Multivariate Linear Cryptanalysis: The Past and Future of PRESENT

Andrey Bogdanov, Elmar Tischhauser, and Philip S. Vejre

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

June 29, 2016

Abstract. Extensions of linear cryptanalysis making use of multiple approximations such as 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 recently extended by Blondeau and Nyberg to also cover an updated wrong key randomization hypothesis, using similar assumptions. As a consequence, the effectiveness of multidimensional linear attacks seems significantly reduced, e.g. to only 24 rounds for PRESENT. It is therefore an important open problem how to take key dependent behaviour for both right and wrong keys into account without introducing other limiting assumptions in the process.
In this paper, we address this issue by proposing *multivariate linear cryptanalysis* as a new technique for using multiple linear approximations. Based on multivariate statistics and featuring a novel distinguishing technique based on quadratic discriminant analysis, it allows more realistic modelling of key dependence, while not relying on the limiting assumptions of previous work. Furthermore, it comes with a flexible signal/noise decomposition approach to allow for a realistic estimation of correlations. As an application of multivariate linear cryptanalysis, we provide attacks on 26 and 27 rounds (the latter marginally faster than exhaustive search) of PRESENT under much more realistic assumptions than previous work.

Keywords: linear cryptanalysis, multivariate, multidimensional cryptanalysis, key variance, PRESENT, key recovery, discriminant analysis, statistical attack.

Acknowledgements: The authors would like to thank Kaisa Nyberg for valuable comments on an earlier version of this work.

1 Introduction

Proposed by Matsui [34, 35] 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 for the first time both theoretically and experimentally. Influential cipher design paradigms such as the wide-trail strategy [22] were specifically developed as a response to the advert of linear and differential cryptanalysis. Nowadays, every newly proposed keyed symmetric primitive such as a block cipher, authenticated encryption scheme, or message authentication code, 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 [31], multiple linear cryptanalysis [4, 30], multidimensional linear cryptanalysis [25, 26, 27], zero-correlation linear cryptanalysis [15] and key-invariant bias attacks [12]. These extensions of linear cryptanalysis have provided the best single-key cryptanalytic results on ciphers such as Serpent [37], PRESENT [17, 49], CLEFIA [13], CAST-256 [48], and LBlock-s [47].

Parallel to the development of these cryptanalytic results, extensive research has been carried out to deepen our understanding of linear cryptanalysis [2] and their extensions [6], especially concerning links between differential and linear cryptanalysis [7] and truncated differential and multidimensional linear techniques [8]. How to provide resistance against these advanced cryptanalysis techniques has been studied in [5, 46].

Background. Already in his linear cryptanalysis of the DES [35], Matsui was making use of not just a single approximation to amplify his attack. Generalizations of this idea, notably multiple linear cryptanalysis [4,30] and multidimensional linear cryptanalysis [25, 26, 27] have proven to be among of the most influential extensions of linear cryptanalysis. In particular, multidimensional linear cryptanalysis has been used to devise the best known key recovery attacks on the ciphers Serpent [37] and especially PRESENT [17, 49].

The block cipher PRESENT. The block cipher PRESENT [14] was proposed at CHES 2007 as a very lightweight cipher. Having withstood extensive cryptanalytic scrutiny [17, 33, 40, 49], it has recently been standardised by ISO/IEC [1]. At CRYPTO 2015, a known-key distinguisher for full 31-round PRESENT was presented [11], which however does not translate to an attack in the secret-key setting.

The evolution of linear cryptanalytic techniques and the analysis of PRESENT. Since its publication, there has been significant interplay between the evolution of the cryptanalysis of PRESENT and the development of extensions to linear cryptanalysis. The statistical saturation attack [19] was devised to attack 15 of PRESENT's 31 rounds experimentally, and up to 24 rounds based on a theoretical estimate. Leander [33] later demonstrated that statistical saturation attacks can actually be interpreted within the framework of multidimensional linear cryptanalysis.

Essentially all further improvements are based on the crucial observation of Ohkuma [40], who demonstrated the presence of a strong linear hull effect in PRESENT, leading to a linear attack on 24 rounds. Following the development of multidimensional linear cryptanalysis [25, 26, 27], Cho [17] combined this with Ohkuma's approximations to obtain attacks on 25 and 26 rounds of PRESENT. Recently, an improvement of Cho's attack to 27 rounds was presented by Zheng and Zhang [49], based on his original model and complexity analysis. At CRYPTO 2015, Huang et al. [28] presented a refined analysis of Cho's multidimensional linear attacks, leading to a slightly increased time complexity for the 26-round attack. Later, Blondeau and Nyberg [9] further refined this analysis, concluding that under more realistic assumptions, multidimensional linear cryptanalysis can only reach up to 24 rounds of PRESENT.

Our problem. In order to evaluate the complexity of a multidimensional linear attack, one usually has to make certain assumptions to simplify the analysis. The original papers [25, 26, 27] show that the complexity depends on the *capacity* of the used linear approximations, which is equal [25] to the sum of the squared correlations of the linear approximations.

Cho's attacks on PRESENT [17] are developed under the assumption that for the right (target) key guess, the *key equivalence hypothesis* holds: namely, that the capacity for any fixed target key will be equal to the average over the key space. This however was demonstrated not to be the case in practice, in particular not for PRESENT [28, 33]. For the wrong key guesses, [17] assumes that the capacity will be identical to zero. This is a simplified interpretation of the *wrong key randomization hypothesis*, which states that wrong key guesses yield essentially random behaviour. It is however well known [21], that for a random permutation, the correlation of a linear approximation is not identically zero, but is itself distributed over the key space. This has been demonstrated to have a significant effect in standard linear cryptanalysis [16].

The refined analysis in [28] removes the limitation imposed by the key equivalence hypothesis, but does not take the distribution for the wrong keys into account, and, as demonstrated in Section 4, in turn needs additional assumptions regarding independence of linear approximations and identical variances that are typically not fulfilled in practice. The extended analysis of [9] does take the wrong key distribution into account, but does not address the issues of independence and identical variances.

Furthermore, all previous work relies on the assumption that the correlation of the used approximations for the right key can be evaluated in an exact manner. For ciphers with a block length of 64 bits or more, this is however computationally infeasible. Instead, one is limited to enumerating a certain amount of trails in the linear hulls defined by the approximations. We summarise the various complexity analyses of multidimensional linear cryptanalysis and their implications for PRESENT in Table 1.

Our problem is now as follows:

	Attack parameters					Limitations					
Rounds	*Success probability*	*#Approximations*	*Time complexity*	*Data complexity*	*Memory complexity*	*L1: Right key*	*L2: Wrong key*	*L3: Independence*	*L4: Identical variances*	*L5: Unknown trails*	**Reference**
24	99%	2295	$2^{59.8}$	$2^{59.8}$	$2^{34.0}$			✗	✗	✗	Blondeau, Nyberg [9]
25	95%	2295	$2^{65.0}$	$2^{62.4}$	$2^{34.0}$	✗	✗			✗	Cho [17]
	95%	2295	$2^{65.0}$	$2^{61.6}$	$2^{34.0}$		✗	✗	✗	✗	Huang, et al. [28]
	Impossible, according to [9]							✗	✗	✗	Blondeau, Nyberg [9]
26	95%	2295	$2^{72.0}$	$2^{64.0}$	$2^{34.0}$	✗	✗			✗	Cho [17]
	80%	2295	$2^{76.0}$	$2^{62.5}$	$2^{34.0}$		✗	✗	✗	✗	Huang, et al. [28][a]
	Impossible, according to [9]							✗	✗	✗	Blondeau, Nyberg [9]
	95%	**189**	$\mathbf{2^{73.0}}$	$\mathbf{2^{63.3}}$	$\mathbf{2^{51.0}}$						**Sect. 8.1**
27	95%	405	$2^{74.0}$	$2^{64.0}$	$2^{70.0}$	✗	✗			✗	Zheng, Zhang [49]
	Impossible, according to [9]							✗	✗	✗	Blondeau, Nyberg [9]
	95%	**189**	$\mathbf{2^{77.0}}$	$\mathbf{2^{63.8}}$	$\mathbf{2^{51.0}}$						**Sect. 8.2**[b]

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

Limitations (Addressed in Section)	**Explanation**
L1: Right key **(Sect. 3)**	Oversimplified key equivalence assumption; the right key linear potential is assumed equal to the average linear potential.
L2: Wrong key **(Sect. 3)**	Oversimplified wrong key randomisation hypothesis; the wrong key linear potential is assumed to be zero.
L3: Independence **(Sect. 4, Fig. 1)**	Simplifying assumptions about the multidimensional probabilities are made that imply independence of the linear correlations over the key space. In general, the linear correlations are not independent over the key space.
L4: Identical variances **(Sect. 5, Fig. 1)**	The variance over the key space of either the correlation values or the multidimensional data values are assumed identical. This is very unlikely to be the case in practice.
L5: Unknown trails **(Sect. 6, Fig. 3)**	Only a part of the hull is analysed – the remaining trails are not considered. This might give a misleading estimate of the linear correlation.

Table 1: *Top*: Linear cryptanalysis of PRESENT. For each attack, the complexities are given, and limitations are highlighted for the models in which the attacks were stated. *Bottom*: Explanations of model limitations.

Can we develop a technique for using multiple linear approximations which simultaneously

1. *incorporates realistic distributions over the right and wrong keys;*
2. *does not rely on simplifying assumptions such as independence or identical variances of the correlations;*
3. *provides a strategy for taking the unknown part of the linear hulls into account; and lastly*
4. *allows us to attack more than 24 rounds of* PRESENT*?*

Our results. In this paper, we answer the above-mentioned questions in the affirmative. To this end, we propose *multivariate linear cryptanalysis* as a technique for linear cryptanalysis with multiple approximations. By developing the underlying model based on multivariate statistics, we are able to incorporate both realistic key equivalence and wrong key randomization hypotheses, as well as to avoid any issues related to dependence of approximations or assuming identical variances. We give an informal preview of the technique in the following model:

Model (Multivariate Linear Cryptanalysis). *Let* (α_i, β_i)*,* $i = 1, \ldots, M$*, be* M *different linear approximations of a cipher, 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 $\mathbf{C}^K \sim \mathcal{N}_M(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *over the key space, for some mean vector* $\boldsymbol{\mu}$ *and covariance matrix* $\boldsymbol{\Sigma}$*.*

In detail, the contributions of our paper are as follows.

Analysis of previous attack models. We investigate the underlying assumptions of previous models for analysing the complexity of multidimensional linear attacks. In particular, we show that all current approaches to incorporating realistic versions of the key equivalence and wrong key randomization hypotheses into the model lead to assumptions equivalent to that of statistical independence – which was one of the issues multidimensional linear cryptanalysis intended to overcome. For an overview of how different limitations are essentially traded off against each other, we refer to Table 1.

Proposal of multivariate linear cryptanalysis. We develop and propose *multivariate linear cryptanalysis* as a new technique for symmetric cryptanalysis using multiple linear approximations. It allows us to incorporate more realistic interpretations of both the key equivalence and wrong key randomization hypotheses, while at the same time avoiding any unrealistic assumptions about the correlation distributions (as illustrated in Table 1). Our concerns about the validity of these assumptions are exemplified in Fig. 1 for the full PRESENT. Unlike previous work, multivariate linear cryptanalysis also comes with a flexible signal/noise decomposition approach to allow for a realistic estimation of correlations based on a number of enumerated known trails (as illustrated in Figure 2).

New distinguishing technique based on QDA. We propose a new statistical technique for distinguishing the multivariate distributions based on quadratic discriminant analysis (QDA). Compared to conventional binary hypothesis testing, it offers greater advantages over exhaustive search for all presented attacks (cf. Sect. 7).

Attacks on 26/27 round PRESENT. As an application of the new multivariate technique, we apply it to the block cipher PRESENT, obtaining an attack on 26 rounds with time/data/memory complexities of $2^{73}/2^{63.3}/2^{51}$ and a marginal attack on 27 rounds with complexities $2^{77}/2^{63.8}/2^{51}$; both under much more realistic assumptions than previous work (see Table 1 and Sect. 8).

Experimental verification. Both the relevance of the limitations of previous models and the advantage of the new multivariate model are supported by extensive experimental evidence on both small-scale and the full PRESENT.

Organization. The remainder of the paper is organized as follows. Sect. 2 introduces the preliminaries. Sect. 3 discusses limitations of the current results on PRESENT. Sect. 4 gives results on independence in the current linear models. Sect. 5 presents multivariate linear cryptanalysis. Sect. 6 shows how to compensate for the unknown part of the hull. Sect. 7 gives two distinguishing methods in the new model. In Sect. 8 we attack PRESENT with our new technique. We conclude in Sect. 9.

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 block size n and key length κ. 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 completely 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*. Usually, a key-schedule is used to expand the κ-bit master key K into $r + 1$ n-bit round keys. By $\bar{K} = k_0 \| k_1 \| \dots \| k_r$ we denote the expanded key, i.e. the concatenation of the round keys.

2.1 Linear Cryptanalysis

Linear cryptanalysis was introduced by Matsui in 1993 [34]. A *linear approximation* of a cipher is a pair $(\alpha, \beta) \in \mathbb{F}_2^n \times \mathbb{F}_2^n \backslash (0, 0)$. 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 the probability is taken over all $x \in \mathbb{F}_2^n$ and $\cdot$ denotes the canonical inner product on $\mathbb{F}_2^n$. Assuming that K is drawn at random from $\mathbb{F}_2^\kappa$, $C_{\alpha,\beta}^K$ is a random variable over the key space. We denote the distribution of this variable over the keys by $C_{\alpha,\beta}$. In [21], Daemen and Rijmen proved that this distribution is approximately $\mathcal{N}(0, 2^{-n})$ for an ideal cipher and any non-trivial linear approximation. Now, if we can find a linear approximation of the target cipher with a distribution $C_{\alpha,\beta}$ sufficiently different from $\mathcal{N}(0, 2^{-n})$, we can use this property to distinguish the cipher from an ideal cipher.

Let (u_i, u_{i+1}), $i = 0, \ldots, r-1$ be a series of one round linear approximations of an iterative block cipher. Such a series of approximations is called a *linear trail*. We will often denote the trail by the concatenation of its masks, i.e. $U = u_0 \| \ldots \| u_r$. 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 such that $u_0 = \alpha$ and $u_r = \beta$ is called the *linear hull* of (α, β). The correlation of (α, β) is then the sum of the correlation contributions of the hull [20, 22]:

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

For a key-alternating cipher, the correlation contribution of a trail can be written as $C_U^K = (-1)^{U \cdot \bar{K} \oplus d_U} C_U$, where $C_U = |C_U^K|$ for any key, and d_U is a sign bit indicated by the sign of C_U^K for the all-zero expanded key [22]. In this case, we can therefore rewrite Equation 1 as

$$C_{\alpha,\beta}^K = \sum_{u_0=\alpha, u_r=\beta} (-1)^{U \cdot \bar{K} \oplus d_U} C_U. \tag{2}$$

Finally, $(C_{\alpha,\beta}^K)^2$ is called the *linear potential (LP)* and $\mathrm{E}((C_{\alpha,\beta}^K)^2)$ the *expected linear potential (ELP)* over keys.

2.2 Statistical Distinguishing

In statistical cryptanalysis, distinguishing is usually a question of determining which distribution an observed value came from – typically one aims to distinguish between the distributions of an ideal and a non-ideal cipher. This value will be computed from the observed cipher data, and we will refer to it in general as the *test statistic* $\mathcal{T}$. In differential cryptanalysis, $\mathcal{T}$ would be the observed probability of a differential, and in classical linear cryptanalysis, $\mathcal{T}$ would be the observed linear correlation of an approximation. We now consider two types of general distinguishing techniques.

Distinguishing as Hypothesis Testing. Assume that $\mathcal{T} \sim \mathcal{D}_I$, if the observed data was generated by an ideal cipher, and $\mathcal{T} \sim \mathcal{D}_N$, otherwise. If the data came from a univariate distribution, we can use binary hypothesis testing to determine which distribution $\mathcal{T}$ came from. Without loss of generality, assume that $\mathrm{E}(\mathcal{D}_I) \leq \mathrm{E}(\mathcal{D}_N)$. Then we conclude that the data came from a non-ideal cipher if $\mathcal{T} \geq \tau$, and from an ideal cipher if $\mathcal{T} < \tau$, for some threshold value τ. In this case, the probability of type I error (false positive) and type II error (false negative) is given by

$$\begin{aligned} P_I &= 1 - F_{\mathcal{D}_I}(\tau), \\ P_{II} &= F_{\mathcal{D}_N}(\tau), \end{aligned}$$

where F_X is the cumulative distribution function of X. The probability of success is then $P_S = 1 - F_{\mathcal{D}_N}(\tau)$. The exact value of τ can then be chosen based on the desired value of these probabilities.

Distinguishing with QDA. If the cipher data comes from a d-variate distribution ($d > 1$), then we can use other methods. Assume that the distribution is a d-variate normal distribution $\mathcal{N}_d(\boldsymbol{\mu}_I, \boldsymbol{\Sigma}_I)$, for an ideal cipher, and $\mathcal{N}_d(\boldsymbol{\mu}_N, \boldsymbol{\Sigma}_N)$, for a non-ideal cipher. In this case, we can use quadratic discriminant analysis (QDA) to distinguish. For QDA with equal misclassification costs and probabilities, we define the test statistic (classically called the discriminant function)

$$\mathcal{T} = (\mathbf{X} - \boldsymbol{\mu}_N)^\top \boldsymbol{\Sigma}_N^{-1} (\mathbf{X} - \boldsymbol{\mu}_N) - (\mathbf{X} - \boldsymbol{\mu}_I)^\top \boldsymbol{\Sigma}_I^{-1} (\mathbf{X} - \boldsymbol{\mu}_I), \tag{3}$$

where $\mathbf{X}$ is the cipher data. We note that the form given here deviates from the discriminant function normally found in the literature by a shift and a scaling factor. This has no impact on our analysis. We decide that the data came from an ideal cipher if $\mathcal{T} \geq \tau$, and from a non-ideal cipher if $\mathcal{T} < \tau$, for some threshold value τ. We can then obtain P_I and P_{II} as

$$\begin{aligned} P_I &= F_{\mathcal{D}_I}(\tau), \\ P_{II} &= 1 - F_{\mathcal{D}_N}(\tau). \end{aligned}$$

However, in this case it is not straightforward to exactly determine $\mathcal{D}_I$ and $\mathcal{D}_N$, and thus P_I, P_{II}, P_S, and τ. (The easiest solution to this problem is to simulate observations from $\mathcal{D}_I$ and $\mathcal{D}_N$ in order to get an estimate of these values. In most cases, this will not be a problem.)

From Distinguishing to Key Recovery. For both methods, we define the advantage $a = -\log_2(P_I)$, a notion first introduced by Selçuk in [44, 45] in the context of key-ranking. Due to recent criticisms of the key-ranking approach, e.g. [42], we will not be considering this method. Luckily, the key-ranking results can be recovered using the hypothesis testing technique above. The advantage plays an important role when we use a distinguisher as a part of a key-recovery attack. To do this for a linear distinguisher, the standard tool is the wrong key randomisation hypothesis.

Algorithm 1 80-bit PRESENT key-schedule.

1: **for** $i = 0$ **to** 31 **do**
2: $K_i = (k_{79}k_{78} \dots k_{17}k_{16})$ ▷ Extract 64 leftmost keybits from K
3: $(k_{79}k_{78} \dots k_1k_0) = (k_{18}k_{17} \dots k_{20}k_{19})$ ▷ Rotate K 61 positions to the left
4: $(k_{79}k_{78}k_{77}k_{76}) = S(k_{79}k_{78}k_{77}k_{76})$ ▷ Apply S-box to leftmost nibble of K
5: $(k_{19}k_{18}k_{17}k_{16}k_{15}) = (k_{19}k_{18}k_{17}k_{16}k_{15}) \oplus (i + 1)$
6: **end for**

Hypothesis 1 (Wrong Key Randomisation). *Consider a linear approximation (α, β) over r rounds of a cipher. Let (P, C) be a plaintext-ciphertext pair encrypted with key K over $r + \ell$ rounds, $\ell > 0$. Assume that we pick a key K', encrypt/decrypt ℓ rounds of (P, C), and measure $C^{K'}_{\alpha,\beta}$. If $K \neq K'$, $C^{K'}_{\alpha,\beta}$ will be distributed as $\mathcal{N}(0, 2^{-n})$.*

Although not stated exactly as above, the general idea for this hypothesis was first given by Harpes et al. in [24] and later stated in the context of linear cryptanalysis by Junod in [29]. The hypothesis, as we have stated it here, was given in [16].

2.3 PRESENT

PRESENT is an ultra-lightweight, key-iterated, 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. We will denote the S-boxes by $S_0, \dots, S_{15}$, with S_0 affecting the 4 least significant bits of the state. The bit permutation sends the bit in position i to position $P(i)$, where P is given by

$$P(i) = \begin{cases} 16 \cdot i \bmod 63 & \text{for } 0 \leq i < 63 \\ 63 & \text{for } i = 63 \end{cases}.$$

An additional round key is added after the last round. The 32 round keys are derived through a key-schedule. The key-schedule for the 80-bit key version of PRESENT is given in Algorithm 1. Here, $K = (k_{79} \dots k_0)$ is initially the 80-bit master key, and K_i is the i'th round key. For further details, we refer to [14].

3 Multidimensional Linear Cryptanalysis of PRESENT

Currently, the best cryptanalytic results on PRESENT have been achieved with multidimensional linear cryptanalysis. In this section, we first give a short overview of multidimensional linear cryptanalysis. We consider the results on PRESENT, and make observations regarding the underlying assumptions made in the models used. We argue that all current attacks are derived using models that exhibit one or more limitations.

Kaliski and Robshaw [30], as well as Biryukov et al. [4], considered using multiple approximations to obtain more powerful linear attacks, under the assumption that the approximations are statistically independent over the *text space*. As Murphy showed in [36], this assumption does not hold in general. Multidimensional linear cryptanalysis tries to work around this problem. It builds on the work done by Baigneres et al. in [3], while most of the main results were given by Hermelin, Cho, and Nyberg in [25, 26, 27]. In multidimensional linear cryptanalysis, an m-dimensional subspace of $\mathbb{F}_2^n \times \mathbb{F}_2^n$ is considered. The distribution of a plaintext-ciphertext pair restricted to this subspace, say (P', C'), is then studied. This distribution can be described by the vector

$$\boldsymbol{\eta}^K = (\eta_0^K, \ldots, \eta_{2^m-1}^K),$$

where $\eta_i^K = \Pr(P' \| C' = i)$. $\boldsymbol{\eta}^K$ is a key-dependent, 2^m-dimensional, discrete probability distribution. The *capacity* is defined as a measure of distance between $\boldsymbol{\eta}^K$ and the uniform distribution:

$$\mathcal{C}^K = \sum_{i=0}^{2^m-1} \frac{(\eta_i^K - 2^{-m})^2}{2^{-m}}. \tag{4}$$

Now, let $(\alpha_1, \beta_1), \ldots, (\alpha_{2^m-1}, \beta_{2^m-1})$ be all possible non-zero linear approximations of the m-dimensional subspace. This collection is called a multidimensional linear approximation. It can be shown [25] that

$$\mathcal{C}^K = \sum_{i=1}^{2^m-1} (C_{\alpha_i,\beta_i}^K)^2 \quad \text{and} \quad \mathrm{E}(\mathcal{C}^K) = \sum_{i=1}^{2^m-1} ELP_i,$$

where ELP_i is the ELP of (α_i, β_i). As for classic linear cryptanalysis, if we can find a multidimensional linear approximation of the target cipher where the distribution of $\mathcal{C}^K$ is sufficiently different from that of an ideal cipher, we can use this property as a distinguisher.

3.1 Cho's Cryptanalysis of PRESENT

The idea of using more than one approximation was considered already by Matsui in 1993. At the time, results in linear cryptanalysis were derived under the assumption that the linear correlations were largely the same for all keys, or that a small number of key-classes partitioned the key space [4, 24, 29, 30]. Thus, only the variance of the observed correlations over the text space was considered. It is therefore in this setting that the first results on multidimensional linear cryptanalysis were developed.

In 2010, Cho gave an attack on 25 and 26 rounds of PRESENT [17]. This attack was derived using the model developed in [26]. As such, the assumption was made that for an ideal cipher, $\eta_i^K = 2^{-m}$, i.e. the uniform distribution, for all keys. Under this assumption, an ideal cipher will have a fixed capacity of zero. In this sense, the attack relies on a wrong key randomisation hypothesis that

deviates from the one we have given in Hypothesis 1, as under this hypothesis the capacity is not fixed, even for an ideal cipher. Moreover, it was assumed that for a non-ideal cipher, all keys have the same multidimensional distribution, i.e. $\boldsymbol{\eta}^K = \boldsymbol{\eta}$ for all K. This implies that the capacity is the same for all keys. The analysis in [17] therefore uses the expected capacity, assuming that this is a good estimate for all (or most) keys. In this setting, if the text pairs are drawn independently at random, Equation 4 is akin to a Pearson χ^2 test statistic, with a known limiting distribution, as shown in [26]. Cho uses this result to derive the complexity of his attack.

Consider the following: It is known that the multidimensional probabilities can be expressed as linear combinations of the correlations [27]. For an ideal cipher [21] shows that the correlations do have a distribution over the key space. Similarly, Equation 2 shows the key-dependence of the correlation values for a non-ideal cipher, and for most practical ciphers we observe that this distribution can have a large variance. Hence, while assuming fixed wrong key or right key correlations and capacity simplifies the analysis considerably, we risk obtaining misleading results. In this work, in particular in 5, we will consider a setting where the key dependent behaviour is taken into account.

In 2015, a 27-round attack on PRESENT was given by Zheng and Zhang [49]. This attack uses the same model as Cho, and is therefore prone to the same limitations. Additionally, both attacks analyse only the 1-bit trails of PRESENT to derive the capacity values and the resulting strength of the attack. As it was argued by Ohkuma in [40], the 1-bit trails of PRESENT do have a much larger correlation contribution than any other trails – and indeed, the best approximations of PRESENT are those that have a large number of these trails. However, the number of trails for any PRESENT approximation is extremely large. Therefore, even if the remaining trails have a small correlation contribution, they can still contribute significantly to the correlation distribution. As such, we would argue that somehow modelling the remainder of the hull will improve the estimate of the correlation distributions.

3.2 Multidimensional Linear Cryptanalysis With Key Dependence

While Cho, Zheng, and Zhang simplify the key dependent behaviour of the capacity, some work has been done to analyse the capacity distribution over the key space. The papers [9,28] both considered multidimensional linear cryptanalysis when the key influence is taken into account. In [28], Huang et al. give a model for the key dependent behaviour of $\boldsymbol{\eta}^K$, and the resulting distribution of the capacity was studied. In particular, the following proposition was given.

Proposition 1 ([28]). *In an m-dimensional linear attack with the probability distributions η_i^K i.i.d. to a normal distribution $\mathcal{N}(2^{-m}, \sigma^2)$, $i \in \mathbb{F}_2^m$, the capacity $\mathcal{C}^K$ follows a Gamma-distribution $\Gamma(\frac{2^m-1}{2}, 2 \cdot 2^m\sigma^2)$.*

We note that Proposition 1 imposes the limitation, that the η_i^K are independent over the key space and all have the same normal distribution. Moreover, the

distribution in the ideal case is not considered in [28]. The same result is given by Blondeau and Nyberg in [9], and additionally the ideal case is considered. The following more general result is also shown.

Theorem 1 ([9]). *In an m-dimensional linear attack, assume that $\eta_i^K \sim \mathcal{N}(\lambda_i, \sigma^2)$, $i \in \mathbb{F}_2^m$. Then $\mathcal{C}^K/(2^m\sigma^2) \sim \chi^2_{2^m-1}(\delta)$, where $\delta = \mathrm{E}(\mathcal{C}^K)/(2^m\sigma^2)$.*

Interestingly, as a consequence of this theorem, the analysis in [9] shows that if the capacity of a multidimensional approximation is lower than $2^{-n}(2^m - 1)$, then a multidimensional attack is not possible. This includes Cho's attack on 25 and 26 rounds of PRESENT, as well as Zheng and Zhang's attack on 27 rounds. However, this would mean that the multidimensional approximation used has one-dimensional correlation values worse than those of an ideal cipher – it is not clear if that can actually happen for a cipher such as PRESENT. We address this issue in Sect. 6.

While not explicitly stated in the assumptions of Theorem 1, the proof in [9] does use the fact that the η_i^K are independent, in order to arrive at the non-central χ^2 distribution. Thus, while both [28] and [9] make valuable corrections to Cho's analysis, simplifying assumptions are needed that limit the applicability of the model, i.e. independence and identical variances. Indeed, we show in the following section that the assumptions of Proposition 1 and Theorem 1 imply that the linear correlations are statistically independent over the key space.

4 Implicit Independence Assumption in Multidimensional Linear Cryptanalysis

As discussed in the previous section, the current state of the art analysis of multidimensional linear cryptanalysis is that given in [9, 28]. This analysis, however, still relies on some limiting assumptions, namely that the multidimensional probabilities are independent and have identical variances over the key space. In this section, we consider what these assumptions imply with regards to the statistical dependence of the linear correlations over the key space.

Joint Normality of Correlations. We start by stating a model that allows for multidimensional distributions with arbitrary parameters, as opposed to the settings of Proposition 1 and Theorem 1.

Model (Multidimensional Linear Cryptanalysis). *In an m-dimensional linear attack, the η_i^K, $i \in \mathbb{F}_2^m$ are independent and distributed as $\mathcal{N}(\lambda_i, \zeta_i^2)$ over the key space.*

By definition of the multivariate normal distribution, the multidimensional model immediately implies that $\boldsymbol{\eta}^K \sim \mathcal{N}_{2^m}(\boldsymbol{\lambda}, \mathbf{Z})$ over the key space, where $\boldsymbol{\lambda} = (\lambda_0 \cdots \lambda_{2^m-1})^\top$ and $\mathbf{Z}$ is a diagonal matrix with the ζ_i^2 in the diagonal. From this observation, we obtain the following theorem.

Theorem 2. *Let $\boldsymbol{\eta}^K$ be the key dependent probability distribution of an m-dimensional linear approximation, and let $\mathbf{C}^K$ be the vector of linear correlations for the corresponding 1-dimensional approximations. Then, in the multidimensional model,*

$$\mathbf{C}^K \sim \mathcal{N}_{2^m}(\mathbf{A}\boldsymbol{\lambda}, \mathbf{AZA}),$$

over the key space, where $\boldsymbol{\lambda} = (\lambda_0 \cdots \lambda_{2^m-1})^\top$, $\mathbf{Z} = \operatorname{diag}(\zeta_0^2, \ldots, \zeta_{2^m}^2)$, and $\mathbf{A}$ is a $2^m \times 2^m$ matrix with $\mathbf{A}_{i,j} = (-1)^{i \cdot j}$.

Proof. As noted, since the η_i are independent normal distributions, the vector $\boldsymbol{\eta}^K = (\eta_0^K \cdots \eta_{2^m-1}^K)^\top$ has multivariate normal distribution $\mathcal{N}_{2^m}(\boldsymbol{\lambda}, \mathbf{Z})$ over the key space, with $\boldsymbol{\lambda}$ and $\mathbf{Z}$ as given in the theorem. Then $\mathbf{C}^K = \mathbf{A}\boldsymbol{\eta}$ is a vector of the corresponding linear correlations since

$$C_{\alpha,\beta}^K = \sum_{i \in \mathbb{F}_2^m} (-1)^{(\alpha \| \beta) \cdot i} \eta_i^K,$$

as shown in [27]. Note that the first element of $\mathbf{C}^K$ is the linear correlation of the trivial approximation $(0, 0)$. Now, let $\mathbf{y} \sim \mathcal{N}_{2^m}(\mathbf{0}, \mathbf{I})$. We can then write,

$$\mathbf{C}^K = \mathbf{ABy} + \mathbf{A}\boldsymbol{\lambda},$$

where $\mathbf{B} = \operatorname{diag}(\zeta_0, \ldots, \zeta_{2^m-1})$. This implies that

$$\mathbf{C}^K \sim \mathcal{N}_{2^m}(\mathbf{A}\boldsymbol{\lambda}, \mathbf{AB}(\mathbf{AB})^\top) = \mathcal{N}_{2^m}(\mathbf{A}\boldsymbol{\lambda}, \mathbf{ABB}^\top\mathbf{A}^\top),$$

and since $\mathbf{BB}^\top = \mathbf{Z}$ and $\mathbf{A}$ is symmetric, this proves the theorem. □

By Theorem 2, the multidimensional model also implies that the vector of correlations, $\mathbf{C}^K$, has a joint normal distribution. However, since $\mathbf{AZA}$ is not necessarily a diagonal matrix, the correlations are not necessarily statistically independent over the key space. Thus, going from the correlation values $\mathbf{C}^K$ to the multidimensional distribution $\boldsymbol{\eta}$ is a matter of transforming a set of dependent random variables to a set of independent random variables using a linear transform. If we let $\mathbf{C}^K \sim \mathcal{N}_{2^m}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, this transform corresponds to diagonalising $\boldsymbol{\Sigma}$. Moreover, this transform is always possible if $\mathbf{C}^K$ has a multivariate normal distribution, as any symmetric matrix is diagonalisable. The following corollary shows us which diagonalisation results in the multidimensional distribution $\boldsymbol{\eta}$.

Corollary 1. *Let $\boldsymbol{\eta}^K$, $\mathbf{C}^K$, and $\mathbf{A}$ be as in Theorem 2 with the notation $\mathbf{C}^K \sim \mathcal{N}_{2^m}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Then $\boldsymbol{\lambda} = 2^{m/2}\mathbf{A}\boldsymbol{\mu}$ and $\mathbf{Z} = 2^m\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}$.*

Proof. We simply note that $\mathbf{A}^{-1} = 2^{m/2}\mathbf{A}$. □

$$\begin{pmatrix} \mathbf{2^{-47.88}} & 0 & 0 & 0 & \cdots \\ 0 & \mathbf{2^{-47.88}} & 0 & 0 & \cdots \\ 0 & 0 & \mathbf{2^{-47.88}} & 0 & \cdots \\ 0 & 0 & 0 & \mathbf{2^{-47.88}} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} \quad \begin{pmatrix} 2^{-47.93} & -2^{-58.14} & -2^{-55.60} & 2^{-54.33} & \cdots \\ -2^{-58.14} & 2^{-47.94} & -2^{-55.05} & -2^{-55.58} & \cdots \\ -2^{-55.60} & -2^{-55.05} & 2^{-47.94} & -2^{-54.67} & \cdots \\ 2^{-54.33} & -2^{-55.58} & -2^{-54.67} & 2^{-47.95} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

(a) Multidimensional model. (b) Measured covariance matrix.

Fig. 1: Partial covariance matrix of the linear correlations for a multidimensional approximation over 17 rounds of PRESENT. The covariance matrix in the multidimensional model was calculated using Ohkuma's results on optimal 1-bit trails [40], while the actual covariance matrix was calculated by enumerating the trails explicitly and calculating their correlation contributions for 5000 random keys.

Independence of Correlations. Now, let us consider the assumptions used in Proposition 1 and Theorem 1. These are equivalent to the multidimensional model with additional restrictions on the λ_i and ζ_i^2 of the distributions. In this case, we get the following result by directly using Theorem 2 and the structure of $\mathbf{A}$.

Corollary 2. *Let $\boldsymbol{\eta}^K$ be as in Theorem 2 with $\zeta_i^2 = \zeta^2$ for $i = 0, \ldots, 2^m - 1$. Then, for the corresponding non-zero approximations, $C^K_{\alpha_i,\beta_i} \sim \mathcal{N}(\mu_i, 2^m\zeta^2)$, $i = 1, \ldots, 2^m - 1$. Additionally, the linear correlations are statistically independent. Furthermore, if $\lambda_i = \lambda$, $i = 0, \ldots, 2^m - 1$, then the μ_i are zero.*

Interestingly, Corollary 2 shows that the assumptions of Proposition 1 and Theorem 1 imply that the linear correlations are independent over the key space. Thus, while multidimensional linear cryptanalysis originally solved the problem of statistical dependence over the text space, the current best results on the behaviour over the key space still needs assumptions equivalent to that of statistical independence of the linear correlations.

As an illustration, see Fig. 1. Here, we consider a multidimensional approximation of PRESENT which contains a number of one-dimensional approximations with a maximal number of 1-bit trails. In the setting of Corollary 2, $\boldsymbol{\Sigma}$ will be a diagonal matrix, and the variances will be the sum of squared trail contributions (since $\mu_i = 0$). The squared correlation contribution of a 1-bit PRESENT trail over r rounds is 2^{-4r}, and Ohkuma gave the maximal number of such 1-bit trails in [40]. From this, we calculate part of $\boldsymbol{\Sigma}$ in the multidimensional model, shown in Fig. 1a. We calculated the same part of $\boldsymbol{\Sigma}$ directly by enumerating the exact same 1-bit trails and then calculating their correlation contribution for 5000 random master keys. The result is shown in Fig. 1b, and clearly deviates from the $\boldsymbol{\Sigma}$ implied by the multidimensional model. The variances are not equal to the sum of squared correlation contributions, indicating that $\mu_i \neq 0$, and the covariances are not zero, showing that the correlations are clearly not statistically

independent. In the following section, we develop a model that expands on the idea of joint normality of the correlation values, while removing assumptions of independence and restrictions on the distribution.

5 Establishing Multivariate Linear Cryptanalysis

As argued in Sect. 4, if the multidimensional probabilities are independent normal distributions over the key space, then the corresponding collection of linear correlations will also be jointly normal, but not necessarily independent. However, if the multidimensional probabilities all have the same variance over the keys, as was assumed in [9, 28], then the correlations will also be independent. Thus, the current best analysis in the area of multidimensional linear cryptanalysis hinges on assumptions that are not necessarily sound for practical ciphers. Moreover, current analysis only considers part of the linear hull. In this section and Sect. 6, we aim to create a model for linear cryptanalysis with multiple approximations that:

- Does not assume independence over the text or key space,
- Does not assume any particular structure of the correlation distributions,
- Takes into account the unknown part of the linear hull.

We will base our model on the idea of the linear correlations having a joint normal distribution. In Theorem 2, the vector $\mathbf{C}^K$ represents the correlations of all linear approximations of a subspace of $\mathbb{F}_2^n \times \mathbb{F}_2^n$. As a generalisation, consider a vector $\mathbf{C}^K$ containing the correlations of *any* M linear approximations of a cipher. We propose the following new model.

Model (Multivariate Linear Cryptanalysis). *Let* (α_i, β_i)*,* $i = 1, \ldots, M$*, be* M *different linear approximations of a cipher, 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 $\mathbf{C}^K \sim \mathcal{N}_M(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *over the key space, for some mean vector* $\boldsymbol{\mu}$ *and covariance matrix* $\boldsymbol{\Sigma}$*.*

While the multidimensional model imposes a certain structure on the distribution of $\mathbf{C}^K$, namely that given in Theorem 2, the multivariate model only assumes joint normality. In this sense, the multivariate model makes a weaker assumption than the multidimensional. Since $\boldsymbol{\Sigma}$ can be arbitrary (as long as it is positive semi-definite), we also do not need any assumption about the statistical dependence of the approximations.

Capacity in the Multivariate Model. We now turn our attention to the distribution of the capacity over the keys. For our analysis, we will need the following concept: For a random d-vector $\mathbf{X}$ and a constant, symmetric, $d \times d$

matrix $\mathbf{Q}$, we say that $Y = \mathbf{X}^\top \mathbf{Q}\mathbf{X}$ is a quadratic form in $\mathbf{X}$. In particular, we can express the capacity as the quadratic form

$$\mathcal{C}^K = (\mathbf{C}^K)^\top \mathbf{I}\mathbf{C}^K = \sum_{i=1}^{M} (C^K_{\alpha_i,\beta_i})^2. \tag{5}$$

We will need the following lemma to determine the mean and variance of $\mathcal{C}$.

Lemma 1 ([43]). *Let $\mathbf{X} \sim \mathcal{N}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\mathbf{Q}$ a symmetric, $d \times d$ matrix. Then*

$$\begin{aligned} \mathrm{E}(\mathbf{X}^\top \mathbf{Q}\mathbf{X}) &= \mathrm{tr}(\mathbf{Q}\boldsymbol{\Sigma}) + \boldsymbol{\mu}^\top \mathbf{Q}\boldsymbol{\mu}, \\ \mathrm{Var}(\mathbf{X}^\top \mathbf{Q}\mathbf{X}) &= 2\mathrm{tr}(\mathbf{Q}\boldsymbol{\Sigma}\mathbf{Q}\boldsymbol{\Sigma}) + 4\boldsymbol{\mu}^\top \mathbf{Q}\boldsymbol{\Sigma}\mathbf{Q}\boldsymbol{\mu}. \end{aligned}$$

Using this lemma and the quadratic form given in Equation 5, we obtain the following result. Note that this result says nothing about the shape of the capacity distribution – in general, this distribution is difficult to determine explicitly when the $C^K_{\alpha_i,\beta_i}$ are not independent.

Theorem 3. *Consider the approximations (α_i, β_i), $i = 1, \ldots, M$ in the multivariate model. Then their capacity has mean and variance over the key space given by*

$$\mathrm{E}(\mathcal{C}^K) = \sum_{i=1}^{M} ELP_i \quad \textit{and} \quad \mathrm{Var}(\mathcal{C}^K) = 2\mathrm{tr}(\boldsymbol{\Sigma}\boldsymbol{\Sigma}) + 4\boldsymbol{\mu}^\top \boldsymbol{\Sigma}\boldsymbol{\mu}.$$

Theorem 3 is useful only if we can measure the exact correlation values. However, in an attack scenario, we typically will not be able to do this, as this would require us to obtain the full codebook. We therefore have to understand how undersampling affects the distribution of $\mathbf{C}^K$.

The Undersampled Distribution. In [36], Murphy showed 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, $\hat{\mathbf{C}}^{K_0} \sim \mathcal{N}_M(\boldsymbol{\mu}^{K_0}, \boldsymbol{\Sigma}^{K_0,N})$, where $\boldsymbol{\mu}_i^{K_0} = C^{K_0}_{\alpha_i,\beta_i}$ 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 $\mathcal{N}_M(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Indeed, $\boldsymbol{\Sigma}^{K_0,N}$ also has a distribution over the key space, but to simplify our analysis, we will only consider the mean of this matrix distribution, namely

$$\boldsymbol{\Sigma}^{N}_{i,j} = \begin{cases} N^{-1} \mathrm{E}(C^K_{\alpha_i \oplus \alpha_j, \beta_i \oplus \beta_j}) & \text{for } i \neq j, \\ N^{-1} & \text{for } i = j. \end{cases} \tag{6}$$

Then, the empirical value of $\mathbf{C}^K$ over keys measured using N random text pairs is approximately given by

$$\hat{\mathbf{C}}^K \sim \mathcal{N}_M(\mathcal{N}_M(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \boldsymbol{\Sigma}^N) = \mathcal{N}_M(\boldsymbol{\mu}, \boldsymbol{\Sigma} + \boldsymbol{\Sigma}^N).$$

With this, we get the following corollary of Theorem 3.

Corollary 3. *Consider the approximations (α_i, β_i), $i = 1, \ldots, M$ in the multivariate model. Then their empirical capacity measured using N random text pairs approximately has mean and variance*

$$\mathrm{E}(\hat{\mathcal{C}}^K) = MN^{-1} + \sum_{i=1}^{M} ELP_i,$$
$$\mathrm{Var}(\hat{\mathcal{C}}^K) = 2\mathrm{tr}\left((\boldsymbol{\Sigma} + \boldsymbol{\Sigma}^N)^2\right) + 4\boldsymbol{\mu}^\top(\boldsymbol{\Sigma} + \boldsymbol{\Sigma}^N)\boldsymbol{\mu},$$

where $\boldsymbol{\Sigma}^N$ is given in Equation 6*.*

Theorem 3, and by extension Corollary 3, gives us a way to estimate the distribution of the capacity over the keys for a completely arbitrary set of approximations. In contrast to Proposition 1 and Theorem 1, no assumptions about independence or the parameters of the involved distributions are required. Additionally, we can pick any M good approximations without having to consider potentially bad approximations to complete a subspace. However, this generality comes at the price of having to estimate at least M^2+M parameters that describe $\mathbf{C}^K$. We will discuss this problem in the following section.

Moreover, if the approximations do not form a full subspace, we potentially have to approximate another M^2 parameters to determine $\boldsymbol{\Sigma}^N$. We note, however, that we can sometimes make a simplifying assumption about $\boldsymbol{\Sigma}^N$. Typically, we use approximations that only active one S-box at the input and one S-box at the output. If the approximations activate different S-boxes, their resulting XOR combined approximations will activate multiple S-boxes, and will therefore have comparatively small correlation values. In this case, we can reasonably simplify Equation 6 to a diagonal matrix with N^{-1} in the diagonal. This simplification does not seem to have any significant impact in practice.

6 Signal/Noise Decomposition for Multivariate Linear Cryptanalysis

Undersampling is not the only uncertainty introduced when applying linear cryptanalysis. In most cases, we will also be unable to calculate the exact distribution of $C^K_{\alpha,\beta}$ for any approximation. While Equations 1 and 2 are promising tools in theory, if the number of trails is very large, we will not be able to compute the full sum. One approach to this problem is to find a set of "dominant trails", i.e. trails with a large contribution to the total correlation [16, 41]. Furthermore, most analysis on the topic only considers the expected correlations or capacity,

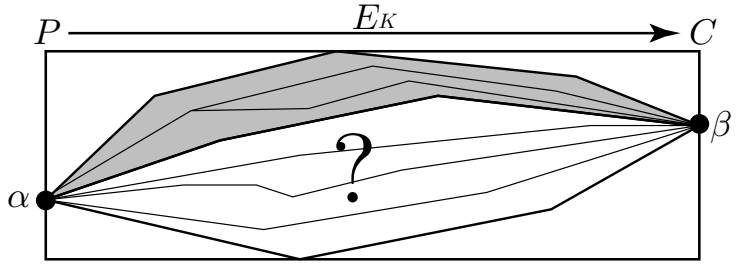


Fig. 2: Illustration of the signal/noise decomposition of a linear hull. The lines from α to β represents the linear trails of the cipher. The grey area shows the signal, i.e. the known trails. The remaining trails are unknown.

as discussed in Sect. 3, which removes the need to consider the key-dependent behaviour of the approximations. It is then sufficient to determine C_U for the dominant trails, and then determine how many of these trails exist. This can be done separately for each approximation.

For the multivariate approach, however, we need to analyse the simultaneous behaviour of all M approximations to estimate the distribution of $\mathbf{C}^K$. This is most easily done for key-alternating ciphers as we can use Equation 2. However, since we need to know how the sign of each trail changes with the key, we need to explicitly know the masks of each trail. If the number of relatively good trails is large, efficient trail enumeration becomes quite important for the multivariate linear model. Depending on the structure of the cipher, this can be quite an obstacle for the analysis. Nevertheless, for PRESENT and similar ciphers, we have been able to develop quite efficient tools for trail enumeration.

We will denote the set $\mathcal{S}$ of enumerated trails of an approximation by the *signal*. We will indicate quantities calculated from the signal by a $\star$. To estimate $\mathbf{C}^K$, we will have to calculate the signal correlations

$$C^{K\star}_{\alpha_i,\beta_i} = \sum_{U\in\mathcal{S}} (-1)^{U\cdot\bar{K}\oplus d_U} C_U,$$

for the M approximations and for a suitably large number of random keys. We can then estimate the joint distribution of the signal correlation $\mathbf{C}^{K\star} \sim \mathcal{N}_M(\boldsymbol{\mu}^\star, \boldsymbol{\Sigma}^\star)$ and use this as an approximation of the true distribution in Corollary 3. Nevertheless, even if we can enumerate a large number of trails, the total number of trails might be extremely large, as is the case for PRESENT. In this case, a large part of the hull is still unknown, as illustrated in Fig. 2. In the following, we discuss how we can deal with this unknown part of the hull.

6.1 Signal/Noise Decomposition

While the signal trails in $\mathcal{S}$ will help to approximate the correlation distributions, there might still be a significant number of trails that we are unable to enumerate. In the following, we discuss how to model the contribution of this unknown part of the hull. We will do so under the following assumption.

Assumption 1. *For an n-bit cipher with a reasonably large key space, if an approximation does not have correlation equal to zero, the variance the correlation is at least 2^{-n}.*

The reasoning behind Assumption 1 is that an approximation of an n-bit permutation has correlation variance 2^{-n} over the space of all n-bit permutations, cf. [21]. A cipher represents a subset of all n-bit permutations the size of its key space. Since we cannot do better than an ideal cipher, the correlation variance must be lower bounded by the ideal case. Nyberg also makes an argument for this lower bound in [38].

Bogdanov and Tischhauser presented one way to model the unknown part of the hull in [16]. As in that work, we will call the unknown part of the hull the *noise*. Then, the correlation distribution of an approximation can be decomposed as $C^K_{\alpha,\beta} = C^{K\star}_{\alpha,\beta} + \text{noise}$. If the contribution from the signal part of the correlation is significant, then the contribution of the noise is probably small, and we therefore model this part of the hull as an ideal cipher. In that case, the full distribution of the linear correlation is $C^K_{\alpha,\beta} = C^{K\star}_{\alpha,\beta} + \mathcal{N}(0, 2^{-n})$. The following assumption formalises this – a similar statement is given in [16].

Assumption 2. *If the variance of the signal part of the linear correlation is larger than 2^{-n}, then the noise part has distribution $\mathcal{N}(0, 2^{-n})$.*

Signal/noise decomposition is reasonable when the contribution from the signal is large, as we do not risk significantly overestimating the strength of the approximation. But what if the variance of our signal is smaller than 2^{-n}? In that case, we make the following, stronger assumption.

Assumption 3 ([16]). *If the number of trails in the unknown part of the hull is large, and their correlation contributions are small, then the correlation of the noise part of the hull is distributed as $\mathcal{N}(0, 2^{-n})$.*

Experimental Verification. To justify this assumption, we conducted experiments on 9-round 32-bit SmallPresent [32]. We picked a set of 255 approximations, namely all non-zero approximations starting at S_5 and ending at S_1. This sets consists of generally weak approximations. We enumerated trails of all approximations having intermediate masks with at most hamming weight 3 – a situation that closely resembles what is possible to enumerate for full present. We calculated the signal correlations for 5000 random keys, and measured the actual correlation for 1000 random keys. The result of adding noise to the measured signal distributions are shown in Fig. 3.

We make the following observations: For these very weak approximations, the signal part of the hulls give a large underestimate of the actual correlation variance – in particular the signal variance is always lower than 2^{-32}, indicating that these approximations would not be useful for distinguishing. However, the actual variances are all larger than 2^{-32}, and so when adding noise, we get a good estimate of the actual value. For only five approximations we get an overestimate of the variance, but the error is not larger than 2^{-36}. This error can be explained

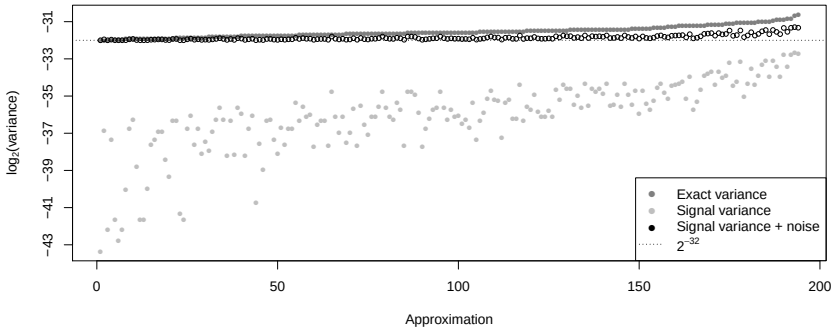


Fig. 3: Variance of correlation values of 9 round approximations of 32-bit SmallPresent starting at S_5 and ending at S_1. Signal and exact variance values are shown along with variance estimates based on signal/noise decomposition. Trails having intermediate masks with Hamming weight up to 3 were enumerated for the signal values. The graph demonstrates that even for approximations with low correlation, and with a signal variance lower than 2^{-n}, it is still reasonable to add the noise distribution $\mathcal{N}(0, 2^{-n})$. In this way, we can compensate for the unknown part of the linear hull, cf. Assumption 3.

by measuring error of the actual correlation values. Thus, we would argue that adding noise to the measured signal distributions, under Assumption 3, give us a much more accurate estimate of the correlation distributions than simply considering the signal. Indeed, after adding noise, we might use an approximation in a linear attack which we would otherwise have discarded based on its signal distribution. This is particularly important in typical attack scenarios where we try to attack as many rounds as possible, and therefore use approximations with correlation values close to that of an ideal cipher.

7 Hypothesis Testing and QDA for the Multivariate Model

Different methods for distinguishing when using many approximations have been proposed. The LLR method was proposed by Baigneres et al. in [3] as an optimal disinguisher and used in [25] in a multidimensional attack against the block cipher Serpent. Both the LLR method and the χ^2 method was studied in [26], where the LLR method was concluded to have better performance. However, as noted by Cho in [17], 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.

In the following, we consider two ways to create a statistical distinguisher using the multivariate model presented in Sect. 5. The first method is equivalent to the χ^2 method. The other method takes advantage of the multivariate distribution of the correlations and uses quadratic discriminant analysis. We first note that for an ideal cipher, $C^K_{\alpha,\beta} \sim \mathcal{N}(0, 2^{-n})$ [29, 39], and the correlations for two different

approximations are independent both over the text-space (by the result in [36]) and the key space (since each bit of the input and output can be considered as independent random variables). Thus, for an ideal cipher,

$$\hat{\mathbf{C}}^K \sim \mathcal{N}_M(\mathbf{0}, (2^{-n} + N^{-1})\mathbf{I}). \tag{7}$$

7.1 Distinguishing with Binary Hypothesis Testing

Consider the binary hypothesis testing framework as given in Sect. 2.2 with the test statistic

$$\mathcal{T} = N \sum_{i=1}^{M} (\hat{C}^K_{\alpha_i,\beta_i})^2. \tag{8}$$

As mentioned, this test statistic has been used frequently in the past, and is usually referred to as the χ^2 test statistic. In the setting of multivariate linear cryptanalysis, this is equal to the scaled quadratic form $\mathcal{T} = N(\hat{\mathbf{C}}^K)^\top \mathbf{I} \hat{\mathbf{C}}^K$. Then, by Corollary 3, the mean and variance of the test statistic distribution for an ideal cipher $\mathcal{D}_I$ are given by

$$\mu_I = M + NM2^{-n} \quad \text{and} \quad \sigma_I^2 = 2MN^2(2^{-n} + N^{-1})^2,$$

while the mean and variance for the non-ideal distribution $\mathcal{D}_N$ are given by

$$\mu_N = M + N \sum_{i=1}^{M} ELP_i, \tag{9}$$

$$\sigma_N^2 = 2N^2 \left(\mathrm{tr}\left((\mathbf{\Sigma} + \mathbf{\Sigma}^N)^2 \right) + 2\boldsymbol{\mu}^\top (\mathbf{\Sigma} + \mathbf{\Sigma}^N)\boldsymbol{\mu} \right). \tag{10}$$

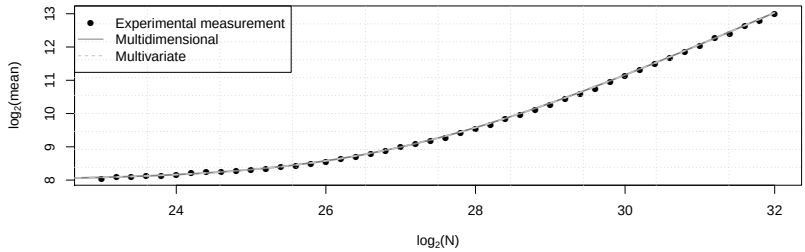
We note that the parameters for the ideal case was also given in [9]. While Corollary 3 says nothing about the shape of $\mathcal{D}_I$ and $\mathcal{D}_N$, we will make the simplifying assumption that they are approximately normal. In that case, the threshold value for a given success probability P_S is given by

$$\tau = \mu_N + \sigma_N \Phi^{-1}(1 - P_S),$$

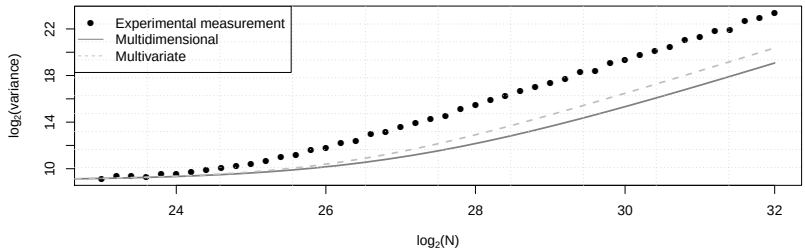
and the resulting probability of false positives for the distinguisher is

$$P_I = 1 - \Phi\left(\frac{\mu_N - \mu_I + \sigma_N \Phi^{-1}(1 - P_S)}{\sigma_I} \right). \tag{11}$$

Experimental Verification Estimates of μ_N and σ_N^2 were also given in [9] in the multidimensional model with $\lambda_i = 1/2^m$ and $\zeta_i^2 = \zeta^2$ for some variance ζ^2. We have conducted experiments on 9-round 32-bit SmallPresent to asses the accuracy of the estimates from [9] and the ones given above. We consider the 255 approximations starting at S_6 and ending at S_6. We measured the exact



(a) $\log_2$ of the mean of the test statistic $\mathcal{T}$, and theoretical estimates.



(b) $\log_2$ of the variance of the test statistic $\mathcal{T}$, and theoretical estimates.

Fig. 4: Exact mean and variance of the hypothesis testing test statistic $\mathcal{T} = N\sum_{i=1}^{M}(\hat{C}^K_{\alpha_i,\beta_i})^2$ for 9 rounds of 32-bit SmallPresent. Approximations starting and ending at S_6 were used. The multidimensional estimate of $\mathrm{E}(\mathcal{T}) = M + N\mathrm{E}(\mathcal{C}^K)$ and $\mathrm{Var}(\mathcal{T}) = \frac{2}{M}(M + N\mathrm{E}(\mathcal{C}^K))^2$ were given in [9]. The multivariate estimate is given in Equations 9 and 10. Using the multivariate model results in an improved estimate of the test statistic variance.

correlation values for 1000 random keys to obtain the distribution of $\mathbf{C}$. We then measured the value of $\mathcal{T}$ for 1000 random keys. The results are shown in Fig. 4.

Fig. 4a shows the test statistic mean μ_N. The estimate given in Equation 9 is identical to the one found in [9], and gives a very good estimate of the actual mean. Fig. 4b shows the test statistic variance σ^2_N. Both the estimate given in Equation 10 and in [9] underestimate the actual variance significantly. However, our estimate using the multivariate model does improve upon the result from [9]. Nevertheless, underestimating the variance is a problem: Since the distribution of $\mathcal{T}$ is fixed for the ideal cipher and a given N, and our estimate of μ_N is good, underestimating σ^2_N means that we will overestimate the power of the distinguisher. In the following, we explore a distinguishing method where both the ideal and non-ideal distribution of $\mathcal{T}$ depends on $\mathbf{C}$.

7.2 Distinguishing with Quadratic Discriminant Analysis.

In order to fully take advantage of the multivariate representation of the linear correlations, we now consider the quadratic discriminant analysis method given in Sect. 2.2. We will use $\hat{\mathbf{C}}^K$ as the cipher data, and by Corollary 3 and Equation 7, we can express Equation 3 as

$$\mathcal{T} = (\hat{\mathbf{C}}^K - \boldsymbol{\mu})^\top (\boldsymbol{\Sigma} + \boldsymbol{\Sigma}^N)^{-1} (\hat{\mathbf{C}}^K - \boldsymbol{\mu}) - \tfrac{N}{2^{-n}N+1} (\hat{\mathbf{C}}^K)^\top \hat{\mathbf{C}}^K. \tag{12}$$

As mentioned in Sect. 2.2, finding the exact distribution of $\mathcal{T}$ is in this case not trivial. However, since sampling from a multivariate normal distribution is easy, we can easily simulate $\mathcal{D}_I$ and $\mathcal{D}_N$, and then calculate the empirical CDFs to get an estimate of τ for the desired success probability and the resulting probability of false positives.

Comparison to Hypothesis Testing. To compare the QDA distinguisher to the hypothesis testing distinguisher, we calculated the advantage of 189 approximations over 22 rounds of PRESENT with $P_S = 0.95$. The approximations are specified in Sect. 8. We enumerated part of the hull of each approximation. All trails with hamming weight 1 masks were enumerated, as well as some trails with hamming weight 2 and 3 masks. An average of 103,482,624 trails were enumerated per hull. The distribution of $\mathbf{C}^{K\star}$ was then estimated using 2000 random master keys. Noise was added to all approximations to compensate for the unknown part of the hull, as described in Sect. 6. The addition of noise is reasonable in this case under Assumption 3, since the number of PRESENT trails is known to be extremely large, and trails with hamming weight larger than 1 have much smaller correlation contribution, as argued in e.g. [17, 40].

The advantage for the hypothesis testing distinguisher was calculated using Equation 11. The advantage for the QDA distinguisher was calculated using 500,000 samples from each normal distribution. We note that due to the structure of the approximations, we can assume that $\boldsymbol{\Sigma}^N$ is a diagonal matrix, as discussed in Sect. 5. This simplifies our analysis considerably, and the results are shown in Fig. 5. For $N < 2^{63.4}$, the advantage of the QDA distinguisher is about 1 bit larger than that of the hypothesis testing distinguisher. However, for $N \geq 2^{63.4}$, the QDA distinguisher achieves the full advantage, whereas the hypothesis testing distinguisher never obtains an advantage larger than 35 bits. We can explain this by the fact that both $\mathcal{D}_I$ and $\mathcal{D}_N$ depends on the distribution of $\hat{\mathbf{C}}^K$, and as this becomes more significant with large N, the two distributions rapidly separate, increasing the advantage.

8 Multivariate Linear Attacks on PRESENT

Under the wrong key randomisation hypothesis, Hypothesis 1 in Sect. 2.2, we can turn our multivariate linear distinguisher into a key-recovery attack. The general idea for the attack is as follows: Collect N text pairs. Guess the k round key

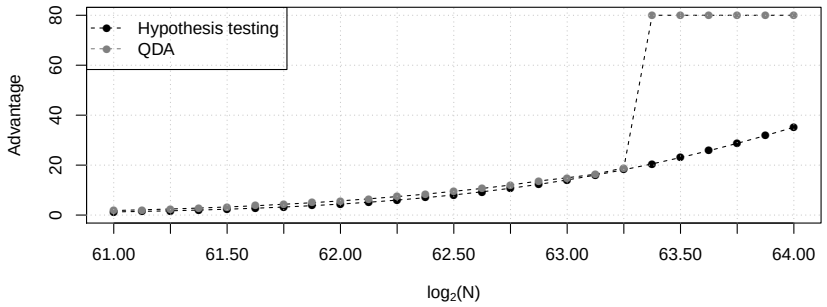


Fig. 5: Advantage for hypothesis testing and QDA distinguishing using 189 approximations of 22-round PRESENT, $P_S = 0.95$. The QDA distinguisher achieves the full advantage when $N \geq 2^{63.4}$. The sudden increase in advantage for the QDA method happens because the distinguishing distributions $\mathcal{D}_I$ and $\mathcal{D}_N$ move away from each other more rapidly than for the hypothesis testing method.

bits required to (partially) encrypt/decrypt ℓ rounds. Apply the distinguisher to the resulting correlations: 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. In the following, we present a new attack on 26 and 27 rounds of PRESENT in the multivariate model.

8.1 Attacking 26 rounds

We aim to recover the master key for r rounds of PRESENT by using a multivariate linear 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. Additionally, due to the effectiveness of the QDA distinguisher, we can afford to use relatively few approximations.

Approximations and Key Guessing. As Ohkuma noted in [40], 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

$$\begin{aligned}
&\alpha = 2^{4i+3}, i \in \{5, 6, 7, 9, 10, 11, 13, 14, 15\}, \\
&\beta = j \cdot 2^{4i+2}, i \in \{5, 6, 7, 9, 10, 11\}, j \in \{1, 2, 3\}, \text{ and} \\
&\beta = 2^{4i+3}, i \in \{13, 14, 15\}.
\end{aligned}$$

These bit positions are highlighted in Fig. 6. Taking all possible combinations of these input and output masks gives us $M = 9 \cdot 6 \cdot 3 + 9 \cdot 3 = 189$ approximations. Fig. 6 shows the S-box positions we need to encrypt/decrypt to calculate the linear correlations of these approximations. The straight forward approach to

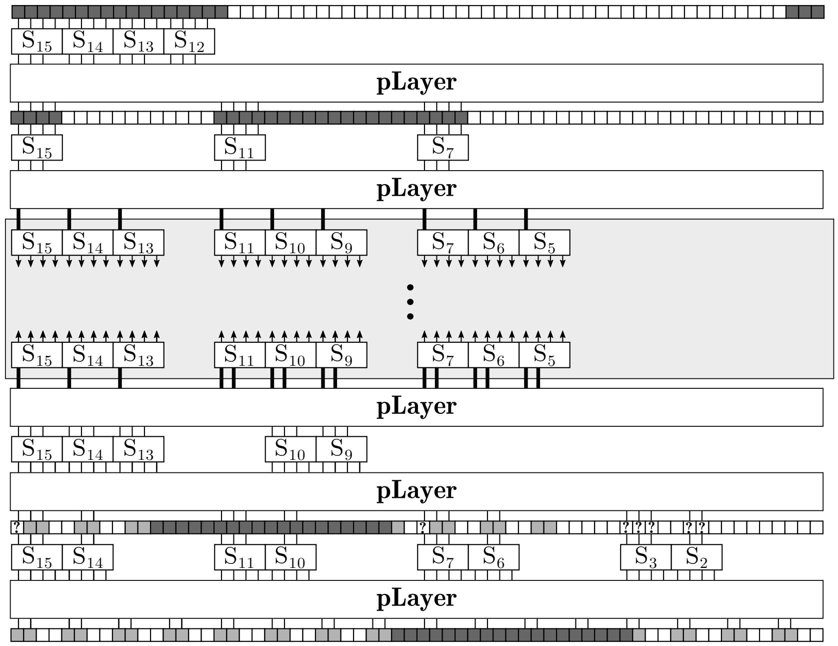


Fig. 6: An outline of the 26-round attack using approximations over 22 rounds. The input/output mask bits are indicated by bold lines. The dark grey squares indicate the round-key bits obtained by guessing 24 bits of the master key. The light grey squares indicate the round-key bits obtained by guessing 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.

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:

$$k_i, i \in [0, 2] \cup [15, 18] \cup [63, 79]. \qquad (13)$$

The 42 round-key bits we obtain from this guess are marked in dark grey in Fig. 6. By guessing the missing 23 bits of K_{26}, we also obtain 8 bits of K_{25}. Finally, we only need to guess an additional 7 bits of K_{25}. In total, we only need to guess 54 bits of key material. Note additionally that each approximation only depends on at most 8 bits of K_{25}.

Attack Description. With the above 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. Fix a 24-bit guess of the master key, K_M.
   (a) For each input mask α, calculate a vector $\mathbf{t}_\alpha^{K_M}$ of size 2^{32}, where
   $$\mathbf{t}_\alpha^{K_M}[j] = \#\{(p_i, c_i) | c_i = j \text{ and } \alpha \cdot \mathcal{E}_{K_M}(p_i) = 0\},$$
   where $\mathcal{E}_{K_M}(p)$ is the partial two-round encryption of p under master key K_M.
   (b) For each output mask β, fix a guess of the relevant bits of K_{25}. Denote the guess K_I. Then calculate the $2^{32} \times 2^{32}$ matrix $\mathbf{A}_\beta^{K_I}$, where
   $$\mathbf{A}_\beta^{K_I}[i, j] = \beta \cdot \mathcal{D}_{K_I}(i \oplus j),$$
   where $\mathcal{D}_{K_I}(c)$ is the partial two-round decryption of c using K_I, but *excluding* the first key XOR.
   (c) Calculate the correlations of all 189 approximations and 2^{32} guesses of the partial K_{26} by calculating the matrix-vector products
   $$\mathbf{C}_{\alpha,\beta} = \tfrac{2}{N}\mathbf{A}_\beta^{K_I}\mathbf{t}_\alpha^{K_M} - 1.$$
   (d) Repeat steps (b) and (c) for all values of K_I, resulting in correlation values for all approximations for at most 2^{40} guesses of the last two round keys. Extract the correlations of at most 2^{30} guesses that agree with K_M.
   (e) Calculate the QDA test statistic $\mathcal{T}$ for each surviving key guess. Save all keys (of 54 bits) with $\mathcal{T} < \tau$.
2. Repeat the above steps for all values of K_M.

Search phase

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

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 $\mathcal{O}(N)$ operations. For the analysis phase, step 1a can be done by iterating over $\mathbf{t}$ once and encrypting two rounds, using $\mathcal{O}(2 \cdot 2^{48})$ operations. Steps 1b and 1c can be performed using the FFT technique given in [18]. Using this technique, we only need to compute the first column of each $\mathbf{A}_\beta^{K_I}$, and then calculate $\mathbf{C}_{\alpha,\beta}$ for a fixed β and all α in time $\mathcal{O}((2 \cdot 9 + 1) \cdot 32 \cdot 2^{32})$. Thus, steps 1b and 1c need a total of $\mathcal{O}(21 \cdot 2^8 \cdot (2 \cdot 2^{32} + (2 \cdot 9 + 1) \cdot 32 \cdot 2^{32})) \approx \mathcal{O}(2^{53.64})$ operations. Step 1d uses $\mathcal{O}(2^{30})$ operations. In step 1e we can calculate the test statistics of 189 keys simultaneously by using two matrix-matrix products. Using Strassen's algorithm for this, we can process 2^{30} keys using $\mathcal{O}(\frac{2^{30}}{189} \cdot 2 \cdot 189^{2.807}) \approx \mathcal{O}(2^{44.67})$ operations. In total, the analysis phase uses $\mathcal{O}(2^{24} \cdot (2^{49} + 2^{53.64} + 2^{30} + 2^{44.67})) \approx \mathcal{O}(2^{77.70})$ operations. If one uses hypothesis testing instead, step 1e will be slightly cheaper, but the final

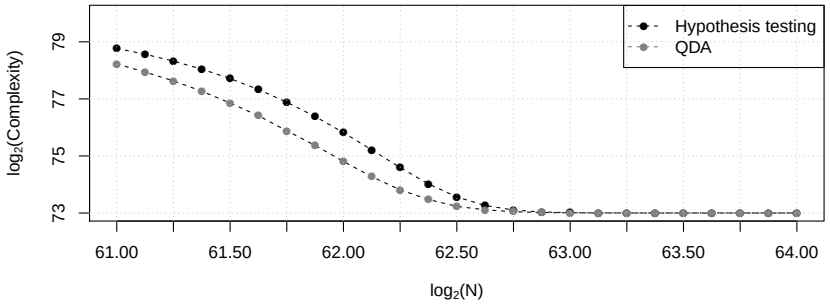


Fig. 7: *Our 26 round attack:* Computational complexity as a function of data complexity for the 26-round attack on PRESENT using 189 approximations over 22 rounds. Non-distinct random texts were used, and $P_S = 0.95$. Note that the QDA distinguisher achieves a better complexity as long as long as the search phase dominates the complexity.

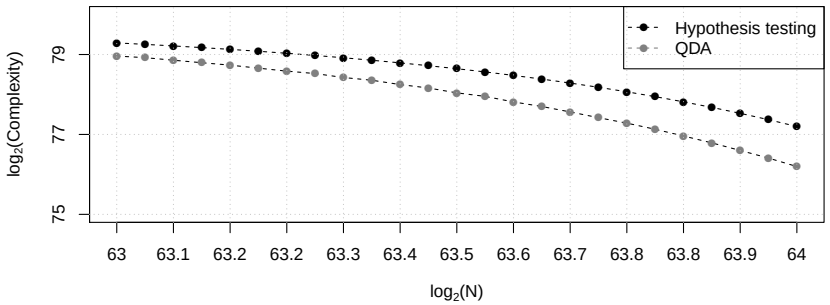


Fig. 8: *Our 27 round attack:* Computational complexity as a function of data complexity for the 27-round attack on PRESENT using 189 approximations over 23 rounds. Distinct random texts were used, and $P_S = 0.95$. Note that the QDA distinguisher generally achieves a better complexity for this attack.

complexity does not change significantly. Finally, the search phase requires $2^{\kappa-54}$ full encryptions of 2^{54-a} candidates keys, using a total of $\mathcal{O}(2^{\kappa-a})$ operations.

The advantage of the attack was determined in Sect. 7. From Fig. 5, we obtain a plot of the computational complexity of the 26-round attack, given in Fig. 7. Here, we have fixed the success probability at 95%. As long as the search phase dominates the complexity, we observe that the QDA distinguisher has an advantage over hypothesis testing. We can highlight two 26 round attacks with different trade-offs. For $N = 2^{63.25}$, we obtain an advantage of 19 bits, and a computational complexity of $\mathcal{O}(2^{77.70}/26) = \mathcal{O}(2^{73.00})$ encryptions. Alternatively, we can decrease the data complexity to $N = 2^{61.75}$, giving and advantage of 4.33 bits, and a computational complexity of $\mathcal{O}(2^{80.58}/26) = \mathcal{O}(2^{75.88})$ encryptions.

8.2 Attacking 27 rounds

The attack described above 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 13, we determine 41 required bits of the round keys. We additionally 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 [9,10]. The resulting attack complexities are shown in Fig. 8. Using the QDA distinguisher with $P_S = 0.95$ and $N = 2^{63.83}$, we obtain an advantage of 3.05 bits and a computational complexity of $\mathcal{O}(2^{76.96})$ encryptions.

9 Conclusions

In this paper, we proposed *multivariate linear cryptanalysis* as a new technique for using multiple linear approximations. It is based on a multivariate statistical model and allows for realistic key equivalence and wrong key randomization hypotheses without introducing additional limiting assumptions about the distributions of the correlations of the used linear approximations. Additionally, it features a signal/noise decomposition approach for a realistic estimation of correlations in the common scenario where only a limited number of trails per hull is known, and we use QDA as a novel distinguishing technique. As an application of this new technique, we have shown attacks on 26 and 27 rounds of PRESENT under significantly more realistic assumptions than previous work.

It remains an interesting open problem to apply multivariate linear cryptanalysis to other ciphers, with Serpent and LBlock being particular natural candidates. Additionally, it would be interesting to investigate the connection between the capacity and the power of the QDA distinguisher.

References

1. ISO/IEC: Information Technology Security Techniques Lightweight cryptography Part 2: Block ciphers., ISO/IEC 29192-2:2012 edn. (2012)
2. Abdelraheem, M.A., Ågren, M., Beelen, P., Leander, G.: On the distribution of linear biases: Three instructive examples. In: Safavi-Naini, R., Canetti, R. (eds.) Advances in Cryptology - CRYPTO 2012. Lecture Notes in Computer Science, vol. 7417, pp. 50–67. Springer (2012)
3. Baigneres, T., Junod, P., Vaudenay, S.: How Far Can We Go Beyond Linear Cryptanalysis? In: Advances in Cryptology-AsiaCrypt 2004, pp. 432–450. Springer (2004)
4. Biryukov, A., De Canniere, C., Quisquater, M.: On Multiple Linear Approximations. In: Advances in Cryptology–CRYPTO 2004. pp. 1–22. Springer (2004)

5. Blondeau, C., Bay, A., Vaudenay, S.: Protecting against multidimensional linear and truncated differential cryptanalysis by decorrelation. In: Leander, G. (ed.) Fast Software Encryption. Lecture Notes in Computer Science, vol. 9054, pp. 73–91. Springer (2015)
6. Blondeau, C., Leander, G., Nyberg, K.: Differential-linear cryptanalysis revisited. In: Cid, C., Rechberger, C. (eds.) Fast Software Encryption. Lecture Notes in Computer Science, vol. 8540, pp. 411–430. Springer (2014)
7. Blondeau, C., Nyberg, K.: New links between differential and linear cryptanalysis. In: Johansson, T., Nguyen, P.Q. (eds.) Advances in Cryptology - EUROCRYPT 2013. Lecture Notes in Computer Science, vol. 7881, pp. 388–404. Springer (2013)
8. Blondeau, C., Nyberg, K.: Links between truncated differential and multidimensional linear properties of block ciphers and underlying attack complexities. In: Nguyen, P.Q., Oswald, E. (eds.) Advances in Cryptology - EUROCRYPT 2014. Lecture Notes in Computer Science, vol. 8441, pp. 165–182. Springer (2014)
9. Blondeau, C., Nyberg, K.: Joint Data and Key Distribution of the Linear Cryptanalysis Test Statistic and Its Impact to Data Complexity Estimates of Multiple/Multidimensional Linear and Truncated Differential Attacks (2015)
10. Blondeau, C., Nyberg, K.: On Distinct Known Plaintext Attacks. In: The Ninth International Workshop on Coding and Cryptography (2015)
11. Blondeau, C., Peyrin, T., Wang, L.: Known-key distinguisher on full PRESENT. In: Gennaro, R., Robshaw, M. (eds.) Advances in Cryptology - CRYPTO 2015. Lecture Notes in Computer Science, vol. 9215, pp. 455–474. Springer (2015)
12. Bogdanov, A., Boura, C., Rijmen, V., Wang, M., Wen, L., Zhao, J.: Key difference invariant bias in block ciphers. In: Sako, K., Sarkar, P. (eds.) Advances in Cryptology - ASIACRYPT 2013. Lecture Notes in Computer Science, vol. 8269, pp. 357–376. Springer (2013)
13. Bogdanov, A., Geng, H., Wang, M., Wen, L., Collard, B.: Zero-correlation linear cryptanalysis with FFT and improved attacks on ISO standards camellia and CLEFIA. In: Lange, T., Lauter, K.E., Lisonek, P. (eds.) Selected Areas in Cryptography. Lecture Notes in Computer Science, vol. 8282, pp. 306–323. Springer (2013)
14. Bogdanov, A., Knudsen, L.R., Leander, G., Paar, C., Poschmann, A., Robshaw, M.J., Seurin, Y., Vikkelsoe, C.: PRESENT: An Ultra-Lightweight Block Cipher. Springer (2007)
15. Bogdanov, A., Rijmen, V.: Linear hulls with correlation zero and linear cryptanalysis of block ciphers. Des. Codes Cryptography 70(3), 369–383 (2014)
16. Bogdanov, A., Tischhauser, E.: On the Wrong Key Randomisation and Key Equivalence Hypotheses in Matsuis Algorithm 2. In: Fast Software EnCryption. pp. 19–38. Springer (2013)
17. Cho, J.Y.: Linear Cryptanalysis of Reduced-Round PRESENT. In: Topics in Cryptology-CT-RSA 2010, pp. 302–317. Springer (2010)
18. Collard, B., Standaert, F.X., Quisquater, J.J.: Improving the Time Complexity of Matsuis Linear Cryptanalysis. In: Information Security and Cryptology-ICISC 2007, pp. 77–88. Springer (2007)
19. Collard, B., Standaert, F.: A statistical saturation attack against the block cipher PRESENT. In: Fischlin, M. (ed.) Topics in Cryptology - CT-RSA 2009,. Lecture Notes in Computer Science, vol. 5473, pp. 195–210. Springer (2009)
20. Daemen, J.: Cipher and Hash Function Design Strategies Based on Linear and Differential Cryptanalysis. Ph.D. thesis, Doctoral Dissertation, March 1995, KU Leuven (1995)
21. Daemen, J., Rijmen, V.: Probability Distributions of Correlation and Differentials in Block Ciphers. Journal of Mathematical Cryptology JMC 1(3), 221–242 (2007)

22. Daemen, J., Rijmen, V.: The Design of Rijndael: AES – The Advanced Encryption Standard. Springer Science & Business Media (2013)
23. Desmedt, Y. (ed.): Advances in Cryptology - CRYPTO '94, Lecture Notes in Computer Science, vol. 839. Springer (1994)
24. Harpes, C., Kramer, G.G., Massey, J.L.: A Generalization of Linear Cryptanalysis and the Applicability of Matsuis Piling-Up Lemma. In: Advances in CryptologyEUROCRYPT95. pp. 24–38. Springer (1995)
25. Hermelin, M., Cho, J.Y., Nyberg, K.: Multidimensional Linear Cryptanalysis of Reduced Round Serpent. In: Information Security and Privacy. pp. 203–215. Springer (2008)
26. Hermelin, M., Cho, J.Y., Nyberg, K.: Multidimensional Extension of Matsuis Algorithm 2. In: Fast Software EnCryption. pp. 209–227. Springer (2009)
27. Hermelin, M., Nyberg, K.: Multidimensional Linear Distinguishing Attacks and Boolean Functions. Cryptography and Communications 4(1), 47–64 (2012)
28. Huang, J., Vaudenay, S., Lai, X., Nyberg, K.: Capacity and Data Complexity in Multidimensional Linear Attack. In: Advances in Cryptology–CRYPTO 2015, pp. 141–160. Springer (2015)
29. Junod, P.: On the Complexity of Matsuis Attack. In: Selected Areas in Cryptography. pp. 199–211. Springer (2001)
30. Kaliski Jr, B.S., Robshaw, M.J.: Linear Cryptanalysis Using Multiple Approximations. In: Advances in CryptologyCrypto94. pp. 26–39. Springer (1994)
31. Langford, S.K., Hellman, M.E.: Differential-linear cryptanalysis. In: Desmedt [23], pp. 17–25
32. Leander, G.: Small Scale Variants of the Block Cipher PRESENT. IACR Cryptology ePrint Archive 2010, 143 (2010)
33. Leander, G.: On linear hulls, statistical saturation attacks, PRESENT and a cryptanalysis of PUFFIN. In: Paterson, K.G. (ed.) Advances in Cryptology - EUROCRYPT 2011. Lecture Notes in Computer Science, vol. 6632, pp. 303–322. Springer (2011)
34. Matsui, M.: Linear Cryptanalysis Method for DES Cipher. In: Advances in CryptologyEUROCRYPT93. pp. 386–397. Springer (1993)
35. Matsui, M.: The first experimental cryptanalysis of the Data Encryption Standard. In: Desmedt [23], pp. 1–11
36. Murphy, S.: The Independence of Linear Approximations in Symmetric Cryptanalysis. IEEE Transactions on Information Theory 52(12), 5510–5518 (2006)
37. Nguyen, P.H., Wu, H., Wang, H.: Improving the algorithm 2 in multidimensional linear cryptanalysis. In: Parampalli, U., Hawkes, P. (eds.) Information Security and Privacy - 16th Australasian Conference, ACISP. Lecture Notes in Computer Science, vol. 6812, pp. 61–74. Springer (2011)
38. Nyberg, K.: Linear Cryptanalysis. `http://mta.ca/sac2015/S3-linear-all.pdf` (August 2015), lecture given at the SAC Summer School 2015
39. O'Connor, L.: Properties of Linear Approximation Tables. In: Fast Software EnCryption. pp. 131–136. Springer (1994)
40. Ohkuma, K.: Weak Keys of Reduced-Round PRESENT for Linear Cryptanalysis. In: Selected Areas in Cryptography. pp. 249–265. Springer (2009)
41. Röck, A., Nyberg, K.: Generalization of Matsuis Algorithm 1 to Linear Hull for Key-Alternating Block Ciphers. Designs, codes and Cryptography 66(1-3), 175–193 (2013)
42. Samajder, S., Sarkar, P.: Another Look at Normal Approximations in Cryptanalysis. Journal of Mathematical Cryptology 10(2), 69–99 (2016)

43. Seber, G.A., Lee, A.J.: Linear Regression Analysis, vol. 936. John Wiley & Sons (2012)
44. Selçuk, A.A.: On Probability of Success in Linear and Differential Cryptanalysis. Journal of Cryptology 21(1), 131–147 (2008)
45. Selçuk, A.A., Biçak, A.: On Probability of Success in Linear and Differential Cryptanalysis. In: Security in Communication Networks, pp. 174–185. Springer (2002)
46. Sun, B., Liu, M., Guo, J., Rijmen, V., Li, R.: Provable security evaluation of structures against impossible differential and zero correlation linear cryptanalysis. In: Fischlin, M., Coron, J. (eds.) Advances in Cryptology - EUROCRYPT 2016. Lecture Notes in Computer Science, vol. 9665, pp. 196–213. Springer (2016)
47. Xu, H., Jia, P., Huang, G., Lai, X.: Multidimensional zero-correlation linear cryptanalysis on 23-round lblock-s. In: Qing, S., Okamoto, E., Kim, K., Liu, D. (eds.) Information and Communications Security. Lecture Notes in Computer Science, vol. 9543, pp. 97–108. Springer (2015)
48. Zhao, J., Wang, M., Wen, L.: Improved linear cryptanalysis of CAST-256. J. Comput. Sci. Technol. 29(6), 1134–1139 (2014)
49. Zheng, L., Zhang, S.w.: FFT-Based Multidimensional Linear Attack on PRESENT Using the 2-Bit-Fixed Characteristic. Security and Communication Networks 8(18), 3535–3545 (2015)