0$ holds. The correlation of the linear approximation

(iM, oM, kM) is defined as $corr(iM, oM, kM) = 2p - 1$. In general, both p and $corr(iM, oM, kM)$ are key-dependent.

A pair (iM, oM) is called connectable if and only if oM can be obtained from iM using the rules of propagation of linear trails introduced in [2, 3]. The pair (iM, oM) is non-connectable otherwise. In this paper only connectable pairs are used.

2.2 Linear hulls and trails for more rounds

A linear trail for r rounds represents a concatenation of linear approximations such that the output mask of the round i equals the input mask of round $i + 1$. Hence, a linear trail may be represented as an $(r + 1)$-length vector $(m_1, m_2, ...m_{r+1})$, where (m_i, m_{i+1}) represents the input and output masks at round i, respectively. The correlation of the linear trail is computed by multiplying the correlation of all single-round linear approximations:

$$corr(m_1, ...m_{r+1}) = \prod_{i=1}^{r} corr(m_i, m_{i+1})$$

A linear hull covering r rounds is a tuple (α, β) and represents the set of all linear trails such as $m_1 = \alpha$ and $m_{r+1} = \beta$, i.e., the input and output masks are the same, but that intermediate masks may be different. The correlation of a linear hull is computed by adding the correlations of all linear trails:

$$corr(\alpha, \beta) = \sum_{m_1=\alpha, m_{r+1}=\beta} corr(m_1, ...m_{r+1})$$

2.3 DES' round function

DES is a block cipher developed by IBM in the early 1970s. It was standardized by the National Bureau of Standards (later NIST) in 1977. The plaintext and the key are 64-bit blocks, even though only 56 out of 64 key-bits are actually used by the algorithm. DES has a Feistel structure where the round function uses a non-linear function f. The overall structure of DES consists of an initial permutation, 16 enciphering rounds and a final permutation.

The input of the round function is a 48-bit round key (denoted by k) and two 32-bit intermediate cipherwords (denoted by x and y).

The round function of DES is given by:

$$R_k(x, y) = (y \oplus f(x, k), x).$$

The f-function consists of four functions:

1. *The expansion function*: the 32-bit input x is expanded to 48-bit output, using the *expansion permutation* described in Table 1; one may notice the fact that, after applying this layer, 16 out of 32 input bits appear twice.

We denote the expansion function by E;

Table 1. The expansion function E. We see that 16 out of the 48 bits appear twice.

32	1	2	3	4	5
4	5	6	7	8	9
8	9	10	11	12	13
12	13	14	15	16	17
16	17	18	19	20	21
20	21	22	23	24	25
24	25	26	27	28	29
28	29	30	31	32	1

2. *Key addition*: the output of the expansion function is XORed with the 48-bit round key;

3. *The substitution layer*: the output of the key addition is divided into eight 6-bit blocks. Each of these blocks is given as input to a different 6-to-4 Sbox, resulting in eight 4-bit outputs. The first two Sboxes used in DES are described in Table 2, while the remaining six Sboxes are described in [10].

The Sboxes of DES are applied as follows: for the input $x_0x_1x_2x_3x_4x_5$, the output after applying the i^{th} Sbox is the value found at the intersection of the row x_0x_5 and the column $x_1x_2x_3x_4$ of the table of S_i. We denote the substitution layer by S.

4. *The permutation layer*: a fixed 32-bit to 32-bit permutation is applied to the result of the substitution layer; the permutation, denoted by P, is described in [10].

Table 2. DES' Sboxes

S_1																
0	14	4	13	1	2	15	11	8	3	10	6	12	5	9	0	7
1	0	15	7	4	14	2	13	1	10	6	12	11	9	5	3	8
2	4	1	14	8	13	6	2	11	15	12	9	7	3	10	5	0
3	15	12	8	2	4	9	1	7	5	11	3	14	10	0	6	13
S_2																
0	15	1	8	14	6	11	3	4	9	7	2	13	12	0	5	10
1	3	13	4	7	15	2	8	14	12	0	1	10	6	9	11	5
2	0	14	7	11	10	4	13	1	5	8	12	6	9	3	2	15
3	13	8	10	1	3	15	4	2	11	6	7	12	0	5	14	9

3 One round hulls in DES

In this section we prove the existence of one-round hulls in DES, which might impact the computation of the correlation on multiple rounds of the cipher.

3.1 One round hulls and trails in DES

In order to describe the behavior of linear trails through one round of DES, it is sufficient to analyze the f-function. This will be the focus of this section.

The permutation layer represents a bijective function so the behavior of a mask through this layer is given by: $oM = P(iM)$, i.e., the bits of the output mask are a shuffle of input mask's bits. The permutation layer does not influence the number of linear trails in the one-round hull, nor the value of the correlations. For simplicity, in the following two subsections, we will ignore the permutation layer.

We use the notation (R, kM, A, B, C) to describe a linear trail of the f-function in DES, where the input mask is $iM = R = (r_i)_{i\in\{1,..8\}}$ and the output mask is $oM = C = (c_i)_{i\in\{1,..8\}}$. We denote by $A = (a_i)_{i\in\{1,..8\}}$ the output of the expansion layer. We denote by $B = (b_i)_{i\in\{1,..8\}}$ and $C = (c_i)_{i\in\{1,..8\}}$ the input and the output masks of the substitution layer, respectively. $kM = (k_i)_{i\in\{1,..8\}}$ represents the key mask. Note that in the remaining of this paper, r_i, a_i, b_i and k_i are represented as sets of position values, depending on the input mask R and the key mask respectively, while c_i's are represented as 4-bit decimal values, being nonlinear functions of R's bits.

Figure 1 depicts the propagation of linear masks through the f-function of DES.

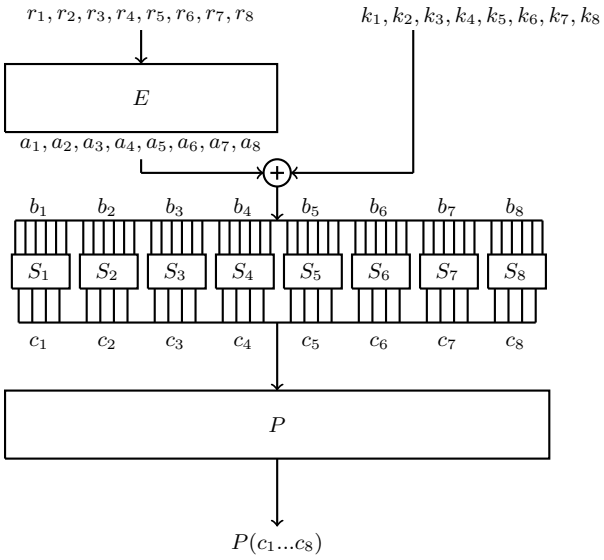

Fig. 1. A linear trail through the f-function of DES

We now study the behavior of linear trails through the f-function of DES, using the rules of propagation of linear trails introduced in [1–3]. The rule of propagation regarding the XOR operation (used in the key addition layer) implies the following constraint:

$$a_i = k_i = b_i, \forall i \in \{1, ...8\}.$$

The rule of propagation of linear masks through the Sboxes is given by the linear approximation tables (LATs) of each Sbox, described in [1, 2].

For the expansion layer we use the rule for linear propagation over the branch operation, as follows:

If R has an active bit in $M_0 = \{2, 3, 6, 7, 10, 11, 14, 15, 18, 19, 22, 23, 26, 27, 30, 31\}$, i.e., those bits that appear only once after the expansion layer, the corresponding bit in A is an active bit.

If a bit of M_0 is not active in R then the corresponding bit in A is not an active bit.

Example If $r_1 = \{2\}$, then $a_1 = \{2\}$.

If R has an active bit in $M_1 = \{1, 4, 5, 8, 9, 12, 13, 16, 17, 20, 21, 24, 25, 28, 29, 32\}$, i.e., those bits that appear twice after the expansion layer, only one of the corresponding bits in A is an active bit.

If a bit of M_1 is not active in R, then either none of the corresponding bits in A are active bits or both of the corresponding bits in A are active bits.

Example If $r_1 = \{4\}$, then
- $a_1 = \{4\}, a_2 = \emptyset$, or
- $a_1 = \emptyset, a_2 = \{4\}$, or
- $a_1 = \{4, 5\}, a_2 = \{5\}$, or
- $a_1 = \{5\}, a_2 = \{4, 5\}$.

Definition If $b_i, c_i \neq \emptyset$, then S_i is called an active Sbox.

Lemma 1. Let (R, C) be a fixed pair of input and output masks and let s denote the number of pairs of active adjacent Sboxes. Then the maximum number of linear trails contained in the hull (R, C) is 2^{2s}.

Let $\oplus$ be defined as an operation between two sets by:

$$A \oplus B = (A \cup B) \setminus (A \cap B)$$

Let S_1 and S_2 be active Sboxes, i.e., $b_1 \neq \emptyset$ and $b_2 \neq \emptyset$. After applying the expansion layer, one may easily notice the fact that any two adjacent Sboxes share two input bits; in the case of S_1 and S_2 the shared bits are the ones in positions 4 and 5 from R. Given the fact that $b_1 \oplus b_2$ is constant then four linear trails can be defined such that the input masks of the Sbox layer are:

- (b_1, b_2)
- $(b_1 \oplus \{4\}, b_2 \oplus \{4\})$
- $(b_1 \oplus \{5\}, b_2 \oplus \{5\})$
- $(b_1 \oplus \{4, 5\}, b_2 \oplus \{4, 5\})$

Figure 2 depicts a linear hull that contains exactly 4 linear trails. Note that, even though a_i and k_i appear to be different, their corresponding bit locations are the same.

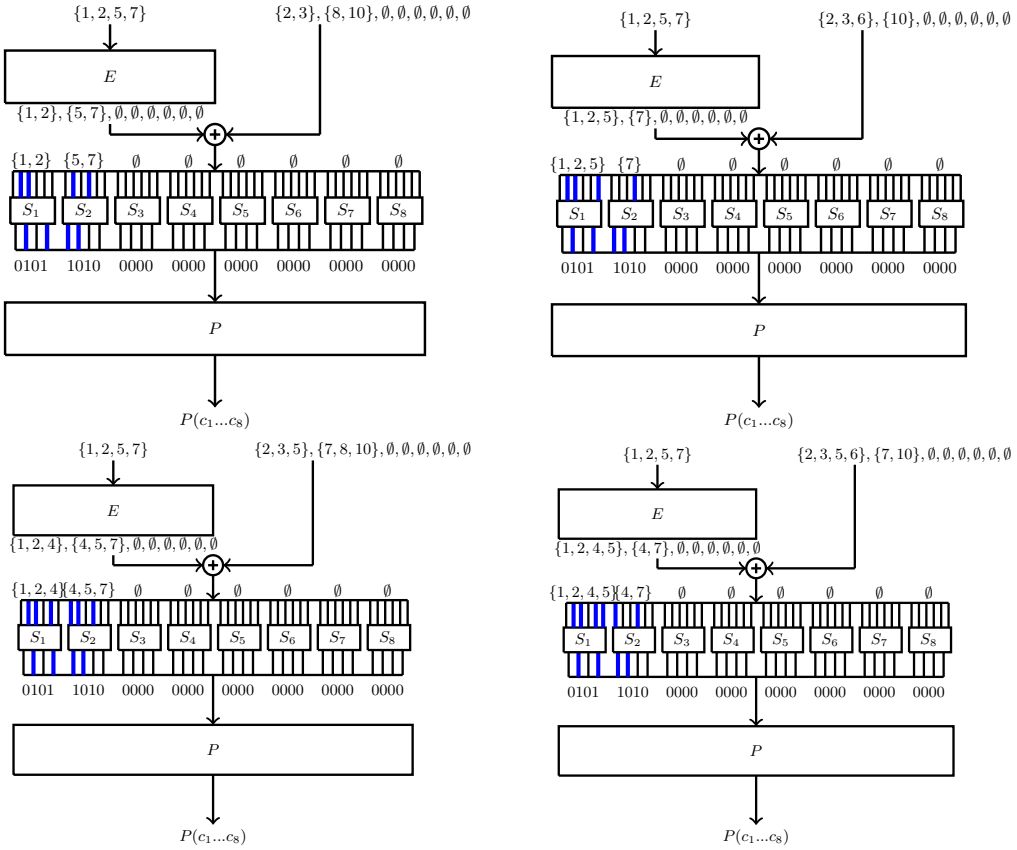

Fig. 2. Four trails of a one-round hull through DES f-function

3.2 Correlations of one-round hulls

We now want to compute the correlations of the trails in Figure 2. The general rule regarding the correlation computation through a composed function is to assume that all the functions act independently and multiply the correlation of each function [9]. The correlation of the linear layers (the expansion layer and the permutation layer) is $corr = 1$. Note that, for the key addition layer, the

correlation is either $corr = 1$ or $corr = -1$, depending on the actual value of the round key bits.

In the case of the substitution layer, the correlation is usually computed by multiplying the correlation of each active Sbox - usually extracted from the Sbox's LAT.

Given the fact that $b_i = k_i, \forall i \in \{1, ..8\}$, for every linear trail in the hull, and given the fact that b_i are different for every trail, the correlations of different trails are influenced by different key bits. The correlation of each linear trail described in Figure 2, up to a sign, together with the corresponding key bits can be found in Table 4.

Table 4. Correlations of linear trails described in Figure 2

Trails	S input masks	S output masks	Key masks	Correlation
Trail 1	$b_1 = \{1, 2\}$	$c_1 = 0101$	$k_1 = \{2, 3\}$	$corr = 0.1875$
	$b_2 = \{5, 7\}$	$c_2 = 1010$	$k_2 = \{8, 10\}$	$corr = 0.0625$
Trail 2	$b_1 = \{1, 2, 5\}$	$c_1 = 0101$	$k_1 = \{2, 3, 6\}$	$corr = -0.1875$
	$b_2 = \{7\}$	$c_2 = 1010$	$k_2 = \{10\}$	$corr = -0.0625$
Trail 3	$b_1 = \{1, 2, 4\}$	$c_1 = 0101$	$k_1 = \{2, 3, 5\}$	$corr = -0.0625$
	$b_2 = \{4, 5, 7\}$	$c_2 = 1010$	$k_2 = \{7, 8, 10\}$	$corr = 0.1875$
Trail 4	$b_1 = \{1, 2, 4, 5\}$	$c_1 = 0101$	$k_1 = \{2, 3, 5, 6\}$	$corr = 0.0625$
	$b_2 = \{4, 7\}$	$c_2 = 1010$	$k_2 = \{7, 10\}$	$corr = -0.1875$

We will now show that the correlation of the hull depends on the actual values of the round key. One may get some information regarding the round key just by looking at the value of the hull's correlation. From Table 4 one may notice the fact that there are 7 key bits that influence the value of the hull's correlation: $k[2], k[3], k[5], k[6], k[7], k[8]$ and $k[10]$. Three of these values, more precisely $k[2], k[3]$ and $k[10]$ influence all the trails' correlation. We will denote their XOR sum by l. Notice the fact that the remaining 4 key bits are exactly the ones that are applied to the two plaintext bits that are shared among S_1 and S_2.

The correlation of the hull is computed by adding the correlation of all the linear trails contained in it:

$$corr_{hull} = corr_{Trail1} + corr_{Trail2} + corr_{Trail3} + corr_{Trail4}$$

Let $c = 0.1875 \cdot 0.0625$. According to Table 4, the correlation of each trail is:

$$corr_{Trail1} = c \cdot (-1)^{k[2] \oplus k[3] \oplus k[8] \oplus k[10]}$$
$$corr_{Trail2} = c \cdot (-1)^{k[2] \oplus k[3] \oplus k[6] \oplus k[10]}$$
$$corr_{Trail3} = -c \cdot (-1)^{k[2] \oplus k[3] \oplus k[5] \oplus k[7] \oplus k[8] \oplus k[10]}$$
$$corr_{Trail4} = -c \cdot (-1)^{k[2] \oplus k[3] \oplus k[5] \oplus k[6] \oplus k[7] \oplus k[10]}$$

Taking into account the fact that c and $(-1)^{k[2]\oplus k[3]\oplus k[10]} = (-1)^l$ are common to all the trails' correlation formulas, the correlation of the hull can be computed as:

$$corr_{hull} = c \cdot (-1)^l \cdot [(-1)^{k[8]} + (-1)^{k[6]} - (-1)^{k[5]\oplus k[7]\oplus k[8]} - (-1)^{k[5]\oplus k[6]\oplus k[7]}]$$

Table 5 considers all possible values of $k[5], k[6], k[7], k[8]$ and the corresponding correlation values of the hull. Note that the value of l will only influence the sign of the correlation.

Table 5. Key-dependent correlation values for the Figure 2 hull

$k[5]$	$k[6]$	$k[7]$	$k[8]$	$corr$
0	0	0	0	0
0	0	0	1	0
0	0	1	0	$4 \cdot c \cdot (-1)^l$
0	0	1	1	0
0	1	0	0	0
0	1	0	1	0
0	1	1	0	0
0	1	1	1	$-4 \cdot c \cdot (-1)^l$

$k[5]$	$k[6]$	$k[1]$	$k[2]$	$corr$
1	0	0	0	$4 \cdot c \cdot (-1)^l$
1	0	0	1	0
1	0	1	0	0
1	0	1	1	0
1	1	0	0	0
1	1	0	1	$-4 \cdot c \cdot (-1)^l$
1	1	1	0	0
1	1	1	1	0

From the table above results the condition of the key that leads to zero-correlation one-round hull for DES:

1. $corr_{hull} = 0 \iff \begin{cases} k[5] = k[7] \text{ or} \\ k[5] \neq k[7] \text{ and } k[6] \neq k[8] \end{cases}$
2. $corr_{hull} \neq 0 \iff k[5] \neq k[7] \text{ and } k[6] = k[8]$

3.3 Key recovery based on the correlation

Looking the other way around, if the correlation of this hull is different from zero, one knows two equalities of the key bits:

$$k[5] = k[7] \oplus 1$$
and
$$k[6] = k[8].$$

Our experiments confirm that the only key bits that influence the absolute value of the correlation are the ones applied on the plaintext bits shared between S_1 and S_2.

Depending on the input and output masks, the percentage of keys that lead to zero-correlation hulls are: 0%, 25%, 50% or 75% - the percentage depends on the number of linear trails in a hull and on the correlations resulting from the Sboxes' LAT.

4 Conclusion

In this paper we extended the knowledge regarding the subject of one-round linear hulls, analyzing this phenomenon on the DES cipher. We proved the fact that the f-function of DES exhibits one-round linear hulls. We also studied the connections between the properties of input and output masks and the existence, the size or the correlation values of one-round linear hulls.

The work described in this paper can be extended in different manners. For example, it will be interesting to identify other block ciphers that exhibit zero-correlation one-round linear hulls and to describe the properties of the round function that led to it. It also remains to be investigated how the properties described in this paper may be extended to more rounds of DES. Future research should also revisit the linear attacks on DES, attacks in which only one trail correlation is used in order to estimate the hull's correlation.

Acknowledgments

The authors would like to thank Vincent Rijmen for all the useful discussions.

References

1. Mitsuru Matsui, *Linear cryptanalysis method for DES cipher*, In: Helleseth, T. (ed.) Advances in Cryptology - EUROCRYPT '93, Workshop on the Theory and Application of Cryptographic Techniques, Lofthus, Norway, May 23-27, 1993, Proceedings. Lecture Notes in Computer Science, vol. 765, pp. 386–397. Springer (1993).
2. Eli Biham, *On Matsui's linear cryptanalysis*, In: Santis, A.D. (ed.): Advances in Cryptology - EUROCRYPT '94, Workshop on the Theory and Application of Cryptographic Techniques, Perugia, Italy, May 9-12, 1994, Proceedings, Lecture Notes in Computer Science, vol. 950. Springer (1995) , pp. 341–355, http://dx.doi.org/10.1007/BFb005344p
3. Florent Chabaud, Serge Vaudenay, *Links between differential and linear cryptanalysis*, In: Santis, A.D. (ed.): Advances in Cryptology - EUROCRYPT '94, Workshop on the Theory and Application of Cryptographic Techniques, Perugia, Italy, May 9-12, 1994, Proceedings, Lecture Notes in Computer Science, vol. 950. Springer (1995) , pp. 356–365, http://dx.doi.org/10.1007/BFb0053450
4. Kaisa Nyberg, *Linear approximation of block ciphers*, In: Santis, A.D. (ed.): Advances in Cryptology - EUROCRYPT '94, Workshop on the Theory and Application of Cryptographic Techniques, Perugia, Italy, May 9-12, 1994, Proceedings, Lecture Notes in Computer Science, vol. 950. Springer (1995) , pp. 439–444, http://dx.doi.org/10.1007/BFb0053460
5. Ray Beaulieu, Douglas Shors, Jason Smith, Stefan Treatman-Clark, Bryan Weeks, Louis Wingers, *Simon and Speck: Block Ciphers for the Internet of Things*, Cryptology ePrint Archive, Report 2013/404 (2013), http://eprint.iacr.org/.
6. Andrey Bogdanov, Vincent Rijmen, *Linear hulls with correlation zero and linear cryptanalysis of block ciphers*. Des. Codes Cryptography 70(3) (2014), pp. 369–383
7. Tomer Ashur, Vincent Rijmen, *On linear hulls and trails*. In: Progress in Cryptology – INDOCRYPT 2016

8. Alex Biryukov, Christophe De Canniére, Michaël Quisquater, *On Multiple Linear Approximations*. In Franklin, M.K., ed.: Advances in Cryptology - CRYPTO 2004, 24th Annual International CryptologyConference, Santa Barbara, California, USA, August 15-19, 2004, Proceedings. Volume 3152 of Lecture Notes in Computer Science., Springer (2004) 1–22
9. Joan Daemen, René Govaerts, Joos Vandewalle: *Correlation matrices*. In: Preneel, B. (ed.) Fast Software Encryption: Second International Workshop. Leuven, Belgium, 14-16 December 1994, Proceedings. Lecture Notes in Computer Science, vol. 1008, pp. 275–285. Springer (1994), http://dx.doi.org/10.1007/3-540-60590-8_21
10. FIPS Publication 46-3, Data Encryption Standard (DES), https://csrc.nist.gov/csrc/media/publications/fips/46/3/archive/1999-10-25/documents/fips46-3.pdf