Nyberg, Kaisa

Affine linear cryptanalysis

Published in:
Cryptography and Communications

DOI:
10.1007/s12095-018-0325-2

Published: 15/05/2019

Document Version
Peer-reviewed accepted author manuscript, also known as Final accepted manuscript or Post-print

Please cite the original version:
Nyberg, K. (2019). Affine linear cryptanalysis. *Cryptography and Communications*, *11*(3), 367-377. https://doi.org/10.1007/s12095-018-0325-2

Noname manuscript No.
(will be inserted by the editor)

Affine Linear Cryptanalysis

Kaisa Nyberg

the date of receipt and acceptance should be inserted later

Abstract In this paper a new variant of the linear cryptanalysis method for block ciphers is proposed. It is based on the existing method of multidimensional linear cryptanalysis, but offers the option of discarding a whole half-space of linear approximations that do not contribute to statistical nonrandomness of the multidimensional linear cryptanalysis, and keep only the information extracted from an affine subspace for statistical inference. Also the connections of the new affine cryptanalysis with conditional linear cryptanalysis and multiple linear cryptanalysis are described and demonstrated in the context of state-of-the-art ciphers.

Keywords: block cipher, linear cryptanalysis, linear approximation, multidimensional linear cryptanalysis, multiple linear cryptanalysis, conditional linear cryptanalysis

1 Introduction

The linear cryptanalysis method is a statistical method used for distinguishing a block cipher from a random family of permutations and can be extended to key recovery attacks in practical ciphers.

Linear cryptanalysis makes use of nonrandom behavior of linear approximations, which are single-bit values obtained by exclusive-or summation of certain input bits and output bits of the block cipher, or some rounds of it, over a large number of plaintexts. How to find bit combinations that exhibit nonrandom behavior is a problem in itself not discussed in this paper. All known algorithms used for finding good linear approximations are based on Matsui's seminal work [8].

A distinguisher can be built as a test between two hypotheses, where the zero hypothesis is that the data is drawn from a random permutation and the alternate hypothesis is that the data originates from the cipher. The distinguishers used in linear cryptanalysis are based on statistical models of the behavior of correlations of the chosen linear approximations over a random permutation, on the one hand,

Aalto University School of Science, Finland
E-mail: kaisa.nyberg@aalto.fi

and over the cipher, on the other hand. The former is used to estimate the level of the test and the latter to determine the power of the test.

The goal of this paper is to contribute to the methods that use multiple linear approximations simultaneously. There is a long history for such methods developed by various authors starting from Matsui himself. For a short overview, see the introduction of [7]. We propose a new chi square distinguisher using correlations of linear approximations that form an affine subspace in the linear space of all linear approximations. We will demonstrate in examples, how the affine method can be used to ignore linear approximations with low correlations and in this manner make it more efficient than the multidimensional linear cryptanalysis.

Correlations of linear approximations over a block cipher with a fixed key are typically not statistically independent when taken as random variables over the data space. Methods that explicitly measure such dependencies, and use them in statistical analysis, have been presented previously by Murphy in [9] and very recently by Biham and Perle [1]. On the other hand, the main motivation of multidimensional linear cryptanalysis is that the dependencies of linear approximations need not be measured explicitly as they are captured by the multidimensional linear test statistic. In this paper, we will present concrete examples to illustrate how this works in practice.

We start by giving a description of the multidimensional linear cryptanalysis. Next we present the new affine method, and subsequently illustrate it for two examples, one originating from the conditional linear cryptanalysis of the Feistel cipher DES and another one for the SPN cipher PRESENT.

2 Multidimensional Linear Cryptanalysis

In the context of linear cryptanalysis, a linear approximation of a transformation F from $\mathbb{F}_2^n$ to $\mathbb{F}_2^m$ is a Boolean function in $\mathbb{F}_2^n$ defined by two vectors $a \in \mathbb{F}_2^n$ and $b \in \mathbb{F}_2^m$ as follows

$$x \mapsto a \cdot x + b \cdot F(x).$$

In the statistical setting, a linear approximation is considered as a binary random variable X over the given space of transformations with a probability density function defined by

$$\Pr(X = 0) = 2^{-n} \#\{x \in \mathbb{F}_2^n \mid a \cdot x + b \cdot F(x) = 0\}.$$

So we can write $X = a \cdot x + b \cdot F(x)$. In the linear-algebraic setting, a linear approximation $a \cdot x + b \cdot F(x)$ is identified with the vector (a, b), called a mask pair, in the linear space $\mathbb{F}_2^n \times \mathbb{F}_2^m$ over $\mathbb{F}_2$.

Each linear approximation of $F(x)$ is a Boolean function and induces a probability distribution on $\{0, 1\}$. Its bias $\varepsilon_{(a,b)}$ is given by

$$\varepsilon_{(a,b)} = \Pr(a \cdot x + b \cdot F(x) = 0) - 1/2$$

and its correlation $c_{(a,b)}$ by

$$c_{(a,b)} = 2\varepsilon_{(a,b)} = \Pr(a \cdot x + b \cdot F(x) = 0) - \Pr(a \cdot x + b \cdot F(x) = 1).$$

The multidimensional linear cryptanalysis method considers a number of linear approximations that form a linear subspace in $\mathbb{F}_2^n \times \mathbb{F}_2^m$. Let t be the dimension

of this subspace. Then a multidimensional linear approximation is a vector-valued Boolean function from $\mathbb{F}_2^n$ to $\mathbb{F}_2^t$. The components of this vector-valued Boolean function are the linear approximations.

Let λ be a vector-valued Boolean function defined by a multidimensional linear approximation which is composed of linear approximations $a \cdot x + b \cdot F(x)$, where $(a, b) \in V$ and V is a linear subspace of $\mathbb{F}_2^n \times \mathbb{F}_2^m$ of dimension t.

The mapping λ induces a probability distribution in $\mathbb{F}_2^t$ with the following probabilities

$$p_\nu = 2^{-n} \#\{x \,|\, \lambda(x) = \nu\}.$$

On the other hand, let us denote by $\mathrm{cor}(\alpha)$ the correlation of the Boolean function $\alpha \cdot \lambda(x)$, that is,

$$\mathrm{cor}(\alpha) = 2^{-n} \left(\#\{x | \alpha \cdot \lambda(x) = 0\} - \#\{x | \alpha \cdot \lambda(x) = 1\}\right).$$

Then it is well known that

$$p_\nu = 2^{-t} \sum_{\alpha \in \mathbb{F}_2^t} (-1)^{\alpha \cdot \nu} \mathrm{cor}(\alpha), \text{ for all } \nu \in \mathbb{F}_2^t, \tag{1}$$

and equivalently,

$$\mathrm{cor}(\alpha) = \sum_{\nu \in \mathbb{F}_2^t} (-1)^{\alpha \cdot \nu} p_\nu, \text{ for all } \alpha \in \mathbb{F}_2^t. \tag{2}$$

Now we observe that each $\alpha \in \mathbb{F}_2^t$ determines a unique component of λ, that is, a mask pair $(a, b) \in V$, and vice versa. We denote by $\alpha(a, b)$ the mask in $\mathbb{F}_2^t$ such that

$$a \cdot x + b \cdot F(x) = \alpha(a, b) \cdot \lambda(x), \text{ for all } x \in \mathbb{F}_2^n. \tag{3}$$

The correspondence $\alpha : (a, b) \mapsto \alpha(a, b)$ defines a linear map from V to $\mathbb{F}_2^t$.

Applying Equation (1) to masks $\alpha = \alpha(a, b)$ we obtain the following relationship between the probability distribution in $\mathbb{F}_2^t$ and the probability distributions of the linear approximations $a \cdot x + b \cdot F(x)$, $(a, b) \in V$.

$$p_\nu = 2^{-t} \sum_{(a,b) \in V} (-1)^{\alpha(a,b) \cdot \nu} c_{(a,b)}, \text{ for all } \nu \in \mathbb{F}_2^t, \tag{4}$$

and conversely, by Equation (2),

$$c_{(a,b)} = \sum_{\nu \in \mathbb{F}_2^t} (-1)^{\alpha(a,b) \cdot \nu} p_\nu, \text{ for all } (a, b) \in V. \tag{5}$$

The strength of a multidimensional linear approximation is measured by its capacity C_λ given as follows

$$C_\lambda = \sum_{(a,b) \in V, (a,b) \neq 0} c_{(a,b)}^2 = 2^t \sum_{\nu \in \mathbb{F}_2^t} (p_\nu - 2^{-t})^2,$$

where the equality between the two expressions follows directly from Equation (4) or (5).

The multidimensional distinguisher is defined by the following test statistic

$$T(D) = N \sum_{(a,b)\in V,(a,b)\neq 0} \hat{c}_{(a,b)}(D)^2 = N2^t \sum_{\nu\in\mathbb{F}_2^t} \left(\hat{p}_\nu(D) - 2^{-t}\right)^2,$$

where

$$\begin{aligned} D &\quad \text{is a sample of } N \text{ plaintexts } x,\\ \hat{c}_{(a,b)}(D) &= N^{-1}\left(\#\{\, x \in D \,|\, a\cdot x + b\cdot F(x) = 0\} - \#\{\, x \in D \,|\, a\cdot x + b\cdot F(x) = 1\}\right),\\ \hat{p}_\nu(D) &= N^{-1}\#\{\, x \in D \,|\, \lambda(x) = \nu\}. \end{aligned}$$

Clearly, Equations (4) and (5) hold also in the case of observed probabilities and correlations, where x is restricted to D instead of running through the entier $\mathbb{F}_2^n$. Hence the equality of the two expressions of $T(D)$ given above holds.

The observed probabilities $\hat{p}_\nu(D)$ are computed from N independently and randomly drawn x. Then $T(D)$ is a Pearson's chi square test statistic with $2^t - 1$ degrees of freedom. For large N and for uniformly distributed data, $T(D)$ follows a central chi square distribution. In the case, where the sample is drawn from a nonuniform distribution, it was argued in [7] based on [6] that $T(D)$ follows a noncentral chi square distribution with noncentrality parameter NC_λ, where C_λ is the nonzero capacity of the nonuniform distribution.

The multidimensional linear cryptanalysis model has been exploited for various cryptanalytic attacks, including distinguishing attacks and key-recovery attacks of block ciphers. The distributions of the attack statistic for random and cipher are used to estimate the minimum size of the data sample necessary to achieve a desired level and power of the test. In key-recovery cryptanalysis, the level of the test is measured by advantage and the power by success probability [11]. Multidimensional cryptanalysis is shown to work well in many experimental works. One of its major advantages is that no assumption about the independence of the involved linear approximations is needed. On the other hand, to form a linear space it is often required to include many weak linear approximations. The affine multidimensional linear cryptanalysis method proposed in this paper has been developed to allow discarding all linear approximations in a half space, if considered too weak to be useful for the attack.

Let us show an example of a typical situation where the affine method would be useful. More concrete examples will be given later. Assume that a multidimensional linear approximation of a cipher is built around a set of mask pairs (a, b), where a is a fixed nonzero mask on the plaintext and the ciphertext masks b vary within a linear subspace B. The least linear subspace to contain all such masks is $\{0, a\} \times B$. Then the correlations of the linear masks in this set of the form $(0, b)$, $b \in B$ have correlation zero, and do not add to the capacity of the multidimensional linear approximation, but just make the linear approximation space larger. Clearly,

$$\{a, 0\} \times B = (\{a\} \times B) \cup (\{0\} \times B).$$

The affine subspace method presented in this paper allows to discard the useless linear approximations in $\{0\} \times B$ and exploit the useful ones in the affine subspace $\{a\} \times B$.

3 Affine Linear Cryptanalysis

Let λ be a multidimensional linear approximation as described in the previous section. We split it into two components λ_1 from $\mathbb{F}_2^n$ to $\mathbb{F}_2$ and λ_2 from $\mathbb{F}_2^n$ to $\mathbb{F}_2^{t-1}$ such that $\lambda(x) = (\lambda_1(x), \lambda_2(x))$. Then λ_1 is a classical one-dimensional linear approximation, we denote its mask pair by (a_1, b_1), and λ_2 is a multidimensional linear approximation of dimension $s = t - 1$. Then all mask pairs for λ can be written in the form $(a_1 + a_2, b_1 + b_2)$, where (a_2, b_2) is a mask pair for λ_2. We denote by U the linear subspace formed by the linear approximations of λ_2 and by α_2 the linear map from U to $\mathbb{F}_2^s$ defined by Equation (3). Then the map α for λ is given as

$$\begin{aligned} \alpha(a_2, b_2) &= (0, \alpha_2(a_2, b_2)), \text{ and} \\ \alpha(a_1 + a_2, b_1 + b_2) &= (1, \alpha_2(a_2, b_2)), \end{aligned}$$

for all linear masks $(a_2, b_2) \in U$.

Given $\eta \in \mathbb{F}_2^s$, it holds that

$$\Pr(\lambda_2(x) = \eta) = \Pr(\lambda(x) = (1, \eta)) + \Pr(\lambda(x) = (0, \eta)) = p_{(1,\eta)} + p_{(0,\eta)},$$

and analogically for the observed probabilities

$$N^{-1}\#\{\, x \in D \,|\, \lambda_2(x) = \eta\} = \hat{p}_{(1,\eta)}(D) + \hat{p}_{(0,\eta)}(D).$$

Let us apply the multidimensional linear model to λ. We denote the test statistic by $T_\lambda(D)$ which is computed as follows

$$\begin{aligned} T_\lambda(D) &= N2^{s+1} \sum_{b=0,1;\eta\in\mathbb{F}_2^s} (\hat{p}_{b,\eta}(D) - 2^{-s-1})^2 \\ &= N \sum_{(a,b)\in V} \hat{c}_{(a,b)}(D)^2. \end{aligned} \tag{6}$$

Let us then apply the multidimensional model to λ_2 on the subspace U to obtain

$$\begin{aligned} T_{\lambda_2}(D) &= N2^s \sum_{\eta\in\mathbb{F}_2^s} (\hat{p}_{(1,\eta)}(D) + \hat{p}_{(0,\eta)}(D) - 2^{-s})^2 \\ &= N \sum_{(a_2,b_2)\in U} \hat{c}_{(a_2,b_2)}(D)^2. \end{aligned} \tag{7}$$

We now define the affine test statistic $T_{\text{aff}}(D)$ as follows

$$T_{\text{aff}}(D) = T_\lambda(D) - T_{\lambda_2}(D) = N \sum_{(a_2,b_2)\in U} \hat{c}_{(a_1+a_2,b_1+b_2)}(D)^2.$$

Lemma 1 *In the setting defined above, we have*

$$T_{\text{aff}}(D) = N2^s \sum_{\eta\in\mathbb{F}_2^s} (\hat{p}_{(1,\eta)}(D) - \hat{p}_{(0,\eta)}(D))^2 \tag{8}$$

Proof Using Equation (7) we obtain

$$\begin{aligned}
&N2^s \sum_{\eta\in\mathbb{F}_2^s} (\hat{p}_{(1,\eta)}(D) - \hat{p}_{(0,\eta)}(D))^2 + T_{\lambda_2}(D)\\
&= N2^s \sum_{\eta\in\mathbb{F}_2^s} \left((\hat{p}_{(1,\eta)}(D) - 2^{-(s+1)}) - (\hat{p}_{(0,\eta)}(D) - 2^{-(s+1)})\right)^2\\
&+N2^s \sum_{\eta\in\mathbb{F}_2^s} \left((\hat{p}_{(1,\eta)}(D) - 2^{-(s+1)}) + (\hat{p}_{(0,\eta)}(D) - 2^{-(s+1)})\right)^2,
\end{aligned}$$

which simplifies to the expression of $T_\lambda(D)$ given by Equation (6).

To compute $T_{\mathrm{aff}}(D)$ according to Equation (8), the independently and randomly chosen x are distributed to 2^s categories according to the value $\lambda_2(x)$. Further within each category the samples x are divided into two subsets according to their value $\lambda_1(x)$. The resulting value in category η is the difference of the sizes of the two subsets. For a uniform distribution on $\{0,1\}\times\mathbb{F}_2^s$ the expected value of this difference is zero. We propose to use Pearson's chi square test for the values obtained in this way in 2^s categories with the test statistic $T_{\mathrm{aff}}(D)$.

We observe that the 2^{s+1} variables $\hat{p}_{(1,\eta)}(D)+\hat{p}_{(0,\eta)}(D)$, $\eta\in\mathbb{F}_2^s$, and $\hat{p}_{(1,\eta)}(D)-\hat{p}_{(0,\eta)}(D)$, $\eta\in\mathbb{F}_2^s$, uniquely determine the distribution $\hat{p}_{b,\eta}(D)$, $b\in\{0,1\}$, $\eta\in\mathbb{F}_2^s$, which has $2^{s+1}-1$ free variables. Since the variables $\hat{p}_{(1,\eta)}(D)+\hat{p}_{(0,\eta)}(D)$, $\eta\in\mathbb{F}_2^s$, determine the distribution on $\mathbb{F}_2^s$, it has 2^s-1 free variables. It follows that the degree of freedom of the set of variables $\hat{p}_{(1,\eta)}(D)-\hat{p}_{(0,\eta)}(D)$, $\eta\in\mathbb{F}_2^s$, must be at least 2^s, and hence is equal to 2^s.

We conclude that for large N, $T_{\mathrm{aff}}(D)$ is chi square distributed with 2^s degrees of freedom. If $p_{b,\eta}(D)$, $b\in\{0,1\}$, $\eta\in\mathbb{F}_2^s$, has a uniform distribution, we have a central chi square distribution with mean 2^s and variance 2^{s+1}. Otherwise, the mean can be computed from the expression

$$T_{\mathrm{aff}}(D) = T_\lambda(D) - T_{\lambda_2}(D)$$

to get

$$\begin{aligned}
\operatorname{Exp} T_{\mathrm{aff}}(D) &= \operatorname{Exp} T_\lambda(D) - \operatorname{Exp} T_{\lambda_2}(D) = 2^{s+1} - 1 + NC_\lambda - (2^s - 1 + NC_{\lambda_2})\\
&= 2^s + N(C_\lambda - C_{\lambda_2}) = 2^s + N2^s \sum_{\eta\in\mathbb{F}_2^s} (p_{(1,\eta)} - p_{(0,\eta)})^2,
\end{aligned}$$

where the last equality is obtained analogically to the proof of Lemma 1. From this we deduce that the noncentrality parameter of the chi square distribution of $T_{\mathrm{aff}}(D)$ for the nonuniform distribution $p_{b,\eta}$ is equal to $N(C_\lambda - C_{\lambda_2})$, and obtain

$$\operatorname{Var} T_{\mathrm{aff}}(D) = 2\left(2^s + 2N(C_\lambda - C_{\lambda_2})\right).$$

Similarly to the multidimensional linear cryptanalysis the derived affine statistical model can be used in cryptanalytic distinguishing and key-recovery attacks. Next we present examples where the affine linear cryptanalysis can improve upon the multidimensional linear cryptanalysis.

4 Examples

4.1 Example of Biham and Perle

Recently, Eli Biham and Stav Perle proposed a new cryptanalysis method called as conditional linear cryptanalysis [1]. The idea is to consider two mutually dependent linear approximations. For example they found two linear approximations in DES, we denote the random variables related to them by X and Y, with the following probability density functions

$$\Pr(X=0)=\frac{1}{2}+\varepsilon \quad \Pr(X=1)=\frac{1}{2}-\varepsilon$$
$$\Pr(Y=0)=\frac{1}{2} \qquad \Pr(Y=1)=\frac{1}{2}$$

and with the following dependency

$$\Pr(X=0|Y=0)=\frac{1}{2}+2\varepsilon, \quad \Pr(X=0|Y=1)=\frac{1}{2},$$
$$\Pr(X=1|Y=0)=\frac{1}{2}-2\varepsilon, \quad \Pr(X=1|Y=1)=\frac{1}{2}.$$

We use this example to illustrate the behavior of three variants of linear cryptanalysis using these two linear approximations.

The multidimensional linear model. The capacity of the 2-dimensional multidimensional linear approximation defined by the linear approximations X and Y is equal to

$$C_{\text{mult}} = c_X^2 + c_Y^2 + c_{X+Y}^2 \tag{9}$$
$$= 4\left((p_{00}-1/4)^2+(p_{01}-1/4)^2+(p_{10}-1/4)^2+(p_{11}-1/4)^2\right). \tag{10}$$

The linear approximation $X+Y$ has the same bias as X, and the bias of Y is equal to zero. By (9) we get $C_{\text{mult}}=8\varepsilon^2$. The expression on the second line of (9) results in the same value, since $p_{00}=1/4+\varepsilon$, $p_{10}=1/4-\varepsilon$, and $p_{01}=p_{11}=1/4$.

Then the multidimensional test statistic

$$T_{\text{mult}} = N\left(\hat{c}_X(D)^2+\hat{c}_Y^2(D)+\hat{c}_{X+Y}^2(D)\right)$$
$$= 4N\left((\hat{p}_{00}(D)-1/4)^2+(\hat{p}_{01}(D)-1/4)^2+(\hat{p}_{10}(D)-1/4)^2+(\hat{p}_{11}(D)-1/4)^2\right)$$

has a noncentral chi square distribution with 3 degrees of freedom and noncentrality parameter equal to $8N\varepsilon^2$.

The affine linear model. Since $c_Y=0$, it does not contribute to the capacity of the multidimensional distribution. To apply the affine linear model we take U to be the subspace $\{0,Y\}$. Then the affine test statistic

$$T_{\text{aff}} = N\left(\hat{c}_X(D)^2+\hat{c}_{X+Y}^2(D)\right) = 2N\left((\hat{p}_{00}(D)-\hat{p}_{10}(D))^2+(\hat{p}_{01}(D)-\hat{p}_{11}(D))^2\right)$$

has chi square distribution with 2 degrees of freedom and noncentrality parameter

$$N\left(c_X^2+c_{X+Y}^2\right) = 2N\left((p_{00}-p_{10})^2+(p_{01}-p_{11})^2\right) = 8N\varepsilon^2.$$

It means that the affine linear test has the same noncentrality parameter but less degrees of freedom than the multidimensional linear test and hence is more efficient.

The conditional linear model. Since $p_{01} - p_{11} = 0$ in this example, it does not contribute to the capacity of the affine test statistic. All nonbalancedness of the pair (X, Y) of linear approximations is therefore measured by the first term

$$(p_{00} - p_{10})^2 = \Pr(Y = 0)^2(\Pr(X = 0 \,|\, Y = 0) - \Pr(X = 1 \,|\, Y = 0))^2,$$

where the difference $\Pr(X = 0 \,|\, Y = 0) - \Pr(X = 1 \,|\, Y = 0)$ is the correlation of the conditional linear approximation $X|_{Y=0}$. We make the following notation

$$c_{X|_{Y=0}} = \Pr(X = 0 \,|\, Y = 0) - \Pr(X = 1 \,|\, Y = 0).$$

Recently, Biham and Perle proposed in [1] the conditional linear cryptanalysis to exploit high conditional correlations. The idea is to use the analogical statistical model as for classical linear cryptanalysis in the context of conditional probabilities and biases by discarding the data that does not satisfy the condition. According to this model the observed number of data $\hat{N}'$ that satisfy $X = 0$ within a sample of N' plaintext-ciphertext pairs that satisfy $Y = 0$ is binomially distributed with probability $\Pr(X = 0 \,|\, Y = 0) = 1/2 + 2\varepsilon$ and sample size N'. The bias of this conditional distribution is 2ε and the correlation is 4ε. Hence the distribution of the observed correlation

$$2\hat{N}'/N' - 1$$

can be approximated by a normal distribution with mean $c_{X|_{Y=0}} = 4\varepsilon$ and variance $1/N'$.

The data complexity estimate obtained from the normal distribution is the same that can be obtain using the chi square distribution obtained from the squared observed correlation [2]. More precisely, the conditional test statistic T_{cond} defined as

$$T_{\text{cond}} = N'(2\hat{N}'/N' - 1)^2 \sim \chi_1^2(\delta)$$

where

$$\delta = N' c_{X|Y=0}^2 = 16N'\varepsilon^2$$

gives the same data complexity estimate as the binomial (normal) test statistic $\hat{N}'/N'$ traditionally used in linear cryptanalysis. Since $\Pr(Y = 0)$, it is estimated that about half of the data is discarded meaning that for the total size N of the required data it holds that $N = 2N'$.

Comparison. When distinguishing the behavior of the cipher from random, the distribution of the test statistic in the random case is assumed to have a central chi square distribution with the same number of degrees of freedom. Hence in each of the three cases discussed above we distinguish between two chi square distributions

$$\chi_k^2 \text{ and } \chi_k^2(8N\varepsilon^2), \tag{11}$$

where $k = 1, 2,$ or 3 for conditional, affine, or multidimensional linear attacks, respectively. The means of the two distributions (11) are k and $k + 8N\varepsilon^2$ and the variances are $2k$ and $2(k + 16N\varepsilon^2)$. In each case, the efficiency of the attack depends on the difference between the means, on the one hand, and on the variances on the other hand: the larger is the difference between the means, or the smaller are the variances, the more efficient is the distinguisher. We can see that in all three cases, the difference between the means is the same, while the variances are the smallest for $k = 1$.

We conclude that from the three statistical models considered in the context of the given example, the conditional linear model of Biham and Perle gives the most efficent statistical distinguisher. The next is the affine linear model, and the least efficient is the distinguisher based on the multidimensional linear approximation.

Multiple conditional linear cryptanalysis (conjecture). In the example above, only one linear approximation is used to condition the observations of the linear approximation X. It means that the image space of λ_2 in the affine linear model has dimension equal to one. We propose to consider a generalisation of the conditional linear cryptanalysis and use multiple values η in the value space $\mathbb{F}_2^s$ of the s-dimensional linear approximation λ_2 as conditions on one linear approximation λ_1. Recall the affine test statistic

$$T_{\text{aff}}(D) = N2^s \sum_{\eta \in \mathbb{F}_2^s} (\hat{p}_{(1,\eta)}(D) - \hat{p}_{(0,\eta)}(D))^2.$$

If for some categories $\eta \in \mathbb{F}_2^s$ the scores $\hat{p}_{(1,\eta)}(D) - \hat{p}_{(0,\eta)}(D)$ are expected to be close to zero, let us discard the data falling in those categories. In other words, we drop the corresponding terms from the test statistic. If N is the total number of data and k values $\eta_1, \ldots, \eta_k$ from $\mathbb{F}_2^s$ remain to be used as conditions, we conjecture that the sum

$$N2^s \sum_{i=1}^{k} (\hat{p}_{1,\eta_i}(D) - \hat{p}_{0,\eta_i}(D))^2,$$

is chi square distributed with k degrees of freedom and noncentrality parameter

$$N2^s \sum_{i=1}^{k} (p_{1,\eta_i} - p_{0,\eta_i})^2.$$

4.2 Dependent linear approximations of PRESENT

The block cipher PRESENT is an iterated SPN cipher. Each round consists of round-key addition, a layer of sixteen parallel 4×4 S-boxes, and a bit permutation. For many years, the best linear attack on PRESENT was a multidimensional attack by Cho [5]. He used the strong linear approximations previously identified by Ohkuma [10]. Cho constructed nine multidimensional linear approximations each composed of all linear masks over the 4-bit input space of an S-box and over an output space of another 4-bit S-box after 24 rounds of encryption. He argued that the nine multidimensional approximations of dimension 8 are independent and defined the test statistic as a sum of the nine multidimensional chi square test statistics. This test statistic has $9(2^8 - 1)$ degrees of freedom.

Recently, Bogdanov, Tischauser and Vejre, proposed the multivariate linear cryptanalysis method which, among other things, allows reducing the degree of freedom and hence improve efficiency [3]. The aim of the multivariate linear cryptanalysis is to take the dependency on the key into account. Since our aim is to illustrate the advantages of the affine linear cryptanalysis we omit consideration of key dependency and consider the test statistic as a function of the data only similarly as in [5] and [7].

The 64-bit input or output data blocks of S-box layers of PRESENT, as well as their linear masks, are often identified with integers written as

$$2^{4k+\ell}$$

where $k = 0, 1, \ldots, 15$ corresponds to the S-box and $\ell = 0, 1, 2, 3$ corresponds to the bit position within one S-box. The linear approximations used in [3] were taken to cover 22 or 23 S-box layers so that the input masks a operate on the input data block to the S-box layer and the output masks b operate on the output data block from the S-box layer. The masks are as follows:

$$\begin{array}{ll} \text{input masks} & a_k = 2^{4k+3},\ k = 5,6,7,9,10,11,13,14,15 \\ \text{output masks} & b_{ij} = \begin{cases} j \cdot 2^{4i+2}, & i = 5,6,7,9,10,11,\ j = 1,2,3 \\ 2^{4i+3}, & i = 13,14,15. \end{cases} \end{array}$$

For each fixed $i = 5, 6, 7, 9, 10, 11$, the three output masks b_{ij}, $j = 1, 2, 3$,

$$4 \cdot 2^{4i},\ 8 \cdot 2^{4i},\ 12 \cdot 2^{4i} \tag{12}$$

are linearly dependent, since any single one of them is an exclusive or of the other two. Due to such linear dependencies, there are statistical dependencies between linear approximations (a_k, b_{ij}) and hence also between their observed correlations, which we denote by $\hat{c}_{(a_k,b_{ij})}(D)$. To handle the dependencies Bogdanov, et al., used the result of Murphy [9] that the observed correlations follow a multivariate normal distribution.

As we saw in the previous example, dependencies between correlations of different linear approximations can also be handled using the affine linear cryptanalysis.

In the case of the current example, we find affine subspaces for each fixed k and $i = 5, 6, 7, 9, 10, 11$ by taking the four mask pairs (a_k, b) where

$$b = 0,\ 4 \cdot 2^{4i},\ 8 \cdot 2^{4i},\ \text{ or } 12 \cdot 2^{4i},$$

that is $b = b_{ij}$, $j = 0, 1, 2, 3$. They form an affine subspace of dimension 2. Then for each input mask a_k and output S-box with index $i = 5, 6, 7, 9, 10, 11$ the statistic

$$T(k,i) = \sum_{j=0,4,8,12} \hat{c}_{(a_k, j \cdot 2^{4i+2})}(D)^2$$

is chi square distributed with 4 degrees of freedom.

Similarly as Cho [5] we can argue that the linear approximations involving different active S-boxes on the first and last round are independent. We form the sum of all $9 \cdot 6$ of such chi square variables with 4 degrees of freedom together with the remaining $9 \cdot 3$ chi-square variables with 1 degree of freedom.

We note that the mask pair with $b = 0$ must be included, which is not optimal, since observed correlations for mask pairs of the form $(a_k, 0)$ have mean zero. By

studying the expected values of the correlations of various mask pairs, a more optimal affine linear subspace might possibly be found.

In a subsequent paper [4], Bogdanov, et al., presented another version of their attack, where the linear approximations form a linearly independent set by removing one mask $b = 12 \cdot 2^{4i}$ from the set (12). Even in such case, linear approximations considered as random variables over the sample space that share the same active S-boxes on the first and the last round are in general statistically dependent. To overcome this problem, an artificial assumption of independence was made in [4]. Next we show that using the affine linear method, such assumption is not needed.

If we take, for each fixed k and $i = 5, 6, 7, 9, 10, 11$, the two mask pairs (a_k, b) obtained by setting

$$b = 4 \cdot 2^{4i}, \text{ or } 8 \cdot 2^{4i}, \tag{13}$$

we get, using the affine linear method, that the sum of the observed squared correlations of these two linear approximations is chi square distributed with 2 degrees of freedom. Note that no assumption about statistical independence of the observed correlations is required.

This example also serves to demonstrate the advantages of the affine method over the multidimensional linear method. The minimum dimension of a linear subspace that contains the two approximations (13) is two, which leads to a chi square variable with three degrees of freedom. The affine method allows to remove the one-dimensional subspace of masks of the form $(0, b)$, which has zero capacity. Hence the one-dimensional affine distinguisher preserves the whole capacity of the two-dimensional distinguisher.

5 Conclusions

A new chi square statistic in the context of linear cryptanalysis was presented and shown to offer a clear advantage over the known multidimensional linear chi square method. The advantage is achieved by removing a linear subspace of weak linear approximations. The noncentrality parameter of the distribution of the affine test statistic can be expressed in terms of a difference between capacities of two multidimensional linear approximations. Hence this parameter can be estimated using the existing methods for estimating capacity of multidimensional linear approximation.

Also the relation between the new conditional linear cryptanalysis, on the one hand, and the multidimensional and affine linear cryptanalysis, on the other hand, was discussed and illustrated by an example from linear cryptanalysis of the DES. We showed that for this example the conditional linear cryptanalysis yields the most efficient test statistic. The difference is only in the degrees of freedom of the chi square test variable.

Encouraged by the good performance of conditional linear cryptanalysis we also sketched without proofs how to build a chi square statistic using one linear approximation under multiple conditions.

Acknowledgements

I wish to thank Eli Biham for discussions related to conditional linear cryptanalysis and Céline Blondeau for suggestions how to improve the presentation. Also the comments by anonymous reviewers are gratefully acknowledged.

References

1. Eli Biham and Stav Perle. Conditional linear cryptanalysis. Presentation at Romanian Cryptology Days, Bucharest, Romania, Sept 18-20,2017, 2017.
2. Céline Blondeau and Kaisa Nyberg. Improved parameter estimates for correlation and capacity deviates in linear cryptanalysis. *IACR Transactions on Symmetric Cryptology*, 2016(2):162–191, 2017.
3. Andrey Bogdanov, Elmar Tischhauser, and Philip S. Vejre. Multivariate Linear Cryptanalysis: The Past and Future of PRESENT. *IACR Cryptology ePrint Archive*, 2016:667, 5 Jul 2016.
4. Andrey Bogdanov, Elmar Tischhauser, and Philip S. Vejre. Multivariate profiling of hulls for linear cryptanalysis. *IACR Transactions on Symmetric Cryptology*, 2018(1):101–125, 2018.
5. Joo Yeon Cho. Linear Cryptanalysis of Reduced-Round PRESENT. In Josef Pieprzyk, editor, *Topics in Cryptology - CT-RSA 2010, The Cryptographers' Track at the RSA Conference 2010, San Francisco, CA, USA, March 1-5, 2010. Proceedings*, volume 5985 of *Lecture Notes in Computer Science*, pages 302–317. Springer, 2010.
6. F.C. Drost, W.C.M. Kallenberg, D.S.Moore, and J.Oosterhoff. Power Approximations to Multinomial Tests of Fit. *Journal of the American Statistican Association*, 84(405):130–141, Mar 1989.
7. Miia Hermelin, Joo Yeon Cho, and Kaisa Nyberg. Multidimensional Extension of Matsui's Algorithm 2. In Orr Dunkelman, editor, *Fast Software Encryption, 16th International Workshop, FSE 2009, Leuven, Belgium, February 22-25, 2009*, volume 5665 of *Lecture Notes in Computer Science*, pages 209–227. Springer, 2009.
8. Mitsuru Matsui. The First Experimental Cryptanalysis of the Data Encryption Standard. In Yvo Desmedt, editor, *Advances in Cryptology - CRYPTO '94, 14th Annual International Cryptology Conference, Santa Barbara, California, USA, August 21-25, 1994*, volume 839 of *Lecture Notes in Computer Science*, pages 1–11. Springer, 1994.
9. Sean Murphy. The independence of linear approximations in symmetric cryptanalysis. *IEEE Transactions on Information Theory*, 52(12):5510–5518, 2006.
10. Kenji Ohkuma. Weak keys of reduced-round present for linear cryptanalysis. In Michael Jacobson Jr, Vincent Rijmen, and Rei Safavi-Naini, editors, *Selected Areas in Cryptography, 16th Annual International Workshop, SAC 2009, Calgary, Alberta, Canada, August 13-14, 2009. Revised Selected Papers*, volume 5867 of *Lecture Notes in Computer Science*, pages 249–265. Springer, 2009.
11. Ali Aydin Selçuk. On Probability of Success in Linear and Differential Cryptanalysis. *J. Cryptology*, 21(1):131–147, 2008.