Capacity and Data Complexity in Multidimensional Linear Attack

Jialin Huang[1,2⋆], Serge Vaudenay[3], Xuejia Lai[2], and Kaisa Nyberg[4]

1. Technische Universität Darmstadt, Germany
2. Shanghai Jiao Tong University, China
3. EPFL, Switzerland
4. Aalto University, Finland

Abstract. Multidimensional linear attacks are one of the most powerful variants of linear cryptanalytic techniques now. However, there is no knowledge on the key-dependent capacity and data complexity so far. Their values were assumed to be close to the average value for a vast majority of keys. This assumption is not accurate. In this paper, under a reasonable condition, we explicitly formulate the capacity as a Gamma distribution and the data complexity as an Inverse Gamma distribution, in terms of the average linear potential and the dimension. The capacity distribution is experimentally verified on the 5-round PRESENT.
Regarding complexity, we solve the problem of estimating the average data complexity, which was difficult to estimate because of the existence of zero correlations. We solve the problem of using the median complexity in multidimensional linear attacks, which is an open problem since proposed in Eurocrypt 2011. We also evaluate the difference among the median complexity, the average complexity and a lower bound of the average complexity – the reciprocal of average capacity. In addition, we estimate more accurately the key equivalent hypothesis, and reveal the fact that the average complexity only provides an accurate estimate for less than half of the keys no matter how many linear approximations are involved.
Finally, we revisit the so far best attack on PRESENT based on our theoretical result.

Keywords: multidimensional linear attack, capacity, data complexity, linear hull effect, linear potential

1 Introduction

Block ciphers are used as basic building primitives in symmetric cryptography for encryption, authentication, construction of hash functions and so on. Evaluation of their practical security has been a hot research issue over the decades, giving rise to different analysis techniques. Statistical attacks exploit non-uniform

⋆ This work was finished when the author affiliated to Shanghai Jiao Tong University and was visiting EPFL.

behaviors of the plaintext-ciphertext data to find information about the key. One of the most prominent statistical attacks is linear cryptanalysis. Previously, linear trails were assumed to behave equally for each key [20,3,17,4]. Then, by considering many trails in one approximation [24,25], the linear hull effect raises interesting discussions about fixed-key behaviors in single linear approximations [22,21]. Daemen et al. gave a fixed-key probability distribution for single linear correlations [13], leading to subsequent works on e.g., fundamental assumptions [9], the effect of key schedules [1] and measures for data complexity [19], all for single linear attacks. However, we still do not understand the situation in multidimensional linear cryptanalysis.

A collection of linear approximations has a *capacity* which measures their bias to the uniform distribution. One important open problem in multidimensional linear cryptanalysis is to estimate the capacity and data complexity when a large number of different keys are considered. In previous work, the capacity was assumed to hold an average value constantly for most of the keys, and the data complexity was usually measured by reciprocal of the average capacity. However, neither is correct. As we know, the key equivalent hypothesis has been questioned for single linear approximations and differential trails [9,5,12]. Now this hypothesis also requires adjustment in multidimensional linear setting.

Also, it has always been difficult to compute average data complexity over the keys in linear cryptanalysis. Using Jensen's inequality, Murphy [22] points out that the Fundamental Theorem [24] can only give a lower bound for the average data complexity when a collection of linear trails in a linear approximation is used. Leander shows that in single linear attacks we should focus on median complexity instead of average complexity since the latter usually turns to infinity [19]. Both Murphy's and Leander's concerns haven't been addressed yet in the scenario of multidimensional linear attacks.

As one of the most powerful variants of linear attacks, multidimensional linear attacks notably benefit the data complexity, both in theory and in practice [10,16,11,23,15]. Moreover, the multidimensional linear distinguisher has been discovered to have connections with other statistical distinguishers, e.g., truncated differential distinguishers [6], statistical saturation distinguishers [19], and integral distinguishers [8]. All the above suggests the importance of multidimensional linear cryptanalysis, hence, the lack of knowledge on fundamental aspects of this attack is especially surprising, and deserves more attention.

Our Contributions. In this paper, we point out that under a reasonable assumption, the distribution of key-dependent capacity can be explicitly formulated with a Gamma distribution, depending on average linear potential and dimension (Sect. 3). This distribution is verified experimentally on the round-reduced PRESENT cipher. Then, we derive the distribution of data complexity, an Inverse Gamma distribution based on the same parameters (Sect. 4). Our results allow a more accurate measurement for multidimensional linear attacks.

With these distributions, in Sect. 5 we discuss three well-known measures when considering the data complexity of multidimensional linear attacks: the reciprocal of average capacity, the average and the (general) median complexity.

The following fundamental questions in single linear attacks are then generalized to multidimensional linear attacks and solved.

Firstly, we consider the standard key equivalence hypothesis. We discover that instead of holding for a majority of keys, the average capacity actually holds for less than half of the keys, no matter how many linear approximations are used. Hence, we modify the hypothesis in a way which is more in line with the practical situation.

Secondly, as we know, the average data complexity of single linear attacks is difficult to calculate, since the linear hull effect may result in zero correlation for some keys. However, we show that the situation changes when multiple linear approximations are involved, and in this case the average data complexity can be easily calculated from the Inverse Gamma distribution. Then, by generalizing Murphy's idea from the case of linear hulls to the case of multiple linear approximations, the reciprocal of average capacity is proved to be only a lower bound of the average data complexity. We also figure out the exact difference between this lower bound and the average data complexity.

Thirdly, we solve the open problem proposed by Leander in [19] by developing the usage of median complexity to multidimensional linear attacks. Finally, all measures of data complexity are compared under different dimensions. An interesting observation is that, the median complexity infinitely approaches to the average one as the dimension increases.

In Sect. 6, we revisit Cho's 25 rounds of multidimensional linear attack on PRESENT [10], which targets the most rounds of PRESENT with data complexity less than the whole codebook. As an application of our theoretical analysis, we can directly estimate the average capacity, instead of making a complex proof like [10]. Our results are very close to Cho's. Moreover, the exact knowledge of the capacity distribution allows us to compute the ratio of weak keys precisely. Using Cho's attack method by changing some parameters in the attack, $2^{123.24}$ weak keys for 26 rounds PRESENT can be recovered with no more than $2^{62.5}$ plaintext-ciphertext pairs.

2 Preliminaries

2.1 Block Ciphers and Linear Cryptanalysis

Let $\mathbb{F}_2$ be the binary field with two elements and $\mathbb{F}_2^n$ be the n-dimensional vector space over $\mathbb{F}_2$. The inner product on $\mathbb{F}_2^n$ is defined by $a \cdot b = \sum_{i=1}^{n} a_i b_i$, where a, $b \in \mathbb{F}_2^n$.

A *block cipher* is a mapping $E : \mathbb{F}_2^n \times \mathbb{F}_2^\kappa \to \mathbb{F}_2^n$ with $E_k(\cdot) \stackrel{def}{=} E(k, \cdot)$ for each $k \in \mathbb{F}_2^\kappa$. If $y = E_k(x)$, x, y and k are referred to as the plaintext, the ciphertext and the master key, respectively. A *key-alternating cipher* is a block cipher consisting of an alternating sequence of unkeyed rounds and simple bitwise key additions.

Linear cryptanalysis uses a linear relation between bits from x, y and k. A *linear approximation* (u, v) is a probabilistic linear relation expressed as a

boolean function of these bits, i.e.,

$$B(k) \stackrel{def}{=} u \cdot x \oplus v \cdot E_k(x), \tag{1}$$

where (u, v) is called the *text mask*. $B(k)$ is a boolean random variable characterized by

$$p(k) \stackrel{def}{=} \Pr_{x \in \mathbb{F}_2^n}(B(k) = 0).$$

We call $c(k) = 2p(k) - 1$ the fixed-key *correlation* of the linear approximation (u, v). The *linear potential* (LP) [24] of approximation (u, v) is defined as $LP(k) = c(k)^2$. Both $c(k)$ and $LP(k)$ vary over different keys, and can be regarded as real-value random variables over the whole key space.

In a linear approximation (u, v), there may be many paths with different intermediate masks, but sharing the same input and output mask (u, v). A path that considers linear relation round by round is called as *linear trail* (or *linear characteristic*). Note that in a key-alternating cipher, the LP of a linear trail[1] is independent of the subkeys.

2.2 Multidimensional Linear Approximations and Data Complexity

Multidimensional linear attacks use m approximations with linearly independent text masks, called *base approximations*, to construct an m-dimensional vectorial boolean function f. Let $p = (p_0, p_1, \ldots, p_{2^m-1})$ be the probability distribution of f. It can be computed by the following lemma.

Lemma 1. *([15, Corollary 1]) Let $f : \mathbb{F}_2^n \mapsto \mathbb{F}_2^m$ be a vectorial boolean function with the probability distribution p. Then, we have*

$$c_a = \sum_{\eta \in \mathbb{F}_2^m} (-1)^{a \cdot \eta} p_\eta, \text{ for all } a \in \mathbb{F}_2^m$$

and

$$p_\eta = 2^{-m} \sum_{a \in \mathbb{F}_2^m} (-1)^{a \cdot \eta} c_a, \text{ for all } \eta \in \mathbb{F}_2^m.$$

Here, c_a is the correlation of the boolean function $a \cdot f$, $a \in \mathbb{F}_2^m$.

In multidimensional linear attack, c_a is indeed the correlation of the approximation that combines the base approximations linearly.

Let $q = (q_0, ..., q_{2^m-1})$ be another discrete probability distribution of an m-bit random variable. Then, the capacity of p and q is defined as follows.

[1] Hereafter, whether the LP is of a linear approximation or of a linear trail will be clear from the context.

Definition 1. *The capacity between two probability distributions p and q is defined by*

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

The capacity of multidimensional linear approximations with probability distribution p is $C(p) = C(p,\theta)$, where θ is the uniform distribution.

Lemma 2. *([15, Corollary 2]) Given an m-dimensional vectorial boolean function f with the probability distribution p, the capacity is*

$$C(p) = \sum_{a\in\mathbb{F}_2^m, a\neq 0} c_a^2.$$

Thus, the capacity of multidimensional linear approximations is computed from m base approximations and other 2^m-1-m approximations that are XOR sum of the m base approximations. These 2^m-1-m approximations, denoted as *combined approximations*, are linearly spanned from the m base approximations.

To estimate the data complexity of multidimensional linear cryptanalysis, the Chernoff information D^* can be considered [2].

Theorem 1. *([2, Theorem 1]) Let $BestAdv_N(p,q)$ be the best advantage for distinguishing probability distribution p from probability distribution q, using N samples. We have*

$$1 - BestAdv_N(p,q) = 2^{-ND^*(p,q)+o(N)}.$$

Hence, the data complexity is $N \approx \frac{1}{D^*(p,q)}$. When q is the uniform distribution and p is close to q, the Chernoff information can be approximated by the capacity $C(p)$, [2, Theorem 7], by

$$D^*(p,q) \simeq \frac{C(p)}{8\ln 2}.$$

In this case, when the optimal distinguisher based on LLR-statistic (or χ^2-statistic) is used, the data complexity is given as $\frac{\lambda}{C(p)}$, where λ depends on the success probability of the distinguisher.

The probability distribution p of an m-dimensional linear approximation actually varies over different keys, so does the capacity (as we will show later). Hereafter, instead of using $C(p(k))$, we use $C(k)$ to represent the variable of key-dependent capacity.

2.3 Related Distributions

Note 3. *Let $\mathcal{N}(\mu,\sigma^2)$ be the normal distribution with mean μ and variance σ^2. Let $\Gamma(\alpha,\theta)$ be the Gamma distribution under the shape-scale parametrization, with mean $\alpha\theta$, the probability density function g and the cumulative distribution function $\mathcal{G}$. If $X \sim \mathcal{N}(0,\sigma^2)$, then $X^2 \sim \Gamma(1/2, 2\sigma^2)$. Inv-Gamma($\alpha$,$\beta$) denotes the inverse-Gamma distribution with mean $\frac{\beta}{\alpha-1}$ for $\alpha > 1$. If $X \sim \Gamma(\alpha,\theta)$, then $\frac{1}{X} \sim$ Inv-Gamma(α, θ^{-1}).*

Daemen et al. give the distribution of the fixed-key LP of linear approximations when linear hull effect is considered [13].

Approximation 4. *[13, Theorem 22] Given a key-alternating cipher with independent round-keys, when the number of linear trails of (u, v) is large enough and their LP are small compared to ELP(u,v), the fixed-key correlation of (u, v), $c(k)$, which is a real-value random variable, follows*

$$c(k) \sim \mathcal{N}(0, ELP(u, v)).$$

The fixed-key $LP(k)$ follows the distribution of $\Gamma(\frac{1}{2}, 2ELP(u, v))$, with mean ELP(u,v) and variance $2ELP(u, v)^2$, where $ELP(\cdot)$ is the average linear potential of the approximation over all keys.

The $ELP(u, v)$ can be denoted as $\overline{c^2}$ and computed by the following proposition for key-alternating ciphers.

Proposition 1. *[24,12] Let E be a key-alternating block cipher and assume that all subkeys are independent. The average LP of a linear approximation is the sum of all LP of the linear trails t_j, LPT(t_j), between the input and output mask of this approximation, i.e.,*

$$ELP(u, v) = \sum_{t_j \in (u,v)} LPT(t_j).$$

3 Key-dependent Capacity in Multidimensional Linear Approximations

In this section, we study the distribution of key-dependent capacity. Let $c(k)$ (resp. $LP(k)$) be a real-value random variable representing the fixed-key correlation (resp. linear potential) of the linear approximation and we can know $c(k)$ and $LP(k)$ from Approximation 4. When multiple linear approximations are used, we use i in the subscript to denote the index of linear approximations, e.g., denote $c_i(k)$ as the fixed-key correlation of the ith linear approximation. W.l.o.g, we use $i = 1, \ldots, m$ to represent the subscript of m base approximations.

In [16], the authors claim that in practical experiments the probability distributions vary a lot over the keys while the capacity remains rather constant. However, in this section we point out that the capacity also varies over different keys from the theoretical point and give experimental verification. We focus on dealing with two cases, both existing in practical block ciphers. The two cases are identified by two different assumptions whose validity should be checked in experiments on reduced round versions of the cipher. These two cases are shown in Proposition 2 and Proposition 3, respectively.

Proposition 2. *Let us assume that in an m-dimensional linear attack using m base approximations the correlations $c_i(k)$ are i.i.d. to $\mathcal{N}(0, \overline{c^2})$ over the keys, where $\overline{c^2}$ is the average LP. If for each fixed key, the binary random variables*

associated to the base approximations are statistically independent, the fixed-key capacity of this m-dimensional linear approximation, $C(k)$, approximately follows Gamma-distribution $\Gamma(\frac{m}{2}, 2\overline{c^2})$.

Proof. Let $f_1(k), \ldots, f_m(k)$ be m linearly independent base approximations to construct the m-dimensional approximation $f(k)$, and $f(k) = (f_1(k), \ldots, f_m(k))$ is an m-dimensional vectorial boolean function with the probability distribution $p(k) = \{p_\eta(k)\}$, where $\eta \in \mathbb{F}_2^m$ and $p_\eta(k)$ is the probability that $f(k) = \eta$. Indeed, $f_i(k)$ is a binary random variable with correlation $c_i(k)$. Since $f_i(k)$ are statistically independent each other for each fixed key k,

$$p_\eta(k) = \prod_{i=1}^{m} (\frac{1}{2} + (-1)^{f_i(k)} \frac{c_i(k)}{2}), \eta \in \mathbb{F}_2^m$$

According to Definition 1,

$$C(k) = \sum_{\eta \in \mathbb{F}_2^m} (p_\eta(k) - 2^{-m})^2 / 2^{-m} = 2^m \sum_{\eta \in \mathbb{F}_2^m} (p_\eta(k) - 2^{-m})^2$$

$$= 2^m \sum_{\eta \in \mathbb{F}_2^m} (\prod_{i=1}^{m} (\frac{1}{2} + (-1)^{f_i(k)} \frac{c_i(k)}{2}) - 2^{-m})^2$$

For each fixed key, $c_i(k) \cdot c_j(k) \ll c_i(k)$,

$$C(k) = 2^m \sum_{\eta \in \mathbb{F}_2^m} [\sum_{i=1}^{m} (-1)^{f_i(k)} \frac{c_i(k)}{2 \cdot 2^{m-1}}]^2$$

$$= 2^m \sum_{\eta \in \mathbb{F}_2^m} [\frac{1}{2^{2m-2}} (\sum_{i=1}^{m} (\frac{c_i(k)}{2})^2 + 2 \sum_{i \neq j} (-1)^{f_i(k) + f_j(k)} \frac{c_i(k)}{2} \frac{c_j(k)}{2})]$$

Since $\sum_{\eta \in \mathbb{F}_2^m} \sum_{i \neq j} (-1)^{f_i(k) + f_j(k)} \frac{c_i(k)}{2} \frac{c_j(k)}{2} = 0$,

$$C(k) = \frac{2^m}{2^{2m-2}} \sum_{\eta \in \mathbb{F}_2^m} \sum_{i=1}^{m} (\frac{c_i(k)}{2})^2 = \sum_{i=1}^{m} c_i(k)^2 = \sum_{i=1}^{m} LP_i(k)$$

Since $c_i(k)$ are i.i.d. to $\mathcal{N}(0, \overline{c^2})$, $LP_i(k)$ are i.i.d to $\Gamma(\frac{1}{2}, 2\overline{c^2})$, $i = 1, \ldots, m$. Thus, $C(k)$ is the sum of m independent Gamma distribution $\Gamma(\frac{1}{2}, 2\overline{c^2})$. Hence, $C(k) \sim \Gamma(\frac{m}{2}, 2\overline{c^2})$. □

Recall that for one-dimensional linear approximations, $\overline{c^2}$ can be calculated by Proposition 1 when the dominant trails in a linear approximation are known.

Proposition 2 considers the scenario where the LP of base approximations are dominant. In this case, we approximate the capacity by summing the LP of base approximations and ignoring the LP of combined approximations (see Lemma

2). To show the reasonableness of this approximated capacity, we also bound the error of our approximation. For this part of analysis, please see Appendix B.

In the other hand, Proposition 3 considers the case that not only m base approximations but also 2^m-1-m combined approximations have non-negligible contribution to the capacity. In this case, the correlations of 2^m-1-m combined approximations are not independent any more. Thus, we derive the capacity in this case under another hypothesis.

Proposition 3. *Let us assume that in an m-dimensional linear attack using the m-dimensional linear approximation the probabilities $p_\eta(k)$ is i.i.d. to a normal distribution $\mathcal{N}(2^{-m}, \sigma^2)$, for all $\eta \in \mathbb{F}_2^m$. Then the fixed-key capacity of this m-dimensional linear approximation, $C(k)$, follows Gamma-distribution $\Gamma(\frac{2^m-1}{2}, 2 \cdot 2^m\sigma^2)$.*

Proof. Since $p_\eta(k)$ are i.i.d. to $\mathcal{N}(2^{-m}, \sigma^2)$,

$$Q = \sum_{\eta=0}^{2^m-1} \frac{(p_\eta(k) - 2^{-m})^2}{\sigma^2} \sim \chi^2(2^m - 1) = \Gamma(\frac{2^m-1}{2}, 2)$$

According to the definition of capacity,

$$C(k) = \sum_{\eta=0}^{2^m-1} \frac{(p_\eta(k) - 2^{-m})^2}{2^{-m}} = 2^m\sigma^2 Q = \Gamma(\frac{2^m-1}{2}, 2 \cdot 2^m\sigma^2)$$

□

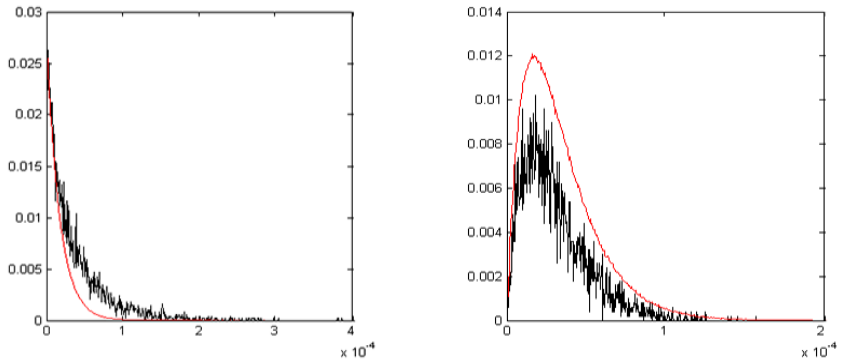
Compared with Proposition 2 which considers only m base approximations with equally dominant correlations, Proposition 3 indeed addresses the situation where the correlation $c_a(k)$ of $2^m - 1$ approximations are identically distributed (for the proof please refer to Appendix A). Thus, the average LP of $2^m - 1$ approximations are equal, denoted as $\overline{c^2}$ again. As we know, the average capacity is the sum of the average LP of involved approximations, i.e., $(2^m - 1) \cdot 2^m\sigma^2 = (2^m - 1)\overline{c^2}$, the distribution of capacity in Proposition 3 can also be represented as $\Gamma(\frac{2^m-1}{2}, 2\overline{c^2})$.

Experimental verification In order to verify that the above analysis reflects the reality with reasonable accuracy, we have experimentally computed the capacity distributions sampled from 5000 randomly chosen keys for 5-round PRESENT. A set of usable one-dimensional linear approximations is discovered in [26], with theoretical average LP computed as $2^{-16.83}$. Thus, the correlation distributions of these approximations are $\mathcal{N}(0, 2^{-16.83})$, and the LP distributions are $\Gamma(\frac{1}{2}, 2^{-15.83})$[2].

[2] For more details about the approximations used in our experiments, please refer to [26].

We can select linearly independent approximations from this set as the base approximations. Here we examine the 2-dimensional and 4-dimensional linear approximations for the case of Proposition 2.

In this case, the base approximations with input masks from different S-boxes in the first round and output masks from different S-boxes in the last round are chosen. According to Proposition 2, the theoretical distribution of 2-dimensional capacity is $\Gamma(1, 2^{-15.83})$ and of 4-dimensional capacity is $\Gamma(2, 2^{-15.83})$. The experimental distributions of 2-dimensional and 4-dimensional capacity sampled over 5000 keys are as (a) and (b) of Fig 1, respectively.



(a) 2-dimension (b) 4-dimension

Fig. 1. Experimental (black) and theoretical (red) distributions of the capacity for the 2 and 4-dimensional approximation of the first case

As illustrated in Fig 1, the experimental distribution of capacity follows the theoretical estimate closely. The scattering of data points occurs due to the fact that we basically use a histogram, and deal with raw data instead of averaging.

4 Distribution of Data Complexity

With the knowledge of capacity distribution, the distribution of data complexity, which approximates to λ times the reciprocal of capacity, can be obtained formally. Hereafter we focus on the case mentioned in Proposition 2. The case of Proposition 3 can be deduced in a similar way.

Corollary 1. *If the fixed-key capacity of the multidimensional linear approximation follows* $C(k) \sim \Gamma(\frac{m}{2}, 2c^2)$*, then the fixed-key data complexity of the corresponding multidimensional attack follows* $N(k) \sim$ *Inv-Gamma*$(\frac{m}{2}, \frac{\lambda}{2c^2})$.

Corollary 1 is derived directly from Proposition 2 (also refer to Note 3), and addresses the case that m correlations of base approximations play a prominent

role in the capacity. Since λ is a constant for any fixed success probability in an attack, w.l.o.g. hereafter we study the above data complexity distribution as Inv-Gamma($\frac{m}{2}$, $\frac{1}{2\overline{c^2}}$). For each key k, $N(k)$ is asymptotically inversely proportional to $C(k)$. The average data complexity over all keys is denoted by N, $N = E_k[N(k)]$, which is proportional to

$$E_k\left[\frac{1}{C(k)}\right] = \frac{1}{|\mathcal{K}|}\sum_{k\in\mathcal{K}}\frac{1}{C(k)},$$

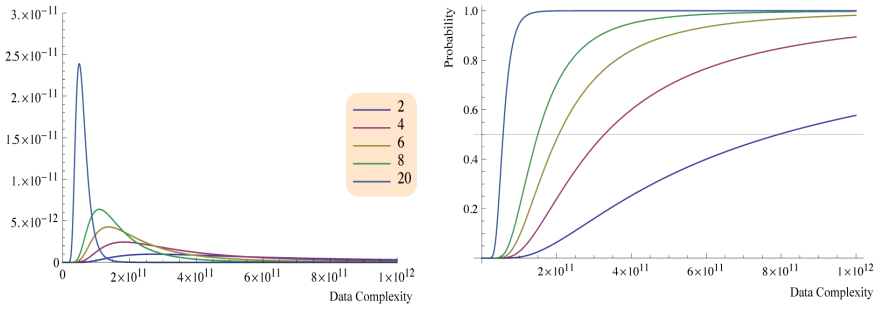
where $\mathcal{K}$ denotes the whole key space, and $E_k(\cdot)$ means an expected value taken over the whole key space. According to Corollary 1 and the mean of inverse Gamma distribution (see Note 3), the average data complexity is $E_k[\frac{1}{C(k)}] = \frac{1}{2\overline{c^2}(m/2-1)} = \frac{1}{m\overline{c^2}-2\overline{c^2}}$.

Remark. The data complexity distribution in Corollary 1 also holds for single linear attacks where $m = 1$. In the case of $m = 1$, the average data complexity is infinite as pointed out by [19][3], which corresponds to the fact that the mean of the distribution Inv-Gamma($\frac{1}{2}$, $\frac{1}{2\overline{c^2}}$) doesn't exist. When m is equal to 2, the mean of the inverse Gamma distribution also doesn't exist because there are always values going to infinite according to the distribution.

Similarly, the average capacity over the keys

$$E_k[C(k)] = \frac{1}{|\mathcal{K}|}\sum_{k\in\mathcal{K}}C(k)$$

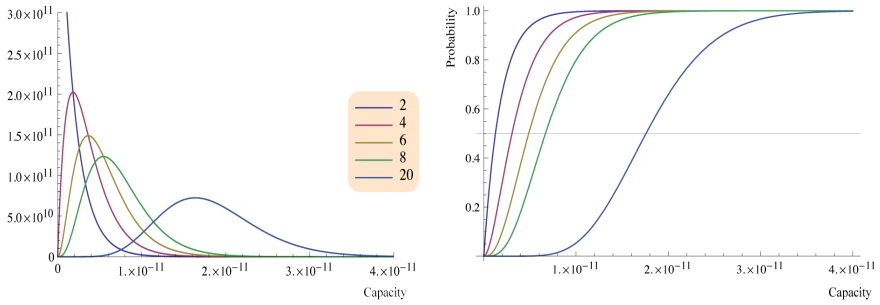
is equal to $m\overline{c^2}$, derived from the mean of the Gamma distribution in Proposition 2 (see Note 3).



(a) Probability density function (b) Cumulative distribution function

Fig. 2. Distributions of the data complexity for $m = 2, 4, 6, 8, 20$.

[3] In fact, the data complexity should be upper-bounded by the size of the codebook.



(a) Probability density function (b) Cumulative distribution function

Fig. 3. Distributions of the capacity for $m = 2, 4, 6, 8, 20$.

Example 5. *For clearer explanation, hereafter a simple example which quite meets real situations in practical ciphers is used in our analysis. We take* $\overline{c^2}$ *as* 2^{-40}*, which roughly equates the case in 15-round PRESENT, and take different* m *as 2, 4, 6, 8, 20 respectively. In this example, the distribution functions of data complexity are shown in Fig 2, and the distribution functions of capacity are shown in Fig 3.*

5 Evaluation of the Data Complexity

In practical attacks, $E_k[\frac{1}{C(k)}]$ and $\frac{1}{E_k[C(k)]}$ are highly related to the evaluation of data complexity. Since $E_k[\frac{1}{C(k)}]$ is hard to estimate, the complexity is usually measured by $\frac{1}{E_k[C(k)]}$. In this section, we firstly propose a refined key equivalent hypothesis for $E_k[C(k)]$ (Sect. 5.1). With the exact description of data complexity distributions, the difficulty of evaluating $E_k[\frac{1}{C(k)}]$ is overcome, and a basic issue about the relation of average capacity and average data complexity is studied (Sect. 5.2). We also extend Leander's idea of exploiting median data complexities [19] to multidimensional linear attacks (Sect. 5.3). Finally, all measures are compared.

5.1 Adjusted Key Equivalence Hypothesis

In regard to the connection between the fixed-key capacity and the average capacity in a multidimensional linear system, the traditional key equivalence hypothesis indicates that the fixed-key capacity does not deviate significantly from its average value [14,18]. This key equivalence hypothesis can be interpreted as follows: $C(k) \approx E_k[C(k)]$, for almost all keys k. As we have shown, the capacity is actually Gamma distributed so that this hypothesis does not hold. Thus, two questions arise: which value is suitable for the evaluation of the attack complexity? Is that average value enough and correct? We start with the following

conjecture to show that the average capacity is far from being able to represent the majority of keys.

Conjecture 1. There are always less than half of the keys having a capacity larger than the average capacity. That is, $|\{k^* \in \mathcal{K} | C(k^*) \geq E_k[C(k)]\}| < \frac{1}{2}|\mathcal{K}|$. Hence, less than half of the right keys can be recovered with a data complexity of $\frac{\lambda}{E_k[C(k)]}$, where $\mathcal{K}$ is the whole key space.

Table 1. The ratio of keys that have a capacity larger than the average capacity

m	2	4	6	8	20
ratio(%)	36.79%	40.6%	42.32%	43.35%	45.79%

This conjecture is illustrated in Table 1 with Example 5. With the increase of m, the ratio of keys that have a capacity larger than the average capacity approximates to $\frac{1}{2}$, but cannot equal $\frac{1}{2}$. This is because, for such a skew Gamma distribution as in Proposition 2, the median value is always smaller than the mean. It can be concluded that, using the number of cipher texts equal to $\frac{\lambda}{E_k[C(k)]}$, more than half of the keys cannot be recovered successfully with a reasonable probability. Thus, the average capacity is not enough to bring a sound estimation of attack complexities for most keys, especially when m is not large enough.

Since the capacity is highly dependent on the choice of the key, we would like to determine the number of plaintext-ciphertext pairs needed so that the multidimensional attack can succeed for a majority of keys. A natural way to adjust the hypothesis is to consider the upper bound of data complexity for, e.g. 90%, of the keys, meaning that for these 90% keys the amount of data texts can guarantee a successful attack with high probability, even for some of these keys this data complexity is overestimated.

Hypothesis 6. (Adjusted Key Equivalence Hypothesis) *If the capacity distribution of an m-dimensional linear attack satisfies Proposition 2, then 90% of the keys in the key space have a capacity no smaller than $\mathcal{G}^{-1}(0.1)$, where $\mathcal{G}$ is the cumulative distribution function of $\Gamma(\frac{m}{2}, 2\overline{c^2})$. Using $\frac{\lambda}{\mathcal{G}^{-1}(0.1)}$ data is enough for recovering 90% of the keys in the key space.*

5.2 On Average Data Complexity

Why the average data complexity is calculable? It is known that in the classical single linear attacks considering linear hull effect, the average data complexity is hard to derive and usually infinite because of the existence of zero correlation. This difficulty now can be solved in the situation of m-dimensional linear attacks, since the average value can be easily derived from the accurate

distributions of data complexity, when m is larger than 2. We will now look at the properties of capacity distributions to explain the reason why the average data complexity is calculable in multidimensional attacks.

In the single linear setting, the keys with zero $C(k)$ may make the average complexity infinite, thus, this part of keys should be focused on. Here, we point out that by taking multiple linear approximations simultaneously into consideration instead of only one, the number of keys with zero capacity can be very tiny so that the average complexity turns out to be computable.

We compare the ratio of keys bringing $C(k)$ between zero and ϵ, where ϵ is a fixed value very close to zero. From (b) of Fig 3, it is obvious that with the increase of m, the ratio of keys with capacity going to zero decreases. This ratio for several fixed ϵ is shown in Table 2. From Table 2 we can see that as the

Table 2. The ratio of keys with capacity close to zero for different m and ϵ

$\epsilon \setminus m$	2	4	6	8	20
10^{-16}	5.5×10^{-5}	1.5×10^{-9}	2.77×10^{-14}	3.8×10^{-19}	6.95×10^{-50}
10^{-20}	5.5×10^{-9}	1.5×10^{-17}	2.77×10^{-26}	3.8×10^{-35}	6.95×10^{-90}
10^{-25}	5.5×10^{-14}	1.5×10^{-27}	2.77×10^{-41}	3.8×10^{-55}	6.95×10^{-140}

increase of m, the ratio of keys with capacity close to zero decreases dramatically. This is because as the number of approximations grows, for each key there is higher probability that at least one approximation brings a non-zero LP, so that a non-zero capacity. Hence, for a fixed ϵ, the more base approximations are used, the less keys exist that have infinite data complexity. When ϵ is small enough and m has a reasonable size, this ratio can be negligible in the whole key space. In this case it is sound to assume that there is no key causing a zero capacity, so that the average data complexity is computable.

A difference between $E_k[\frac{1}{C(k)}]$ and $\frac{1}{E_k[C(k)]}$. The problem discussed here was first pointed out in the context of linear hull effect by Murphy [22]. We extend it to multidimensional linear attacks and make further investigation.

In some attack analysis, e.g. [10], the reduction in data complexity given by multiple approximations is based on the assertion that the data complexity N is proportional to $\frac{1}{E_k[C(k)]}$. Like the effectiveness issue of linear hull effect studied in [22], there is also a difference between $\frac{1}{E_k[C(k)]}$ and the actual average data complexity. According to Jensen's Inequality and the fact that reciprocal of positive real numbers is a convex function, we have

$$E_k\left[\frac{1}{C(k)}\right] \geq \frac{1}{E_k[C(k)]}.$$

Thus, the $\frac{1}{E_k[C(k)]}$ can only be used to give a lower bound to the average data complexity.

Jensen's Inequality gives a general comparison without considering the details of the variables. When the distributions of both $C(k)$ and $\frac{1}{C(k)}$ are known, $E_k[\frac{1}{C(k)}]$ and $\frac{1}{E_k[C(k)]}$ can be derived as in Sect. 4. Their difference is formulated as $\frac{1}{mc^2-2c^2} - \frac{1}{mc^2} = \frac{2}{m(m-2)c^2}$. Therefore, in fact the equality will never hold for m larger than 2, i.e., $E_k[\frac{1}{C(k)}]$ is always larger than $\frac{1}{E_k[C(k)]}$. The difference can be ignored only when m is large enough. Fig. 4 shows the difference for $m = 4$ and $m = 20$. For small m the difference is much more non-negligible, and $\frac{1}{E_k[C(k)]}$

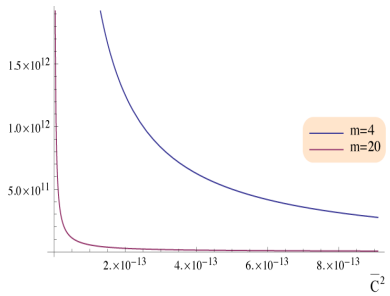


Fig. 4. The difference between $E_k[\frac{1}{C(k)}]$ and $\frac{1}{E_k[C(k)]}$ with $\overline{c^2}$ ranging from 2^{-60} to 2^{-40}.

does not reflect the real average data complexity. As more approximations are involved, the difference has a quicker trend to be small. For a fixed m, the smaller is the average LP, the larger is the difference. That is, as $\overline{c^2}$ decreases, which is a typical case since cryptanalysts always try to break as many rounds of the cipher as possible, the difference between $E_k[\frac{1}{C(k)}]$ and $\frac{1}{E_k[C(k)]}$ turns to be huge.

5.3 On Median Data Complexity

Leander proposed a way to overcome the problem of infinite data complexities for single linear attacks [19]. Namely, instead of studying the average complexity, he studied the median complexities $\widetilde{N}$ such that for half of the keys the data complexity of an attack is less than or equal to $\widetilde{N}$. So far the usage of median complexity in multidimensional linear attacks remains unsolved, which we will discuss in this section. A general definition of N_p is as follows, where $\widetilde{N} = N_{1/2}$.

Definition 2. *([19, Definition 1]) N_p is defined as the complexity such that the probability that for a given key the attack complexity is lower than N_p, is p.*

Although Leander gave this general definition, he focused on the case of $N_{1/2}$ in single linear attacks. With the knowledge of accurate distributions of

data complexity, we generalize Leander's Theorem 2 in [19] not only under the multidimensional linear model but also from $N_{1/2}$ to N_p.

Theorem 2. *Assuming independent subkeys in an m-dimensional linear attack using m base approximations with the i.i.d. LP that is $\Gamma(\frac{1}{2}, 2\overline{c^2})$, p percent of the keys yield to a capacity of at least $\mathcal{G}^{-1}(1-p)$, where $\mathcal{G}$ is the cumulative distribution function of $\Gamma(\frac{m}{2}, 2\overline{c^2})$. Thus, the complexity of this m-dimensional linear attack is less than $\frac{\lambda}{\mathcal{G}^{-1}(1-p)}$ with the probability p.*

Leander's Theorem 2 is a special case of Theorem 2 taking m as 1 and p as $\frac{1}{2}$, when the noisy linear trails are ignored in the linear hull effect (If the noisy trails are considered, the ratio of keys reduces by a factor of 2). If we explain Leander's Theorem 2 in our context, we use the fact that $\mathcal{G}^{-1}(1/2) = 0.46\overline{c^2}$, where $\mathcal{G}$ is the cumulative distribution function of $\Gamma(\frac{1}{2}, 2\overline{c^2})$ (see [19] for more details).

As illustrated in (b) of Fig 3, for the Y-axis at 1/2, the median capacity increases with the increment of m. That is, when the LP of base approximations are i.i.d., the more approximations we use, the lower data complexity we require for the same ratio of weak keys. Given a fixed capacity (so that a fixed data complexity), the ratio of keys causing a larger capacity than the fixed one increases when more base approximations are used. Thus, the ratio of weak keys resulting in a data complexity lower than the fixed one also increases.

Considering Example 5 again, we take different p, and fix the same λ (as 1 w.l.o.g.) for each m. The highest data complexity required for different m-dimensional linear attacks for p percent of keys is shown in Table 3.

Table 3. The highest data complexity for different m and different ratios of keys

m	2	4	6	8	20
$\log_2(N_{1/3})$	38.864	37.805	37.22	36.813	35.532
$\log_2(N_{1/2})$	39.529	38.253	37.581	37.123	35.727
$\log_2(N_{2/3})$	40.302	38.75	37.974	37.457	35.931

When the general median complexity N_p is applied, there is such a question: which p is more suitable for measuring and comparing the strength of a linear attack. Obviously, it is meaningless to compare $N_{1/3}$ and $N_{2/3}$ directly. A natural and simple way is to consider the value of $\frac{N_p}{p}$ because the division of p can unify the disparity for different N_p to a reasonably great extent. For example, if the attack complexity is lower than $N_{1/3}$ with probability 1/3, then the attack requires to be repeated 3 times for a sufficiently sound success rate. This should be equivalently compared with the case that, let's say, an attack with complexity lower than $N_{1/2}$ has to be repeated twice. By confirming the existence of the

minimal $\frac{N_p}{p}$, we can evaluate different multidimensional linear attacks with the value of $\min_p \frac{N_p}{p}$. The results are shown in Table 4.

Table 4. Comparison of the average data complexity, the median data complexity, the reciprocal of average capacity, and $\min_p \frac{N_p}{p}$.

m	2	4	6	8	20
$\log_2(E_k[\frac{1}{C(k)}])$	∞	39	38	37.41	35.83
$\log_2(N_{1/2})$	39.529	38.253	37.581	37.123	35.727
$\log_2(\frac{1}{E_k[C(k)]})$	39	38	37.41	37	35.68
$\log_2(\min_p \frac{N_p}{p})$	40.44	39.25	38.55	38.04	36.46

Moreover, comparing $E_k[\frac{1}{C(k)}]$, $\frac{1}{E_k[C(k)]}$ and the median complexity, we observe that the average complexity is always larger than the median one, and the median complexity is always larger than the reciprocal of average capacity. As m increases, the difference between these three values decreases. When m is large enough, these values are approximately equal (see Table 4), since the Gamma and Inverse Gamma distribution turn to be normal distributions.

6 Application to Cho's Multidimensional Attack on PRESENT

6.1 Cho's Attack on 25-round PRESENT

The structure of PRESENT [7] makes it vulnerable for a multidimensional attack: there are several strong one-dimensional approximations. The linear hull of each such approximation with non-negligible correlations consists of several equally strong single-bit trails, whose intermediate masks have Hamming weight one. The average LP $\overline{c^2}$ of all such approximations are $2^{2(-2r)}L(r)$ [26], where $L(r)$ is the number of r-round trails in each approximation. The so far best result for PRESENT is proposed by Cho aiming to 25 rounds [10]. Nine 23-round m-dimensional linear approximations are used simultaneously, and each of them has the dimension $m = 8$ 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. They recover 16 bits of key in the first round and 16 bits of key in the last round. Please refer to [10] for more details of this attack. Cho proved that the average capacity is $2^{-52.77}$, and gave the formula of data complexity as in [10]:

$$N = (\sqrt{advantage \cdot 4 \cdot M} + 4(\Phi^{-1}(2P_s - 1))^2)/C = \lambda/C \qquad (2)$$

where Φ is the cumulative distribution function of the normal distribution, P_s is the success probability, $C(p)$ is the capacity, M is the number of linear approximations used in the attack. In Equation (2), if the advantage is equal to a

bits, then the right key candidate should be within the position of $2^{\ell-a}$, where ℓ is the number of targeted key bits. Cho chose the $\lambda = 2^{9.08}$ (*advantage* is 32 bits, $M = 9 \cdot (2^8 - 1)$, $P_s = 0.95$)[4] , and estimated the average data complexity about $2^{61.85}$.

6.2 Our Investigation on Cho's Attack

We give a simpler but close estimation on the capacity and data complexity of Cho's attack. The authors in [16] claimed that Cho observes in practical experiments that the probability distribution of multidimensional linear approximations varies a lot with the keys, while the capacity remains rather constant. We have shown that the capacity also varies for different keys from theoretical and experimental viewpoints.

In order to attack 25-round PRESENT, 23-round approximations are used, thus $r = 23$. According to [26], $L(23) = 367261713$, thus $\overline{c^2} = 2^{-63.55}$. With Proposition 2 and Proposition 3, the fixed-key capacity of 9 8-dimensional approximations is estimated to be $\Gamma(9 \cdot \frac{2^8-1}{2}, 2^{-62.55})$. Hence, the average capacity is $2^{-52.39}$. With the same λ as Cho, we obtain the data complexity $N = \frac{2^{9.08}}{C(k)} \sim \text{Inv-Gamma}(9 \cdot \frac{2^8-1}{2}, 2^{71.63})$. The average data complexity is $2^{61.47}$. This result is very close to the estimate in Cho's attack, but easier to compute.

In the same way, we see that the capacity distribution which was used by Cho in his key-recovery attack on 26-round PRESENT, is distributed approximately as $\Gamma(9 \cdot \frac{2^8-1}{2}, 2^{-65.16})$. With the knowledge of distributions, we can derive the exact number of weak keys corresponding to different attack scenarios. Using Cho's attack method by taking $\lambda = 2^{7.58}$ (*advantage* is 4 bits, $P_s = 0.8$), there are now $2^{123.24}$ (3.7% in the whole key space) weak keys with capacity larger than $2^{-54.92}$. That means, for $2^{123.24}$ keys out of 2^{128} keys, 26-round PRESENT can be attacked using less than $2^{62.5}$ plaintext/ciphertext pairs, with success probability 0.8.

7 Conclusion and Further Work

In this paper, we deal with the multidimensional linear attacks using m base approximations with i.i.d. correlations (linear potential). We focus more on the case where the base linear approximations can be regarded as statistically independent. In this case, we point out that the capacity of multidimensional linear approximations satisfies a Gamma distribution, which also leads to an exact Inverse Gamma distribution for the data complexity. Both distributions are parametrized by the dimension and the average linear potential of each approximation. These theoretical results have been verified by experiments on PRESENT. We establish an explicit connection between the fixed-key behaviour

[4] This result is slightly different from [10], since Eq. (2) is slightly corrected in [16] and our computation uses the corrected formula.

and the average behaviour. Based on the distributions, several fundamental issues are discussed in more detail. Multidimensional linear attacks not only benefit from data complexity, but also offer more convenience for measuring the average data complexity due to the fact that the ratio of keys with capacity going to zero decreases with the increase of dimension. The relation of the median and average data complexity, as well as the inverse of average capacity is derived. When the dimension is large enough, these three values are infinitely close. We also propose a modified key equivalent hypothesis that is more suitable for practical situations. Finally, the multidimensional linear attack on 25- and 26-round PRESENT is analyzed based on our theoretical result.

In future work, more complicated cases about the relations of LP distributions should be studied, which may bring more precise evaluation on multidimensional attacks. The measure of $\frac{N_p}{p}$ can be extended to single linear attacks. Moreover, given the close relation between statistical saturation attacks and multidimensional linear attacks, our results may allow a clearer understanding for the capacity of statistical saturation attacks, whose key-dependent performance still lacks accurate measurement.

References

1. Mohamed Ahmed Abdelraheem, Martin Ågren, Peter Beelen, and Gregor Leander. On the Distribution of Linear Biases: Three Instructive Examples. In *CRYPTO*, pages 50–67, 2012.
2. Thomas Baignères and Serge Vaudenay. The Complexity of Distinguishing Distributions (Invited Talk). In *Information Theoretic Security*, volume 5155 of *LNCS*, pages 210–222. Springer Berlin Heidelberg, 2008.
3. Eli Biham. On Matsui's Linear Cryptanalysis. In *EUROCRYPT*, pages 341–355, 1994.
4. Alex Biryukov, Christophe De Cannière, and Michaël Quisquater. On Multiple Linear Approximations. In *Crypto*, volume 3152, pages 1–22. Springer-Verlag, 2004.
5. Céline Blondeau, Andrey Bogdanov, and Gregor Leander. Bounds in Shallows and in Miseries. In *CRYPTO*, pages 204–221, 2013.
6. Céline Blondeau and Kaisa Nyberg. New Links between Differential and Linear Cryptanalysis. In *EUROCRYPT*, pages 388–404, 2013.
7. Andrey Bogdanov, Lars R. Knudsen, Gregor Leander, Christof Paar, Axel Poschmann, Matthew J. B. Robshaw, Yannick Seurin, and C. Vikkelsoe. PRESENT: An Ultra-Lightweight Block Cipher. In *CHES*, pages 450–466, 2007.
8. Andrey Bogdanov, Gregor Leander, Kaisa Nyberg, and Meiqin Wang. Integral and Multidimensional Linear Distinguishers with Correlation Zero. In *ASIACRYPT*, pages 244–261, 2012.
9. Andrey Bogdanov and Elmar Tischhauser. On the Wrong Key Randomisation and Key Equivalence Hypotheses in Matsui's Algorithm 2. *FSE 2013*, 2013.
10. Joo Yeon Cho. Linear Cryptanalysis of Reduced-Round PRESENT. In *Topics in Cryptology - CT-RSA 2010*, volume 5985 of *LNCS*, pages 302–317. Springer Berlin Heidelberg, 2010.
11. Joo Yeon Cho, Miia Hermelin, and Kaisa Nyberg. A New Technique for Multidimensional Linear Cryptanalysis with Applications on Reduced Round Serpent. In *ICISC*, pages 383–398, 2008.

12. Joan Daemen and Vincent Rijmen. *The Design of Rijndael: AES - The Advanced Encryption Standard*. Information Security and Cryptography. Springer, 2002.
13. Joan Daemen and Vincent Rijmen. Probability Distributions of Correlation and Differentials in Block Ciphers. In *Journal of Mathematical Cryptology*, volume 1, pages 221–242. 2007.
14. Carlo Harpes, Gerhard G. Kramer, and James L. Massey. A Generalization of Linear Cryptanalysis and the Applicability of Matsui's Piling-Up Lemma. In *EUROCRYPT*, pages 24–38, 1995.
15. Miia Hermelin, Joo Yeon Cho, and Kaisa Nyberg. Multidimensional Linear Cryptanalysis of Reduced Round Serpent. In *ACISP*, pages 203–215, 2008.
16. Miia Hermelin and Kaisa Nyberg. Linear cryptanalysis using multiple linear approximations. In Pascal Junod and Anne Canteaut, editors, *Advanced Linear Cryptanalysis of Block and Stream Ciphers*. IOS Press, 2011.
17. Burton S. Kaliski Jr. and Matthew J. B. Robshaw. Linear Cryptanalysis Using Multiple Approximations. In *CRYPTO*, pages 26–39, 1994.
18. Xuejia Lai and James L. Massey. Markov Ciphers and Differentail Cryptanalysis. In *EUROCRYPT*, pages 17–38, 1991.
19. Gregor Leander. On Linear Hulls, Statistical Saturation Attacks, PRESENT and a Cryptanalysis of PUFFIN. In *EUROCRYPT*, volume 6632 of *LNCS*, pages 303–322. Springer Berlin Heidelberg, 2011.
20. Mitsuru Matsui. Linear Cryptoanalysis Method for DES Cipher. In *EUROCRYPT*, pages 386–397, 1993.
21. Sean Murphy. The Independence of Linear Approximations in Symmetric Cryptanalysis. In *Information Theory, IEEE Transactions on*, volume 52, pages 5510–5518. 2006.
22. Sean Murphy. The Effectiveness of the Linear Hull Effect. In *Journal of Mathematical Cryptology*, volume 6, pages 137–148. 2012.
23. Phuong Ha Nguyen, Lei Wei, Huaxiong Wang, and San Ling. On Multidimensional Linear Cryptanalysis. In *ACISP*, pages 37–52, 2010.
24. Kaisa Nyberg. Linear Approximation of Block Ciphers. In *EUROCRYPT*, volume 950 of *LNCS*, pages 439–444. Springer Berlin Heidelberg, 1995.
25. Kaisa Nyberg. Correlation Theorems in Cryptanalysis. In *Discrete Applied Mathematics*, volume 111, pages 177–188, 2001.
26. Kenji Ohkuma. Weak Keys of Reduced-Round PRESENT for Linear Cryptanalysis. In *SAC*, pages 249–265, 2009.

A Appendix - Proof of Lemma 7

Lemma 7. *For an m-dimensional linear approximation with the probability distribution $p_\eta(k)$ i.i.d. to the normal distribution $\mathcal{N}(2^{-m}, \sigma^2)$, $\eta = 0,\ldots,2^m - 1$, the correlations $c_a(k)$ ($a \in \mathbb{F}_2^m$, $a \neq 0$) of the involved $2^m - 1$ approximations are identically distributed.*

Proof. According to Lemma 1, for $a \neq 0$,

$$c_a(k) = \sum_{\eta \in \mathbb{F}_2^m} (-1)^{a \cdot \eta} p_\eta(k)$$

$$= \sum_{\eta \in \mathbb{F}_2^m} (-1)^{a \cdot \eta} (p_\eta(k) - 2^{-m} + 2^{-m})$$

$$= \sum_{\eta \in \mathbb{F}_2^m} (-1)^{a \cdot \eta} (p_\eta(k) - 2^{-m}) + \sum_{\eta \in \mathbb{F}_2^m} (-1)^{a \cdot \eta} 2^{-m}$$

As $p_\eta(k)$ are i.i.d. to the normal distribution $\mathcal{N}(2^{-m}, \sigma^2)$, $p_\eta(k) - 2^{-m}$ are i.i.d. to $\mathcal{N}(0, \sigma^2)$. Thus,

$$\sum_{\eta \in \mathbb{F}_2^m} (-1)^{a \cdot \eta} (p_\eta(k) - 2^{-m}) \sim \mathcal{N}(0, 2^m \sigma^2)$$

As $\sum_{\eta \in \mathbb{F}_2^m} (-1)^{a \cdot \eta} 2^{-m}$ is equal to 0, $c_a(k)$ are identically distributed to the normal distribution $\mathcal{N}(0, 2^m \sigma^2)$, where $a \in \mathbb{F}_2^m$ and $a \neq 0$. □

B Appendix - Error Bound of Proposition 2

In Proposition 2, the binary random variables associated to the base approximations are statistically independent, for each fixed key. According to Piling-up Lemma, the LP of combined approximations is equal to the multiplication of the corresponding base LPs. Thus, the accurate capacity is the summation of LP of all base and combined approximations (see Lemma 2):

$$C(k) = LP_1(k) + \cdots + LP_m(k) + LP_1(k) \times LP_2(k) + \cdots + LP_1(k) \times LP_2(k) \times \cdots \times LP_m(k)$$

$$= \prod_{i=1}^{m} (LP_i(k) + 1) - 1,$$

while our approximated capacity in Proposition 2 is $\sum_{i=1}^{m} LP_i(k)$. Their difference is

$$\prod_{i=1}^{m} (LP_i(k) + 1) - 1 - \sum_{i=1}^{m} LP_i(k)$$

$$< (\frac{\sum_{i=1}^{m} LP_i(k) + m}{m})^m - 1 - \sum_{i=1}^{m} LP_i(k)$$

In practical attacks, $LP_i(k) \ll 1$ is natural and reasonable. Denote $\sum_{i=1}^{m} LP_i(k)$ as A, and $A \ll 1$. The above formula can be written as

$$(\frac{A+m}{m})^m - 1 - A = (\frac{A}{m} + 1)^m - 1 - A = 1 + \sum_{i=1}^{m} \frac{C_m^i}{m^i} A^i - 1 - A$$

$$= \sum_{i=2}^{m} \frac{C_m^i}{m^i} A^i < \sum_{i=2}^{m} A^i < (m-1) A^2$$

In our case, A is a random variable distributed to $\Gamma(\frac{m}{2}, 2\overline{c^2})$. The expected value of A, $E(A)$, is $m\overline{c^2}$. The variance of A, $D(A)$, is $m/2 \times (2\overline{c^2})^2$. The expected value of A^2, $E(A^2)$, is equal to $D(A) + [E(A)]^2$, i.e.,

$$E(A^2) = D(A) + [E(A)]^2$$

$$= m(m+2)(\overline{c^2})^2$$

Thus, the expected value of the error is less than $(m-1)m(m+2)(\overline{c^2})^2$, which is reasonably smaller than the expected value of our approximated capacity, $m\overline{c^2}$. As we target towards attacking more and more rounds of the cipher, in average $\overline{c^2}$ tends to be close to the inverse of the message space, for example, 2^{-64}, meaning that the error is negligible in this case.