Linear Cryptanalysis Using Multiple Linear Approximations

Miia HERMELIN [a], Kaisa NYBERG [b]
[a] *Finnish Defence Forces*
[b] *Aalto University School of Science and Nokia*

Abstract. In this article, the theory of multidimensional linear attacks on block ciphers is developed and the basic attack algorithms and their complexity estimates are presented. As an application the multidimensional linear distinguisher derived by Cho for the block cipher PRESENT is discussed in detail.

Keywords. linear cryptanalysis, multidimensional cryptanalysis, Matsui's algorithm, block cipher, key recovery, linear hull effect, PRESENT

Introduction

A natural idea for enhancing linear cryptanalysis is using multiple approximations instead of one. Matsui was the first to suggest this enhancement. In 1994 he proposed to using two approximations simultaneously [20]. In the same year, Kaliski and Robshaw used several approximations in an attempt to reduce the data complexities of Matsui's algorithms [17]. As a different approach, Johansson and Maximov presented an idea of a multidimensional distinguishing attack against the stream cipher Scream [16].

Biryukov, et al., [2] used multiple approximations for finding several bits of the secret key with reduced data complexity in 2004. However, the theoretical foundations of the methods by Kaliski and Robshaw and Biryukov, et al., both depend on assumptions about the statistical properties of the one-dimensional linear approximations. In particular, they assumed that the one-dimensional linear approximations are statistically independent. Murphy pointed out that the assumption may not hold in a general case [21].

Baignères, et al., presented in 2004 a linear distinguisher that does not suffer from this limitation [1]. The distinguisher has also another advantage over the previous approaches: it is based on a well established statistical theory of log-likelihood ratio (LLR), but remained theoretical without an efficient way for determining the probability distribution that is needed in their method. Englund and Maximov presented computational methods for determining the distribution directly [9], but they are in general not feasible for handling distributions of larger than 32-bit values.

We showed in [13] how one-dimensional approximations can be used for efficient construction of the multidimensional approximation and its probability dis-

tribution. This method does not rely on the assumption about statistical independence of the one-dimensional approximations. We then considered a multidimensional Matsui's Alg. 1 in [10]. We showed that it is indeed advantageous to use multiple approximations instead of just one. Moreover, our method gives a more accurate estimate of the data complexity, which is always smaller than estimated by Biryukov, et al., where only linearly independent approximations are used.

We considered in [11] using the LLR for determining the key in Alg. 1. We proposed another Alg. 1. method, called the convolution method in [14]. We also gave a proper statistical framework for Alg. 1 and showed how different methods can be compared. We showed that under certain conditions, which hold for practical ciphers, the convolution method and the other Alg. 1 methods we considered in [10] and [11] have the same data complexities. The empirical tests done on Serpent verified the theoretical results. Since the convolution method had the smallest time complexity, we conclude that it is the most efficient of these methods in practice. In this paper, we give a short "cookbook" description of the LLR and convolution methods. For details of the statistical analysis, we refer to our previous work, especially [14].

In [12] we considered extending Matsui's Alg. 2 to multiple dimensions. We considered two methods based on different test statistics: the LLR-method and χ^2-method. We describe the use of these methods in this paper. Selçuk presented the concept of advantage for measuring the efficiency of one-dimensional Alg. 2 [24]. We extended the theory to multiple dimensions to compare the different methods and showed that the LLR-method is more efficient than the χ^2-method. We proposed applying the convolution method for Alg. 2 in [15]. We made no practical experiments, but based on the results with Alg. 1 and the theoretical calculations we claim that the convolution method has smaller time complexity than the other methods and the same data complexity.

Usually the Piling up lemma [19] is used for combining the correlations over several rounds of a block cipher. In some cases, such as DES and SERPENT, this approach gives good estimates, as there is only one linear trail with a non-negligible correlation and fixed input and output masks through the cipher. In multiple dimensions it means that the expected probability distribution of the approximation is approximately the same for all keys. It is then possible to use Alg. 1.

Daemen [7] and Nyberg [22] noted that, in the general case, several approximation trails exist. This makes Alg. 1 impossible, since there is no key independent expectation of the correlation. However, as noted by Nyberg [22], all these linear trails contribute to the magnitude of the correlation and therefore distinguishing distributions from uniform exploited in Alg. 2 is still possible. Cho studied a practical application of both multidimensional method and linear hull effect on block cipher PRESENT [4]. We study the theory behind his attack and show how the linear hull effect makes the attack more efficient.

The structure of this paper is as follows: In Section 1 we introduce the necessary mathematical background and notation. Some statistical concepts are introduced in Section 2. We consider the problem of determining the correlation of the one-dimensional approximations and the linear hull effect in Section 3. Section 4 then shows how the multidimensional linear approximation is constructed.

The realisations of multidimensional Alg. 1 and Alg. 2 are studied in Sections 5 and 6, respectively. We do not go into statistical details, rather we just show how the different methods are used. We also discuss the suitability of the methods in different situations. In Section 7 we present and discuss Cho's attack on the block cipher PRESENT as an application of the multidimensional method and the linear hull effect.

1. Probability Distributions and Boolean Functions

The linear space of n-dimensional binary vectors is denoted by $\mathbb{Z}_2^n$. The sum modulo 2 is denoted by $\oplus$. The inner product for $a = (a^1, \ldots, a^n), b = (b^1, \ldots, b^n) \in \mathbb{Z}_2^n$ is defined as $a \cdot b = a^1 b^1 \oplus \cdots \oplus a^n b^n$. Then the vector a is called the (linear) mask of b. The binary vector $a \in \mathbb{Z}_2^n$ can be identified with a unique integer $b \in \{0, 1, \ldots, 2^n - 1\}$ such that

$$b = \sum_{i=1}^{n} a^i 2^{i-1}.$$

A Galois field $\mathrm{GF}(2^n)$ is a linear space. Hence, depending on the context, a is used interchangeably to notate a binary vector, the corresponding integer and an element of the finite field.

A function $f : \mathbb{Z}_2^n \mapsto \mathbb{Z}_2$ is called a Boolean function. A linear Boolean function is a mapping $x \mapsto u \cdot x$. The correlation between a Boolean function $f : \mathbb{Z}_2^n \mapsto \mathbb{Z}_2$ and zero is

$$c(f) = 2^{-n} \left(\#\{x \in \mathbb{Z}_2^n : f(x) = 0\} - \#\{x \in \mathbb{Z}_2^n : f(x) \neq 0\} \right)$$

and it is also called the correlation of f.

We denote random variables by capital boldface letters $\mathbf{X}, \mathbf{Y}, \ldots$, their domains by $\mathcal{X}, \mathcal{Y}, \ldots$ and their realisations by low case letters $x \in \mathcal{X}, y \in \mathcal{Y}, \ldots$. Let $\mathbf{X}$ be a random variable taking on values in $\mathcal{X} = \{0, 1, \ldots, M\}$. We call the vector $p = (p_0, \ldots, p_M)$ that satisfies $\Pr(\mathbf{X} = \eta) = p_\eta$, for all $\eta \in \mathcal{X}$ the discrete probability distribution (p.d.) of $\mathbf{X}$. We denote the uniform p.d. by θ.

A function $f : \mathbb{Z}_2^n \mapsto \mathbb{Z}_2^m$ with $f = (f_1, \ldots, f_m)$, where f_i are Boolean functions, is called a vector Boolean function of dimension m. A linear Boolean function from $\mathbb{Z}_2^n$ to $\mathbb{Z}_2^m$ is represented by an $m \times n$ binary matrix U. The m rows of U are denoted by $u_1, \ldots, u_m$, where each u_i is a linear mask. Let $f : \mathbb{Z}_2^n \mapsto \mathbb{Z}_2^m$ and $\mathbf{X}$ be uniformly distributed with $\mathcal{X} = \{0, 1, \ldots, 2^n - 1\}$. If $\mathbf{Y} = f(\mathbf{X})$, then the p.d. of $\mathbf{Y}$ is called the p.d. of f.

A proof of the following lemma can be found in [13]:

Lemma 1.1. *Let $f : \mathbb{Z}_2^n \mapsto \mathbb{Z}_2^m$ be a Boolean function with p.d. p. Then*

$$p_\eta = 2^{-m} \sum_{a \in \mathbb{Z}_2^m} (-1)^{a \cdot \eta} c(a \cdot f), \quad \textit{for all} \quad \eta \in \mathbb{Z}_2^m.$$

This result is also known as the Cramér-Wold theorem [6] that states that the p.d. is uniquely determined by its Fourier-Stieltjes transforms, i.e, the correlations $c(a \cdot f)$, $a \in \mathbb{Z}_2^m$.

Let $p = (p_0, \ldots, p_M)$ and $q = (q_0, \ldots, q_M)$ be some discrete p.d.'s of random variables with domain $\mathcal{X} = \{0, 1, \ldots, M\}$. The Kullback-Leibler distance between p and q is defined as follows:

Definition 1.2. The *relative entropy* or *Kullback-Leibler distance* between p and q is

$$\mathrm{D}\left(p \parallel q\right) = \sum_{\eta=0}^{M} p_\eta \log \frac{p_\eta}{q_\eta}, \tag{1}$$

with the conventions $0 \log 0/b = 0$, if $b \neq 0$, and $b \log b/0 = \infty$.

We say that p is *is close to* q, if there exists $\epsilon, 0 < \epsilon < 1/2$, such that

$$|p_\eta - q_\eta| \leq \epsilon q_\eta, \quad \text{for all} \quad \eta \in \mathcal{X}. \tag{2}$$

If p is close to q, their Kullback-Leibler distance can be approximated using Taylor series [1] such that

$$\mathrm{D}\left(p \parallel q\right) = C(p, q)/2 + \mathcal{O}(\epsilon^3),$$

where ϵ is the parameter in (2) and the capacity $C(p, q)$ of p and q is defined as follows:

Definition 1.3. The *capacity* between two p.d.'s p and q is

$$C(p, q) = \sum_{\eta=0}^{M} (p_\eta - q_\eta)^2 q_\eta^{-1}.$$

If q is the uniform distribution, then $C(p, q)$ is denoted by $C(p)$ and called the capacity of p.

From Lemma 1.1 we have the following result for any $f : \mathbb{Z}_2^n \mapsto \mathbb{Z}_2^m$ and its p.d. p [13]:

$$C(p) = \sum_{a=1}^{2^m - 1} c(a \cdot f)^2. \tag{3}$$

2. Statistics

Consider a sequence of N independent and identically distributed (i.i.d.) random variables with domain $\mathcal{X} = \{0, 1, \ldots, M\}$. Let the realisation of the random sample be $x_1, \ldots, x_N$. We compute the empirical distribution q of the sample by

$$q_\eta = N^{-1}\#\{t = 1, \ldots, N : x_t = \eta\},\ \eta \in \mathcal{X}. \tag{4}$$

Given two p.d.'s p and p', $p' \neq p$, assume we wish to decide whether the given data is drawn from p or p'. We can solve this problem by using a suitable test statistic. The optimal test statistic minimises the error of choosing the wrong p.d. with given amount of data N and it is given by the log-likelihood ratio (LLR) defined as

$$\mathrm{LLR}(x_1, \ldots, x_N; p, p') = \mathrm{LLR}(q; p, p') = \sum_{\eta \in \mathcal{X}} N q_\eta \log \frac{p_\eta}{p'_\eta}. \tag{5}$$

We decide p (p') if $\mathrm{LLR}(q; p, p') > 0$ (< 0).

In a goodness-of-fit problem we wish to decide whether the data $x_1, \ldots, x_N$ is drawn from one given p.d. p or not. A basic tool for solving this type of problem is given by the χ^2 goodness-of-fit test. Using q, the χ^2-test statistic is calculated by

$$\chi^2(q; p) = N \sum_{\eta \in \mathcal{X}} \frac{(q_\eta - p_\eta)^2}{p_\eta}. \tag{6}$$

Large values of $\chi^2(q; p)$ imply that the sample is not drawn from p. We decide p (not p) if $\chi^2(q) \leq \tau$ ($\chi^2(q) \geq \tau$), where τ is a threshold that depends on the probability of rejecting p when it is the right p.d.

LLR is used if accurate estimates of both p and p', $p' \neq p$ are available. If only p can be estimated accurately and the only information available about p' is that it is different from p, then χ^2 must be used.

The multidimensional linear cryptanalysis method discussed in this article is based on probability distributions related to linear approximations, and constructed in practice from correlations according to Lemma 1.1. Similarly as in classical (one-dimensional) linear cryptanalysis, obtaining accurate information about the expected correlation is essential and will be studied next.

3. Estimating Correlation of Linear Approximation of Block Cipher

For the purposes of linear cryptanalysis a block cipher is considered as a vector Boolean function

$$f : \mathbb{Z}_2^n \times \mathbb{Z}_2^\ell \to \mathbb{Z}_2^n \times \mathbb{Z}_2^\ell \times \mathbb{Z}_2^n,\ f(x, K) = (x, K, \mathcal{E}_K(x)),$$

where $\mathcal{E}_K(x)$ is the block cipher encryption of plaintext $x \in \mathbb{Z}_2^n$ with key $K \in \mathbb{Z}_2^\ell$. A linear approximation of a block cipher with mask $(u, v, w) \in \mathbb{Z}_2^{2n+\ell}$ is a Boolean function defined as

$$(x, K) \mapsto u \cdot x \oplus v \cdot K \oplus w \cdot \mathcal{E}_K(x). \tag{7}$$

The most difficult task in linear cryptanalysis is finding linear approximations with correlation of large absolute value, and in particular, determining an ade-

quate estimate of the correlation. Let us now assume that the block cipher is an iterated block cipher with round function $G(x, K_i)$ where x is the data input and K_i is the key input to the round. With a fixed key K the iterated block cipher is a composition of a number, say R, of round functions. Then the correlation $c(u \cdot x \oplus w \cdot \mathcal{E}_K(x))$ can be calculated as

$$c(u \cdot x \oplus w \cdot \mathcal{E}_K(x)) = \sum_{\substack{u_2, \ldots, u_R \\ u_1 = u,\, u_{R+1} = w}} \prod_{i=1}^{R} c(u_i \cdot x \oplus u_{i+1} \cdot G(x, K_i)).$$

The sequences $u_1 = u, u_2, \ldots, u_R, u_{R+1} = w$, over which the summation is taken, are called linear (approximation) trails from u to w and the product of the round correlations $c(u_i \cdot x \oplus u_{i+1} \cdot G(x, K_i))$, $i = 1, \ldots, R$, is called the trail-correlation of the trail. The goal of classical linear cryptanalysis, as first proposed by Matsui [19], is to find masks u and w such that for almost all keys K this correlation is large in absolute value. In general, the situation is difficult to handle, but something more can be said in the case of key-alternating block ciphers, for which the round function is of the form $G(x, K_i) = g(x \oplus K_i)$. Then

$$c(u_i \cdot x \oplus u_{i+1} \cdot G(x, K_i)) = (-1)^{u_i \cdot K_i} c(u_i \cdot x \oplus u_{i+1} \cdot g(x)),$$

that is, only the sign of the correlation over the round function depends on the key, and we have proved the following theorem, which we call the Correlation theorem.

Theorem 3.1. *([8], [23]) Let g be the round function of an R-round key-alternating iterated block cipher $\mathcal{E}_K$ with round keys $(K_1, K_2, \ldots, K_R)$. Then for any $u \in \mathbb{Z}_2^n$ and $w \in \mathbb{Z}_2^n$ it holds that*

$$c(u \cdot x \oplus w \cdot \mathcal{E}_K(x)) = \sum_{\substack{u_2, \ldots, u_R \\ u_1 = u,\, u_{R+1} = w}} (-1)^{u_1 \cdot K_1 \oplus \cdots \oplus u_R \cdot K_R} \prod_{i=1}^{R} c(u_i \cdot x \oplus u_{i+1} \cdot g(x)).$$

The goal of Alg. 1 is to determine the bit $v \cdot K$ of information of the key K based on the sign of the observed correlation $c(u \cdot x \oplus w \cdot \mathcal{E}_K(x))$. This will succeed under two conditions. First, the observed correlation $c(u \cdot x \oplus w \cdot \mathcal{E}_K(x))$ for the fixed unknown key K must be large, and secondly a good theoretical estimate of the sign of the correlation $c(u \cdot x \oplus v \cdot K \oplus w \cdot \mathcal{E}_K(x))$ must be available. Let us now investigate the average behaviour of the latter correlation. Similarly, as in Theorem 3.1 we first write this correlation as follows:

$$\begin{aligned} & c(u \cdot x \oplus v \cdot K \oplus w \cdot \mathcal{E}_K(x)) \\ = & \sum_{\substack{u_2, \ldots, u_R \\ u_1 = u,\, u_{R+1} = w}} (-1)^{u_1 \cdot K_1 \oplus \cdots \oplus u_R \cdot K_R \oplus v \cdot K} \prod_{i=1}^{R} c(u_i \cdot x \oplus u_{i+1} \cdot g(x)). \end{aligned} \tag{8}$$

By averaging over the keys we get

$$
\begin{aligned}
&\mathrm{E}_{\mathbf{K}}\left[c(u\cdot x\oplus v\cdot\mathbf{K}\oplus w\cdot\mathcal{E}_{\mathbf{K}}(x))\right]\\
&=\sum_{\substack{u_2,\ldots,u_R\\ u_1=u,\,u_{R+1}=w}} 2^{-\ell}\sum_{K}(-1)^{u_1\cdot K_1\oplus\cdots\oplus u_R\cdot K_R\oplus v\cdot K}\prod_{i=1}^{R}c(u_i\cdot x\oplus u_{i+1}\cdot g(x))\\
&=\sum_{\substack{u_2,\ldots,u_R\\ u_1=u,\,u_{R+1}=w}} c(u_1\cdot K_1\oplus\cdots\oplus u_R\cdot K_R\oplus v\cdot K)\prod_{i=1}^{R}c(u_i\cdot x\oplus u_{i+1}\cdot g(x)).
\end{aligned}
$$

The first correlations are related to the key scheduling function. It is a vector Boolean function, which, given key K as input, outputs the round keys $K_1,\ldots,K_R$. If the key scheduling is a linear function, these correlations take only on values 1 or 0 depending on whether $u_1\cdot K_1\oplus\cdots\oplus u_R\cdot K_R\oplus v\cdot K$ is equal to zero for all keys, or not. Under the assumption that the round keys are independent, that is, if $K=(K_1,\ldots,K_R)$, only one linear trail $v=(u_1,\ldots,u_R)$ remains, and we have

$$
\mathrm{E}_{\mathbf{K}}\left[c(u\cdot x\oplus v\cdot\mathbf{K}\oplus w\cdot\mathcal{E}_{\mathbf{K}}(x))\right]=\prod_{i=1}^{R}c(u_i\cdot x\oplus u_{i+1}\cdot g(x)),
$$

where as before, $u_1=u$ and $u_{R+1}=w$. The right hand side of the equation is the trail-correlation of v, which in this manner can be represented as the average correlation of the linear approximation (7) taken over the keys. The trail-correlation is commonly used as an estimate of the correlations $c(u\cdot x\oplus v\cdot K\oplus w\cdot\mathcal{E}_K(x))$ and was justified by the Piling up lemma by Matsui [19].

How good is this estimate in general? From (8) we see that it can be very inaccurate if the linear approximation is composed of more than one linear approximation trails with large correlations (in absolute value). To illustrate this phenomenon let us borrow an example from [8]. In this example, it is assumed that the correlation expressed in Theorem 3.1 takes the form $c(u\cdot x\oplus w\cdot\mathcal{E}_K(x))=(-1)^{\gamma\cdot K}c_\gamma+(-1)^{\lambda\cdot K}c_\lambda$ where c_γ and c_λ are the correlations of the linear trails γ and λ, and $c_\gamma\approx c_\lambda$. Let γ be the trail selected to be used in the analysis. Then $\mathrm{E}_{\mathbf{K}}\left[c(u\cdot x\oplus\gamma\cdot\mathbf{K}\oplus w\cdot\mathcal{E}_{\mathbf{K}}(x))\right]=\mathrm{E}_{\mathbf{K}}\left[c_\gamma+(-1)^{(\lambda\oplus\gamma)\cdot\mathbf{K}}c_\lambda\right]=c_\gamma$. But this gives a useful estimate only for a half of the keys. Those are the keys K for which $(\lambda\oplus\gamma)\cdot K=0$. For such keys, $c(u\cdot x\oplus\gamma\cdot K\oplus w\cdot\mathcal{E}_K(x))=c_\gamma+c_\lambda$, and the sign of the correlation will be predicted correctly . For the other half of the keys we have $c(u\cdot x\oplus\gamma\cdot K\oplus w\cdot\mathcal{E}_K(x))=c_\gamma-c_\lambda\approx 0$, and hence no adequate estimate of the sign of the correlation can be achieved.

Due to the ambiguity of the trail correlations described above, Alg. 1 is applied only for ciphers that admit correlations with a single dominant trail $v=(u_1,\ldots,u_R)$. Then the correlation of the linear trail v is a valid estimate of its true correlation, whatever key has been used in encryption. More precisely,

$$
\begin{aligned}
\mathrm{E}_{\mathbf{K}}\left[c(u\cdot x\oplus v\cdot\mathbf{K}\oplus w\cdot\mathcal{E}_{\mathbf{K}}(x))\right]&=\prod_{i=1}^{R}c(u_i\cdot x\oplus u_{i+1}\cdot g(x))\\
&\approx c(u\cdot x\oplus v\cdot K\oplus w\cdot\mathcal{E}_K(x)),
\end{aligned}
$$

for all keys K.

When applying Alg. 2 for block ciphers only the quantity of the correlation matters, not the sign. Therefore often, and in particular in the context of multidimensional linear cryptanalysis of a block cipher, the strength of the linear approximation (7) is evaluated in terms of the squared correlation. It was observed in [22] that the average value of the square of this correlation taken over the keys is under some conditions independent of the linear trail v and, moreover, the correlations of all linear trails contribute to the expected value. This result, often called as the "linear hull effect" is given in the next theorem.

Theorem 3.2. *([8], [22], [23]) Let g be the round function of an R-round key-alternating iterated block cipher $\mathcal{E}_K$ with key $K = (K_1, K_2, \ldots, K_R)$. Then for any $u \in \mathbb{Z}_2^n$, $v \in \mathbb{Z}_2^{Rn}$ and $w \in \mathbb{Z}_2^n$ it holds that*

$$\begin{aligned}\mathrm{E}_{\mathbf{K}}\left[c(u \cdot x \oplus v \cdot \mathbf{K} \oplus w \cdot \mathcal{E}_{\mathbf{K}}(x))^2\right] &= \mathrm{E}_{\mathbf{K}}\left[c(u \cdot x \oplus w \cdot \mathcal{E}_{\mathbf{K}}(x))^2\right] \\ &= \sum_{\substack{u_2, \ldots, u_R \\ u_1 = u,\, u_{R+1} = w}} \prod_{i=1}^{R} c(u_i \cdot x \oplus u_{i+1} \cdot g(x))^2.\end{aligned}$$

In practice, an estimate of the average squared correlation can be calculated by first finding as many trails from u to w with non-zero correlation as possible and then summing up their squared trail-correlations. This value gives a lower bound of the squared correlation $c(u \cdot x \oplus w \cdot \mathcal{E}_K(x))^2$ that one is expected to observe from the data on the average over the keys. As noted above, the more about equally strong linear trails are present the more these correlations vary with the key K used in $\mathcal{E}_K(x)$. If the data requirement is based on the average squared correlation, the distinguishing step of Alg. 2 is likely to fail for a key K which results in a smaller (squared) correlation than the average. On the other hand, a large proportion of the keys will have correlations greater than the average, and the distinguishing phase of Alg. 2 will succeed. Since linear cryptanalysis uses estimates based on average behaviour, the performance of any practical attack designed on a specific cipher should be verified empirically, and estimates of the trade-off between success probability and data complexity should be presented.

The set of keys, which have squared correlations less than the average, depends typically on the input and output masks u and w. Multidimensional linear cryptanalysis uses several different input and output masks and hence each key is more likely to result in a high correlation with respect to some input and output masks. Then the multidimensional linear distinguisher typically works as predicted for almost all keys. This phenomenon will be further elaborated in the context of block cipher PRESENT later in this article.

4. Multidimensional Linear Approximation of a Block Cipher

Let us study a block cipher with block size n. Let x be the plaintext, y the output of the cipher after R rounds and K the expanded key, that is, a vector consisting of all the (fixed) round key bits used in the R rounds. Then an m-dimensional

linear approximation of the block cipher can be considered as a vector Boolean function

$$\mathbb{Z}_2^n \times \mathbb{Z}_2^n \to \mathbb{Z}_2^m,\ (x, y) \mapsto Ux \oplus VK \oplus Wy, \tag{9}$$

where U and W are $m \times n$ binary matrices. The matrix V has also m rows and it divides the expanded keys, and therefore also the keys, to at most 2^m equivalence classes $z = VK$, $z \in \mathbb{Z}_2^m$.

We use Shannon's model of a secrecy system with the three random variables: plaintext $\mathbf{X}$, ciphertext $\mathbf{Y}$ and the key $\mathbf{K}$, where $\mathbf{X}$ and $\mathbf{K}$ are independent. Since linear cryptanalysis is targeted on ciphers without obvious weaknesses it is assumed that $\mathbf{Y}$ and $\mathbf{K}$ are uniformly distributed. Also, in the basic linear cryptanalysis, it is assumed that the plaintext $\mathbf{X}$ is also uniformly distributed. This assumption is usually extended to input data of all rounds. In particular it means that if the input data to a round is split into disjoint blocks, for example, inputs to parallel S-boxes, then the inputs are assumed to be independent and uniformly distributed.

Assume now that the key is fixed and consider one-dimensional linear approximations of the form $u \cdot \mathbf{X} \oplus w \cdot \mathbf{Y}$. In general, it is difficult to determine whether two such linear approximations are statistically dependent or not [21]. Usually, we can assume that the approximations are independent if the input masks involve disjoint sets of the input data bits and they remain disjoint over at least one round of the cipher, and the output masks are similarly disjoint. On the other hand, statistically dependent linear approximations are common. A typical example is the case when two linear approximations with nonzero correlations share the same input mask. If such approximations would be independent, then by the Piling up lemma, their sum, where the inputs cancel, should have a correlation, which is the product of the correlations of the linear approximations. But since the sum involves only output data bits, its correlation is zero by the assumption about uniform distribution of the data bits.

The problem is now to determine the p.d. p of the approximation. We noted in [10] that Lemma 1.1 can be used for determining p if the correlations $c(a \cdot (Ux \oplus VK \oplus Wy))$, $a \in \mathbb{Z}_2^m$ for all the one-dimensional linear approximations $a \cdot (Ux \oplus VK \oplus Wy)$ are available. If the correlations hold for most keys and can be determined accurately, we get an estimate for p that also holds for most keys. On the other hand, if the correlations vary with the key or we are not able to determine them with a satisfactory accuracy, then our knowledge of p is also limited and we must modify our method accordingly.

We denote by p^z the p.d. of $Ux \oplus Wy$, a fixed permutation of p determined by z. Then all the p.d.'s p^z, $z \in \mathbb{Z}_2^m$, are each other's permutations, and in particular,

$$p^z_{\eta \oplus a} = p^{z \oplus a}_{\eta}, \quad \text{for all} \quad z, \eta, a \in \mathbb{Z}_2^m. \tag{10}$$

From this it follows, for example, that $C(p) = C(p^z)$, for all $z \in \mathbb{Z}_2^m$.

Output: empirical p.d. q
initialise 2^m counters $q_\eta, \eta \in \mathbb{Z}_2^m$;
for $t = 1, \ldots, N$ **do**
  draw (x_t, y_t) from cipher ;
  for $i = 1, \ldots, m$ **do**
    calculate bit $\eta_i = u_i \cdot x_t \oplus w_i \cdot y_t$;
  end
  increment counter $q_\eta = \#\{t : Ux_t \oplus Wy_t = \eta\}$, where η is the vector $(\eta_1, \ldots, \eta_m)$ interpreted as an integer;
end
output q/N;

Figure 1. Algorithm 1: Computing empirical p.d q in the on-line phase

5. Key Recovery with Algorithm 1

In this section we assume that we are given a strong multidimensional approximation of the form (9) and a good estimate p of its p.d. In [14] we studied the different ways of generalising Matsui's one-dimensional Alg. 1 to multiple dimensions. We will now briefly recall these methods.

First, we obtain the empirical distribution $q = (q_0, \ldots, q_{2^m-1})$ of the multidimensional approximation $Ux \oplus Wy$ using N plaintext-ciphertext pairs (x_t, y_t), $t = 1, \ldots, N$ as follows:

$$q_\eta = N^{-1}\#\{t = 1, \ldots, N : Ux_t \oplus Wy_t = \eta\}, \quad \text{for all} \quad \eta \in \mathbb{Z}_2^m. \tag{11}$$

We call this the on-line phase of Alg. 1, see also Fig. 1. Next we rank the key classes using a suitable real-valued statistic. Each key class z is given a mark which is the realisation of the statistic using the data q. We order the keys according to their marks, and the right key class should be the first in this ordering.

The theoretically most efficient method is based on the log-likelihood ratio, LLR. We compute for each z the mark

$$l(z) = \mathrm{LLR}(q; p^z, \theta) = \sum_{\eta \in \mathbb{Z}_2^m} q_\eta \log \frac{p_\eta^z}{2^{-m}}. \tag{12}$$

Then we order the marks according to their magnitude and the right key should have the largest mark, i.e., the z that maximises $l(z)$ is selected. The algorithm for the LLR-method is described in Fig. 2. The data complexity N of the LLR-method is proportional to

$$m/C_{\min}(p),$$

where $C_{\min}(p) = \min_{z \neq 0} C(p^z, p)$.

An alternative method for finding the key class is depicted in Fig. 3. In this method, the mark is given by

Input: empirical p.d. q and theoretical p.d.'s p^z
Output: key class z'
for *key classes* $z = 0, \ldots, 2^m - 1$ **do**
    compute $l(z) = \sum_{\eta \in \mathbb{Z}_2^m} q_\eta \log p_\eta^z$;
end
find z' that maximises $l(z)$;
output z';

Figure 2. Algorithm 1: Finding the key class with LLR

Input: empirical p.d. q and theoretical p.d. p
Output: key class z'
compute $p * q$ using FFT;
find mode z' of $p * q$;
output z';

Figure 3. Algorithm 1: Finding the key class with convolution method

$$g(z) = (q * p)_z, \tag{13}$$

where $q * p$ is the convolution of p and q. We choose the z that maximises g. In other words, since $q * p$ is also a p.d., we have $z = \text{mode}(q * p)$. As we noted in [14], there is no practical difference between the data complexities of these methods in Alg. 1. However, this so-called convolution method is more efficient in practice, since it has time complexity $m2^m$ whereas for the LLR-method the time complexity is 2^{2m}.

Note that Alg. 1 can only be used if we have a good estimate of p. If the correlations $c(a \cdot (Ux \oplus VK \oplus Wy))$ are inaccurate, for example, if they have different signs than predicted, the Alg. 1 methods tend to fail in identifying correctly the key class. We consider next the multidimensional Alg. 2.

6. Key Recovery with Algorithm 2

Consider a block cipher with R+1 rounds, depicted in Fig. 4. Let x be the plaintext and y' be the ciphertext after R+1 rounds. Let the part of the last round key to be recovered be $k \in \mathbb{Z}_2^\ell$. Alg. 2 uses a strong linear approximation (9) over R rounds such that the bits of y involved in (9) can be computed by partially decrypting y' using the key part k. We denote this partial decryption as $\mathcal{D}_k(y')$. Analogically to the one-dimensional case, Alg. 2 is divided into four phases: distillation, analysis, ranking (or marking as we will call it), and, finally, the search phase, where the remaining key bits are searched and the correctness of the ranking result is verified. We concentrate on the first three phases.

In the distillation phase, N plaintext-ciphertext pairs (x_t, y'_t), $t = 1, \ldots, N$, are obtained. In the analysis phase, we compute, for each round key candidate k,

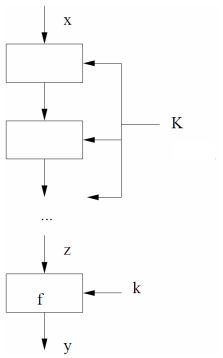

Figure 4. The linear approximation of an $R+1$-round block cipher for Alg 2. Notation: plaintext x, ciphertext y', input to the last round y, key data in R-rounds K, ℓ bits to be recovered from last round key k.

the empirical distribution q^k,

$$q^k_\eta = N^{-1}\{t = 1, \ldots, N : Ux_t \oplus W\mathcal{D}_k(y') = \eta\}, \eta \in \mathbb{Z}_2^m. \tag{14}$$

Before going to the marking phase, let us examine the complexity of the analysis phase. For a generic partial last round decryption, the function $\mathcal{D}_k(y')$ needs to be evaluated N times for each of the 2^ℓ key candidates, that is, the time complexity is $N2^\ell$.

Matsui observed in [20] that if the partial decryption function admits the form $\mathcal{D}_k(y') = d(y'' \oplus k)$, where y'' is an ℓ-bit sub-block of y', then it suffices to compute only the 2^ℓ decryptions $d(a), a \in \mathbb{Z}_2^\ell$, which, moreover, can be done off-line. Then the time complexity of the analysis phase in the classical one-dimensional case can be reduced from $N2^\ell$ to $N + 2^{2\ell}$. Collard, et al., showed in [5] how to further reduce it to $N + \ell 2^\ell$.

Matsui's observation can be exploited also in the multidimensional case to show that the complexity $N2^\ell$ can be reduced to $N + 2^{2\ell+m}$. When collecting the data (x_t, y'_t), $t = 1, \ldots, N$, we count the frequencies

$$T(i, j) = \#\{t : Ux_t = i,\ y''_t = j\},\ i \in \mathbb{Z}_2^m,\ j \in \mathbb{Z}_2^\ell$$

and store them in a $2^m \times 2^\ell$-table $T = T(i, j)$, where i denotes the row and j the column. This step takes time N. Similarly, as in Matsui's case, assume that we have the values $Wd(a)$, $a \in \mathbb{Z}_2^\ell$, computed and stored in an array $T_d(a)$. Given a key candidate k the table T_d is permuted in such a way that the $(j \oplus k)^{\text{th}}$ entry $T_d(j \oplus k) = Wd(j \oplus k)$ will be in the j^{th} position. Then each column of table $T(i, j)$ is permuted so that the i^{th} entry will be in the row $\eta = i \oplus Wd(j \oplus k)$.

Now summing up the entries in each row $\eta \in \mathbb{Z}_2^m$ gives the values

$$\#\{t : Ux_t \oplus Wd(y''_t \oplus k) = \eta\}.$$

Sorting the $2^m \times 2^\ell$-table T takes time $2^{\ell+m}$. Hence, the total time to compute the empirical distribution q^k for all key candidates is $N + 2^{2\ell+m}$. It is not clear

if the trick proposed by Collard, et al., in [5] can be used in multiple dimensions. Nevertheless, usually $N \gg 2^{2\ell+m}$ and the time complexity is dominated by N.

Let us now assume that we have computed the empirical p.d.'s q^k defined by (14). The remaining task is to mark the keys in the marking phase. Similarly as for Alg. 1, the marks are given by a statistic that is computed using the data represented by the empirical p.d.'s q^k. A classical assumption is the wrong-key randomisation hypothesis that is needed in Alg. 2. We state it as follows:

Assumption 6.1 (Wrong-key Hypothesis)**.** *There are two p.d.'s $\mathcal{D}$ and $\mathcal{D}'$, $\mathcal{D} \neq \mathcal{D}'$ such that for the right key k_0, the data is drawn from $\mathcal{D}$ and for a wrong key $k \neq k_0$ the data is drawn from $\mathcal{D}' \neq \mathcal{D}$.*

Usually $\mathcal{D}'$ is the uniform distribution. The wrong-key hypothesis states that for each wrong round key candidate, deciphering with the wrong key produces uniformly distributed data that is statistically independent for different keys whereas the data derived using k_0 is not uniformly distributed.

We consider three different methods for marking the key candidates in Alg. 2: the LLR-method, the convolution method and the χ^2-method. A convenient way for comparing the efficiency of these methods is the "advantage". Selçuk proposed to using it for measuring the time complexity of the search phase in the one-dimensional Alg. 2 attack [24]. We define it as follows:

Definition 6.2. We say that a key recovery attack for an ℓ-bit key achieves an advantage of a bits over exhaustive search, if the marking phase puts the correct key among the top $r = 2^{\ell-a}$ out of all 2^ℓ key candidates.

We can now compare the different methods by finding an expression between the advantage and the data complexity. We showed in [12] that for fixed data complexity N, the advantage of the LLR-method is larger than the advantage of the χ^2-method. Hence, the LLR-method is more efficient. Next we describe briefly the realisation of the three methods.

In the LLR- method, for each round key candidate, the mark is given by

$$L_k = \max_{z \in \mathbb{Z}_2^m} \mathrm{LLR}(q^k; p^z, \theta). \tag{15}$$

The keys are then ranked in decreasing order according to their marks and the right key candidate k_0 should have the largest mark L_{k_0}. In this way, we get both z and k. The LLR-method is depicted in Fig. 5. We obtain the following result:

Theorem 6.3. *Suppose the cipher satisfies Assumption 6.1 where $\mathcal{D}' = \theta$ and the p.d.'s p^z, $z \in \mathbb{Z}_2^m$ and θ are close to each other. Then the advantage of the LLR-method for finding the last round key k_0 is given by*

$$a_{\mathrm{LLR}} = (\sqrt{NC(p)} - \Phi^{-1}(P_{12}))^2/2 - m \approx NC(p) - m. \tag{16}$$

Here N is the amount of data used in the attack, $P_{12}\,(> 0.5)$ is the probability of success, Φ is the cumulative distribution function of the normed normal distribution and $C(p)$ and m are the capacity and the dimension of the linear approximation (9)*, respectively.*

Input: table of empirical p.d.'s q^k, $k = 0, \ldots, 2^l - 1$, table of theoretical p.d.'s $p^z, z = 0, \ldots, 2^m - 1$
Output: store mark L_k and key class z for each $k \in \mathbb{Z}_2^l$
for $k = 0, \ldots, 2^l - 1$ **do**
  for $z = 0, \ldots, 2^m - 1$ **do**
    $L(k, z) = \mathrm{LLR}(q^k; p^z, \theta)$; `/* takes time` 2^m `*/`
  end
  store $L_k = \max_z L(k, z)$ and the maximising class $z(k)$;
end

Figure 5. Marking phase of Alg. 2 using LLR-method: The permutations p^z, $z \in \mathbb{Z}_2^m$ are stored and used for determining the mark L_k. The outputs k and z can be determined simultaneously.

Input: table of empirical p.d.'s q^k, $k = 0, \ldots, 2^l - 1$, and the theoretical p.d. p
Output: store mark G_k and key class z for each $k \in \mathbb{Z}_2^l$
for $k = 0, \ldots, 2^l - 1$ **do**
  compute $q^k * p$ using FFT;
  store mark $G_k = \max_z (q^k * p)_z$ and $z(k) = \mathrm{mode}(q^k * p)$;
end

Figure 6. Marking phase of Alg. 2 using the convolution method: One p.d. p is stored and used for determining the mark G_k. The outputs k and z can be determined simultaneously.

We now propose using the convolution method also for Alg. 2. It has, at least in theory, the same advantage as the LLR-method but smaller time complexity in the marking phase. The mark is

$$G_k = \max_{z \in \mathbb{Z}_2^m} (q^k * p)_z. \tag{17}$$

The marks are again ranked in decreasing order such that the mark G_{k_0} of the right key candidate should be largest. We obtain both k and z simultaneously. The marking phase with convolution method is depicted in Fig. 6.

The χ^2-test is theoretically the weakest method. The marks are given by

$$S_k = 2^m N \sum_{\eta \in \mathbb{Z}_2^m} (q_\eta^k - 2^{-m})^2. \tag{18}$$

Again, the right key k_0 should have the largest mark. The method is depicted in Fig. 7. After determining the last round key candidate k it is possible to use Alg. 1 for determining z, provided that the p.d. q^k is saved with the mark S_k. We obtained the following result in [12].

Input: table of empirical p.d.'s $q^k, k = 0, \ldots, 2^l - 1$
Output: store mark S_k and possibly the corresponding p.d. q^k for each $k \in \mathbb{Z}_2^l$
for $k = 0, \ldots, 2^l - 1$ **do**
  compute $S_k = \sum_{\eta=0}^{2^m-1} (q_\eta^k - 2^{-m})^2$;
  if *wish to recover* z **then**
    store (S_k, q^k);
  else
    store S_k;
  end
end

Figure 7. Marking phase of Alg. 2 using χ^2-method: The mark S_k is determined and stored for each k. Also q^k is stored with the mark, if z needs to be recovered.

Theorem 6.4. *Suppose the cipher satisfies Assumption 6.1 where $\mathcal{D}' = \theta$ and the p.d.'s p^z, $z \in \mathbb{Z}_2^m$ and θ are close to each other. Then the advantage of the χ^2-method using statistic* (18) *is given by*

$$a_{\chi^2} = \frac{(NC(p) - 4\varphi)^2}{4M}, \; \varphi = \Phi^{-2}(2P_S - 1), \; M = 2^m - 1, \tag{19}$$

where P_S (> 0.5) is the probability of success, N is the amount of data used in the attack and $C(p)$ and m (≥ 5) are the capacity and the dimension of the linear approximation (9)*, respectively.*

Some practical experiments with χ^2 and LLR on reduced round SERPENT can be found in [10]. In these experiments the χ^2-method seems indeed to be weaker than the LLR-method: the data complexity for finding the last round key with given probability and advantage is larger for χ^2 than for LLR. On the other hand, the LLR-method and the convolution method are practically equal with respect to the data complexity. Hence, if p is known then we propose using the convolution method. However, if we have a cipher where p cannot be determined with reasonable accuracy, or it varies significantly depending on the key, the only available method is the χ^2-method. Cho observed that this seems to be the case for example for block cipher PRESENT [4]. Cho's multidimensional χ^2-attack on reduced round PRESENT is presented in the next section.

7. Using Multiple Linear Approximations in PRESENT

7.1. PRESENT

The goal of the attack is to use the multidimensional Alg. 2. for determining key bits from the first and the last rounds. In this section, we use mostly the notation of [4]. PRESENT is a key alternating SPN block cipher consisting of 31 rounds. Each round consists of three layers: AddRoundKey, SboxLayer and pLayer. The

AddRoundKey layer is an XOR-operation with a round key. The non-linearity is provided by the 16 S-boxes $S_0, \ldots, S_{15}$ used in the SboxLayer. Each S-box S_i is the same non-linear bijective mapping $S : \mathbb{Z}_2^4 \mapsto \mathbb{Z}_2^4$. The pLayer is a bit-by-bit permutation P of the set $\{0, 1, \ldots, 63\}$. The complete description of PRESENT is given in [3]. We denote by $\mathcal{E}_K^{(n)}(x)$ the encryption of x over n rounds, where the permutation is not used in the n^{th} round.

The structure of PRESENT makes it possible to realise a multidimensional attack: there are several strong one-dimensional approximations. The linear hull of each such approximation consists of several equally strong approximation trails. Therefore, instead of Piling up lemma, the correlation over several rounds must be calculated using the Correlation theorem, Theorem 3.1. The properties of PRESENT makes it possible to do these calculations sufficiently accurately. In the following, we first show how to determine all the strongest one-dimensional approximations over $n+4$ rounds. We then describe the multidimensional approximation and the attack proposed by Cho, and discuss its theoretical foundations.

7.2. One-Dimensional Approximation

The calculation of an estimate for a correlation of a one-dimensional approximation over multiple rounds of PRESENT is based on the following observation: only the so-called single-bit linear trails have non-negligible trail-correlations. A single-bit linear trail is an approximation trail with intermediate masks of Hamming weight one. Therefore, the sum over all the middle masks can be estimated using the sum over masks with Hamming weight one. Next we show how to reduce the number of possible trails even more.

Denote by $\mathcal{S} = \{S_5, S_6, S_7, S_9, S_{10}, S_{11}, S_{13}, S_{14}, S_{15}\}$ and $\mathcal{B} = \{4i + 1, 4i + 2, 4i+3 : 0 \leq i \leq 15,\ S_i \in \mathcal{S}\}$ a subset of the S-boxes and a subset of possible input and output bits of the SboxLayer, respectively. Cho showed that the set $\mathcal{B}$ is closed in the following sense: for any linear approximation starting and ending at S-box in $\mathcal{S}$, all single-bit approximation trails that have non-zero trail-correlations have intermediate masks in $\mathcal{B}$. We denote by $\rho(\alpha, \beta)$ the correlation of an approximation over an S-box with input and output masks α and β, respectively.

Cho made the following observations about the single-bit input and output masks α and β:

- For $\alpha,\ \beta \in \{2, 4, 8\}$, $\rho(\alpha, \beta) = \pm 2^{-2}$, except $\rho(8, 4) = 0$
- For $\alpha \in \{1, 2, 4, 8\}$, $\rho(\alpha, 1) = \rho(1, \alpha) = 0$.

Consider a linear approximation starting from the s^{th} bit position and ending in the t^{th} bit position, where $s, t \in \mathcal{B}$. Denote by $\theta^{(n)}(s, t; K)$ the correlation of the linear approximation over n rounds with user-supplied key K. By the Correlation theorem, one can obtain an estimate of $\theta^{(n)}(s, t; K)$ using the following recursive formula:

$$\theta^{(r)}(s, t; K) = \sum_{i=1}^{3} (-1)^{K_r[\nu]} \rho(2^i, 2^{P^{-1}(t) \bmod 4}) \theta^{(r-1)}(s, \nu; K),\ r = 1, \ldots, n, \quad (20)$$

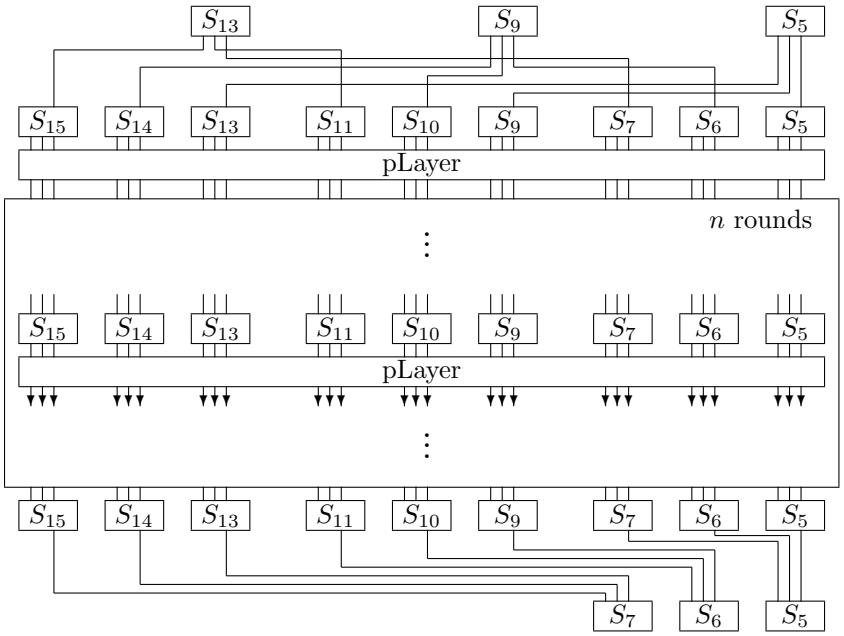

Figure 8. Single-bit trails of the linear hull over $n+4$ rounds (adapted from [4])

where $\nu = 4\lfloor P^{-1}(t)/4 \rfloor + i$ and $K_r[\nu]$ is the ν^{th} bit of the r^{th} round key. Defining $\theta^{(0)}(s,t;K) = 1$, this recursive formula allows as to compute $\theta^{(n)}$ for all $n \geq 1$ and for all keys K.

Consider now a linear approximation over $n+4$ rounds starting at one of the S-boxes S_i, $i = 5, 9$ or 13 and ending at one of the S-boxes S_j, $j = 5, 6$ or 7. All single-bit trails with non-zero correlation are depicted in Fig. 8.

Fix i and j and consider a one-dimensional approximation with input and output masks α and β. We denote by $A_i \in \mathcal{B}$ the set of the bit positions that can be reached from the bits $4i+1$, $4i+2$ or $4i+3$ (output bits of S_i) using single-bit trails over a pLayer, an SboxLayer and one more pLayer, for any $i = 5, 9$ or 13. Hence, for given α, there are nine single-bit trails that lead from α to A_i, each of which has correlation $2^{-2}\rho(\alpha, 2^u)$ for a unique $u \in \{1, 2, 3\}$ that is determined by the end bit position $s \in A_i$.

Similarly, let $B_j \in \mathcal{B}$ be the set of the bit positions from where one can reach $4j+1$, $4j+2$ or $4j+3$ using single-bit trails over a pLayer, an SboxLayer and one more pLayer, where $j = 5, 6$ or 7. Each of the nine single-bit trails leading from B_j to β have correlation $2^{-2}\rho(2^v, \beta)$, for a unique $v \in \{1, 2, 3\}$ that is determined by the start bit position $t \in B_j$.

Hence, by the Correlation theorem, the correlation over $n+4$ rounds with a given key K can be estimated, by considering single-bit trails only, as

$$c(\alpha \cdot x \oplus \beta \cdot \mathcal{E}_K^{(n+4)}(x)) = 2^{-4} \sum_{s \in A_i} \sum_{t \in B_j} (-1)^{K_p} \rho(\alpha, 2^u) \theta^{(n)}(s,t;K) \rho(2^v, \beta), \quad (21)$$

where K_p denotes the parity of the relevant round key bits.

7.3. Expected Capacity of the Multidimensional Approximation

Consider now a multidimensional approximation over $n+4$ rounds starting at one of the S-boxes S_i, $i = 5, 9$ or 13 and ending at one of the S-boxes S_j, $j = 5, 6$ or 7. The input and output masks α and β in (21) both span a four-dimensional subspace. Therefore, $m = 8$. An estimate of the average capacity of such an approximation, denoted by $C(i,j)$, is given by the following result:

Theorem 7.1. *Assume that the round keys of PRESENT are statistically independent. For a positive integer n, the expected capacity over the keys is for any $i = 5, 9, 13$ and $j = 5, 6, 7$*

$$C(i,j) = 2^{-8} \sum_{s \in A_i} \sum_{t \in B_j} \mathrm{E}_{\mathbf{K}} \left[\theta^{(n)}(s,t;\mathbf{K})^2 \right], \quad (22)$$

where the average squared correlations $\mathrm{E}_{\mathbf{K}} \left[\theta^{(n)}(s,t;\mathbf{K})^2 \right]$ are given by the following recursive formula:

$$\mathrm{E}_{\mathbf{K}} \left[\theta^{(r)}(s,t;\mathbf{K})^2 \right] = \sum_{i=1}^{3} \rho(2^i, 2^{P^{-1}(t) \bmod 4})^2 \, \mathrm{E}_{\mathbf{K}} \left[\theta^{(r-1)}(s,\nu;\mathbf{K})^2 \right], \; r = 1, \ldots, n,$$

where $\nu = 4\lfloor P^{-1}(t)/4 \rfloor + i$ and $\mathrm{E}_{\mathbf{K}} \left[\theta^{(0)}(s,t;\mathbf{K})^2 \right] = 1$.

The proof is given in [4] as a part of the proof of Theorem 2 but we give it here for completeness.

Proof. The round keys of PRESENT are not statistically independent as they are all derived from the same relatively short initial key. Nevertheless, assuming the round keys to be independent even when they are not is common in linear cryptanalysis and Cho relies on it, too. His practical tests show that the attack can be realised successfully using the assumption. By the assumption that the round keys are independent, it follows from Theorem 3.2 and (20) that

$$\mathrm{E}_{\mathbf{K}} \left[\theta^{(r)}(x,y;\mathbf{K})^2 \right] = \sum_{i=1}^{3} \rho(2^i, 2^{P^{-1}(y) \bmod 4})^2 \, \mathrm{E}_{\mathbf{K}} \left[\theta^{(r-1)}(x,\nu;\mathbf{K})^2 \right],$$

where $\nu = 4\lfloor P^{-1}(y)/4 \rfloor + i$.

Since we define in (20) that $\theta^{(0)}(s,t;K) = 1$ for all K, then $\mathrm{E}_{\mathbf{K}} \left[\theta^{(0)}(s,t;\mathbf{K})^2 \right] = 1$.

Since the round keys are statistically independent, we have $\mathrm{E}_{\mathbf{K}}\left[(-1)^{\mathbf{K}_p}(-1)^{\mathbf{K}'_p}\right] = 1$ if and only if $\mathbf{K}_p = \mathbf{K}'_p$ and otherwise the expected value is zero. Taking expectation of the square of the correlation (21) gives

$$\mathrm{E}_{\mathbf{K}}\left[c(\alpha \cdot x \oplus \beta \cdot \mathcal{E}_K^{(n+4)}(x); \mathbf{K})^2\right] = 2^{-8}\,\mathrm{E}_{\mathbf{K}}\left[\sum_{s\in A_i}\sum_{t\in B_j} \rho(\alpha, 2^u)^2\theta^{(n)}(s,t;\mathbf{K})^2\rho(2^v,\beta)^2\right].$$

Consider now Formula (3) that gives the capacity as the sum of all the correlations as α and β vary. Using Parseval's theorem by summing over the masks α and β gives the expected capacity for given S-boxes S_i and S_j (for S_j we actually apply Parseval's theorem to the inverse of S_j):

$$\begin{aligned} C(i,j) &= \sum_{\alpha,\beta\in\mathbb{Z}_2^4} \mathrm{E}_{\mathbf{K}}\left[c(\alpha \cdot x \oplus \beta \cdot \mathcal{E}_K^{(n+4)}(x); \mathbf{K})^2\right] \\ &= 2^{-8}\,\mathrm{E}_{\mathbf{K}}\left[\sum_{s\in A_i}\sum_{t\in B_j} \theta^{(n)}(s,t;\mathbf{K})^2 \sum_{\alpha\in\mathbb{Z}_2^4}\rho(\alpha,2^u)^2 \sum_{\beta\in\mathbb{Z}_2^4}\rho(2^v,\beta)^2\right] \\ &= 2^{-8}\sum_{s\in A_i}\sum_{t\in B_j} \mathrm{E}_{\mathbf{K}}\left[\theta^{(n)}(s,t;\mathbf{K})^2\right]. \end{aligned}$$

□

7.4. Realisation of the Attack

The attack is performed over R rounds of PRESENT in an Alg. 2 type attack. In [4], the nine multidimensional linear approximations, explained in the preceding section, were used simultaneously over $R-2$ rounds. The inputs to each of the three S-boxes S_5, S_9 and S_{13} depend on the same 16-bit part of the first round key. This 16-bit part of the key is denoted by k_e. Similarly, the outputs of the S-boxes S_5, S_6 and S_7 on the last but second round are determined by a 16-bit part of the last round key, denoted by k_d. We encrypt with each key candidate k_e over one round and decrypt with each candidate k_d over one round. Then we proceed as in Section 6 with $k = (k_e, k_d)$.

Let us first calculate an estimate for the total capacity C of a multidimensional approximation from the input of three S-boxes S_5, S_9 and S_{13} to the output of the three S-boxes S_5, S_6 and S_7. Recall that single-bit trails outperform all other approximation trails. Hence, it suffices to take into consideration only those output and output masks that involve bits from one S-box only. We denote by α_i and β_j the 12-bit masks that have non-zero components only at the input positions of the S-box S_i and output positions of the S-box S_j, respectively. We can estimate

$$
\begin{aligned}
C &= \mathrm{E}_{\mathbf{K}}\left[\sum_{\alpha,\beta\in\mathbb{Z}_2^{12}} c(\alpha\cdot x \oplus \beta\cdot \mathcal{E}_K^{(R-2)})^2\right] \\
&\approx \mathrm{E}_{\mathbf{K}}\left[\sum_{i=5,9,13}\sum_{j=5,6,7}\sum_{\alpha_i,\beta_j\in\mathbb{Z}_2^{12}} c(\alpha_i\cdot x \oplus \beta_j\cdot \mathcal{E}_K^{(R-2)})^2\right] \\
&\approx \sum_{i=5,9,13}\sum_{j=5,6,7} C(i,j) \\
&= 2^8 \sum_{s,t\in\mathcal{B}} \mathrm{E}_{\mathbf{K}}\left[\theta^{(R-6)}(s,t;\mathbf{K})^2\right].
\end{aligned}
$$

The last equality follows from Theorem 7.1 taking into account the property that $A_5 \cup A_9 \cup A_{13} = \mathcal{B}$ and $B_5 \cup B_6 \cup B_7 = \mathcal{B}$ and the three sets A_i, $i = 5, 9, 13$, and similarly B_j, $j = 5, 6, 7$, are mutually disjoint.

As the aim is to distinguish the right key candidate from the wrong ones either LLR (or convolution) or χ^2 method can be used. The dimension of the multidimensional linear approximation is $m = 24$. Given the desired advantage a the estimated data complexity N_{LLR} of the LLR-method is by Theorem 6.3

$$
N_{\mathrm{LLR}} \approx \frac{a+m}{C} = 56C^{-1}
$$

for the full 32-bit key recovery and setting $P_S = 0.95$. By Theorem 6.4 the corresponding estimate for the χ^2-method is

$$
N_{\chi^2} \approx \frac{\sqrt{a\cdot(2^{24}-1)}}{C} = 2^{14.5}C^{-1}, \tag{23}
$$

which is significantly larger.

Unfortunately, the LLR-method cannot be used due to the lack of accurate estimate of the p.d. of the multidimensional approximation. Cho observed in practical experiments that the p.d. varies a lot with the keys, while the capacity remains rather constant. Hence, the χ^2-method remains the only possibility and should work about equally well for all keys.

Instead of (23), Cho used the following formula[1]:

$$
N \approx 2\sqrt{a\cdot 9(2^8-1)}/C, \tag{24}
$$

for the data complexities N, which also agreed with the experimental results. Here the coefficient $M' = 9(2^8 - 1)$ is the number of one-dimensional approximations used in the attack and it is significantly smaller than the number of the degrees of freedom $2^{24} - 1$ of a 24-dimensional p.d. used in (23).

Mathematically, if we assume that the approximations from S_i to S_j are statistically independent, we can obtain the χ^2-statistic over the whole 24-dimensional system as a sum of the χ^2-statistics of the 8-dimensional approxima-

[1] There should be coefficient 4 and not 8 in the formula (2) of [4]

tions. That is, we consider the problem of the linear combination of independent χ^2-tests on the same hypothesis, which was studied for example by Koziol and Perlman [18]. Since the number of degrees of freedom for each 8-dimensional approximation is $2^8 - 1$, the sum of nine χ^2-distributed and independent random variables is χ^2-distributed with $M' = 9 \cdot (2^8 - 1)$ degrees of freedom and we obtain (24).

Let us take a closer look at the linear approximations used in Cho's attack to investigate the assumption about statistical independence. As noted in Section 3, linear approximations sharing the same input mask and having non-zero correlations cannot be statistically independent. For each S-box S_5, S_9 and S_{13} there are several input masks α with nonzero correlation to different single-bit output masks of the S-box. It follows that the 8-dimensional approximations cannot, strictly speaking, be statistically independent, if they share the same input S-box or the same output S-box. Hence, only a subset of three of the nine χ^2-statistics are potentially statistically independent. We conclude that the heuristic approach used by Cho in combining the nine multidimensional approximations using a sum of χ^2-statistics seems to work very well in practice, but its theoretical justification remains an open question.

7.5. Complexity of the Attack and an Extension Using Related Keys

The multidimensional approximation presented above can be used for launching linear attacks on PRESENT with different strategies and estimated complexity parameters. The attack presented by Cho in [4] on PRESENT reduced to R rounds aims at discovering 16 bits of the first and the last round key. The estimated capacity of the approximation over 24 rounds is 2^{-55} which results in data complexity of 2^{64} for the full advantage $a = 32$. Hence, using the full code book of 26-round PRESENT, 32 bits of the key can be recovered using this attack.

In [25] Shamir suggested to collecting more data using related keys. As observed by Cho in the experiment, the probability distributions of the multidimensional linear approximations vary with different keys while their capacities remain about the same. We denote this capacity by C. Hence it is not realistic to assume that we can collect information of data in the same distribution for different keys but for each key a separate distribution must be observed. However, it is realistic to assume the these distributions are statistically independent. Hence given data from a number, say B different keys, we get B independent χ^2-statistics each computed from a distribution with capacity C. By taking the sum of these χ^2-statistics the data requirement is

$$N \approx 2\frac{\sqrt{aB9(2^8 - 1)}}{BC}.$$

Due to the nonlinear key schedule of PRESENT any key relation is expected to become obscure at the last round. However, simple and natural relations, which allow recovering the same secret key bits on the first round can be defined in different ways. For example, one can consider a set of keys that have the same key bits in the positions to be recovered, or keys that differ from each other by known constants. Hence to attack the full 31-round PRESENT using such a related key

attack, a 30-round linear approximation must be used. The capacity of a 30-round approximation was estimated to be $C = 2^{-71}$ in [4]. Using the above formula, we obtain that the required number of plaintexts, from which data is observed and these B separate distributions are computed, is

$$N \approx 2\frac{\sqrt{16B9(2^8-1)}}{BC} \approx 2^{79.6}/\sqrt{B}.$$

From this we see that the full code book of $N = 2^{64}$ of plaintexts is sufficient to carry out this attack if $B = 2^{31.2}$. The total time complexity for handling these distributions is $BN = 2^{95.2}$. We conclude that such a related key attack breaks the full 31-round PRESENT with 128-bit key, but not, if the key length is 80 bits.

8. Conclusions

Multidimensional linear cryptanalysis uses several linear approximations for recovering information about the secret key used in a block cipher. For that, we have presented multidimensional generalisations of Matsui's Alg. 1 and Alg. 2 and demonstrated in theory and in practice how to realise them.

For Alg. 1, all the methods are practically equal. The theory and practical experiments suggest that the convolution method is most efficient. However, the use of Alg. 1 is based on the assumption that a good estimate of the p.d. of the approximation is available. If the p.d. varies significantly with the key or is otherwise ambiguous, Alg. 1 can fail.

In this sense, Alg. 2 is not as vulnerable to the p.d. Provided that the p.d. has a non-negligible capacity, it is still possible to realise the attack with χ^2-method. On the other hand, if the p.d. is available, we suggest using the convolution or LLR method for Alg. 2, also.

We also considered the linear hull effect, the relevance of Piling up lemma and Correlation theorem. We showed how the linear hull effect contributes to the correlation so that distinguishing attacks – on the average over the keys – enhanced by it. We used Cho's multidimensional Alg. 2 attack on the block cipher PRESENT [4] to demonstrate the theoretical and practical effects of the linear hull effect in multiple dimensions. The attack clearly benefitted from both the use of multiple approximations in the context of multidimensional linear cryptanalysis and the linear hull effect.

Acknowledgements

We thank Joo Yeon Cho for helpful discussions concerning Section 7. The first author's work on this article was supported by Matine project number 776.

References

[1] Thomas Baignères, Pascal Junod, and Serge Vaudenay. How Far Can We Go Beyond Linear Cryptanalysis? In Pil Joong Lee, editor, *Advances in Cryptology – ASIACRYPT*

'04, 10th International Conference on the Theory and Application of Cryptology and Information Security, Jeju Island, Korea, December 5–9, 2004. Proceedings, volume 3329 of *Lecture Notes in Computer Science*, pages 432–450, Berlin/Heidelberg, 2004. Springer.

[2] Alex Biryukov, Christophe De Cannière, and Michal Quisquater. On Multiple Linear Approximations. In Matt Franklin, editor, *Advances in Cryptology – CRYPTO '04, 24th Annual International Cryptology Conference, Santa Barbara, California, USA, August 15–19, 2004. Proceedings*, volume 3152 of *Lecture Notes in Computer Science*, pages 1–22, Berlin/Heidelberg, 2004. Springer.

[3] A. Bogdanov, L.R. Knudsen, G. Leander, C. Paar, A. Poschmann, M.J.B. Robshaw, Y. Seurin, and C. Vikkelsoe. PRESENT: An Ultra-Lightweight Block Cipher. In Pascal Paillier and Ingrid Verbauwhede, editors, *Cryptographic Hardware and Embedded Systems - CHES 2007 9th International Workshop, Vienna, Austria, September 10–13, 2007. Proceedings*, volume 4727, pages 450–466. Springer, 2007.

[4] Joo Yeon Cho. Linear cryptanalysis of reduced-round PRESENT. In *In Topics in Cryptology – CT-RSA 2010, The Cryptographers' Track at the RSA Conference 2010, San Francisco, CA, USA, March 1–5, 2010.*, volume 5985 of *Lecture Notes in Computer Science*, pages 302–317. Springer, 2010.

[5] Baudoin Collard, F. X. Standaert, and Jean-Jacques Quisquater. Improving the Time Complexity of Matsui's Linear Cryptanalysis. In Kil-Hyun Nam and Gwangsoo Rhee, editors, *Information Security and Cryptology – ICISC 2007, 10th International Conference, Seoul, Korea, November 29–30, 2007.*, volume 4717 of *Lecture Notes in Computer Science*, pages 77–88, Berlin/Heidelberg, 2007. Springer.

[6] H. Cramèr and H. Wold. Some theorems on distribution functions. *J. London Math. Soc.*, s1-11(4):290–295, Oct 1936.

[7] J. Daemen, R. Govaerts, J. Vandewalle: Correlation matrices. In Preneel, B., ed.: Fast Software Encryption Second International Workshop Leuven, Belgium, December 14–16, 1994 Proceedings. Volume 1008 of Lecture Notes in Computer Science., Berlin/Heidelberg, Springer (1995) 275–285

[8] Joan Daemen and Vincent Rijmen. *The Design of Rijndael: AES - The Advanced Encryption Standard (Information Security and Cryptography)* Springer, Berling-Heidelberg, 2002.

[9] H. Englund and A. Maximov. Attack the Dragon. In Subhamoy Maitra and C.E. Veni Madhavan, editors, *Progress in Cryptology – INDOCRYPT '05, 6th International Conference on Cryptology in India, Bangalore, India, December 10–12, 2005. Proceedings*, volume 3797 of *Lecture Notes in Computer Science*, pages 130–142, Berlin/Heidelberg, 2005. Springer.

[10] Miia Hermelin, Kaisa Nyberg, and Joo Yeon Cho. Multidimensional Linear Cryptanalysis of Reduced Round Serpent. In Jennifer Seberry Yi Mu, Willy Susilo, editor, *Information Security and Privacy, 13th Australasian Conference, ACISP 2008, Wollongong, Australia, July 7–9, 2008. Proceedings*, volume 5107 of *Lecture Notes in Computer Science*, pages 203–215, Berlin/Heidelberg, 2008. Springer.

[11] Miia Hermelin, Joo Yeon Cho and Kaisa Nyberg: Statistical Tests for Key Recovery Using Multidimensional Extension of Matsui's Algorithm 1. EUROCRYPT '09 - POSTER SESSION (2009)

[12] 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 Revised Selected Papers*, volume 5665 of *Lecture Notes in Computer Science*, pages 209–227, Berlin/Heidelberg, 2009. Springer.

[13] Miia Hermelin and Kaisa Nyberg. Multidimensional Linear Distinguishing Attacks and Boolean Functions. In *Fourth International Workshop on Boolean Functions: Cryptography and Applications*, 2008.

[14] Miia Hermelin and Kaisa Nyberg. Dependent Linear Approximations – The Algorithm of Biryukov and Others Revisited. 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*, volume 5985 of *Lecture Notes in Computer Science*, pages 318–333.

Springer 2010.
[15] Miia Hermelin. Multidimensional Linear Cryptanalysis. PhD thesis, Aalto University School of Science and Technology (2010)
[16] Thomas Johansson and Alexander Maximov. A Linear Distinguishing Attack on Scream. *IEEE Transactions on Information Theory*, 53(9):3127 – 3144, 2007. Previously appeared in ISIT 2003. Yokohama, Japan, June 29 – July 4,2003.
[17] Jr. Burton, S. Kaliski and M. J. B. Robshaw. Linear Cryptanalysis Using Multiple Approximations. In Yvo G. Desmedt, editor, *Advances in Cryptology – CRYPTO '94, 14th Annual International Cryptology Conference Santa Barbara, California, USA August 21–25, 1994 Proceedings*, volume 839 of *Lecture Notes in Computer Science*, pages 26–39, Berlin/Heidelberg, 1994. Springer.
[18] James A. Koziol and Michael D. Perlman. Combining Independent Chi-Squared Tests. *Journal of the American Statistical Association*, 73(364):753–763, Dec 1978.
[19] Mitsuru Matsui Linear Cryptanalysis Method for DES Cipher. In Tor Helleseth, editor, *Advances in Cryptology – EUROCRYPT '93, Workshop on the Theory and Application of Cryptographic Techniques Lofthus, Norway, May 2327, 1993 Proceedings*, volume 765 of *Lecture Notes in Computer Science*, pages 386–397, Berlin/Heidelberg, 1994. Springer.
[20] Mitsuru Matsui. The First Experimental Cryptanalysis of the Data Encryption Standard. In Yvo G. Desmedt, editor, *Advances in Cryptology – CRYPTO '94, 14th Annual International Cryptology Conference Santa Barbara, California, USA August 21–25, 1994 Proceedings*, volume 839 of *Lecture Notes in Computer Science*, pages 1–11, Berlin/Heidelberg, 1994. Springer.
[21] S. Murphy. The Independence of Linear Approximations in Symmetric Cryptology. *IEEE Transactions on Information Theory*, 52(12):5510–5518, Dec 2006.
[22] Kaisa Nyberg. Linear approximation of block ciphers. In *Advances in Cryptology - EUROCRYPT'94*, volume 950 of Lecture Notes in Computer Science. Springer-Verlag, 1995.
[23] Kaisa Nyberg. Correlation theorems in cryptanalysis. *Discrete Applied Mathematics*, 111(1–2):177–188, July 2001.
[24] A. A. Selçuk. On probability of success in linear and differential cryptanalysis. Journal of Cryptology **21**(1) (January 2008) 131–147
[25] Adi Shamir. Comment at CT RSA 2010, Linear Cryptanalysis session, March 4, 2010
[26] Serge Vaudenay. An Experiment on DES Statistical Cryptanalysis. In *CCS '96: Proceedings of the 3rd ACM Conference on Computer and Communications Security*, pages 139–147, New York, NY, USA, 1996. ACM.