Multivariate Profiling of Hulls for Linear Cryptanalysis

Andrey Bogdanov, Elmar Tischhauser and Philip S. Vejre

Technical University of Denmark, Denmark
{anbog,ewti,psve}@dtu.dk

Abstract. Extensions of linear cryptanalysis making use of multiple approximations, such as multiple and multidimensional linear cryptanalysis, are an important tool in symmetric-key cryptanalysis, among others being responsible for the best known attacks on ciphers such as Serpent and PRESENT. At CRYPTO 2015, Huang et al. provided a refined analysis of the key-dependent capacity leading to a refined key equivalence hypothesis, however at the cost of additional assumptions. Their analysis was extended by Blondeau and Nyberg to also cover an updated wrong key randomization hypothesis, using similar assumptions. However, a recent result by Nyberg shows the equivalence of linear dependence and statistical dependence of linear approximations, which essentially invalidates a crucial assumption on which all these multidimensional models are based.

In this paper, we develop a model for linear cryptanalysis using multiple linearly independent approximations which takes key-dependence into account and complies with Nyberg's result. Our model considers an arbitrary multivariate joint distribution of the correlations, and in particular avoids any assumptions regarding normality. The analysis of this distribution is then tailored to concrete ciphers in a practically feasible way by combining a signal/noise decomposition approach for the linear hulls with a profiling of the actual multivariate distribution of the signal correlations for a large number of keys, thereby entirely avoiding assumptions regarding the shape of this distribution.

As an application of our model, we provide an attack on 26 rounds of PRESENT which is faster and requires less data than previous attacks, while using more realistic assumptions and far fewer approximations. We successfully extend the attack to present the first 27-round attack which takes key-dependence into account.

Keywords: linear cryptanalysis · multivariate · multidimensional cryptanalysis · key-dependence · PRESENT · key recovery · statistical attack

1 Introduction

Proposed by Matsui [Mat93, Mat94b] in the early 1990s, linear cryptanalysis has proven to be a seminal cryptanalytic technique for symmetric-key cryptography. Most prominently, linear cryptanalysis was successfully applied to the former U.S. encryption standard DES, breaking it experimentally for the first time. Influential cipher design paradigms, such as the wide-trail strategy [DR02], were specifically developed as a response to the advert of linear and differential cryptanalysis. Today, every newly proposed keyed symmetric primitive is expected to be accompanied by strong evidence of resistance against this attack.

In the last two decades, a number of advanced variants of linear cryptanalysis have been developed, among others differential-linear cryptanalysis [LH94], multiple linear cryptanalysis [JR94, BCQ04], multidimensional linear cryptanalysis [HCN08, HCN09, HN12],

zero-correlation linear cryptanalysis [BR14], and key-invariant bias attacks [BBR+13]. These extensions of linear cryptanalysis have provided the best single-key cryptanalytic results on ciphers such as Serpent [NWW11], PRESENT [Cho10, ZZ15], CLEFIA [BGW+13], CAST-256 [ZWW14], and LBlock-s [XJHL15].

Parallel to the development of these cryptanalytic results, extensive research has been carried out to deepen our understanding of linear cryptanalysis [AÅBL12] and its extensions [BLN14], e.g. concerning links between differential and linear cryptanalysis [BN13] and truncated differential and multidimensional linear techniques [BN14]. How to provide resistance against these advanced cryptanalysis techniques has been studied in [BBV15, SLG+16].

Key-dependence in Linear Cryptanalysis. Linear cryptanalysis relies on identifying linear relations between the input and output bits of a cipher which exhibit large linear correlations. The correlation can be viewed as a random variable over the space of inputs as well as over the space of encryption keys. A central question in linear cryptanalysis is therefore this: *What is the stochastic behaviour of the linear correlation?*

While early analysis assumed that this behaviour was largely identical for all keys [Mat93, Mat94b, Jun01, Sel08, Bih94, ZWW14], and so only depends on the randomness of the plaintexts, several works have demonstrated that this is not true in general [AÅBL12, Lea11], and models have been developed for the key-dependent behaviour of the correlation of a single linear approximation [DR07, BT13]. These models assert that the linear correlation follows a normal distribution, both in the case of a random permutation and specific block ciphers.

Even though we have a good understanding of the key-dependent behaviour of *single* approximations, it is only recently that the key-dependent behaviour of *multiple* approximations has been studied, despite the relatively large number of multiple and multidimensional linear attacks in the literature. In this work, we consider the three principal papers on this topic and reflect on the precise assumptions used by the models developed by them. We then develop a new model which aims to remove many of these assumptions in order to obtain more accurate estimates of the power of linear attacks.

State of the Art and Problems. There are three principal works considering key-dependence in the context of multiple and multidimensional linear cryptanalysis. First, [HVLN15] by Huang et al. considers the key-dependent behaviour of the multiple and multidimensional capacity and develops a model in which this follows a gamma distribution under the assumption that the individual correlations are independently and identically distributed. Second, [BN17] by Blondeau and Nyberg relaxes the assumptions of [HVLN15] such that the correlation distributions need not have identical means, which results in a model that describes the capacity as a scaled, non-central χ^2-distribution. However, this model assumes an accurate estimate of the parameters of the correlation distributions. Blondeau and Nyberg relaxed this assumption in [BN16] by incorporating the signal/noise decomposition from [BT13] into the model. Although the models developed in these works are a step on the way towards accurate assessments for multiple and multidimensional attacks, we identify the following main problems with the approaches:

- **Independence assumptions:** Multidimensional linear cryptanalysis was originally introduced to solve the requirement for statistically independent approximations, but recently Nyberg showed [Nyb17] that under reasonable assumptions about pairwise statistical independence, linear independence and statistical independence of approximations are equivalent concepts. Multidimensional linear cryptanalysis uses many linearly dependent approximations, but the models described above often assume these to be statistically independent for a random permutation. Moreover,

the models are typically derived in a setting with independent round keys – a setting that does not strictly reflect most actual ciphers.

- **Restricted approximation choice:** The models described above restrict which approximations can be used. In the case of multiple linear cryptanalysis, equal correlation variances are required, and so we cannot necessarily freely choose the approximations that best facilitate an attack, as they might have different distributions. Ideally, a cryptanalyst would like to be able to pick the best trade-off between strong approximations and approximations that make the attack efficient to perform. For multidimensional linear cryptanalysis, models are given in which a set of dominant approximations are present and the rest of the approximations are treated as noise. The advantage of the multidimensional approach then seems to stem from the fact that the attacker can sometimes get a few rounds for free, if the resulting approximations still allow for efficient key guessing.
- **Parameter estimation:** As mentioned, the models of [HVLN15, BN17] require an accurate knowledge of the correlation distributions or multidimensional probability distributions. Obtaining this is extremely difficult for most reasonable block and key sizes. This problem is mostly solved in [BN16] by applying the signal/noise decomposition, but this approach is still quite computationally expensive if simplifying assumptions, such as independent round keys, are not used. In general, this problem seems to be quite difficult to avoid.

Our Results. The results of [Nyb17] poses a problem for any model of linear cryptanalysis with multiple approximations that uses linearly dependent approximations, including multidimensional linear attacks. This paper therefore revisits multiple linear cryptanalysis in the case where all approximations are linearly independent.

We first investigate the joint correlation distribution of such a set of approximations. We find that this distribution can be assumed to be jointly normal for a long-key cipher, in accordance with theory, but that this is not the case for other key-schedules. We therefore propose *multivariate linear cryptanalysis*. This model:

- Does not assume a specific key-schedule,
- Does not assume statistical independence of the correlations,
- Is able to model any arbitrary (not necessarily normal) joint correlation distribution,
- Uses signal/noise decomposition to practically obtain accurate attack estimates.

The model expresses the joint correlation distribution of M approximations as a general M-variate probability distribution. While the multivariate model relaxes many assumptions used by previous models, it comes at the cost of a larger effort during the off-line analysis of the cipher. In particular, the more accurate an estimate of the signal distribution the cryptanalyst can obtain the better. This only affects the amount of effort she has to put into the analysis, and not the effectiveness of the resulting attack. We confirm the accuracy of our model through experiments on 32-bit SmallPresent.

As a result, we are able to present new attacks on present (with an 80-bit key), which at the same time avoid the above modeling problems. Crucially, our analysis model is in accordance with [Nyb17]. We identify a very sparse set of 135 approximations over 22 rounds, and use these to attack 26 rounds of present. The computational complexity of this attack is $2^{68.6}$, while the data complexity is $2^{63.0}$. Interestingly, this attack is about 11 times faster than Cho's original attack on the same number of rounds, and uses half the data, all the while using far fewer approximations and more realistic assumptions. This demonstrates that a multidimensional linear attack is not necessarily stronger than a

Table 1: Comparison of attacks on PRESENT. The attacks of [Cho10] and [ZZ15] do not take the key-dependence into account. All models, except the one presented in this work, use assumptions that contradict the equivalence of linear independence and statistical independence of linear correlations shown in [Nyb17].

Rounds	Success probability	#Approximations	Time complexity	Data complexity	Memory complexity	F1: Key-dependent	**F2: Complies with [Nyb17]**	T1: Signal/noise	T2: Profiling	Reference
25	95%	2295	$2^{65.0}$	$2^{62.4}$	$2^{34.0}$		N/A			[Cho10]
	95%	2295	$2^{65.0}$	$2^{61.6}$	$2^{34.0}$	✓				[HVLN15]
	74%	2295	$2^{72.0}$	$2^{61.0}$	$2^{34.0}$	✓				[BN16]
26	95%	2295	$2^{72.0}$	$2^{64.0}$	$2^{34.0}$		N/A			[Cho10]
	80%	2295	$2^{76.0}$	$2^{62.5}$	$2^{34.0}$	✓				[HVLN15][a]
	51%	2295	$2^{72.0}$	$2^{63.8}$	$2^{34.0}$	✓		✓		[BN16]
	95%	**135**	$\mathbf{2^{68.6}}$	$\mathbf{2^{63.0}}$	$\mathbf{2^{48.0}}$	✓	✓	✓	✓	**Section 6.2**
27	95%	405	$2^{74.0}$	$2^{64.0}$	$2^{70.0}$		N/A			[ZZ15]
	95%	**135**	$\mathbf{2^{77.3}}$	$\mathbf{2^{63.8}}$	$\mathbf{2^{48.0}}$	✓	✓	✓	✓	**Section 6.3**[b]

Feature/Technique	Explanation
F1: Key-dependent	The model accounts for the fact that the linear correlation of an approximation varies over the key space.
F2: Complies with [Nyb17]	The model does not assume that linearly dependent approximations of a random permutation are statistically independent. Doing so contradicts [Nyb17].
T1: Signal/noise	The model uses the signal/noise decomposition of [BT13] to obtain accurate estimates of the correlation distributions.
T2: Profiling	The model measures the actual multivariate distribution of the signal for a large number of keys to avoid assumptions of the shape of this distribution.

[a] For 3.7% of the key space.
[b] Uses distinct texts. All other attacks use non-distinct texts.

multiple linear attack. We extend the attack to 27 rounds, resulting in a computational complexity of $2^{77.3}$ and a data complexity of $2^{63.8}$. This is the first attack on 27 rounds of PRESENT in a model that accounts for key-dependence. Our attacks are compared to previous attacks on PRESENT in Table 1.

2 Preliminaries

We consider a block cipher $E(P, K) : \mathbb{F}_2^n \times \mathbb{F}_2^\kappa \to \mathbb{F}_2^n$ with a block size of n bits and key length of κ bits. For each key $K \in \mathbb{F}_2^\kappa$, $E_K := E(\cdot, K)$ is a permutation on $\mathbb{F}_2^n$. If a block cipher picks a permutation uniformly at random from the space of all $(2^n)!$ permutations for each key, we say that it is *ideal*.

Most modern block ciphers are *iterative block ciphers* where the encryption function is a composition of r key-dependent round functions. If each round function can be described as a key-independent transformation followed by an XOR of the round key, we call the cipher a *key-alternating cipher*. Additionally, an initial key is XOR'ed to the input before

the first round. Usually, a *key-schedule* is used to expand the κ-bit master key K into the required $r+1$ n-bit round keys. We denote the expanded key by $\bar{K} = k_0 \| k_1 \| \dots \| k_r$, i.e. the concatenation of the round keys. If all round keys are chosen uniformly and independently, i.e. $\kappa = (r+1)n$ and $K = \bar{K}$, we call the cipher a *long-key cipher*.

2.1 Linear Cryptanalysis

Linear cryptanalysis was introduced by Matsui in 1993 [Mat93] and considers one or more *linear approximations* of a cipher. A linear approximation is a pair $(\alpha, \beta) \in \mathbb{F}_2^n \times \mathbb{F}_2^n \backslash (0,0)$, where α is called the *input mask* and β the *output mask*. The key-dependent *linear correlation* of the approximation is defined as $C_{\alpha,\beta}^K = 2\Pr(\alpha \cdot x = \beta \cdot E_K(x)) - 1$, where "$\cdot$" denotes the canonical inner product on $\mathbb{F}_2^n$, and the probability is taken over all $x \in \mathbb{F}_2^n$. Assuming that K is drawn uniformly at random, $C_{\alpha,\beta}^K$ is a random variable over the key space. If an estimate of $C_{\alpha,\beta}^K$ is calculated using N plaintext-ciphertext pairs, we denote this value by $C_{\alpha,\beta}^{K,N}$, which is a random variable over both the key and text space, where the latter is of size N. The goal of linear cryptanalysis is to find pairs (α, β) such that the probability distribution of the correlation for the block cipher in question is distinguishable from the correlation distribution of an ideal cipher.

Let (u_i, u_{i+1}), $i = 0, \dots, r-1$, be a series of one round linear approximations of an iterative block cipher. Such a series of approximations is called a *linear trail*. We can also denote the trail by the concatenation of its masks, i.e. $U = u_0 \| \dots \| u_r$. Then the *correlation contribution* of trail is defined by $C_U^K = \prod_{i=0}^{r-1} C_{u_i,u_{i+1}}^K$. The collection of all trails with $u_0 = \alpha$ and $u_r = \beta$ is called the *linear hull* of (α, β). Moreover, the correlation of (α, β) is the sum of the correlation contributions of all trails in the hull [Dae95, DR02]:

$$C_{\alpha,\beta}^K = \sum_{u_0=\alpha, u_r=\beta} C_U^K. \tag{1}$$

A useful concept is that of the *expected linear potential (ELP)*, defined by $\mathrm{E}((C_{\alpha,\beta}^K)^2)$. For a long-key cipher, it can be shown that $ELP = \sum (C_U^K)^2$, and that $(C_U^K)^2$ is independent of the key [DR02].

2.2 Statistical Distinguishing

In cryptanalysis of block ciphers, a first step towards more powerful attacks is often to build a *distinguisher*. A distinguisher aims to determine whether some observed data is the output of a specific block cipher or an ideal cipher. In statistical cryptanalysis, a distinguisher consists of performing a statistical test which distinguishes between two probability distributions. Typically, the test computes a value from the data, which we refer to as the *test statistic* $\mathcal{T}$. Note that the test statistic is a random variable. Let $\mathcal{T}_I$ be the random variable if the observed data was produced by an ideal cipher, and let $\mathcal{T}_N$ be the random variable if the observed data was produced by a specific block cipher. Assume that $\mathcal{T}_I$ and $\mathcal{T}_N$ follow univariate distributions. Then a simple and often used statistical test is to check the value of $\mathcal{T}$ against some *threshold value* τ. Without loss of generality, assume that $\mathrm{E}(\mathcal{T}_I) \leq \tau \leq \mathrm{E}(\mathcal{T}_N)$. If $\mathcal{T} \geq \tau$, we conclude that $\mathcal{T}$ was drawn from the distribution of $\mathcal{T}_N$, otherwise we conclude that $\mathcal{T}$ was drawn from the distribution of $\mathcal{T}_I$. It may be the case that we need to compare against multiple threshold values – for a discussion of this case, we refer to [BN17]. Note that we can define several different tests of the type described above, namely by calculating the test statistic $\mathcal{T}$ in different ways. We consider a commonly used test statistic in Section 6.1, namely the χ^2 test statistics.

When assessing the efficiency of a threshold test, we are mainly interested in two parameters: the *success probability* and the *advantage*. Let F_X denote the cumulative

distribution function of the random variable X. We define the probability of success as

$$P_S = 1 - F_{\mathcal{T}_N}(\tau),$$

i.e. the probability that $\mathcal{T}_N \geq \tau$. The advantage, a notion first introduced by Selçuk in [Sel08, SB02] in the context of key-ranking, is in turn defined by

$$a = -\log_2(1 - F_{\mathcal{T}_I}(\tau)),$$

and relates to the number of false successes that arises from the threshold test. This number is important when we want to use a distinguisher as part of a key recovery attack. In order to assess these quantities, we need to know the distributions of $\mathcal{T}_I$ and $\mathcal{T}_N$, and the question of determining these is therefore central to the study of linear cryptanalysis.

2.2.1 From Distinguishing to Key Recovery.

It is possible to turn a distinguisher over r rounds of an iterative block cipher into a key recovery attack over $r' > r$ rounds in a generic way. Consider the case $r' = r + 1$ as an example. Denote by E^r the r-round encryption function, and let F_k denote the last round function such that $E^{r'} = F_k \circ E^r$. Let $\bar{E}^r$ be the truncation of E^r such that only the bits required to calculate the test statistic $\mathcal{T}$ are output.

The attacker obtains some data from $E^{r'}$, and guesses the parts of k required to partially invert F_k and calculate the output of $\bar{E}^r$. The attacker then calculates the test statistic $\mathcal{T}$ and runs the distinguisher. If the attacker guessed the partial key k correctly, the distinguisher should indicate that $\mathcal{T}$ was drawn from the distribution of $\mathcal{T}_N$ with probability P_S. If not, the hypothesis is that the distinguisher will behave as if $\mathcal{T}$ was drawn from the distribution of $\mathcal{T}_I$. The reasoning here is that for a wrong key guess, the attacker is basically observing data from a cipher with $r + 2$ rounds, which should behave more like an ideal cipher than a cipher with r rounds. This idea was first formally stated by Harpes et al. [HKM95] and later stated in the context of linear cryptanalysis by Junod [Jun01]. Once all candidates for the partial key k have been tested, the attacker has to guess the remaining bits of the master key K, discarding any wrong guesses by trial encryption. By definition of the advantage, the attacker has to try $2^{\kappa-a}$ candidates.

2.3 PRESENT

PRESENT is an ultra-lightweight, key-alternating, block cipher. It is an SPN cipher with 31 rounds, a block size of 64-bit, and a key size of either 80 bit or 128 bit. Each round consists of an XOR with a round key, a layer of 16 parallel 4-bit S-boxes, and bit permutation. An additional round key is added after the last round. The 32 round keys are derived through a key-schedule. For details on the bit permutation and the key-schedule, we refer to [BKL+07]. Due to the choice of S-box, PRESENT exhibits some interesting linear properties [Ohk09]. It is therefore a common target for new linear cryptanalysis techniques. We consider new attacks on PRESENT in Section 6.

3 Survey of Previous Work

As discussed in Section 2.2, it is of primary interest to determine the distributions of $\mathcal{T}_I$ and $\mathcal{T}_N$ for a given statistical test. For linear cryptanalysis, the test statistic is derived from the observed correlation of one or more linear approximations. An equivalent question in this context is therefore what the distribution of the correlation $C^N_{\alpha,\beta}$, for a given approximation or set of approximations, looks like, both for a specific block cipher and for an ideal cipher. Starting with [Nyb94], this topic has been extensively investigated in the literature. In the

following, we consider a series of models that have been proposed since the introduction of linear cryptanalysis, and reflect on their assumptions and requirements. We divide the models into two main categories: models that assume that $C^K_{\alpha,\beta}$ is approximately equal for all keys, and models that include the influence of the key.

3.1 Models Without Key Influence

Matsui introduced linear cryptanalysis in [Mat93, Mat94b] as a means to attack DES. The approximations used for this attack exhibit a single *dominant trail* each, i.e. there exists a trail U such that $|C^K_U| \gg |C^K_{U'}|$ for any $U' \neq U$. Then by Equation 1, $C^K_{\alpha,\beta} \approx C^K_U$ for all keys. Moreover, it can be shown that for key-alternating ciphers (or ciphers that can be expressed as such, e.g. DES) the correlation contribution is given by $C^K_U = (-1)^{U\cdot\bar{K}}|C^K_U|$, where $|C^K_U|$ is independent of the key [DR02]. Thus, Matsui asserts that for DES, $C^K_{\alpha,\beta} \approx \pm|C^K_U|$ for all keys. This leads to the concept of *right-key equivalence*:

Hypothesis 1 (Right-Key Equivalence – Matsui)**.** *If a linear approximation (α, β) has a single dominant trail U, then the absolute value of the linear correlation is approximately equal for all keys, with $|C^K_{\alpha,\beta}| \approx |C^K_U|$.*

Similarly, Matsui assumed that for a wrong key guess, the correlation would be approximately zero for all keys, leading to the concept of *wrong-key randomisation*:

Hypothesis 2 (Wrong-Key Randomisation – Matsui)**.** *During a key recovery attack, the linear correlation of a linear approximation (α, β) is approximately equal to zero for all wrong keys, i.e. $C^K_{\alpha,\beta} = 0$.*

Under Hypotheses 1 and 2 the distribution of $C^{K,N}_{\alpha,\beta}$ only depends on the number N of observed plaintext-ciphertext pairs. Using a normal approximation to the binomial distribution, it can be shown that

$$C^{K_R,N}_{\alpha,\beta} \sim \mathcal{N}(\pm|C^K_U|, N^{-1}) \quad \text{and} \quad C^{K_W,N}_{\alpha,\beta} \sim \mathcal{N}(0, N^{-1}) \tag{2}$$

where K_R and K_W represents a right and wrong key guess, respectively. This and similar models have been used extensively in the literature, first in classical linear cryptanalysis [Mat93, Mat94b, Jun01, Sel08, Bih94, ZWW14], and later in its extensions *multiple linear cryptanalysis* [JR94, BCQ04] and *multidimensional linear cryptanalysis* [HCN08, HCN09, Ohk09, Cho10, HN12, ZZ15]. Notably, the best attacks on the block cipher PRESENT (both multidimensional), the 26-round attack by Cho [Cho10] and the 27-round attack by Zheng and Zhang [ZZ15], both use this model.

3.2 Models Incorporating the Key

3.2.1 Single Approximations

While the idea of identical behaviour for all keys simplifies analysis, it does not reflect the behaviour of most modern ciphers. Indeed, if the number of trails with a significant correlation contribution is large, then by Equation 1 the correlation $C^K_{\alpha,\beta}$ will take on many values over the key space. Dubbed the *linear hull effect*, this phenomenon was first discussed by Nyberg in [Nyb94]. Ohkuma later pointed out that for PRESENT this effect is very strong, as the number of trails with the same best correlation contribution is large [Ohk09]. The situation is similar for most other modern ciphers designed with resistance to linear cryptanalysis in mind. Thus, Hypothesis 1 is not true for most ciphers of interest.

Although the correlation $C^K_{\alpha,\beta}$ is a random variable over the key space, it is not immediately clear what distribution it follows. For a long-key cipher, it can be shown that the distribution is normal with mean zero and variance equal to ELP [DR05]. For other

key-schedules, the distribution has been studied in several works [BT13, Lea11, AÅBL12], and have been found to be close to normal – however, the key-schedule can have an impact on the parameters of the distribution, invalidating the veracity of Hypothesis 1. This leads to the following revised right-key hypothesis, which has been used several times in the literature [HVLN15, BN17, BN16].

Hypothesis 3 (Right-Key Randomisation – Single [DR07, BT13])**.** *The linear correlation $C^K_{\alpha,\beta}$ of a linear approximation (α, β) of a block cipher, which does not have a single dominant trail, is a random variable over the key space with distribution $C^K_{\alpha,\beta} \sim \mathcal{N}(\mu, \sigma^2)$.*

Note that by the definition of ELP and variance, we can write $\sigma^2 = ELP - \mu^2$. Moreover, for a long-key cipher, $\mu = 0$ [DR07, DR02]. For the wrong-key, the situation is a little simpler. In [DR07], Daemen and Rijmen show that the correlation distribution of an ideal cipher is normal with mean zero and variance 2^{-n}. Thus, we obtain the following hypothesis in this case.

Hypothesis 4 (Wrong-Key Randomisation – Single)**.** *During a key recovery attack, for a wrong key guess, the linear correlation $C^K_{\alpha,\beta}$ of a linear approximation (α, β) is a random variable with distribution $C^K_{\alpha,\beta} \sim \mathcal{N}(0, 2^{-n})$.*

While the picture seems clear in the case of a single approximation, moving to extensions that use multiple approximations simultaneously in order to extract more information seems to complicate matters considerably.

3.2.2 Multiple Linear Cryptanalysis

Kaliski and Robshaw first proposed the use of multiple approximations simultaneously in [JR94]. The idea was extended by Biryukov et al. in [BCQ04], where they also defined the *capacity* of a set of linear approximations as a measure of the strength of this set. For a set of M approximations $(\alpha_1, \beta_1), \ldots, (\alpha_M, \beta_M)$, the capacity is defined as

$$\mathcal{C}^K = \sum_{i=1}^{M} (C^K_{\alpha_i,\beta_i})^2. \tag{3}$$

Similar to the correlations, we denote an estimate of the capacity based on N plaintext-ciphertext pairs by $\mathcal{C}^{K,N}$. The main problem with this approach is that the linear approximations are not in general statistically independent, making the analysis of the capacity very difficult. Indeed, statistical independence was assumed in [JR94, BCQ04]. This approach is commonly referred to as *multiple linear cryptanalysis*.

3.2.3 Multidimensional Linear Cryptanalysis

To avoid the problem of independence, Hermelin et al. proposed *multidimensional linear cryptanalysis* in [HCN08, HCN09], based on the work done by Baignères et al. in [BJV04]. It considers an m-dimensional subspace of $\mathbb{F}_2^n \times \mathbb{F}_2^n$ and studies the distribution of a plaintext-ciphertext pair $(\bar{x}, \bar{E}_K(x))$ restricted to this subspace, which can be described by the vector $\boldsymbol{\eta}^K = (\eta_0^K, \ldots, \eta_{2^m-1}^K)$, where $\eta_i^K = \Pr(\bar{x} \| \bar{E}_K(x) = i)$. $\boldsymbol{\eta}^K$ is a key-dependent, 2^m-dimensional, discrete probability distribution. It can then be shown that the capacity of the set of all linear approximations in the subspace can be calculated from $\boldsymbol{\eta}^K$.

Theorem 1. *[HCN08] Consider an m-dimensional subspace of $\mathbb{F}_2^n \times \mathbb{F}_2^n$, and denote the multidimensional probabilities by η_i^K. The capacity of all linear approximations in this subspace can be calculated as*

$$\mathcal{C}^K = \sum_{i=1}^{2^m-1} (C^K_{\alpha_i,\beta_i})^2 = \sum_{i=0}^{2^m-1} \frac{(\eta_i^K - 2^{-m})^2}{2^{-m}},$$

The main advantage of multidimensional linear cryptanalysis is that it can be shown that the amount of data needed for a multidimensional distinguisher (with a fixed success probability) is inversely proportional to the capacity, regardless of statistical dependence of the associated approximations [BJV04].

While the influence of the key on the correlation of a single approximation has been studied for some time, it is only recently that versions of Hypotheses 3 and 4 have been developed for multiple and multidimensional linear cryptanalysis. In the following, we give a short summary of the contributions of the three main works in this area, and in Section 4 we consider their results in depth.

Huang et al., CRYPTO'15 [HVLN15] To the best of our knowledge, this is the first work to study the key-dependent distribution of the multidimensional capacity, although the wrong-key capacity is not considered. Under some assumptions on the one-dimensional approximations, it is shown that the capacity follows a gamma distribution. Two cases are considered giving the following results.

Result 1 ([HVLN15], Proposition 2)**.** *Consider an m-dimensional linear approximation where m linearly independent base approximations have dominant ELPs. Moreover, let the correlations of these base approximations, $C^K_{\alpha_1,\beta_1},\ldots,C^K_{\alpha_m,\beta_m}$, be i.i.d as $\mathcal{N}(0,ELP)$. Then $\mathcal{C}^K \sim \Gamma(\frac{m}{2},2ELP) = ELP\cdot\chi^2_m$.*

Result 2 ([HVLN15], Proposition 3)**.** *Consider an m-dimensional linear approximation with probability distribution $\boldsymbol{\eta}^K = (\eta^K_0,\ldots,\eta^K_{2^m-1})$. Assume that the multidimensional probabilities η^K_i are i.i.d as $\mathcal{N}(2^{-m},\sigma^2)$. Then $\mathcal{C}^K \sim \Gamma(\frac{2^m-1}{2},2^{m+1}\sigma^2) = 2^m\sigma^2\cdot\chi^2_{2^m-1}$.*

Blondeau and Nyberg, DCC'17 [BN17] This work improves upon [HVLN15] in several ways. First, both the key and data dependence are included in the models, as opposed to [HVLN15] that only consideres the exact distribution of capacity. Moreover, both sampling of the texts with and without replacement is considered; here, we will only cover the case without replacement, and refer to [BN17] and [BN15] for further details.

A model for the wrong-key is derived by using Hypothesis 4 and Theorem 1, under the assumption that approximations of ideal ciphers are statistically independent.

Result 3 ([BN17], Theorem 6)**.** *Consider a multiple or multidimensional attack using M approximations and N text pairs. Then, for a wrong key guess, $\mathcal{C}^{K,N} \sim (N^{-1}+2^{-n})\chi^2_M$.*

For the right-key, [BN17] considers a more general case where the mean of the correlations is not necessarily zero. Let $\chi^2_\ell(k)$ be the non-central χ^2-distribution with ℓ degrees of freedom and non-centrality parameter k. The following result is given.

Result 4 ([BN17], Theorem 7 and 8)**.** *Consider a multiple or multidimensional attack using M approximations and N text pairs. For a multiple attack, assume that the linear correlations of the approximations, $C^K_{\alpha_i,\beta_i}$, are independently distributed as $\mathcal{N}(\mu_i,\sigma^2)$, $i=1,\ldots,M$. For a multidimensional attack, assume that the multidimensional probabilities η^K_i are normally distributed with equal variances and that each set of M probabilities are statistically independent. Let μ_i be the mean of the correlation of the related approximation, $i=1,\ldots,M$. Then*

$$\mathcal{C}^{K,N} \sim \Delta\chi^2_M\left(\frac{N\sum\mu_i^2}{N\Delta}\right) \quad \textit{where} \quad \Delta = N^{-1}+M^{-1}\sum(ELP_i-\mu_i^2).$$

For the multidimensional probabilities, note that the assumption of statistical indepedence of sets of size M arises since $\sum\eta^K_i = 1$.

Blondeau and Nyberg, ToSC'16 [BN16] While [BN17] derives the capacity distributions under some assumptions, Result 4 requires that the cryptanalyst can get accurate estimates of the distribution parameters of the one-dimensional correlations or the multidimensional probabilities. Obtaining these is left as an open problem. [BN16] aims to solve this problem by utilising the signal/noise decomposition technique developed in [BT13].

The idea of the signal/noise decomposition is to first get an estimate of the correlation distribution by computing a part of the linear hull, i.e. some (significant) terms of Equation 1. We call this set of known trails the *signal*, denoted by $\mathcal{S}$. Then, the unknown part of the hull, i.e. the trails not in $\mathcal{S}$, are modeled as *noise* with the distribution $\mathcal{N}(0, 2^{-n})$. We will take a closer look at this method in Section 5.2. Using the signal/noise decomposition, the following result is given for the right-key distribution of capacity. Note that [BN16] uses the wrong-key result given in Result 3.

Result 5 ([BN16], Theorem 4)**.** *Given M linear approximations, assume that a signal $\mathcal{S}$ is known for ℓ approximations, and that the noise of these ℓ approximations, as well as the correlations of the remaining $M - \ell$ approximations, are statistically independent. Let $C_{\mathcal{S}} = \sum_{i=1}^{\ell} \sum_{U \in \mathcal{S}_i} (C_U^K)^2$ be the signal capacity. Then, for a long-key cipher,*

$$\mathrm{E}(\mathcal{C}^K) = C_{\mathcal{S}} + M2^{-n}, \quad \text{and} \quad \mathrm{Var}(\mathcal{C}^K) = 2\sum_{i=1}^{\ell} \left(\sum_{U \in \mathcal{S}_i} (C_U^K)^2 \right)^2 + C_{\mathcal{S}} 2^{2-n} + M2^{1-2n}.$$

4 Limitations of Current Models

The results described in Section 3 use one or more assumptions about the linear correlation distributions. Moreover, the results are not as general as a cryptanalyst might want, i.e, the situations in which they can be used are restricted in some way. In the following, we consider the validity of these assumptions and describe some of these restrictions.

4.1 Independence Assumptions

Dealing with statistical independence has long been a problem for linear cryptanalysis. Indeed, the very reason for the introduction of multidimensional linear cryptanalysis was to avoid this issue. When trying to incorporate the key-dependence in the models, however, it seems difficult to avoid assumptions on the statistical behaviour of the approximations. We note that Results 1 to 5 all use some assumptions on the statistical independence of (some of) the approximations. Recently, Nyberg proved the following theorem:

Theorem 2 ([Nyb17])**.** *Let A be a set of pair-wise statistically independent linear approximations. Then the correlations of the linear approximations in A are statistically independent if and only if they are linearly independent.*

While it is an open problem to formally prove when two approximations are statistically independent, for all practical intents and purposes, assuming pair-wise statistical independence seems reasonable in the case of random permutations of the block size used in practice. With this assumption in mind, let us consider a general set of M linear approximations, (α_i, β_i), $i = 1, \ldots, M$. We denote the vector of their correlations by $\mathbf{C}^K = (C^K_{\alpha_1,\beta_1} \cdots C^K_{\alpha_M,\beta_M})^\top$. Under the wrong-key hypothesis, Hypothesis 4, $C^K_{\alpha_i,\beta_i} \sim \mathcal{N}(0, 2^{-n})$, $i = 1, \ldots, M$. In this case, if the approximations are linearly independent, Theorem 2 asserts that $\mathbf{C}^K \sim \mathcal{N}_M(\mathbf{0}, \mathrm{diag}(2^{-n}))$. But this is not the case if the linear approximations are linearly dependent, which poses an interesting problem for the multidimensional models. In particular, not all the one-dimensional approximations are linearly independent, and so by Theorem 2, they cannot be statistically independent. The consequence for Result 3 is that it is unknown whether the capacity is χ^2-distributed in a

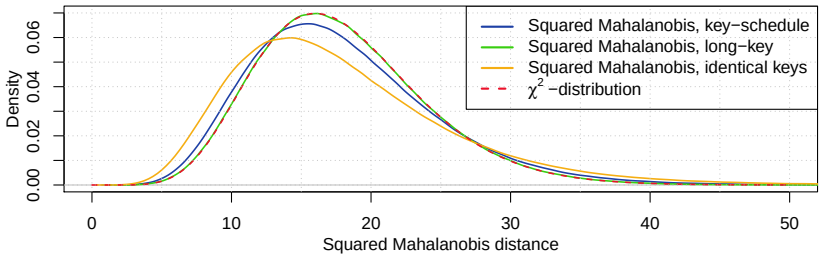

Figure 1: The densities of the squared Mahalanobis distance of the joint correlation distribution for 18 approximations over 9 rounds of 32-bit SmallPresent for three different key-schedules. The plot show a connection between dependence between the round keys, and how much the correlation distribution deviates from joint normality.

multidimensional linear attack. For a multiple linear attack the result still holds if the approximations are linearly independent.

For the right-key models, Theorem 2 has the biggest impact on Result 5. When adding noise to the model, the assumption is that the noise distributions behave as for a random permutation and are independent, but this cannot be the case for a multidimensional approximation. For Results 2 and 4, it is assumed that the multidimensional probabilities are independent, and thus Theorem 2 does not affect these models. Whether this assumption is sound is an open problem.

Finally, we note that an often used assumption when deriving these models is that the cipher is a long-key cipher, where pair-wise statistical independence might also be a reasonable assumption in practice. In this case, we could choose linearly independent approximations, and then by Theorem 2 and [DR05], $\mathbf{C}^K \sim \mathcal{N}_M(\mathbf{0}, \mathrm{diag}(ELP_i))$. However, most ciphers do not actually have independent round keys. If a key-schedule is used, we can no longer use Theorem 2 to equate linear independence with statistical independence. Moreover, we cannot even guarantee that the distribution is jointly normal. We take a close look at the key-schedule influence in the following.

4.1.1 Non-Normality of Linearly Independent Approximations

In light of Theorem 2, the joint correlation distribution of multiple linear approximations of an ideal cipher is currently unknown. Since knowledge of this distribution is crucial to linear cryptanalysis, it seems safer to consider sets of linearly independent approximations. But how do these behave for a specific block cipher that does not have independent round keys? To investigate this, we consider a set of 18 linearly independent approximations over 9 rounds of 32-bit SmallPresent [Lea10]. The input and output masks are given by

$$\alpha = 2^{4i+3}, i \in 5,6,7, \quad \text{and} \quad \beta = 2^{4i+j}, i \in 5,6,7, j \in 2,3.$$

We note that these approximations have the same form as those we will later use to attack present in Section 6. We consider three different key-schedules: long-key, identical round keys, and a 40-bit key-schedule described in Appendix A. For each key-schedule, we calculated the linear correlation of each approximation for the full code-book and 2 000 000 randomly chosen keys. Let $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ be the mean vector and covariance matrix of each of the data sets, respectively. To assess how much the distribution of $\mathbf{C}^K$ deviates from joint normality, we consider the squared Mahalanobis distance, defined by

$$d^2 = (\mathbf{C}^K - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{C}^K - \boldsymbol{\mu}).$$

Note that if $\mathbf{C}^K \sim \mathcal{N}_{18}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then $d^2 \sim \chi^2_{18}$. Figure 1 shows the density of d^2 for the three data sets against the density of the χ^2-distribution.

We make the following observations: For the long-key, the joint distribution of $\mathbf{C}^K$ is very close to the multivariate normal distribution $\mathcal{N}_{18}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. When we switch to a key-schedule with dependent round keys, we observe a deviation from normality. The most drastic effect is seen in case of the strongest dependence between the round keys, namely for identical round keys. Here, the distribution of d^2 is heavier towards zero, but also has a heavier tail towards infinity, compared to the χ^2_{18}-distribution. For such a key-schedule, it does not seem reasonable to approximate the distribution of $\mathbf{C}^K$ by a multivariate normal distribution. For the 40-bit key-schedule, the distribution of d^2 also deviates somewhat from χ^2_{18}. The 40-bit key-schedule we have used here is a scaled down version of the 80-bit key-schedule used in PRESENT, and so it might be natural to assume that the cipher behaves as a long-key cipher, in order to simplify analysis. However, there is still quite some overlap of the bits in consecutive round keys, which seems to have a non-negligible influence on the shape of the joint correlation distribution. It would then seem that, strictly speaking, joint normality is not a fair assumption, even for good key-schedules.

4.2 Restricted Approximation Choices

The right-key models of [HVLN15, BN17] set certain requirements for the set of approximations used. The primary requirement is on the parameters of either the correlation or multidimensional probability distributions. For Results 1 and 2, the assumption is that all the distributions are identical. For Result 4, the assumption is that the distributions have identical variances. Although it might be possible to find sets of approximations such that these assumptions are satisfied, it does restrict the ability of the cryptanalyst to freely choose a set of approximations that can optimally facilitate an attack. This can for example make it hard to do efficient key-guessing, and so would result in a worse attack than if the cryptanalyst could choose approximations freely.

While the use of the multidimensional probability distribution in Result 2 is promising, it seems that there are more works that analyse the correlation distributions directly – perhaps because the distribution of these is more well understood. For models that use the correlation distributions directly, it seems that these are currently either multiple (Result 4) or multidimensional with similar restrictions to the multiple case (Results 1 and 5). For Result 1, a set of (linearly independent) dominant base approximations are required, and so the combined approximations derived from these cannot by assumption contribute significantly to the attack. For Result 5, the noise part of the ℓ known approximations are modelled as approximations of a random permutation and must be independent, and so by Theorem 2 and Appendix B, they must be linearly independent. Additionally, the remaining approximations only contribute with noise.

4.3 Parameter Estimation

As noted by [BN16], one major challenge when trying to apply Results 1, 2 and 4 is to get an accurate estimate of the various distribution parameters. For single approximations, this problem was identified in [BT13] and the signal/noise decomposition was proposed. This approach was nicely applied in [BN16], and was shown to give more accurate results. However, [BN16] uses the long-key assumption to avoid considering the actual distribution of the signal, instead only considering the signal ELP. Extending the discussion of Section 4.1.1, this might not be accurate for other key-schedules. In this case, the cryptanalyst would have to get an estimate of the actual signal distribution.

To estimate the parameters of the signal, one could find a set of trails with large correlation contribution, and calculate part of the sum in Equation 1 for a significant number of randomly chosen keys. Doing this can be a significant challenge, especially for PRESENT-like ciphers where the number of good trails is extremely large. Various methods for finding good trails of a cipher have been proposed, e.g. the branch-and-bound

method [Mat94a] and sparse correlation matrices [Abd12], but it can still be quite the computational challenge to obtain good parameters for the signal. In Section 6, we use a method similar to that of [Abd12] and significant computational power to obtain estimates for a set of PRESENT approximation.

While it might be possible to avoid the other issues discussed in this section, if we abandon the long-key assumption, parameter estimation seems like a challenge that is difficult to avoid. Indeed, the model we propose in Section 5 in some sense trades assumptions for increased computational effort. As such, efficient algorithms for computation of the signal trails seems like an increasingly important research topic. In connection to this, note that while an estimate of the parameters of the correlation distributions can be obtained by the above method, we are not aware of any such method to estimate the parameters of the multidimensional probability distributions.

5 Multivariate Linear Cryptanalysis

As argued in Section 4.1, when a cipher uses round keys that exhibit some dependence between them, the joint distribution $\mathbf{C}^K$ of linear correlations for a set of linearly independent approximations can deviate from the joint normality we would expect from a long-key cipher. Indeed, it seems very difficult to describe the exact joint distribution in this case. On a lower level, the marginal distributions do not necessarily have identical variances, as was assumed in [HVLN15, BN17]. Additionally, as discussed in Section 4.2, the current models for multidimensional linear cryptanalysis do not seem to fully use most of the approximations in the chosen subspace of $\mathbb{F}_2^n \times \mathbb{F}_2^n$, and so by using the multidimensional approach, the attacker has to consider approximations that only add noise. What is worse, it seems that we are not able to formulate a wrong-key hypothesis in the multidimensional case that fully agrees with Theorem 2. Thus, the need for a wrong-key model forces us to consider the case of multiple, linearly independent approximations. It is therefore our aim to create a more powerful model for this setting which: models the behaviour of any set of linearly independent approximations; does not assume statistical independence of approximations or round keys; does not assume the shape of the joint correlation distribution; and takes into account the unknown part of the linear hull.

In the following we propose *multivariate linear cryptanalysis*. In Section 5.1 we present the main right- and wrong-key hypotheses the model relies on. This model in some sense trades assumptions for computational effort during the off-line analysis. In Section 5.2 we incorporate the signal/noise decomposition of [BT13] into the model, similar to [BN16], in order to make the model practically usable. In Section 5.3 we describe the model as used in a key-recovery attack where the attacker does not have access to the full codebook.

5.1 The Main Model: Arbitrary Right-Key Distribution

The first part of our model is very general, and simply expresses the fact that the correlations of a set of M linear approximations follow *some* multivariate probability distribution. Consider the vector $\mathbf{C}^K$ containing the correlations of M linear approximations with linearly independent masks. We propose the following right-key and wrong-key models.

Model 1 (Right-key – Multiple)**.** *Let (α_i, β_i), $i = 1, \ldots, M$, be M different linear approximations of a block cipher with linearly independent masks, and let $\mathbf{C}^K = (C^K_{\alpha_1,\beta_1} \cdots C^K_{\alpha_M,\beta_M})^\top$ be a vector containing the linear correlations. Then $\mathbf{C}^K \sim \mathcal{D}_M$ over the key space, for some M-variate probability distribution $\mathcal{D}_M$.*

Hypothesis 5 (Wrong-key – Multiple)**.** *Let $\mathbf{\Sigma}^\delta = \mathrm{diag}(2^{-n})$. During a key recovery attack, for a wrong key guess, the linear correlation vector $\mathbf{C}^K$ of M linear approximations with linearly independent masks is a random vector with distribution $\mathbf{C}^K \sim \mathcal{N}_M(\mathbf{0}, \mathbf{\Sigma}^\delta)$.*

The wrong-key hypothesis is inspired by Theorem 2 and the result of [DR07], and the veracity of the hypothesis therefore relies on the assumption of pair-wise statistical independence of linear approximations of a random permutation. We take some steps towards validating Hypothesis 5 in Appendix B. For the right-key, this model allows the attacker to pick any set of linearly independent approximations, but requires that she can somehow estimate the shape of the distribution $\mathcal{D}_M$. While this at first does not seem very useful, as determining this distribution seems like a very hard problem in general, we propose a way to do this in the following by applying the signal/noise decomposition. We note that, interestingly, Model 1 could be extended to any arbitrary set of approximations, but it is currently unknown how to express Hypothesis 5 in this setting. It is therefore a very interesting open problem to derive the distribution of linearly dependent approximations of an ideal cipher.

5.2 The Practical Model: Signal/Noise Decomposition

The model presented requires the cryptanalyst to somehow obtain the distribution $\mathcal{D}_M$ for the right-key distribution. In most cases, we will be unable to calculate the exact distribution of $C^K_{\alpha,\beta}$ for any single approximation, and we therefore have to estimate $\mathcal{D}_M$. In order to do this, we take a similar approach to [BT13, BN16]. Let $\mathcal{S}$ be the set of known *signal trails* for an approximation (α, β). Then we define the signal correlation as

$$C^{K\star}_{\alpha,\beta} = \sum_{U \in \mathcal{S}} C^K_U. \tag{4}$$

The signal correlation $C^{K\star}_{\alpha,\beta}$ will itself follow some probability distribution – we denote this by $\mathcal{D}^\star_{\alpha,\beta}$. We then assume that the unknown trails, the *noise*, behave as for a random permutation, i.e. their correlation is distributed as $\mathcal{N}(0, 2^{-n})$. Then we can approximate the full correlation with the distribution

$$C^K_{\alpha,\beta} \sim \mathcal{D}^\star_{\alpha,\beta} + \mathcal{N}(0, 2^{-n})$$

However, we still have the problem that $\mathcal{D}^\star_{\alpha,\beta}$ is unknown. This problem can be solved computationally. By computing Equation 4 for a large number of keys, we obtain a set of values drawn from $\mathcal{D}^\star_{\alpha,\beta}$. Whenever we need to randomly sample from $\mathcal{D}^\star_{\alpha,\beta}$, as we will need to do to estimate the strength of an attack, we simply sample from this data set. The same can be done for multiple approximations by calculating the signal correlations simultaneously for all M approximations for a randomly chosen set of keys. In this way, we trade any assumptions on the shape of the distribution $\mathcal{D}^\star_M$ for a potentially large computational effort. However, this computational effort is only required during the off-line analysis, and so has no influence on the computational complexity of an attack.

Under the assumption that the noise behaves as for a random permutation, the noise of linearly independent approximations will also be statistically independent, by Theorem 2 and Appendix B. Then we can make the following generalisation of the signal/noise decomposition to several approximations. Note that compared to [BN16], we here consider the distribution of the signal over the keys, as opposed to only the ELP of the approximations.

Model 2. *Let $\mathbf{\Sigma}^\delta = \mathrm{diag}(2^{-n})$. If the distribution, $\mathcal{D}^\star_M$, of the signal $\mathbf{C}^{K\star}$ is known, then the distribution of $\mathbf{C}^K$ in Model 1 is closely approximated by $\mathbf{C}^K \sim \mathcal{D}^\star_M + \mathcal{N}_M(\mathbf{0}, \mathbf{\Sigma}^\delta)$.*

5.2.1 Experimental Verification.

In order to verify Model 2, we again consider the set of 18 approximation over 9 rounds of 32-bit SmallPresent defined in Section 4.1.1. We considered the version with the

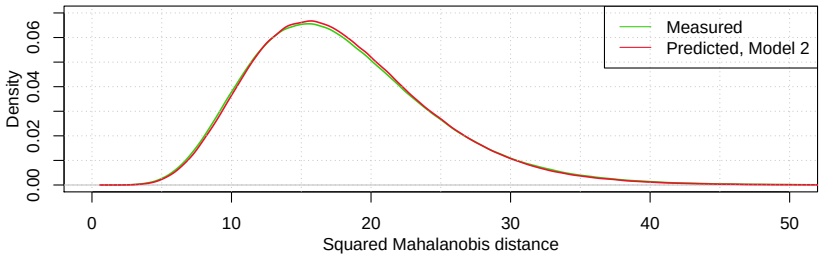

Figure 2: A density of the squared Mahalanobis distance for the joint distribution of linear correlation for 18 approximations over 9 rounds of 32-bit SmallPresent using a 40-bit key-schedule. The plot compares the density measured using the full codebook to a prediction made using Model 2.

40-bit key-schedule, and enumerated part of the hull of each approximation, by using an approach very similar to the sparse correlation matrix method in [Abd12]. In this way, we obtain a set of signal trails that includes all trails having intermediate masks with hamming weight at most four in each round. We did this simultaneously for all 18 approximations and 500 000 randomly chosen keys, in order to get an estimate of the distribution $\mathcal{D}^{\star}_{18}$. Furthermore, we measured the actual correlation values of the cipher for 2 000 000 randomly chosen keys. We then applied Model 2 to our signal estimate, and calculated the squared Mahalanobis distance of the two resulting data sets. The result is shown in Figure 2. The figure shows that Model 2 gives us a very close estimate of the actual distribution.

5.3 The Attack Model: Dealing with Undersampling

Even though Model 2 provides a way to get a good estimate of the multivariate correlation distribution, we would often like to avoid using the full codebook in a key-recovery attack. Thus, we also need to be able to express the distribution of the undersampled correlation, $\mathbf{C}^{K,N}$. Using a result due to Murphy, we develop such a model next.

Murphy showed [Mur06] that the joint distribution over the text space of the empirical correlations, measured using N randomly drawn text pairs for a *fixed* key K_0, has a multivariate normal distribution, $\mathbf{C}^{K_0,N} \sim \mathcal{N}_M(\boldsymbol{\mu}^{K_0}, \boldsymbol{\Sigma}^{K_0,N})$, where $\boldsymbol{\mu}^{K_0}_i = C^{K_0}_{\alpha_i,\beta_i}$ and

$$\boldsymbol{\Sigma}^{K_0,N}_{i,j} = \begin{cases} N^{-1} C^{K_0}_{\alpha_i \oplus \alpha_j, \beta_i \oplus \beta_j} & \text{for } i \neq j, \\ N^{-1} & \text{for } i = j. \end{cases}$$

When taken as a random variable over the key space, we note that $\boldsymbol{\mu}^{K_0} = \mathbf{C}^K$ and therefore has distribution $\mathcal{D}_M$. Indeed, $\boldsymbol{\Sigma}^{K_0,N}$ also has a distribution over the key space, making the distribution over both the key and text space extremely difficult to analyse. However, as Murphy notes, it is often the case that the combined approximations $(\alpha_i \oplus \alpha_j, \beta_i \oplus \beta_j)$ are extremely weak, e.g. in the case where (α_i, β_i) and (α_j, β_j) activate different S-boxes at the input and output. In this case, $N^{-1} C^{K_0}_{\alpha_i \oplus \alpha_j, \beta_i \oplus \beta_j} \ll N^{-1}$, and we can set these covariances to zero. As Murphy says, in this case the fixed-key correlations are "approximately statistically independent" over the text space, in the sense that any contribution by the covariances is negligible. Under this assumption, we obtain the following theorem.

Theorem 3. *Let* $\boldsymbol{\Sigma}^N = \mathrm{diag}(N^{-1})$. *Consider a set of* M *approximations as given in Model 1. Assume that the correlation of any combination of two such approximations is zero. Then the empirical correlation vector of these approximations, measured with* N *randomly drawn plaintext-ciphertext pairs, has distribution* $\mathbf{C}^{K,N} \sim \mathcal{D}_M + \mathcal{N}_M(\mathbf{0}, \boldsymbol{\Sigma}^N)$. *For the wrong-key scenario of Hypothesis 5,* $\mathbf{C}^{K,N} \sim \mathcal{N}_M(\mathbf{0}, \boldsymbol{\Sigma}^{\delta} + \boldsymbol{\Sigma}^N)$.

Proof. From [Mur06] we have that $\mathbf{C}^{K_0,N} \sim \mathcal{N}_M(\boldsymbol{\mu}^{K_0}, \boldsymbol{\Sigma}^{K_0,N})$ for a fixed key K_0. By assumption, we further have that $\boldsymbol{\Sigma}^{K_0,N} = \boldsymbol{\Sigma}^N = \mathrm{diag}(N^{-1})$, and so is independent of the key. The distribution of $\mathbf{C}^{K_0,N}$ over keys is therefore $\mathcal{N}_M(\mathcal{D}_M, \boldsymbol{\Sigma}^N) = \mathcal{D}_M + \mathcal{N}_M(\mathbf{0}, \boldsymbol{\Sigma}^N)$. For the wrong-key, $\mathcal{D}_M = \mathcal{N}_M(\mathbf{0}, \boldsymbol{\Sigma}^\delta)$, finishing the proof. □

By applying Model 2 to this theorem, we obtain the following corollary.

Corollary 1. *For a set of M approximations as in Theorem 3, if the distribution, $\mathcal{D}^\star_M$, of the signal $\mathbf{C}^{K\star}$ is known, then the distribution of $\mathbf{C}^{K,N}$ is closely approximated by $\mathbf{C}^{K,N} \sim \mathcal{D}^\star_M + \mathcal{N}_M(\mathbf{0}, \boldsymbol{\Sigma}^N + \boldsymbol{\Sigma}^\delta)$.*

As an interesting observation, this result shows how the original model by Matsui, Equation 2, can misleadingly give accurate results when N is relatively small, as is the case for the attack on DES. In this case, and as long as $\mathcal{D}_M$ does not deviate too much from joint normal distribution, N^{-1} will dominate the variance terms of $\mathrm{Cov}(\mathcal{D}_M)$ and $\boldsymbol{\Sigma}^\delta$, making the key-variance undetectable. This also shows that conducting experiments for a low number of rounds with low data complexity can not necessarily confirm a model.

Corollary 1 gives us a way to estimate the distribution of the correlation vector over the keys for a set of linearly independent approximations. In contrast to Results 1, 2, 4 and 5, no assumptions about independence or the parameters of the involved distributions are required, and we do not assume independent round keys. This generality of course comes with a cost: the approximations have to be linearly independent (although we are not forced to consider weak approximations), and we have to estimate the distribution $\mathcal{D}^\star_M$. We have partially discussed the latter issue in Section 5.2, and we will discuss how we have done this for PRESENT in Section 6.

6 Multivariate Linear Attacks on PRESENT

Different methods for distinguishing when using many approximations have been proposed. The LLR method was proposed by Baignères et al. in [BJV04] as an optimal distinguisher and used in [HCN08] in a multidimensional attack against the block cipher Serpent. Both the LLR method and the χ^2 method were studied in [HCN09], where the LLR method was concluded to have better performance. However, as noted by Cho in [Cho10], the LLR method is often not practical to use, as it requires an accurate knowledge of the key-dependent behaviour of the multidimensional probability distribution. For this reason, the χ^2 method is more commonly used. We now present a new attack on 26 and 27 rounds of PRESENT using this method and the improved multiple linear model of Section 5.

6.1 Determining the Advantage

The χ^2 method has been widely used as a distinguisher in various attacks. For this method, the test statistic is defined as

$$\mathcal{T}_{\chi^2} = N \sum_{i=1}^{M} (C^{K,N}_{\alpha_i,\beta_i})^2.$$

In the following, we describe how to determine the advantage of the χ^2 distinguisher using the theory developed in this paper. The approximations used are chosen based on the observations made by Ohkuma in [Ohk09]: the best approximations of PRESENT are those that start and end with the S-boxes S_i with $i \in \{5, 6, 7, 9, 10, 11, 13, 14, 15\}$. For our attack, we consider the input and output masks

$$\alpha = 2^{4i+3}, i \in \{5, 6, 7, 9, 10, 11, 13, 14, 15\},$$
$$\beta_1 = 2^{4i+3}, i \in \{5, 6, 7, 9, 10, 11, 13, 14, 15\}, \quad \beta_2 = 2^{4i+2}, i \in \{5, 6, 7, 9, 10, 11\}.$$

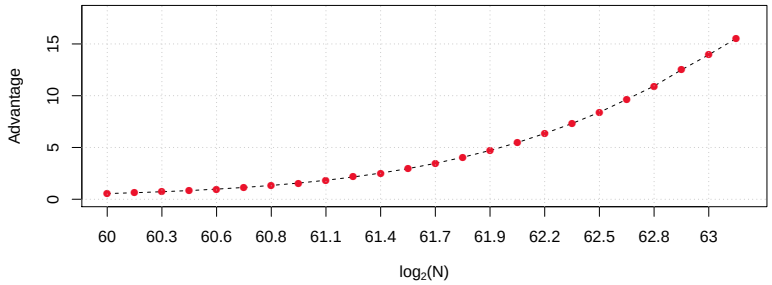

Figure 3: Advantage of the χ^2 distinguisher using 135 approximations of 22-round PRESENT, with $P_S = 0.95$. At half the codebook, $N = 2^{63}$, the advantage is 14.5 bits.

Taking all possible combinations of these input and output masks gives us $M = 135$ approximations. These approximations are chosen to facilitate efficient key-guessing over a large number of rounds, as will become evident in Section 6.2. We note that due to the structure of the approximations, it can be assumed that the undersampling matrix $\mathbf{\Sigma}^N$ is a diagonal matrix, as discussed in Section 5.3. This does not imply independence of the approximations, but simplifies our analysis considerably.

With this choice, we obtain the advantage in the following way. By using a signal that includes all trails having intermediate masks with hamming weight at most four in each round, and a technique similar to that of [Abd12], we obtain a data set of observations from the signal distribution $\mathcal{D}^{\star}_{135}$. We used 217 100 random master keys to generate these observations. We now simulate observations from $\mathbf{C}^{K,N}$ in the following way: We fix a sample size for the simulation, say k. For the right key, we randomly sample k observations of $\mathcal{D}^{\star}_{135}$ (with replacement, if k is larger than the number of observations we have collected) from our data set. We then sample k random observations from the normal distribution $\mathcal{N}_M(\mathbf{0}, \mathbf{\Sigma}^N + \mathbf{\Sigma}^{\delta})$. These two samples are then added together, following Corollary 1. The wrong-key distribution is simulated by randomly sampling k times from the normal distribution $\mathcal{N}_M(\mathbf{0}, \mathbf{\Sigma}^N + \mathbf{\Sigma}^{\delta})$, according to Theorem 3.

We note for comparison to previous works that the expected right-key capacity obtained from these simulations is $\mathrm{E}(\mathcal{C}^K) = 2^{-55.01}$ with a variance of $\mathrm{Var}(\mathcal{C}^K) = 2^{-115.59}$, whereas the wrong-key capacity has $\mathrm{E}(\mathcal{C}^K) = 2^{-56.92}$ and $\mathrm{Var}(\mathcal{C}^K) = 2^{-119.92}$.

We can now calculate the empirical CDFs of the simulated right-key and wrong-key distributions. For a fixed success probability P_S, we use the right-key CDF to obtain a threshold τ, as described in Section 2.2. The advantage is finally calculated using the wrong-key CDF and τ, as defined in Section 2.2. Figure 3 shows the result of applying this procedure for $k = 2^{22}$, $P_S = 0.95$, and varying values of N. We note that we need to set k fairly high to obtain sufficient resolution of the empirical CDFs. For the chosen k, we can detect probabilities down to 2^{-22}, allowing us in turn to detect advantages of up to 22 bits. At half the codebook, $N = 2^{63}$ we obtain an advantage of 14.5 bits.

6.2 Attacking 26 rounds

Under the wrong key randomisation hypothesis, Hypothesis 5, we can turn our multivariate linear distinguisher into a key-recovery attack, as described in Section 2.2.1. That is, the attack proceeds as follows: Collect N plaintext-ciphertext pairs. Guess the bits of the outer round keys required to (partially) encrypt/decrypt the desired number of rounds. Apply the χ^2 distinguisher to the resulting correlations, and save the key guess if the distinguisher indicates a non-ideal cipher. Repeat for all guesses of the round key bits. For each saved key we can find the master key by exhaustively guessing the remaining bits and verifying by trial encryption.

We aim to recover the master key for r rounds of PRESENT-80 by using a multiple linear

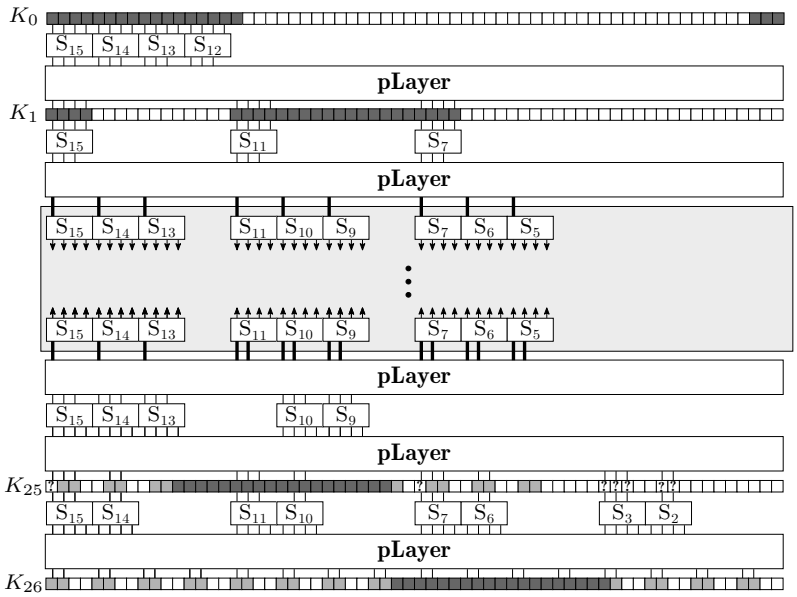

Figure 4: An outline of the 26-round attack using 22 round approximations. The input/output mask bits are indicated by bold lines. The dark grey squares indicate the round key bits obtained by guessing 24 bits of the master key. The light grey squares indicate the round key bits obtained by guessing 23 bits of the last round key. The squares indicated by **?** are extra bits of the second to last round key that need to be guessed.

approximation over $r-4$ rounds. Because of the large number of outer rounds we need to bypass, the approximations are chosen such that the involved round key bits are sparse. We consider the set of 135 approximations described above. The bit positions of the input and output masks are highlighted in Figure 4. Figure 4 shows the S-box positions we need to encrypt/decrypt to calculate the linear correlations of these approximations. The straightforward approach to partially encrypting/decrypting these positions would require guessing 80 key bits across the four round keys. By considering the key-schedule, we can dramatically improve this. We first guess the following 24 bits of the master key:

$$k_i, i \in [0,2] \cup [15,18] \cup [63,79]. \tag{5}$$

The round key bits we obtain from this guess are marked in dark grey in Figure 4, as well as 42 additional bits needed by the attack. By guessing the missing 23 bits of K_{26}, we also obtain 13 bits of K_{25}. Finally, we only need to guess an additional 7 bits of K_{25}. In total, we need to guess 54 key bits. Note that each approximation only depends on 4 bits of K_{25} and 16 bits of K_{26}. With these considerations in mind, the attack proceeds as follows.

Distillation phase

1. Obtain N partial text pairs (p_i, c_i), where p_i is 16 bits and c_i is 32 bits.
2. Generate a vector $\mathbf{t}$ of size 2^{48} where $\mathbf{t}[s\|t] = \#\{i \mid p_i = s \text{ and } c_i = t\}$.

Analysis phase

1. For each 24-bit guess of the partial master key, K_M, perform these steps:
   (a) For each input mask α, calculate two vectors $\mathbf{t}_{\alpha_1}^{K_M}$ and $\mathbf{t}_{\alpha_2}^{K_M}$ of size 2^{16}, where

$$\mathbf{t}_{\alpha_x}^{K_M}[j] = \#\{(p_i, c_i) | G_x(c_i) = j \text{ and } \alpha \cdot \mathcal{E}_{K_M}(p_i) = 0\},$$

where $\mathcal{E}_{K_M}(p)$ is the partial two-round encryption of p under key K_M, and G_x selects the bits of c_i required to calculate the output masks of β_x, $x \in \{1, 2\}$.

(b) For each output mask β, fix a guess of the relevant 4 bits of K_{25}. Denote the guess K_I. Then calculate the $2^{16} \times 2^{16}$ matrix $\mathbf{A}_\beta^{K_I}$, where

$$\mathbf{A}_\beta^{K_I}[i, j] = \beta \cdot \mathcal{D}_{K_I}(i \oplus j),$$

and $\mathcal{D}_{K_I}(c)$ is the partial two-round decryption of the 16-bit value c using K_I, but *excluding* the first key XOR.

(c) Calculate the correlations of all 135 approximations and 2^{16} guesses of the partial K_{26} by calculating the matrix-vector products

$$\mathbf{C}_{\alpha,\beta} = \tfrac{2}{N}\mathbf{A}_\beta^{K_I}\mathbf{t}_{\alpha_x}^{K_M} - 1.$$

(d) Repeat steps (b) and (c) for all values of K_I, resulting in correlation values for all approximations for at most 2^{36} guesses of the last two round keys.

(e) Extract the correlations of at most 2^{30} guesses that agree with K_M.

(f) Calculate the χ^2 test statistic $\mathcal{T}_{\chi^2}$ for each surviving key guess. Save all keys (of 54 bits) with $\tau < \mathcal{T}_{\chi^2}$.

Search phase

1. For each key candidate, perform trial encryption to find the remaining $80 - 54 = 26$ bits of the master key.

6.2.1 Attack Complexity.

We now consider the computational complexity of the attack. We consider the number of single round encryption equivalent operations performed.

- The distillation phase requires N operations.
- For the analysis phase:
  - Step 1a can be done by iterating over $\mathbf{t}$ once and encrypting two rounds, using $2 \cdot 2^{48}$ operations.
  - Steps 1b and 1c can be performed using the FFT technique given in [CSQ07]. Using this technique, we only need to compute the first column of each $\mathbf{A}_\beta^{K_I}$, at a cost of $2 \cdot 2^{16}$ operations, and then calculate $\mathbf{C}_{\alpha,\beta}$ for a fixed β and all α in time $(2 \cdot 9 + 1) \cdot 16 \cdot 2^{16}$.
  - There are 2^4 values of K_I and 15 output masks. Thus, steps 1d needs a total of $15 \cdot 2^4 \cdot (2 \cdot 2^{16} + (2 \cdot 9 + 1) \cdot 16 \cdot 2^{16}) \approx 2^{32.16}$ operations.
  - Step 1e uses 2^{30} operations.
  - Step 1f takes roughly $2 \cdot 135 \cdot 2^{30} = 2^{38.08}$ operations.
  - In total, this phase uses $2^{24} \cdot (2^{49} + 2^{32.16} + 2^{30} + 2^{38.08}) \approx 2^{73.00}$ operations.
- Finally, the search phase requires $2^{\kappa-54}$ full encryptions of 2^{54-a} candidate keys, using a total of $26 \cdot 2^{\kappa-a}$ operations.

From Figure 3, we obtain a plot of the computational complexity of the 26-round attack, given in Figure 5. Here, we have fixed the success probability at 95%. As long as the search phase dominates, we can increase the number of texts to decrease to computational complexity. We can highlight two 26 round attacks with different trade-offs. For $N =$

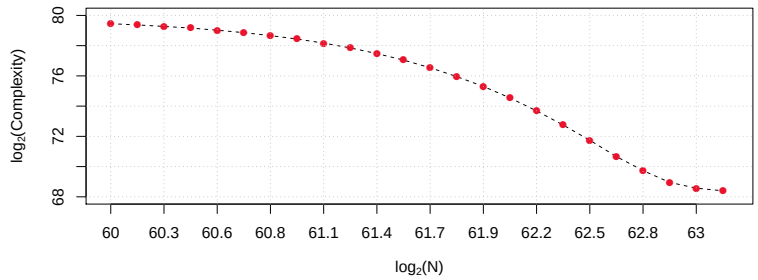

Figure 5: *Our 26 round attack:* Computational complexity as a function of data complexity for the 26-round attack on PRESENT using 135 approximations over 22 rounds. Non-distinct random texts were used, and $P_S = 0.95$. Note that the complexity reaches a lower limit close to $N = 2^{63}$ when the advantage becomes sufficiently large.

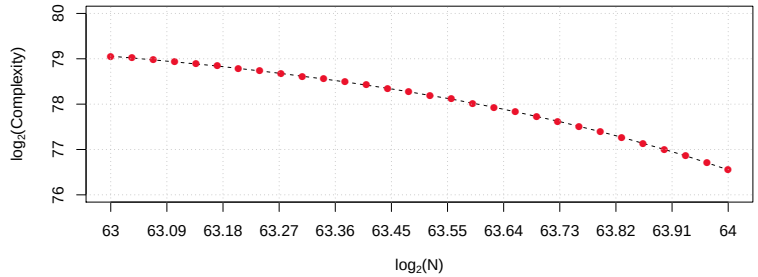

Figure 6: *Our 27 round attack:* Computational complexity as a function of data complexity for the 27-round attack on PRESENT using 135 approximations over 23 rounds. Distinct random texts were used, and $P_S = 0.95$.

$2^{63.0}$, the advantage is 14.0 bits, and the computational complexity is $2^{73.27}/26 = 2^{68.57}$ encryptions. Interestingly, this multiple attack uses far fewer approximations than Cho's multidimensional attack [Cho10], but at half the data complexity and a computational complexity that is 11 times smaller, all the while needing far fewer assumptions. Compared to the reevaluation of Cho's attack in [BN16] (which has the same computational complexity as the original attack), our attack uses less data, and has a higher success probability. Alternatively, we can decrease the data complexity to $N = 2^{61.9}$, giving an advantage of 4.7 bits, and a computational complexity of $2^{80.00}/26 = 2^{75.30}$ encryptions. While being slower than Cho's attack, to the best of our knowledge, this attack has the lowest data complexity of any 26-round attack on PRESENT presented in the literature.

6.3 Attacking 27 rounds

The attack can be extended to 27 rounds by using the same approximations over 23 rounds. By guessing the bits of the master key given in Equation 5, we determine 41 required bits of the round keys. We then have to guess 25 bits of K_{27} and 6 bits of K_{26}, for a total of 55 bits of key material. Due to the way we carry out the attack, the complexity calculation is not affected by this – only the lower advantage has an influence. However, if we use non-distinct random texts for the attack, the advantage is too low. If we instead use distinct random texts, we obtain a better advantage. This scenario is in some sense a chosen plaintext attack, and has been studied in [BN15, BN17]. The only change to our model is that $\mathbf{\Sigma}^N = \mathrm{diag}(\frac{2^n - N}{N \cdot (2^n - 1)})$ in Corollary 1. The distribution of $\mathbf{C}^{K\star}$ was again estimated using 217 100 random keys as for the 26 round attack, and we obtain $\mathrm{E}(\mathcal{C}^K) = 2^{-56.38}$ and $\mathrm{Var}(\mathcal{C}^K) = 2^{-118.73}$ for the right-key. The resulting attack complexities are shown in Figure 6. Using the χ^2 distinguisher with $P_S = 0.95$ and $N = 2^{63.83}$, we obtain an advantage of 2.73 bits and a computational complexity of $2^{77.27}$ encryptions.

References

[AÅBL12] Mohamed Ahmed Abdelraheem, Martin Ågren, Peter Beelen, and Gregor Leander. On the Distribution of Linear Biases: Three Instructive Examples. In *Advances in Cryptology - CRYPTO 2012*, pages 50–67, 2012.

[Abd12] Mohamed Ahmed Abdelraheem. Estimating the Probabilities of Low-Weight Differential and Linear Approximations on PRESENT-Like Ciphers. In *Information Security and Cryptology - ICISC 2012 - 15th International Conference, Seoul, Korea, November 28-30, 2012, Revised Selected Papers*, pages 368–382, 2012.

[BBR+13] Andrey Bogdanov, Christina Boura, Vincent Rijmen, Meiqin Wang, Long Wen, and Jingyuan Zhao. Key Difference Invariant Bias in Block Ciphers. In *Advances in Cryptology - ASIACRYPT 2013*, pages 357–376, 2013.

[BBV15] Céline Blondeau, Asli Bay, and Serge Vaudenay. Protecting Against Multidimensional Linear and Truncated Differential Cryptanalysis by Decorrelation. In *Fast Software Encryption*, pages 73–91, 2015.

[BCQ04] Alex Biryukov, Christophe De Cannière, and Michaël Quisquater. On Multiple Linear Approximations. In *Advances in Cryptology - CRYPTO 2004, 24th Annual International CryptologyConference, Santa Barbara, California, USA, August 15-19, 2004, Proceedings*, pages 1–22, 2004.

[BGW+13] Andrey Bogdanov, Huizheng Geng, Meiqin Wang, Long Wen, and Baudoin Collard. Zero-Correlation Linear Cryptanalysis with FFT and Improved Attacks on ISO Standards Camellia and CLEFIA. In *Selected Areas in Cryptography*, pages 306–323, 2013.

[Bih94] Eli Biham. On matsui's linear cryptanalysis. In *Advances in Cryptology - EUROCRYPT '94, Workshop on the Theory and Application of Cryptographic Techniques, Perugia, Italy, May 9-12, 1994, Proceedings*, pages 341–355, 1994.

[BJV04] Thomas Baignères, Pascal Junod, and Serge Vaudenay. How Far Can We Go Beyond Linear Cryptanalysis? In *Advances in Cryptology - ASIACRYPT 2004, 10th International Conference on the Theory and Application of Cryptology and Information Security, Jeju Island, Korea, December 5-9, 2004, Proceedings*, pages 432–450, 2004.

[BKL+07] 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 *Cryptographic Hardware and Embedded Systems - CHES 2007, 9th International Workshop, Vienna, Austria, September 10-13, 2007, Proceedings*, pages 450–466, 2007.

[BLN14] Céline Blondeau, Gregor Leander, and Kaisa Nyberg. Differential-Linear Cryptanalysis Revisited. In *Fast Software Encryption*, pages 411–430, 2014.

[BN13] Céline Blondeau and Kaisa Nyberg. New Links between Differential and Linear Cryptanalysis. In *Advances in Cryptology - EUROCRYPT 2013*, pages 388–404, 2013.

[BN14] Céline Blondeau and Kaisa Nyberg. Links between Truncated Differential and Multidimensional Linear Properties of Block Ciphers and Underlying Attack Complexities. In *Advances in Cryptology - EUROCRYPT 2014*, pages 165–182, 2014.

[BN15] Céline Blondeau and Kaisa Nyberg. On distinct known plaintext attacks. In *The Ninth International Workshop on Coding and Cryptography*, 2015.

[BN16] Céline Blondeau and Kaisa Nyberg. Improved Parameter Estimates for Correlation and Capacity Deviates in Linear Cryptanalysis. *IACR Trans. Symmetric Cryptol.*, 2016(2):162–191, 2016.

[BN17] Céline Blondeau and Kaisa Nyberg. Joint Data and Key Distribution of Simple, Multiple, and Multidimensional Linear Cryptanalysis Test Statistic and Its Impact to Data Complexity. *Design, Codes and Cryptography*, 82(1-2):319–349, 2017.

[BR14] Andrey Bogdanov and Vincent Rijmen. Linear Hulls with Correlation Zero and Linear Cryptanalysis of Block Ciphers. *Des. Codes Cryptography*, 70(3):369–383, 2014.

[BT13] Andrey Bogdanov and Elmar Tischhauser. On the Wrong Key Randomisation and Key Equivalence Hypotheses in Matsui's Algorithm 2. In *Fast Software Encryption - 20th International Workshop, FSE 2013, Singapore, March 11-13, 2013. Revised Selected Papers*, pages 19–38, 2013.

[Cho10] Joo Yeon Cho. Linear Cryptanalysis of Reduced-Round PRESENT. In *Topics in Cryptology - CT-RSA 2010, The Cryptographers' Track at the RSA Conference 2010, San Francisco, CA, USA, March 1-5, 2010. Proceedings*, pages 302–317, 2010.

[CSQ07] Baudoin Collard, François-Xavier Standaert, and Jean-Jacques Quisquater. Improving the Time Complexity of Matsui's Linear Cryptanalysis. In *Information Security and Cryptology - ICISC 2007, 10th International Conference, Seoul, Korea, November 29-30, 2007, Proceedings*, pages 77–88, 2007.

[Dae95] Joan Daemen. *Cipher and Hash Function Design Strategies Based on Linear and Differential Cryptanalysis.* PhD thesis, Doctoral Dissertation, March 1995, KU Leuven, 1995.

[DR02] Joan Daemen and Vincent Rijmen. *The Design of Rijndael: AES - The Advanced Encryption Standard.* Information Security and Cryptography. Springer, 2002.

[DR05] Joan Daemen and Vincent Rijmen. Probability distributions of correlation and differentials in block ciphers. *IACR Cryptology ePrint Archive*, 2005:212, 2005.

[DR07] Joan Daemen and Vincent Rijmen. Probability Distributions of Correlation and Differentials in Block Ciphers. *J. Mathematical Cryptology*, 1(3):221–242, 2007.

[HCN08] Miia Hermelin, Joo Yeon Cho, and Kaisa Nyberg. Multidimensional Linear Cryptanalysis of Reduced Round Serpent. In *Information Security and Privacy, 13th Australasian Conference, ACISP 2008, Wollongong, Australia, July 7-9, 2008, Proceedings*, pages 203–215, 2008.

[HCN09] Miia Hermelin, Joo Yeon Cho, and Kaisa Nyberg. Multidimensional Extension of Matsui's Algorithm 2. In *Fast Software Encryption, 16th International Workshop, FSE 2009, Leuven, Belgium, February 22-25, 2009, Revised Selected Papers*, pages 209–227, 2009.

[HKM95] Carlo Harpes, Gerhard G. Kramer, and James L. Massey. A Generalization of Linear Cryptanalysis and the Applicability of Matsui's Piling-Up Lemma. In *Advances in Cryptology - EUROCRYPT '95, International Conference on the Theory and Application of Cryptographic Techniques, Saint-Malo, France, May 21-25, 1995, Proceeding*, pages 24–38, 1995.

[HN12] Miia Hermelin and Kaisa Nyberg. Multidimensional Linear Distinguishing Attacks and Boolean Functions. *Cryptography and Communications*, 4(1):47–64, 2012.

[HVLN15] Jialin Huang, Serge Vaudenay, Xuejia Lai, and Kaisa Nyberg. Capacity and Data Complexity in Multidimensional Linear Attack. In *Advances in Cryptology - CRYPTO 2015 - 35th Annual Cryptology Conference, Santa Barbara, CA, USA, August 16-20, 2015, Proceedings, Part I*, pages 141–160, 2015.

[JR94] Burton S. Kaliski Jr. and Matthew J. B. Robshaw. Linear Cryptanalysis Using Multiple Approximations. In *Advances in Cryptology - CRYPTO '94, 14th Annual International Cryptology Conference, Santa Barbara, California, USA, August 21-25, 1994, Proceedings*, pages 26–39, 1994.

[Jun01] Pascal Junod. On the Complexity of Matsui's Attack. In *Selected Areas in Cryptography, 8th Annual International Workshop, SAC 2001 Toronto, Ontario, Canada, August 16-17, 2001, Revised Papers*, pages 199–211, 2001.

[Lea10] Gregor Leander. Small Scale Variants Of The Block Cipher PRESENT. *IACR Cryptology ePrint Archive*, 2010:143, 2010.

[Lea11] Gregor Leander. On Linear Hulls, Statistical Saturation Attacks, PRESENT and a Cryptanalysis of PUFFIN. In *Advances in Cryptology - EUROCRYPT 2011*, pages 303–322, 2011.

[LH94] Susan K. Langford and Martin E. Hellman. Differential-Linear Cryptanalysis. In *Advances in Cryptology - CRYPTO '94, 14th Annual International Cryptology Conference, Santa Barbara, California, USA, August 21-25, 1994, Proceedings*, pages 17–25, 1994.

[Mat93] Mitsuru Matsui. Linear Cryptanalysis Method for DES Cipher. In *Advances in Cryptology - EUROCRYPT '93, Workshop on the Theory and Application of of Cryptographic Techniques, Lofthus, Norway, May 23-27, 1993, Proceedings*, pages 386–397, 1993.

[Mat94a] Mitsuru Matsui. On correlation between the order of s-boxes and the strength of DES. In *Advances in Cryptology - EUROCRYPT '94, Workshop on the Theory and Application of Cryptographic Techniques, Perugia, Italy, May 9-12, 1994, Proceedings*, pages 366–375, 1994.

[Mat94b] Mitsuru Matsui. The First Experimental Cryptanalysis of the Data Encryption Standard. In *Advances in Cryptology - CRYPTO '94, 14th Annual International Cryptology Conference, Santa Barbara, California, USA, August 21-25, 1994, Proceedings*, pages 1–11, 1994.

[Mur06] S. Murphy. The Independence of Linear Approximations in Symmetric Cryptanalysis. *IEEE Trans. Information Theory*, 52(12):5510–5518, 2006.

[NWW11] Phuong Ha Nguyen, Hongjun Wu, and Huaxiong Wang. Improving the Algorithm 2 in Multidimensional Linear Cryptanalysis. In *Information Security and Privacy - 16th Australasian Conference, ACISP*, pages 61–74, 2011.

[Nyb94] Kaisa Nyberg. Linear approximation of block ciphers. In *Advances in Cryptology - EUROCRYPT '94, Workshop on the Theory and Application of Cryptographic Techniques, Perugia, Italy, May 9-12, 1994, Proceedings*, pages 439–444, 1994.

[Nyb17] Kaisa Nyberg. Statistical and linear independence of binary random variables. *IACR Cryptology ePrint Archive*, 2017:432, 2017.

[Ohk09] Kenji Ohkuma. Weak Keys of Reduced-Round PRESENT for Linear Cryptanalysis. In *Selected Areas in Cryptography, 16th Annual International Workshop, SAC 2009, Calgary, Alberta, Canada, August 13-14, 2009, Revised Selected Papers*, pages 249–265, 2009.

[SB02] Ali Aydin Selçuk and Ali Biçak. On Probability of Success in Linear and Differential Cryptanalysis. In *Security in Communication Networks, Third International Conference, SCN 2002, Amalfi, Italy, September 11-13, 2002. Revised Papers*, pages 174–185, 2002.

[Sel08] Ali Aydin Selçuk. On Probability of Success in Linear and Differential Cryptanalysis. *J. Cryptology*, 21(1):131–147, 2008.

[SLG+16] Bing Sun, Meicheng Liu, Jian Guo, Vincent Rijmen, and Ruilin Li. Provable Security Evaluation of Structures Against Impossible Differential and Zero Correlation Linear Cryptanalysis. In *Advances in Cryptology - EUROCRYPT 2016*, pages 196–213, 2016.

[XJHL15] Hong Xu, Ping Jia, Geshi Huang, and Xuejia Lai. Multidimensional Zero-Correlation Linear Cryptanalysis on 23-Round LBlock-s. In *Information and Communications Security*, pages 97–108, 2015.

[ZWW14] Jingyuan Zhao, Meiqin Wang, and Long Wen. Improved Linear Cryptanalysis of CAST-256. *J. Comput. Sci. Technol.*, 29(6):1134–1139, 2014.

[ZZ15] Lei Zheng and Shao-Wu Zhang. FFT-Based Multidimensional Linear Attack on PRESENT Using the 2-Bit-Fixed Characteristic. *Security and Communication Networks*, 8(18):3535–3545, 2015.

A 40-bit Key-Schedule for SmallPresent

We define a 40-bit key-schedule for 32-bit SmallPresent, which is a scaled down version of the 80-bit PRESENT key-schedule. Let $K = k_{39}k_{38}\dots k_1k_0$ be a 40-bit key register, initialised to the master key. At round i, the round key is extracted as the 32 most significant bits of K, i.e.

$$K_i = k_{39}k_{38}\dots k_9k_8. \tag{6}$$

Then, K is updated as follows:

- K is rotated 9 bits to the right,
- The PRESENT S-box is applied to $k_{39}k_{38}k_{37}k_{36}$,
- A round counter is xor'ed to the least significant bits.

The round counter starts at 1 and is incremented by 1 for each round.

B Pair-Wise Independence of Linear Correlations

The wrong-key hypothesis presented here, Hypothesis 5, follows from Theorem 2 and [DR07], assuming that linear approximations of random permutations can be considered pair-wise independent. While it seems difficult to show when this assumption is true, we here take some steps towards verifying Hypothesis 5 experimentally. We first note that the normality of the marginal distributions of $\mathbf{C}^K$ for a random permutation is proven in [DR07]. Moreover, it seems unlikely that the joint distribution would deviate much from a multivariate normal distribution for most sets of approximations. Thus, if we can demonstrate that pairs of correlation distributions are independent, we can be confident that Hypothesis 5 is reasonable. To this end, we performed the following experiment:

- Fix a size of the permutation, say 2^n,
- Pick two random linear approximations,
- Generate 10 000 random permutations of the given size and measure the exact correlation of both approximations for each permutation using the full code-books,
- Perform Pearson's χ^2 test of independence between the two correlation distributions and record the p-value,
- Repeat the above process the desired number of times.

We note that when performing Pearson's χ^2 test of independence, the null hypothesis is that the two observed distributions are independent, and thus a p-value larger than e.g. 0.05 would indicate independence at the 95% significance level.

We performed the above experiment for varying sizes of the permutations, and the results are shown in Table 2. Here, we observe that for a 16-bit permutation, one out of 20 000 pairs of permutations had a significant p-value of 0.021. However, already for the slightly larger 20-bit permutation, the lowest p-value was 0.975; in other words, even in the worst case, there was only 2.5% chance that the two correlation distributions were dependent. For a 24-bit permutation, this results are even clearer, with the lowest p-value being extremely close to 1.

Additionally, we repeated the experiments for the 20-bit permutations, but this time using approximations that only differed in a single bit. Even for these very similar approximations, we observed the exact same results as for randomly chosen pairs of approximations. In light of these experimental results, it thus seems quite reasonable that correlations of 64- or 128-bit permutations would be independent for all practical intents and purposes.

Table 2: Results of the Pearson χ^2 test of independence for various permutation sizes. A p-value larger than 0.05 indicates that the correlations of two linear approximations are statistically independent at the 95% significance level.

Size	**Experiments**	**% of Experiments with p-value > 0.05**	**Smallest observed p-value**
2^{16}	20000	99.995	0.021
2^{20}	20160	100.00	0.975
2^{24}	15342	100.00	1.000