J Cryptol (2017) 30:859–888
DOI: 10.1007/s00145-016-9237-5

Journal of
CRYPTOLOGY

Differential-Linear Cryptanalysis Revisited *

Céline Blondeau
Department of Computer Science, Aalto University School of Science, Espoo, Finland
celine.blondeau@aalto.fi

Gregor Leander
Faculty of Electrical Engineering and Information Technology, Ruhr Universität Bochum,
Bochum, Germany
gregor.leander@rub.de

Kaisa Nyberg
Department of Computer Science, Aalto University School of Science, Espoo, Finland
kaisa.nyberg@aalto.fi

Communicated by Vincent Rijmen.

Received 27 January 2015 / Revised 15 June 2016
Online publication 7 October 2016

Abstract. The two main classes of statistical cryptanalysis are the linear and differential attacks. They have many variants and enhancements such as the multidimensional linear attacks and the truncated differential attacks. The idea of differential-linear cryptanalysis is to apply first a truncated differential attack and then a linear attack on different parts of the cipher and then combine them to a single distinguisher over the cipher. This method is known since 1994 when Langford and Hellman presented the first differential-linear cryptanalysis of the DES. Recently, in 2014, Blondeau and Nyberg presented a general link between differential and linear attacks. In this paper, we apply this link to develop a concise theory of the differential-linear cryptanalysis. The differential-linear attack can be, in the theoretical sense, considered either as a multidimensional linear or a truncated differential attack, but is for both types an extreme case, which is best measured by the differential-linear bias. We give an exact expression of the bias in a closed form under the sole assumption that the two parts of the cipher are independent. Unlike in the case of ordinary differentials and linear approximations, it is not granted that restricting to a subset of characteristics of a differential-linear hull will give a lower bound on the absolute value of the bias. Given this, we revisit the previous treatments of differential-linear bias by Biham et al. in 2002–2003, Liu et al. in 2009, and Lu in 2012, and formulate assumptions under which a single differential-linear characteristic gives a close estimate of the bias. These results are then generalized by considering a subspace of linear approximations over the second part of the cipher. To verify the assumptions made, we present several experiments on a toy-cipher.

* The work of Gregor Leander was funded by the BMBF UNIKOPS Project

Keywords. Block cipher, Differential cryptanalysis, Linear cryptanalysis, Truncated differential, Multidimensional linear approximation, Bias of differential-linear approximation.

1. Introduction

The use of cryptography primitives in daily life plays an increasingly crucial role. Among the different primitives, block ciphers are arguably the most widely used ones.

Great progress has been made in designing and analyzing block ciphers, especially with the introduction of the AES, but also more recently with many block ciphers appearing in the area of lightweight cryptography. However, there is still research on fundamental aspects of these ciphers going on and important questions are still not understood. For instance, we are not able to assess the security of a block cipher as such, but only its security with respect to known attacks. The problem of precisely assessing the security of a given primitive becomes more important for lightweight primitives. Those ciphers often aim at achieving tighter trade-off between security and efficiency which in turn sets higher requirements for the accuracy of cryptanalysis.

The focus of this paper is on statistical cryptanalysis of block ciphers. The two main classes to be considered here are linear and differential attacks and their variants.

1.1. *Differential Cryptanalysis*

The first type of statistical attacks that is applicable to a large set of block ciphers is the differential attack introduced by Biham and Shamir in [8]. Since its invention in the early nineties several variants, tweaks and generalizations have been proposed. In 1994, Knudsen introduced so-called *truncated differentials attacks* [30]. This relaxation of classical differential attacks has since then been applied to many block ciphers. In the same paper, Knudsen furthermore introduced the concept of higher-order differentials, an attack vector based on initial consideration by Lai in [32]. Another variant of differential cryptanalysis (again by Knudsen) is the *impossible differential* cryptanalysis which uses differentials with probability zero. This concept, introduced in [31], has later been successfully applied numerously, e.g., to (almost) break the cipher Skipjack [3][1]. In 1999, Wagner introduced the boomerang attack, which allows to concatenate any two, not necessarily coinciding, differentials over parts of a cipher. This attack allowed, among others, breaking the cipher COCONUT98 [48]. Later, the boomerang attack itself has been generalized to amplified boomerang attack [29] and the rectangle attack [4].

1.2. *Linear Cryptanalysis*

The second generally applicable attack on block ciphers is Matsui's linear attack [41]. Similarly to differential attacks, since its introduction many extensions and improvements have been made, and we mention a selection here. More precise estimates for the success probability and the data complexity are given by Selçuk [46]. The effect of using

[1] The term *impossible differential* appeared first in [3].

more than one linear trail, referred to as linear hulls, has been introduced by Nyberg [45]; see also Daemen and Rijmen [22]. Multidimensional linear attacks have been studied by Hermelin, Cho, and Nyberg [27] as a way to further reduce the data complexity of the basic attack. These approaches have been used for example by Cho [20]. More recently, the zero-correlation attacks introduced by Bogdanov and Rijmen in [16,17] have become popular. These attacks, which can be seen as the natural counterpart of impossible differential attacks, are based on linear approximations with probability exactly $1/2$. A further generalization of zero-correlation attacks, namely attacks based on key-invariant biases, was presented in [14].

1.3. *Theoretical Links Between Linear and Differential Cryptanalysis*

Most of the work has been done independently for linear and differential cryptanalysis, and there are examples of ciphers that appear to be more vulnerable to one type than under the other. However, the concepts are closely related. A first fundamental link between them was already given in 1994 by Chabaud and Vaudenay (see [19]), where it was shown that the probability of a differential can be expressed in terms of a sum of correlations of linear approximations. Interestingly, this link was for a long time not used in practice due to its large computational complexity. Only in 2013, Blondeau and Nyberg used the link in [12] to compute the probability of a differential given the correlations of many linear approximations and proved the equivalence of the existence of a zero-correlation multidimensional linear distinguisher and an impossible truncated differential. More recently, this link was extended by Sun et al. to cover also integral attacks [47] and by Blondeau and Nyberg to cover general multidimensional linear approximations and truncated differentials [13]. The latter paper also relates the different statististical attack settings of these properties.

1.4. *Differential-Linear Cryptanalysis*

On the cryptanalytical side, differential and linear attacks have been used jointly for the first time by Langford and Hellman [35]. The basic idea of *differential-linear cryptanalysis* is to split the cipher under consideration into a composition of two parts. The split should be such that for the first part of the cipher there exists a strong truncated differential and for the second part there exists a strongly biased linear approximation. In [35], the particular case where the differential over the first part holds with probability one has been introduced. Later on, Biham et al. [5,34] generalized this attack using a probabilistic truncated differential on the first rounds of the distinguisher.

More recently in 2012 [39,40], Lu studied the validity of the model proposed by Biham et al. with the aim of minimizing the assumptions needed for the validity of the attack.

Wagner presented ideas toward a unified view of statistical block cipher cryptanalysis [49]. While concentrating on structural similarities between different attacks in a Markov setting, he relied, albeit with some doubts, on the previously made heuristic assumptions under which the differential-linear attacks had been claimed to work.

It is very remarkable that in none of the previous work on differential-linear cryptanalysis, the theoretical link presented in [19] between linear and differential attacks is used to model and understand better the general behavior of differential-linear cryptanalysis.

Recently in [23,37], differential-linear attacks on Ascon and Chaskey have been proposed.

1.5. *Our Contribution*

In this paper, we take the natural step and apply the theoretical link between linear and differential cryptanalysis to differential-linear cryptanalysis. This has a couple of interesting consequences.

To the best of our knowledge, we are, for the first time, able to exactly express the bias of a differential-linear approximation by a closed expression. The formula is exact under the sole assumption that the two parts of the cipher are independent. In particular, it is exact when averaging over all round-keys.

Evaluating this exact expression, the differential-linear hull is computationally infeasible, and in practice only a part of it or just a single term of it is evaluated. Unlike in the case of ordinary differentials and linear hulls, we are not able to theoretically justify that by doing this, we will obtain a lower bound of the absolute value of the bias. We found, however, that it is possible to state some explicit assumptions under which a single differential-linear characteristic gives a good estimate of the bias. We then revisit the previous treatments of differential-linear bias by Biham et al. [5,6], Liu et al. [38], and Lu in [39,40] and show how instead of the piling-up lemma the given bias estimates can be derived from the differential-linear hull.

Moreover, given this exact expression of the bias and along with this a deeper understanding of differential-linear attacks allows us to substantially generalize the attack vector. In particular, we study the possibility to take into consideration the hull of a differential-linear approximation and introduce a multidimensional generalization of differential-linear cryptanalysis which is defined for multiple input differences and multidimensional linear output masks.

Note that we do not propose new concrete attacks. Rather we provide a sound framework for previous and future work on differential-linear cryptanalysis.

1.6. *Organization of the Paper*

In Sect. 2, we fix our notations and state several general results on differential and linear cryptanalysis. The related work is summarized in Sect. 3. In Sect. 4, we develop the exact expression for the bias of the differential-linear distinguisher (cf. Theorem 2) and outline its meaning with an example using the block cipher Serpent. Furthermore, we elaborate more on the comparison with previous work. In Sect. 5, we derive conditions on how and if it is possible to obtain good and practical estimations of the exact expression. We back up our assumption with experiments using small-scale variants of the cipher PRESENT. Finally, in Sect. 6, we generalize the concept of differential-linear cryptanalysis to the case of multiple differentials and multiple linear approximations and derive expressions for the biases and the attack complexities for this generalization. Section 7 concludes the paper.

2. Basics of Linear and Differential Cryptanalysis

2.1. *Linear Correlation and Differential Probability*

In differential cryptanalysis [8], the attacker is interested in identifying and exploiting non-uniformity in occurrences of plaintext and ciphertext differences. Given a vectorial Boolean function $F : \mathbb{F}_2^n \rightarrow \mathbb{F}_2^n$, a differential is given by a pair (δ, Δ) of an input difference $\delta \in \mathbb{F}_2^n$ and an output difference $\Delta \in \mathbb{F}_2^n$ and its probability is defined as

$$\Pr[\delta \overset{F}{\rightarrow} \Delta] = 2^{-n} \#\{x \in \mathbb{F}_2^n \mid F(x) + F(x + \delta) = \Delta\}.$$

Linear cryptanalysis [41] uses a linear relation between bits from plaintexts, corresponding ciphertexts and encryption key. A linear relation of bits of data $x \in \mathbb{F}_2^n$ is determined by a mask $a \in \mathbb{F}_2^n$ and is given as a Boolean function $f(x) = a \cdot x$ where "$\cdot$" is the natural inner product of the vectors a and x in $\mathbb{F}_2^n$. The strength of a linear relation is measured by its correlation. The correlation of a Boolean function $f : \mathbb{F}_2^n \rightarrow \mathbb{F}_2$ is defined as

$$\mathrm{cor}(f) = \mathrm{cor}(f(x)) = 2^{-n} \Big[\# \big\{ x \in \mathbb{F}_2^n | f(x) = 0 \big\} - \# \big\{ x \in \mathbb{F}_2^n | f(x) = 1 \big\} \Big],$$

where the quantity within brackets corresponds to the Fourier coefficient of f at zero and can be computed using the Walsh transform of f, see e.g. [18].

In this paper, a block cipher, or a part of it, with a fixed key and block size n is considered as a bijective vector-valued Boolean function $F : \mathbb{F}_2^n \rightarrow \mathbb{F}_2^n$. In the general model of the differential-linear cryptanalysis to be built in this paper, we consider a set of input differences to the cipher that form a linear subspace of $\mathbb{F}_2^n$. Given a subspace U of $\mathbb{F}_2^n$, let us denote by $U^\perp$ the orthogonal subspace of U with respect to the inner product of $\mathbb{F}_2^n$. That is,

$$U^\perp = \{v \in \mathbb{F}_2^n \mid u \cdot v = 0, \quad \text{for all } u \in U\}.$$

Let us denote by $0_\ell \in \mathbb{F}_2^\ell$ the all-zero string of length ℓ. If $U = \mathbb{F}_2^s \times \{0_t\}$, for some positive integers s and t, where $s + t = n$, then $U^\perp = \{0_s\} \times \mathbb{F}_2^t$. In this manner, we obtain a splitting of $\mathbb{F}_2^n$ to two mutually orthogonal subspaces, whose intersection is $\{0_n\}$. However, it is not true in general that the intersection of U and $U^\perp$ is always $\{0_n\}$. Another type of example of orthogonal subspaces is obtained for $U = \{(0, 0), (1, 1)\} \times \{0_{n-2}\}$. Then $U^\perp = \{(0, 0), (1, 1)\} \times \mathbb{F}_2^{n-2}$, in which case $U \subset U^\perp$. However, in any case the dimensions of U and $U^\perp$ add up to n.

A truncated differential [30] over a vectorial Boolean function $F : \mathbb{F}_2^n \rightarrow \mathbb{F}_2^n$ is a set of ordinary differentials (δ, Δ). In this paper, we restrict to the case where $\delta \in U^\perp$ and $\Delta \in V^\perp$ and U and V are linear subspaces of $\mathbb{F}_2^n$. In this way, the truncated differential is determined by a pair of linear spaces U and V. The strength of a truncated differential is measured by the number of solutions $(x, \delta, \Delta) \in \mathbb{F}_2^n \times (U^\perp \setminus \{0\}) \times V^\perp$ to the equation

$$F(x + \delta) + F(x) = \Delta. \tag{1}$$

To facilitate the derivations, we will use, in this paper, a different but closely related quantity that allows the zero difference in the input. It is straightforward to show that the number of solutions $(x, \delta, \Delta) \in \mathbb{F}_2^n \times U^\perp \times V^\perp$ of Eq. (1) can be computed as

$$\sum_{\delta \in U^\perp, \Delta \in V^\perp} \#\{x \in \mathbb{F}_2^n \mid F(x+\delta) + F(x) = \Delta\}.$$

We denote by $\Pr[U^\perp \xrightarrow{F} V^\perp]$ the probability that a pair of inputs $(x, x+\delta)$, where x is picked uniformly at random in $\mathbb{F}_2^n$ and δ uniformly at random in $U^\perp$, gives an output difference $\Delta \in V^\perp$.

Proposition 1. *Let $F : \mathbb{F}_2^n \to \mathbb{F}_2^n$ be a vectorial Boolean function and U and V be linear subspaces of $\mathbb{F}_2^n$. Then*

$$\begin{aligned} \Pr[U^\perp \xrightarrow{F} V^\perp] &= \frac{1}{2^n |U^\perp|} \# \left\{ (x, \delta, \Delta) \in \mathbb{F}_2^n \times U^\perp \times V^\perp \mid F(x+\delta) + F(x) = \Delta \right\} \\ &= \frac{1}{|U^\perp|} \sum_{\delta \in U^\perp,\, \Delta \in V^\perp} \Pr[\delta \xrightarrow{F} \Delta]. \end{aligned} \tag{2}$$

The probability $\Pr[U^\perp \xrightarrow{F} V^\perp]$ which can be expressed in the two different ways shown in Proposition 1 will be called the truncated differential probability.

Let us denote by $\Pr[U^\perp \setminus \{0\} \xrightarrow{F} V^\perp]$ the probability for the truncated differential derived analogically as above but without allowing the zero input difference. Then, we have the following relation:

$$|U^\perp| \cdot \Pr[U^\perp \xrightarrow{F} V^\perp] = 1 + (|U^\perp| - 1) \cdot \Pr[U^\perp \setminus \{0\} \xrightarrow{F} V^\perp]. \tag{3}$$

In particular, for the ordinary differential probability, we have

$$\Pr[\delta \xrightarrow{F} \Delta] = 2 \cdot \Pr[sp(\delta) \xrightarrow{F} \Delta] - 1$$

for all $\delta,\ \Delta \in \mathbb{F}_2^n$, $\delta \neq 0$. Here, as well as later in the paper, we use the notation $\mathrm{sp}(a)$ to denote the vector subspace $\{0, a\} \subset \mathbb{F}_2^n$ spanned by a.

If F is bijective, then the probability of a single differential is symmetric, that is,

$$\Pr[\delta \xrightarrow{F} \Delta] = \Pr[\Delta \xrightarrow{F^{-1}} \delta].$$

The truncated differential probability is not symmetric, except in the case when $|U| = |V|$. In general, we have

$$|U^\perp| \cdot \Pr[U^\perp \xrightarrow{F} V^\perp] = |V^\perp| \cdot \Pr[V^\perp \xrightarrow{F^{-1}} U^\perp].$$

The main tool used in this paper to relate the differentials and linear approximations is the link between the differential probabilities and the squared correlations of linear approximations of vectorial Boolean functions presented by Chabaud and Vaudenay [19]. We state it as follows.

$$\Pr[\delta \xrightarrow{F} \Delta] = 2^{-n} \sum_{u \in \mathbb{F}_2^n} \sum_{v \in \mathbb{F}_2^n} (-1)^{u \cdot \delta + v \cdot \Delta} \mathrm{cor}^2\left(u \cdot x + v \cdot F(x)\right), \tag{4}$$

where $F : \mathbb{F}_2^n \to \mathbb{F}_2^n$ is a vectorial Boolean function, and $(\delta,\ \Delta) \in \mathbb{F}_2^n \times \mathbb{F}_2^n$. By applying this link for all $\delta \in U^\perp$ and $\Delta \in V^\perp$ in Eq. (2), we obtain the following result which is a generalization of [12,13].

Theorem 1. *The probability of a truncated differential with input differences in $U^\perp$ and output differences in $V^\perp$ can be computed as a sum of squared correlations with input masks in U and output masks in V as*

$$\Pr[U^\perp \xrightarrow{F} V^\perp] = \frac{1}{|V|} \sum_{u \in U, v \in V} \mathrm{cor}^2\left(u \cdot x + v \cdot F(x)\right).$$

Proof. If for $u \in \mathbb{F}_2^n$ we have $u \cdot \delta = 1$ for some $\delta \in U^\perp$, then the linear function $\delta \mapsto u \cdot \delta$ is nonzero and hence balanced on $U^\perp$. Thus, in this case $\sum_{\delta \in U^\perp} (-1)^{u \cdot \delta} = 0$. This is not the case exactly if we have $u \in U$ and then $\sum_{\delta \in U^\perp} (-1)^{u \cdot \delta} = |U^\perp|$. Then, applying the same reasoning for all $v \in \mathbb{F}_2^n$ gives the claim. □

The following corollary concerning a special case will be used later.

Corollary 1. *For all $w \in \mathbb{F}_2^n \setminus \{0\}$ and $\Delta \in \mathbb{F}_2^n \setminus \{0\}$ we have*

$$\Pr[\Delta \xrightarrow{F} \mathrm{sp}(w)^\perp] = \sum_{v \in \mathrm{sp}(\Delta)^\perp} \mathrm{cor}^2(v \cdot x + w \cdot F(x)).$$

Proof. From Eq. (3) and Theorem 1, we have

$$\begin{aligned}\Pr[\Delta \xrightarrow{F} \mathrm{sp}(w)^\perp] &= 2 \cdot \Pr[\mathrm{sp}(\Delta) \xrightarrow{F} \mathrm{sp}(w)^\perp] - 1 \\ &= 2 \cdot \frac{1}{2} \cdot \sum_{v \in \mathrm{sp}(\Delta)^\perp, b \in \mathrm{sp}(w)} \mathrm{cor}^2(v \cdot x + b \cdot F(x)) - 1 \\ &= \sum_{v \in \mathrm{sp}(\Delta)^\perp} \mathrm{cor}^2(v \cdot x + w \cdot F(x)).\end{aligned}$$

□

2.2. *Round Independence*

Differential probabilities or linear correlations over an iterated cipher are often computed assuming that the rounds of the cipher behave independently. The notion of independence is defined precisely in the following definition.

Definition 1. Two parts E_0 and E_1 of an n-bit block cipher $E = E_1 \circ E_0$ are said to be *differentially round independent* if for all $(\delta, \Omega) \in \mathbb{F}_2^n \times \mathbb{F}_2^n$ the following holds

$$\Pr[\delta \overset{E}{\to} \Omega] = \sum_{\Delta \in \mathbb{F}_2^n} \Pr[\delta \overset{E_0}{\to} \Delta] \Pr[\Delta \overset{E_1}{\to} \Omega].$$

Analogically, the parts E_0 and E_1 are said to be *linearly round independent* if for all $(u, w) \in \mathbb{F}_2^n \times \mathbb{F}_2^n$ the following holds

$$\mathrm{cor}^2(u \cdot x + w \cdot E(x)) = \sum_{v \in \mathbb{F}_2^n} \mathrm{cor}^2(u \cdot x + v \cdot E_0(x))\mathrm{cor}^2(v \cdot y + w \cdot E_1(y)).$$

It was proved in [2] that the rounds of a Markov cipher [33] are both differentially and linearly round independent. Next, we show that differential and linear round independence are equivalent concepts for any cipher.

Proposition 2. *Two parts E_0 and E_1 of an n-bit block cipher $E = E_1 \circ E_0$ are differentially round independent if and only if they are linearly round independent.*

Proof. Let us start by stating Eq. (4) in the following equivalent form

$$\sum_{\delta \in \mathbb{F}_2^n} (-1)^{u \cdot \delta} \Pr[\delta \overset{F}{\to} \Delta] = \sum_{v \in \mathbb{F}_2^n} (-1)^{v \cdot \Delta} \mathrm{cor}^2 \left(u \cdot x + v \cdot F(x)\right).$$

This is obtained by applying the inverse Fourier transform to the input difference. By applying it to the output difference, another equivalent form can be given where the first summation is taken over Δ and the second summation over u. We refer to these equations as partial inverses of Eq. (4). A further variant is obtained by applying the inverse Fourier transform on both differences. We call it the inverse of Eq. (4).

Let us now assume that the parts of the cipher are differentially round independent. Then using the inverse of Eq. (4) and the assumption of differential round independence, we get

$$\begin{aligned} \mathrm{cor}^2 \left(u \cdot x + w \cdot E(x)\right) &= 2^{-n} \sum_{\delta \in \mathbb{F}_2^n} \sum_{\Omega \in \mathbb{F}_2^n} (-1)^{u \cdot \delta + w \cdot \Omega} \Pr[\delta \overset{E}{\to} \Omega] \\ &= 2^{-n} \sum_{\Delta \in \mathbb{F}_2^n} \sum_{\delta \in \mathbb{F}_2^n} (-1)^{u \cdot \delta} \Pr[\delta \overset{E_0}{\to} \Delta] \sum_{\Omega \in \mathbb{F}_2^n} (-1)^{w \cdot \Omega} \Pr[\Delta \overset{E_1}{\to} \Omega]. \end{aligned}$$

Then using the both partial inverses of Eq. (4), we obtain

$$\begin{aligned}
&\mathrm{cor}^2\left(u \cdot x + w \cdot E(x)\right)\\
&\quad = 2^{-n}\sum_{\Delta\in\mathbb{F}_2^n}\sum_{v\in\mathbb{F}_2^n}(-1)^{v\cdot\Delta}\mathrm{cor}^2(u\cdot x+v\cdot E_0(x))\sum_{v'\in\mathbb{F}_2^n}(-1)^{v'\cdot\Delta}\mathrm{cor}^2\left(v'\cdot y+w\cdot E_1(y)\right)\\
&\quad = 2^{-n}\sum_{v\in\mathbb{F}_2^n}\sum_{v'\in\mathbb{F}_2^n}\mathrm{cor}^2(u\cdot x+v\cdot E_0(x))\,\mathrm{cor}^2\left(v'\cdot y+w\cdot E_1(y)\right)\sum_{\Delta\in\mathbb{F}_2^n}(-1)^{(v+v')\cdot\Delta}.
\end{aligned}$$

The sum over Δ is nonzero if and only if $v = v'$ and the value of this sum, 2^n, cancels with the factor 2^{-n}. Thus, we get

$$\mathrm{cor}^2\left(u \cdot x + w \cdot E(x)\right) = \sum_{v\in\mathbb{F}_2^n}\mathrm{cor}^2\left(u\cdot x+v\cdot E_0(x)\right)\mathrm{cor}^2\left(v\cdot y+w\cdot E_1(y)\right),$$

which is exactly the condition of linear round independence as defined above. The converse proof is analogous. □

Few ciphers satisfy round independence in the strict sense of Definition 1. On the other hand, n-bit ciphers of the form $E_K(x) = E_1(E_0(x) + K)$ with n-bit key K are round independent on average over the key. For simplicity, the results given in this paper will be stated in terms of strict round independence, but can be reformulated using average round independence for such ciphers.

3. Differential-Linear Distinguisher

In this section, we recall the basic principle of the differential-linear distinguisher and discuss previous treatments of how to estimate its efficiency. This will be the starting point to study the theoretical justification of the combination of linear and differential distinguisher. In particular to examine the underlying assumptions for such a combination of two distinguishers of different type.

3.1. *Differential-Linear Bias*

In statistical attacks, one can distinguish between the offline analysis consisting in detecting a weak property of the cipher and the online analysis consisting in extracting information on the encryption key. In this respect, the differential-linear attacks introduced in [5,35] do not differ from the classical differential or linear attacks. The differences found between these attacks are mostly caused by the method used to detect the weak property of the cipher. For instance, in the differential context under the Markov cipher assumption (or the differential independence as defined above), the probability of a differential characteristic is obtained by multiplying the probabilities over the different rounds of the cipher. Then, methods such as Branch-and-Bound algorithms or matrix methods have been proposed to search for differentials with high probabilities. For some ciphers with strong diffusion, however, these methods soon meet their limits. Either the

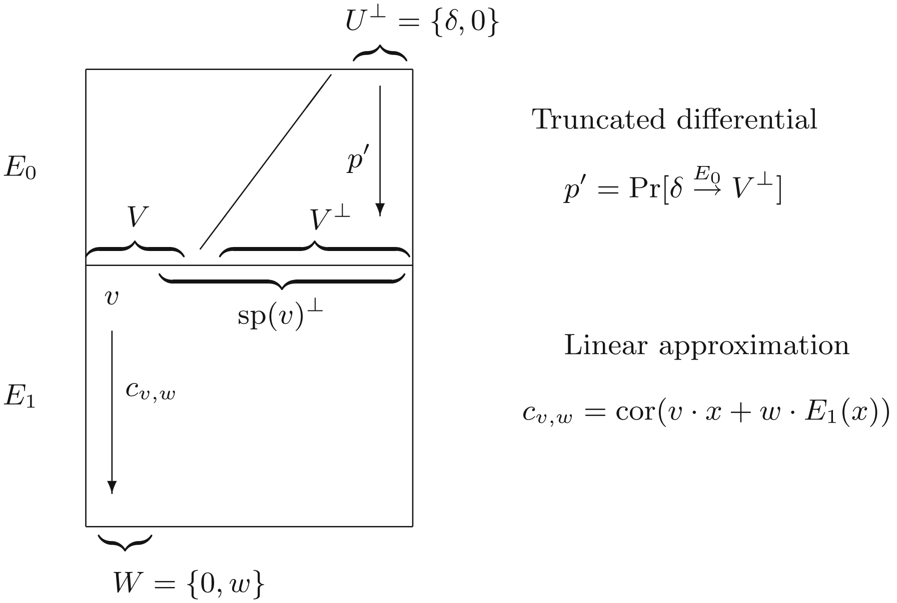
$U^\perp = \{\delta, 0\}$
E_0
p'
V
$V^\perp$
v
$\mathrm{sp}(v)^\perp$
E_1
$c_{v,w}$
$W = \{0, w\}$
Truncated differential
$p' = \Pr[\delta \xrightarrow{E_0} V^\perp]$
Linear approximation
$c_{v,w} = \mathrm{cor}(v \cdot x + w \cdot E_1(x))$

Fig. 1. The setting of a differential-linear distinguisher with input difference δ, output mask w and intermediate mask $v \in V$.

size of the search tree or the matrix grows exponentially as the number of studied rounds increases. The idea of differential-linear attacks is to find a strong differential-linear distinguisher by combining a strong truncated differential with strong linear approximations.

Let $E : \mathbb{F}_2^n \to \mathbb{F}_2^n$ be an encryption function of a block cipher with a fixed key. When applying the technique of differential-linear cryptanalysis, the cipher is presented as a composition $E = E_1 \circ E_0$ of two parts. The first part E_0 is chosen in such a way that there is some strong truncated differential over E_0.

The typical setting of a differential-linear property is depicted in Fig. 1. Over the first part E_0, the space U is selected so that $U^\perp$ is one-dimensional. The output difference space $V^\perp$ is usually larger. For the second part E_1, it is then assumed that there is a strong linear approximation (v, w) over E_1, where $v \in V$, which means that $v \cdot \Delta = 0$ for all $\Delta \in V^\perp$.

Then, the bias of the differential-linear approximation is defined as

$$\mathcal{E}_{\delta,w} = \Pr[w \cdot (E(x + \delta) + E(x)) = 0] - \frac{1}{2}. \tag{5}$$

By combining a truncated differential $(\delta, V^\perp)$ with high probability p' and a strong linear approximation (u, v) with $v \in V$ and a correlation $c_{v,w}$ of high absolute value, the attacker is expecting to get a differential-linear approximation with bias of high absolute value.

In this definition of the bias, neither the input mask v nor the output difference space $V^{\perp}$ does appear. In practice, however, when deriving estimates of the differential-linear bias they play a crucial role as we will see in this paper.

3.2. *Attack Algorithm and Statistics*

The attacker targets on a linear relation expressed by the linear mask w of the bits of each output difference obtained for pairs of plaintexts with input difference δ. Algorithm 1 presents the differential-linear distinguisher (δ, w) given a sample of pairs of plaintexts of size N_S.

Alg. 1 Differential-linear distinguisher

```
Set a counter $T$ to 0
**for** $N_S$ plaintext pairs $(x, x+\delta)$ **do**
    Increment $T$ if $w \cdot \Big(E(x) + E(x \oplus \delta)\Big) = 0$
**if** $T$ deviates enough from $N_S/2$ **then**
    The data is drawn from the cipher $E$.
```

In the previous literature, the data complexity of the differential-linear attack has usually been estimated as follows [6]. Let N be the number of pairs used in the attack, and let T_w and T_0 be two random variables following the binomial distributions $\mathcal{B}(N, 1/2)$ and $\mathcal{B}(N, 1/2 + \mathcal{E}_{\delta,w}))$, respectively. As in the linear context [41,46], these binomial distributions are approximated by normal distributions. Therefore, we can assume that $T_w/N - 1/2$ and $T_0/N - 1/2$ follow the normal distributions $\mathcal{N}(0, \frac{1}{4N})$ and $\mathcal{N}(\mathcal{E}_{\delta,w}, \frac{1}{4N})$, respectively.

Unlike in [46], there is no need to use the folded normal distribution. This fact corresponds to the observation made in [6] that, in the case of positive expected bias, it suffices to test whether $T/N - 1/2 > \varepsilon$ rather than $|T/N - 1/2| > \varepsilon$. It is even claimed in [6] that the sign of the bias "is unaffected by any key bit (as all the affected key bits are used twice and thus cancelled)." While this is clearly the reason for the differential-linear probability to be a nonnegative value, this probability may, for some keys, be less than 1/2, even if the average probability taken over the keys is greater than 1/2. In any case, the main difference between the linear context and the differential-linear context is that the expected bias in the latter case is a nonzero value. In the linear context, only the absolute value of the expected bias is nonzero, while the expectation of the bias is typically zero due to the effect of the key bits.

With this modification, which allows saving one bit in the advantage, the framework of [46] is then adapted to the differential-linear context to give the success probability of a key recovery attack as

$$P_S = \Phi\left(2\sqrt{N}\mathcal{E}_{\delta,w} - \Phi^{-1}(1 - 2^{-a})\right),$$

where Φ is the cumulative distribution function of the standard normal distribution and a is the advantage of the attack in bits as defined in [46]. From this estimate, the data complexity of the differential-linear attack is then deduced.

Lemma 1. *Given the bias $\mathcal{E}_{\delta,w}$ of a differential-linear approximation as defined in Eq.* (5)*, the data complexity of a key-recovery attack with advantage a and success probability P_S can be given as*

$$N = \frac{\left(\Phi^{-1}(P_S) + \Phi^{-1}(1-2^{-a})\right)^2}{4\mathcal{E}_{\delta,w}}.$$

3.3. *Previous Treatments*

In the previous treatments [5,35,39], the bias

$$\mathcal{E}_{\delta,w} = \Pr[w \cdot (E(x+\delta) + E(x)) = 0] - \frac{1}{2}$$

is evaluated by decomposing the Boolean variable $w \cdot (E(x+\delta) + E(x))$ as a sum of three variables

$$\begin{aligned} w \cdot (E(x+\delta) + E(x)) = {} & v \cdot E_0(x+\delta) + w \cdot E(x+\delta) \\ & + v \cdot (E_0(x+\delta) + E_0(x)) \\ & + v \cdot E_0(x) + w \cdot E(x), \end{aligned} \tag{6}$$

where v is the input mask to the strong linear approximation over the second part of the cipher. Then, these three terms are assumed to be statistically independent as x varies and the bias is evaluated using the piling-up lemma [41].

By using the following notation for the involved biases

$$\begin{aligned} \epsilon_{v,w} &= \Pr[v \cdot y + w \cdot E_1(y) = 0] - \frac{1}{2} \\ \varepsilon_{\delta,v} &= \Pr[v \cdot (E_0(x+\delta) + E_0(x)) = 0] - \frac{1}{2} = \Pr[\delta \overset{E_0}{\to} \mathrm{sp}(v)^{\perp}] - \frac{1}{2}, \end{aligned} \tag{7}$$

the piling-up lemma gives

$$\mathcal{E}_{\delta,w} = 4\varepsilon_{\delta,v}\epsilon_{v,w}^2. \tag{8}$$

It remains to determine $\varepsilon_{\delta,v}$ given the truncated differential probability $p' = \Pr[\delta \overset{E_0}{\to} V^{\perp}]$. That is, one has to deal with the probability of differentials where the output difference is contained in $\mathrm{sp}(v)^{\perp}$ but not in $V^{\perp}$.

This is where the previous studies differ. In [35], $\Pr[\delta \overset{E_0}{\to} V^{\perp}] = 1$, in which case $\Pr[\delta \overset{E_0}{\to} \mathrm{sp}(v)^{\perp}] = 1$, since $v \in V$. The general case where $\Pr[\delta \overset{E_0}{\to} V^{\perp}] < 1$ was considered first in [34] and then by Biham et al. [5] (see also Biham et al. [7]). Then assuming that when $\Delta \notin V^\perp$ the parities of $v \cdot \Delta$ are balanced they obtain the estimate

$$\Pr[\delta \overset{E_0}{\to} \mathrm{sp}(v)^\perp] \approx p' + (1-p')\frac{1}{2},$$

where equality holds if $p' = 1$. If $p' < 1$, equality does not hold, for the simple reason that if a linear function $v \cdot y$ vanishes in $V^\perp$, it cannot be balanced outside $V^\perp$. The estimate is better, if $V^\perp$ is small, which is the case studied in [5]. It becomes worse, however, as $V^\perp$ increases. The extreme case is $\mathrm{sp}(v) = V$. Then, $v \cdot y = 1$, for all $y \notin V^\perp$.

This problem was observed by Lu [39]. In his study, the assumption about uniform distribution of parities is not needed, since the output difference space of the truncated differential is taken equal to the hyperplane $\mathrm{sp}(v)^\perp$.

As in practice, $V^\perp$ is often smaller than a zero space of the linear approximation, we have that $\Pr[\delta \overset{E_0}{\to} V^\perp]$ is less than or equal to $\Pr[\delta \overset{E_0}{\to} \mathrm{sp}(v)^\perp]$. As the inequality may be strict, replacing the latter by the former in the estimation of the bias Eq. (7) may lead to a wrong result for $\mathcal{E}_{\delta,w}$. Biham et al. suggest that the other output differences $\Delta \in \mathrm{sp}(v)^\perp \setminus V^\perp$ may occur with high probability and affect their approximation. Therefore, they stress the importance of experimental verification.

Note that it would be possible to improve the assumption by Biham et al. by correcting the probability of zero parity outside $V^\perp$ from $\frac{1}{2}$ to $(2^{n-1} - |V^\perp|)/(2^n - |V^\perp|)$. We will not proceed in this direction, since in Sect. 5.1 we present a new approach to justify the bias estimate of Biham et al. without the assumption of the piling-up lemma.

In [38], the authors mention the possibility of using multiple linear approximations in order to improve the complexity of a differential-linear distinguisher. Their study, which is based on the differential-linear model of Biham et al. [6] and on the multiple linear model of Biryukov et al. [9], assumes that the distinguisher is built from the combination of only one truncated differential with independent linear approximations.

While differential-linear cryptanalysis has been applied successfully in many previous work, the estimates on the attack complexities have been derived under heuristic assumptions and presented in an informal manner. The goal of this paper is to present a rigorous analysis what is happening in the intermediate layer of the differential-linear approximation and take into account not only more high-probability output differences from E_0 but also more, not necessarily independent, linear approximations over E_1. Based on this analysis, we then present alternative approaches for getting accurate estimates of the strength of differential-linear approximations.

4. Differential-Linear Hull

The basic tool in examining the intermediate layer between E_0 and E_1 is the following theorem. We use the notation $\mathcal{E}_{\delta,w}$ and $\varepsilon_{\delta,v}$ introduced in the preceding section, and denote the correlation of the linear approximation $v \cdot y + w \cdot E_1(y)$ by $c_{v,w}$. Then, $c_{v,w} = 2\epsilon_{v,w}$ in relation to the notation used in the preceding section.

Theorem 2. *Assume that the parts E_0 and E_1 of the block cipher $E = E_1 \circ E_0$ are independent. Using the notation previously defined, we have*

$$\mathcal{E}_{\delta,w} = \sum_{v \in \mathbb{F}_2^n} \varepsilon_{\delta,v} c_{v,w}^2 \tag{9}$$

for all $\delta \in \mathbb{F}_2^n \setminus \{0\}$ and $w \in \mathbb{F}_2^n \setminus \{0\}$.

Proof. First, we apply the assumption of independence to the probability $\Pr[\delta \overset{E}{\to} \mathrm{sp}(w)^{\perp}]$ and then the link given by Corollary 1 to the differential probability over E_1.

$$\begin{aligned}
\Pr[\delta \overset{E}{\to} \mathrm{sp}(w)^{\perp}] &= \sum_{\Delta \in \mathbb{F}_2^n} \Pr[\delta \overset{E_0}{\to} \Delta] \Pr[\Delta \overset{E_1}{\to} \mathrm{sp}(w)^{\perp}] \\
&= \sum_{\Delta \in \mathbb{F}_2^n} \Pr[\delta \overset{E_0}{\to} \Delta] \sum_{v \in \mathrm{sp}(\Delta)^{\perp}} \mathrm{cor}^2(v \cdot y + w \cdot E_1(y)) \\
&= \sum_{v \in \mathbb{F}_2^n} \sum_{\Delta \in \mathrm{sp}(v)^{\perp}} \Pr[\delta \overset{E_0}{\to} \Delta] \mathrm{cor}^2(v \cdot y + w \cdot E_1(y)) \\
&= \sum_{v \in \mathbb{F}_2^n} \Pr[\delta \overset{E_0}{\to} \mathrm{sp}(v)^{\perp}] \mathrm{cor}^2(v \cdot y + w \cdot E_1(y)),
\end{aligned}$$

where changing the order of summation is possible since

$$\{(v, \Delta) \mid \Delta \in \mathbb{F}_2^n,\ v \in \mathrm{sp}(\Delta)^{\perp}\} = \{(v, \Delta) \mid v \in \mathbb{F}_2^n,\ \Delta \in \mathrm{sp}(v)^{\perp}\}.$$

Now by subtracting $\frac{1}{2}$ from both of the sides of the obtained equality and using Parseval's theorem gives the result. □

We call the expression Eq. (9) the differential-linear hull of $E = E_1 \circ E_0$. The differential-linear method has been previously applied in cases, where only one correlation $c_{v,w}$ has been identified to have a large absolute value, but the output differential space of the truncated differential is a strict subspace of the zero space of v. Consequently, more than one linear approximation trail must be taken into account when estimating the bias of the differential-linear approximation. We illustrate this in the context of an attack on the Serpent cipher [1].

4.1. *Example on Serpent*

Differential-linear cryptanalysis [6,24] which has been applied to many ciphers remains, together with the multidimensional linear cryptanalysis [42,44], the most powerful attack on the Serpent cipher [1]. In this section, we summarize in our notation the distinguisher proposed in [6] on nine rounds of Serpent and examine its derivation. Later, in [24], another similar distinguisher on Serpent was provided. The only difference was that a new and stronger truncated differential over the three rounds of E_0 was used.

To be used in a key-recovery attack, this distinguisher was defined to start from the second round of the cipher. First, a truncated differential is defined on three rounds of Serpent . In this attack, only one input difference is taken into consideration meaning that $U^{\perp}$ is one-dimensional. The output space of the truncated differential consists of all differences which have the bits number 1 and 117 equal to zero. Hence, it is the orthogonal of the two-dimensional space V spanned by the bits (taken as basis vectors) number 1 and 117. The strong linear approximation over the six following rounds has input mask $\nu \in V$ where both bits number 1 and 117 are equal to 1. The output mask is denoted by w. The bias of this linear approximation is estimated to $\epsilon_{\nu,w} = 2^{-27}$.

The resulting differential-linear relation spans over nine rounds of Serpent. In [6], its bias was estimated using a formula $2pq^2$, where p is such that $\frac{1}{2} + p$ is the probability of the truncated differential, and $q = \epsilon_{\nu,w}$ is the bias of the linear approximation. But instead of taking $\frac{1}{2} + p = \Pr[\delta \overset{E_0}{\to} V^{\perp}]$, they observe in experiments that, in addition to the differences where the two bits 1 and 117 are equal to zero, also "other differentials predict the difference in the bits" of the input mask ν and then they sum all these differentials to estimate p. In Sect. 3, we showed that in the piling-up lemma approach an estimate of $\Pr[v \cdot (E_0(x+\delta) + E_0(x)) = 0] = \Pr[\delta \overset{E_0}{\to} \mathrm{sp}(v)^{\perp}]$ is needed. It seems that in [6] such an estimate was obtained experimentally and its value was $\frac{1}{2} + 2^{-7}$. Then substituting $p = 2^{-7}$ and $q = 2^{-27}$ gives an estimate of the $2pq^2 = 2^{-60}$ for the bias of the differential-linear distinguisher in [6].

Let us analyze next the bias of the same distinguisher by considering the differential-linear hull. From Theorem 2, we deduce that $\mathcal{E}_{\delta,w}$ can be computed as

$$\mathcal{E}_{\delta,w} = \sum_{v \in V} \varepsilon_{\delta,v} c_{v,w}^2 + \sum_{v \in \mathbb{F}_2^n \setminus V} \varepsilon_{\delta,v} c_{v,w}^2. \tag{10}$$

We observe that for the two masks $v \in V$, for which only one bit, either number 1 or 117, is equal to 1, the correlations $c_{v,w}$ are equal to zero. Then, it follows that the first sum on the right side of Eq. (10) is, indeed, equal to $\varepsilon_{\delta,\nu} c_{\nu,w}^2$, that is, the same value as we would obtain by the direct application of Eq. (8).

Note that the terms in the summations Eq. (10) can be positive or negative. So strictly speaking, restricting to a partial sum may not give a lower bound. On the other hand, this may well happen in practice, which motivates the investigations in the next section.

4.2. *Recent Differential-Linear Attacks*

Recently in 2015, a differential-linear attack [28] on the authenticated cipher ICEPOLE [43] has been presented. While the authors used classical methods to find potentially good input differences and output masks, the bias of the differential-linear characteristics was so large that it was possible for the authors to estimate this one experimentally, meaning that no assumptions on the differential-linear hull was made in this estimate.

The block cipher CTC2 [21] has been proven to be weak against differential-linear attacks. While in previous attacks [25,39,40], the space V was one dimensional, using the differential-linear hull the authors of [26] improved the best attack on this cipher.

The estimation of the differential-linear bias is based on the theory we developed [11] and which is presented in the next section.

5. Estimating the Bias Using the Intermediate Space

5.1. *Previous Models*

In the few available analyses in the differential-linear context recalled in Sect. 3, the differential-linear bias is estimated by considering the differential-linear approximation as a combination of one strong truncated differential with one strong linear approximation under the assumption that the piling-up lemma holds. In addition, as pointed out by Lu, also other assumptions may be needed. Now that we have the exact expression of the bias available, the question arises whether the piling-up lemma could be avoided and replaced by other reasonable assumptions in the previous attack models. Answering this question would have also practical relevance. The cryptographers would know which properties of the cipher remain to be validated in simulations.

Biham et al. [5] and Lu [39] treat the case of one strong linear approximation over the second part of the cipher. Next, we show that in this case it is possible to replace the piling-up lemma by the assumption that the bias is equally small for all other input masks on the intermediate layer.

Theorem 3. *Suppose that there is one strong linear approximation with masks* (ν, w) *over the second part of the cipher and that the correlations of all other approximations are of equal absolute value. Then*

$$\mathcal{E}_{\delta,w} \approx \varepsilon_{\delta,\nu} c^2_{\nu,w}.$$

Proof. Let us denote $\rho = c_{\nu,w}$. By the assumption and Parseval's theorem we have

$$c^2_{v,w} = \frac{1-\rho^2}{2^n-2}, \;\; v \neq \nu.$$

Then

$$\begin{aligned}\Pr[\delta \overset{E}{\to} \mathrm{sp}(w)^\perp] &= \sum_{v \in \mathbb{F}_2^n} \Pr[\delta \overset{E_0}{\to} \mathrm{sp}(v)^\perp] c^2_{v,w} \\ &= \Pr[\delta \overset{E_0}{\to} \mathrm{sp}(\nu)^\perp]\rho^2 + \sum_{v \neq \nu,\, 0} \Pr[\delta \overset{E_0}{\to} \mathrm{sp}(v)^\perp] \frac{1-\rho^2}{2^n-2}. \qquad (11)\end{aligned}$$

From the fact

$$\sum_{v \neq 0} \Pr[\delta \overset{E_0}{\to} \mathrm{sp}(v)^\perp] = \sum_{v \neq 0} \sum_{\Delta \in \mathrm{sp}(v)^\perp, \Delta \neq 0} \Pr[\delta \overset{E_0}{\to} \Delta] = 2^{n-1} - 1$$

we get the following inequalities

$$\frac{1}{2} \geq \frac{1}{2^n - 2} \sum_{v \neq \nu,\, 0} \Pr[\delta \overset{E_0}{\to} \mathrm{sp}(v)^{\perp}] \geq \frac{1}{2} \cdot \frac{2^{n-1} - 2}{2^{n-1} - 1},$$

which justify the estimate

$$\frac{1}{2^n - 2} \sum_{v \neq \nu,\, 0} \Pr[\delta \overset{E_0}{\to} \mathrm{sp}(v)^{\perp}] \approx \frac{1}{2}.$$

Substituting it to Eq. (11) gives the claim. □

The model of Biham et al. [5] is a special case of the model of Lu, where the probability of the truncated differential $\Pr[\delta \overset{E_0}{\to} \mathrm{sp}(\nu)^{\perp}]$ is estimated based on one strong differential $\delta \overset{E_0}{\to} \Delta$.

Corollary 2. *Let us consider the setting as described in Theorem 3 with one output difference Δ. If then the differential probabilities $\Pr[\delta \overset{E_0}{\to} \Omega]$ for $\Omega \neq \Delta$ are equally small then*

$$\mathcal{E}_{\delta,w} \approx \frac{\Pr[\delta \overset{E_0}{\to} \Delta]}{2} c^2_{\nu,w}.$$

Proof. Let us denote $p = \Pr[\delta \overset{E_0}{\to} \Delta]$. Then

$$\Pr[\delta \overset{E_0}{\to} \Omega] = \frac{1-p}{2^n - 2},$$

for $\Omega \neq \Delta$. From this we get

$$\Pr[\delta \overset{E_0}{\to} \mathrm{sp}(\nu)^{\perp}] = \sum_{\Omega \in \mathrm{sp}(v)^{\perp}} \Pr[\delta \to \Omega] = p + \frac{1-p}{2^n - 2}(2^{n-1} - 2) \approx \frac{p}{2} + \frac{1}{2},$$

and the claim follows from Theorem 3. □

5.2. *Supporting Subset*

By Theorem 2, computation of the exact value of the bias $\mathcal{E}_{\delta,w}$ of a differential-linear approximation requires the knowledge of the correlations over E_1 for all input masks $v \in \mathbb{F}_2^n$, which may be impossible to obtain in practice for many ciphers. In this subsection, we discuss the possibility of using the differential-linear hull more efficiently by restricting the masks v to a subset of the intermediate layer.

In the previous subsection, we already saw that particular assumptions allow focusing on a single strong linear approximation over the second part of the cipher. In fact, it is straightforward to generalize Theorem 3 to a larger set of intermediate masks v. On the other hand, such assumptions are difficult to verify in practice, and they may not even be necessary as we only need a lower bound to the magnitude of the bias to get an upperbound to the data complexity of the attack. This can be guaranteed if the differential-linear approximation admits a supporting subset in the sense of the following definition [cf. Eq. (10)].

Definition 2. Given a differential-linear approximation with input difference δ and output mask w, a subset V in the intermediate layer is called supporting if

$$\left| \sum_{v \in V, v \neq 0} \varepsilon_{\delta,v} c_{v,w}^2 \right| \leq \left| \mathcal{E}_{\delta,w} \right| .$$

When the parts of a block cipher E_0 and E_1 are round independent, then any differential probability over the cipher is underestimated, if we take into consideration only a subset of the differential characteristics relative to the differential. The same holds for the squared correlation of a linear approximation. In terms of Definition 2, we can state that for differentials and linear approximations all subsets are supporting.

Is it possible to achieve the same nice state of affairs for differential-linear hull? This is a difficult problem and left open in this paper. Since the terms of the sum of the differential-linear hull in Eq. (10) may be positive or negative, including more terms in the summation may decrease the estimated bias. Currently, the only way to get some evidence that a subset is supporting is to experimentally compute the bias of a differential-linear approximation over a reduced number of rounds of the cipher. In [5,6], the authors conducted this kind of experiments to check the validity of their results.

5.3. *Breaking Up at Intermediate Layer*

If the supporting set V is large, it may be infeasible to compute the biases $\varepsilon_{\delta,v}$ over E_0 or the correlations $c_{v,w}$ over E_1 for all $v \in V$, while it may be feasible to obtain the probability of the truncated differential over E_0 and the capacity of the multidimensional linear approximation over E_1. Next, we show that under an additional assumption that certain probabilities over E_0 are equal, the sum $\sum_{v \in V} \varepsilon_{\delta,v} c_{v,w}^2$ can be estimated by the product of a truncated differential probability and a multidimensional linear capacity.

Theorem 4. *Let V be a subspace in the intermediate layer of a differential-linear approximation with input difference δ and output linear mask w. Let $\varepsilon_{\delta,V} = \Pr[\delta \overset{E_0}{\to} V^{\perp}] - \frac{1}{|V|}$ be the bias of a truncated differential with one nonzero input difference δ and output differences in $V^{\perp}$. Further, we denote by $C_{V,w} = \sum_{v \in V, v \neq 0} c_{v,w}^2$ the capacity of the multidimensional linear approximation with all input masks v in V and one output mask $w \neq 0$. If then, for all $\Delta \notin V^{\perp}$, the differential probabilities $\Pr[\delta \overset{E_0}{\to} \Delta]$ are equal,* *we have*

$$\sum_{v \in V} \varepsilon_{\delta,v} c_{v,w}^2 = \frac{1}{2} \frac{|V|}{|V|-1} \varepsilon_{\delta,V} C_{V,w}. \tag{12}$$

Proof. For a purpose of clarity, let us denote $Q = \Pr[\delta \overset{E_0}{\to} V^\perp]$. We denote by p the common value of the probabilities $\Pr[\delta \overset{E_0}{\to} \Delta]$ for $\Delta \notin V^\perp$. Then by $\sum_{\Delta \in \mathbb{F}_2^n} \Pr[\delta \overset{E_0}{\to} \Delta] = 1$ we deduce that $p = \dfrac{1-Q}{2^n - |V^\perp|}$.
Since $V^\perp \subset \mathrm{sp}(v)^\perp$ holds for all $v \in V$, we have

$$\begin{aligned}\Pr[\delta \overset{E_0}{\to} \mathrm{sp}(v)^\perp] &= \Pr[\delta \overset{E_0}{\to} V^\perp] + \sum_{\Delta \in \mathrm{sp}(v)^\perp,\ \Delta \notin V^\perp} \Pr[\delta \overset{E_0}{\to} \Delta] \\ &= Q + (2^{n-1} - |V^\perp|) \cdot \frac{1-Q}{2^n - |V^\perp|}.\end{aligned}$$

Therefore, for all $v \in V$, we have

$$\begin{aligned}\varepsilon_{\delta,v} &= \Pr[\delta \overset{E_0}{\to} \mathrm{sp}(v)^\perp] - \frac{1}{2} = Q + \left(2^{n-1} - |V^\perp|\right) \frac{1-Q}{2^n - |V^\perp|} - \frac{1}{2} \\ &= \frac{1}{2} \cdot \frac{2^n Q - |V^\perp|}{2^n - |V^\perp|} = \frac{1}{2} \cdot \frac{Q - |V|^{-1}}{1 - |V|^{-1}} \\ &= \frac{1}{2} \cdot \frac{|V|}{|V|-1} \left(Q - \frac{1}{|V|}\right) = \frac{1}{2} \cdot \frac{|V|}{|V|-1} \varepsilon_{\delta,V}.\end{aligned}$$

And we deduce

$$\sum_{v \in V} \varepsilon_{\delta,v} c_{v,w} = \frac{1}{2} \frac{|V|}{|V|-1} \varepsilon_{\delta,V} \sum_{v \in V} c_{v,w}^2 = \frac{1}{2} \frac{|V|}{|V|-1} \varepsilon_{\delta,V} C_{V,w}.$$

□

Let us note that if $|V| = 2$, we have $\sum_{v \in V} \varepsilon_{\delta,v} c_{v,w}^2 = \varepsilon_{\delta,V} C_{V,w}$. The larger the size of $|V|$, the closer to $\frac{1}{2} \varepsilon_{\delta,V} C_{V,w}$ we get.

5.4. *Experiments*

The experiments of this section have been performed on a 32-bit scaled version of PRESENT [15,36] called SmallPresent-[8]. The differential-linear approximations are defined for one input difference δ and one output mask w. To limit the number of assumptions, the bias $\varepsilon_{\delta,v}$ and the correlations $c_{v,w}$ are computed experimentally using 2^{30} plaintexts and averaged over 200 keys. When using Theorem 2, round independence is only required between E_0 and E_1.

The purpose of these experiments was to investigate supporting subsets in the sense of Definition 2. In each of the figures of this section, we plotted as a reference the experimental bias

$$\mathcal{E}_{\delta,w} = \Pr[\delta \to \mathrm{sp}(w)^{\perp}] - \frac{1}{2}, \tag{13}$$

over eight rounds of SmallPresent-[8], and given a space V, compare it with

$$\sum_{v \in V} \varepsilon_{\delta,v} c_{v,w}^{2}. \tag{14}$$

While experiments have been performed for many differential-linear approximations on eight rounds of SmallPresent-[8], we present results for the input difference $\delta = \texttt{0x1}$ and the output mask $w = \texttt{0x80000000}$. The bias of this differential-linear approximation is positive and we hope to have supporting subsets V such that Eq. (14) gives an underestimate of the actual bias. In Fig. 2, resp. in Fig. 3, the differentials are taken over three rounds, resp. four rounds, and the correlations are taken over 5 rounds, resp. 4 rounds of SmallPresent-[8]. The set V is chosen to be a linear subspace.

As the accuracy of these approximations depends mostly on the size of the subspace, we study the evolution of Eq. (14) in regard to $\log(|V|)$.

Result of the different experiments shows that in the case of SmallPresent-[8], Eq. (14) gives as expected an underestimate of the actual bias $\mathcal{E}_{\delta,w}$. In most of the cases by increasing the size of the supporting space V, we have a better estimate of the bias (in this experiments, the initial spaces V are subset of the larger ones). Nevertheless as the second sum of Eq. (10) is not always positive, we observe that this gain can be somewhat relative. When experiments are conducted for a fixed key instead of averaged over keys, we strictly observe that Eq. (14) is not an increasing function of $|V|$.

In Theorem 4, based on the assumption that for all $\Delta \notin V^{\perp}$, the probabilities $\Pr[\delta \xrightarrow{E_0} \Delta]$ are equal, we propose an estimate of Eq. (14). This one is relatively easier to compute, since, independently of the size of V, only one truncated differential probability and one capacity need to be computed. The blue curves (lowest curves for V of large size) in Figs. 2 and 3 correspond to the computation of the expression on the right side of Eq. (12). While this expression seems to be a correct estimate of Eq. (14) for V of small size, the assumption that for all $\Delta \notin V^{\perp}$, the probabilities $\Pr[\delta \xrightarrow{E_0} \Delta]$ are equal, seems to be getting less realistic as the size of V increases.

The differential-linear approximations simulated in this section also act as examples of truncated differentials composed of two truncated differentials. In these experiments, the true probability of the truncated differential is always larger than the product of the probabilities of its parts. As shown in [10], this is not always the case as sometimes erroneously assumed, but the true probability can also be much smaller. We conclude that the assumption made in Theorem 4 does not always hold in practice and therefore the bias estimate proposed in this theorem must be studied carefully in experiments for each particular cipher.

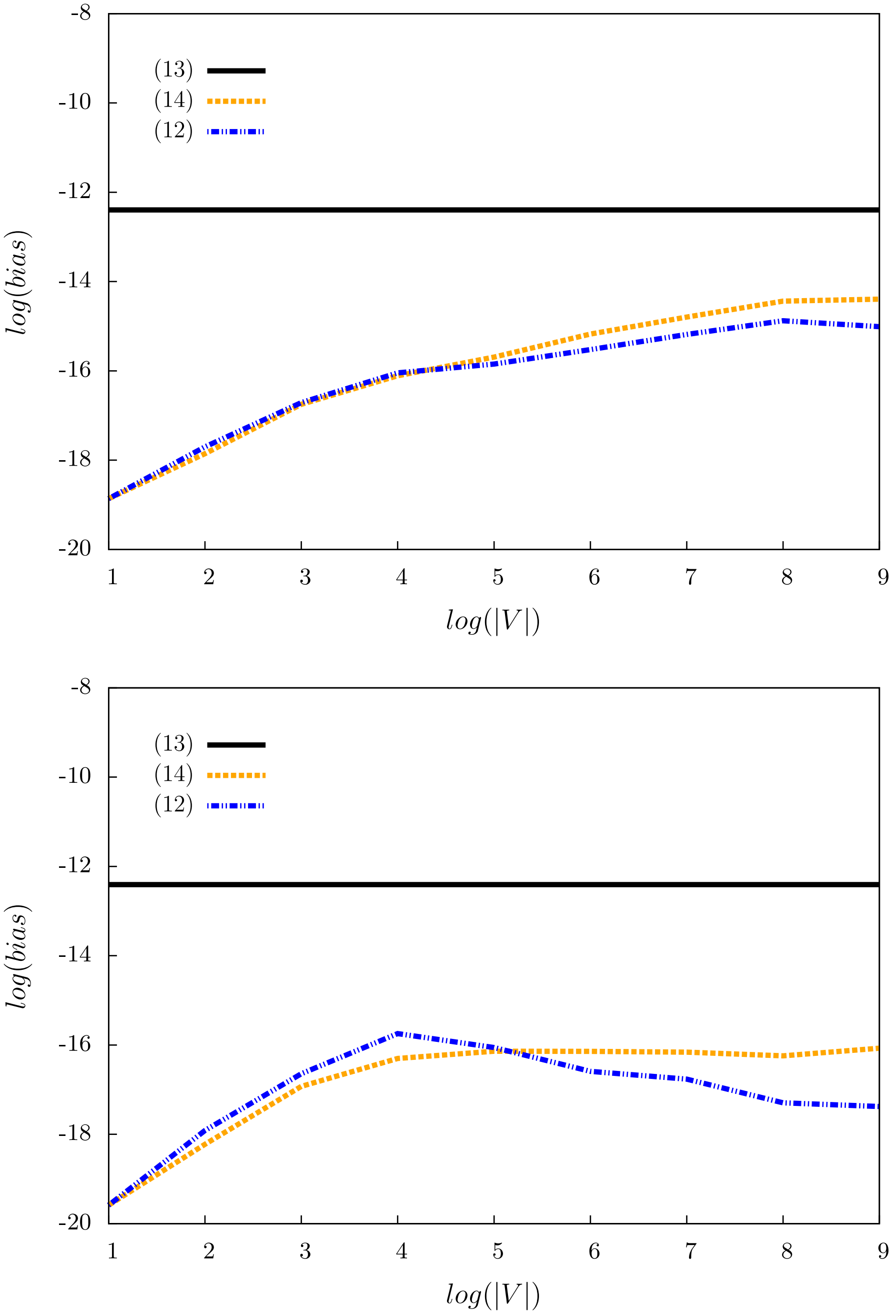

Fig. 2. Estimation of the bias a differential-linear approximation on $3 + 5$ rounds of SmallPresent-[8] for two different chains of intermediate spaces.

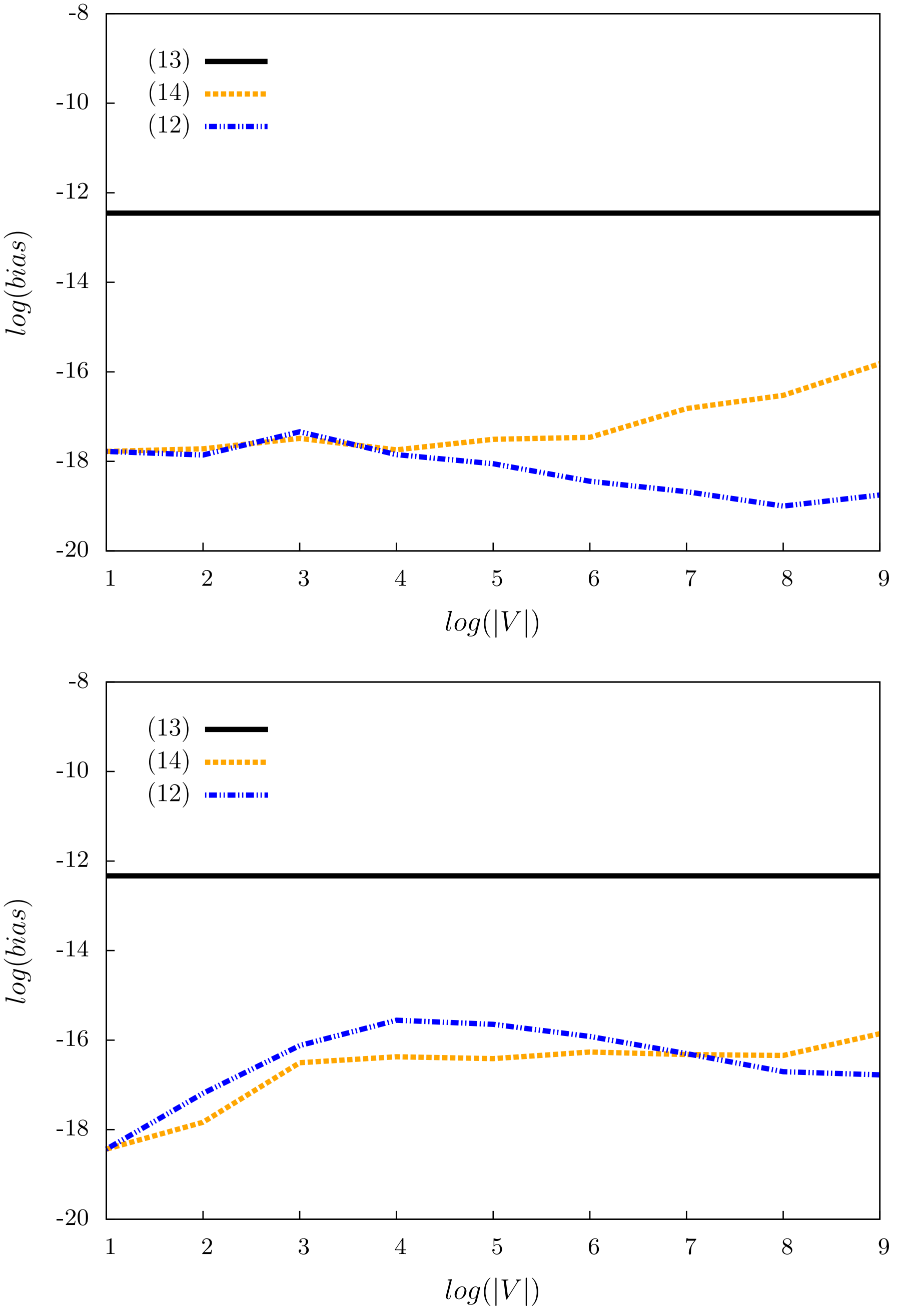

Fig. 3. Estimation of the bias of a differential-linear approximation on $4 + 4$ rounds of SmallPresent-[8] for two different chains of intermediate spaces.

6. Multidimensional Differential-Linear Distinguisher

6.1. *The Model*

The idea of taking advantage of multiple differentials or multiple linear approximations is widely spread out in the cryptographic community. To generalize the results of Sect. 4, let us now consider the case where the space $U^\perp$ of possible input differences is an arbitrary subspace of $\mathbb{F}_2^n$. The linear approximation over E_1 is assumed to be multidimensional such that the output masks form a linear subspace W of $\mathbb{F}_2^n$. We denote its orthogonal space by $W^\perp$.

The conditions on which it would be possible to combine such a truncated differential and multidimensional linear approximation to a strong truncated differential over the full cipher are similar to the ones in the one-dimensional case expressed in Sect. 5.

We express here the generalization of Theorem 2 to compute the bias

$$\mathcal{E}_{U^\perp,W} = \Pr[U^\perp \setminus \{0\} \xrightarrow{E} W^\perp] - \frac{1}{|W|}, \tag{15}$$

of a multidimensional differential-linear approximation.

Theorem 5. *Let $\mathcal{E}_{U^\perp,W}$ be as defined in Eq. (15). Assume that the parts E_0 and E_1 of the block cipher $E = E_1 \circ E_0$ are independent. Then*

$$\mathcal{E}_{U^\perp,W} = \frac{2}{|W|} \sum_{v\in\mathbb{F}_2^n, v\neq 0} \varepsilon_{U^\perp,v} C_{v,W}, \tag{16}$$

where $\varepsilon_{U^\perp,v} = \Pr[U^\perp \setminus \{0\} \xrightarrow{E_0} \mathrm{sp}(v)^\perp] - 1/2$, and $C_{v,W} = \sum_{w\in W, w\neq 0} \mathrm{cor}^2(v\cdot y + w\cdot E_1(y))$, is for $v \neq 0$, the capacity of the multidimensional linear approximation with input mask v and all nonzero output masks w in W.

Proof. First, let us state the following generalization of Corollary 1. Given a bijective function $F : \mathbb{F}_2^n \to \mathbb{F}_2^n$, a subspace $U \subset \mathbb{F}_2^n$ and a mask vector $v \in \mathbb{F}_2^n$, we have

$$2\cdot\Pr[U^\perp \xrightarrow{F} \mathrm{sp}(v)^\perp] - 1 = \sum_{u\in U} \mathrm{cor}^2(u\cdot x + v\cdot F(x)).$$

Using Theorem 1 to write the truncated differential probability in terms of squared correlations, we apply this result together with Proposition 2 to obtain

$$\begin{aligned}
&\Pr[U^\perp \xrightarrow{E} W^\perp] \\
&\quad = \frac{1}{|W|} \sum_{u\in U, v\in\mathbb{F}_2^n, w\in W} \mathrm{cor}^2(u\cdot x + v\cdot E_0(x))\mathrm{cor}^2(v\cdot y + w\cdot E_1(y)) \\
&\quad = \frac{1}{|W|} \sum_{v\in\mathbb{F}_2^n} \left(2\Pr[U^\perp \xrightarrow{E_0} \mathrm{sp}(v)^\perp] - 1\right) \sum_{w\in W} \mathrm{cor}^2(v\cdot y + w\cdot E_1(y)).
\end{aligned}$$

The next step consists at removing the zero from the possible input differences. To use relation of Eq. (3), we multiply the probabilities on the first and second line by $|U|$ and then subtract $1 = \frac{1}{|W|} \sum_{w \in W} \sum_{v \in \mathbb{F}_2^n} \mathrm{cor}^2(v \cdot y + w \cdot E_1(y))$ to get

$$\begin{aligned}
&(|U| - 1) \Pr[U^\perp \setminus \{0\} \overset{E}{\rightarrow} W^\perp] \\
&\quad = \frac{1}{|W|} \sum_{v \in \mathbb{F}_2^n} \left(2|U| \Pr[U^\perp \overset{E_0}{\rightarrow} \mathrm{sp}(v)^\perp] - |U| - 1 \right) C_{v,W} \\
&\quad = \frac{1}{|W|} \sum_{v \in \mathbb{F}_2^n} \left(2(|U| \Pr[U^\perp \overset{E_0}{\rightarrow} \mathrm{sp}(v)^\perp] - 1) - |U| + 1 \right) C_{v,W} \\
&\quad = \frac{1}{|W|} \sum_{v \in \mathbb{F}_2^n} \left((|U| - 1)(2 \Pr[U^\perp \setminus \{0\} \overset{E_0}{\rightarrow} \mathrm{sp}(v)^\perp] - 1) \right) C_{v,W}.
\end{aligned}$$

We obtain the claim by dividing the first and last expressions in this chain of equalities by $|U| - 1$ and then observing that the term for $v = 0$ in the last expression is equal to $1/|W|$. □

6.2. *Bias Estimation*

In this section, we investigate under which conditions the multidimensional differential-linear hull given by Eq. (16) can be used to compute an estimate of the bias of a multidimensional differential-linear approximation.

The approach is similar to the one of Sect. 5. Given a set V, the sum in Eq. (16) can be decomposed into two sums:

$$\mathcal{E}_{U^\perp, W} = \frac{2}{|W|} \sum_{v \in V, v \neq 0} \varepsilon_{U^\perp, v} C_{v,W} + \frac{2}{|W|} \sum_{v \notin V} \varepsilon_{U^\perp, v} C_{v,W}. \tag{17}$$

Practical computation of the bias of a multidimensional differential-linear approximation relies on the fact that computing only the first partial sum gives us an underestimate of the absolute bias $|\mathcal{E}_{U^\perp, W}|$. This leads us to the following definition.

Definition 3. Given a multidimensional differential-linear approximation with input difference space $U^\perp$ and output mask space W, a subset V in the intermediate layer is called supporting if

$$\left| \frac{2}{|W|} \sum_{v \in V, v \neq 0} \varepsilon_{U^\perp, v} C_{v,W} \right| \leq \left| \mathcal{E}_{U^\perp, W} \right|.$$

It follows that if a subset is supporting for all $\delta \in U^\perp$ and $w \in W$ in the sense of Definition 2, then it is supporting for the multidimensional differential-linear approximation with input difference space $U^\perp$ and output mask space W. But the converse may not be true.

To compute an estimate of the bias more efficiently, we can break up the differential-linear approximations at the intermediate layer. Similarly as in Theorem 4, an exact equality is obtained under an additional assumption.

Corollary 3. *Let* $\mathcal{E}_{U^\perp,V} = \Pr[U^\perp \setminus \{0\} \overset{E_0}{\to} V^\perp] - \frac{1}{|V^{|}}$ *be the bias of a truncated differential with nonzero input differences in* $U^\perp$ *and output differences in* $V^\perp$*, where* V *is a linear subspace. Further, we denote by* $C_{V,W} = \sum_{w \in W, w \neq 0} \sum_{v \in V} c_{v,w}^2$ *the capacity of the multidimensional linear approximation.*
If then, for all $\Delta \notin V^\perp$ *the truncated differential probabilities* $\Pr[U^\perp \setminus \{0\} \overset{E_0}{\to} \Delta]$ *are equal, we can compute the first sum on the right side of Eq. (17) as*

$$\frac{2}{|W|} \sum_{v \in V} \varepsilon_{U^\perp,v} C_{v,W} = \frac{1}{|W|} \frac{|V|}{|V|-1} \varepsilon_{U^\perp,V} C_{V,W}.$$

To test the usefulness of the results presented in this section, we conducted similar experiments than the ones presented in Sect. 5.4 on SMALLPRESENT[8]. Conclusions of these experiments are similar to the ones in the one-dimensional case. We found that supporting subspaces exist. In the case of the PRESENT cipher, all subspaces V in the experiments satisfied

$$\left| \frac{2}{|W|} \sum_{v \in V} \varepsilon_{U^\perp,v} C_{v,W} \right| < \left| \mathcal{E}_{U^\perp,V} \right|.$$

Similarly as in the test cases described in Sect. 5.4, we observed that the accuracy of the computational result gets worse as the size of V increases. This behavior indicates that the assumption in Corollary 3 about the equality of the probabilities $\Pr[U^\perp \setminus \{0\} \overset{E_0}{\to} \Delta]$ for $\Delta \notin V^\perp$ does not hold.

6.3. *Data Complexity of Multidimensional Differential-Linear Attack*

When the differential-linear approximation is characterized by only one output mask w, as recalled in Lemma 1, the data complexity is inversely proportional to the square of the bias $\mathcal{E}_{\delta,w}$, meaning that larger its absolute value is, less costly the underlying distinguishing attack is.

When using multiple output masks, the differential-linear probability should be distinguishable from the uniform probability $1/|W|$ and the data complexity of the differential-linear distinguisher depends of the number $|W|$ of output masks. Then by repeating the derivations given in Sect. 3.2 with binomial probabilities $1/|W|$ and $1/|W| + \mathcal{E}_{U^\perp,W}$ yields to the following generalization of the number of data pairs when $\mathcal{E}_{U^\perp,W}$ is small compared to $1/|W|$.

Proposition 3. *Given a success probability* P_S *and an advantage* a*, the number of input data pairs* N_S *required in a multidimensional differential-linear attack with nonzero* *input differences in $U^{\perp}$, output masks in W, and bias $\mathcal{E}_{U^{\perp},W}$ as defined in Eq.* (15) *is*

$$N_S = \frac{[\Phi^{-1}(P_S) + \Phi^{-1}(1-2^{-a})]^2 \left(1/|W| - 1/|W|^2\right)}{\mathcal{E}^2_{U^{\perp},W}}. \tag{18}$$

Proof. Assuming that, compared to $1/|W|$, $\mathcal{E}_{U^{\perp},V}$ is small, we can safely use a normal approximation of the binomial distribution to estimate the data complexity of a differential-linear attack. Using the notation of [46], let us denote by X_0 and X_w two random variables that follow the standard normal distributions with means $\mu_0 = \mathcal{E}_{U^{\perp},V}$ and $\mu_w = 0$ and variances $\sigma_0^2 \approx \frac{1}{N_S}\left(\frac{1}{|W|} - \frac{1}{|W|^2}\right)$ and $\sigma_w^2 = \frac{1}{N_S}\left(\frac{1}{|W|} - \frac{1}{|W|^2}\right)$, respectively. The success probability P_S of the attack is estimated as

$$P_S = \Phi\left(\frac{\mathcal{E}_{U^{\perp},W}\sqrt{N_S}}{\sqrt{1/|W| - 1/|W|^2}} - \Phi^{-1}\left(1-2^{-a}\right)\right),$$

from where the claim follows. □

For attacks using multiple input differences, it is useful to sample the input data N_S in structures, which are sets of the form $x + Z$ where $x \in U$ and Z is a fixed subset of $U^{\perp}$. When structures are used, the total data complexity N^{DL} of a differential-linear attack is two times the required number of data pairs divided by the size $|Z|$ of the structure.

6.4. *Comparison with Truncated Differential and Multidimensional Linear Attack*

The relation between the truncated differential and multidimensional linear attacks was investigated in [13]. Let us briefly recall those results. Let U be a linear subspace of the input space and W a linear subspace of the output space of the cipher E. Then, the probability of the truncated differential over the cipher E with input differences in $U^{\perp}$ and output differences in $W^{\perp}$ is defined as

$$P_{U^{\perp},W^{\perp}} = \Pr[U^{\perp} \overset{E}{\to} W^{\perp}].$$

The capacity of the related multidimensional linear approximation is equal to

$$C_{U,W} = \sum_{u \in U\setminus\{0\}, w \in W} \operatorname{cor}^2(u \cdot x + w \cdot E(x)),$$

and we have the following relation [13]

$$P_{U^{\perp},W^{\perp}} = \frac{1}{|W|}(C_{U,W} + 1).$$

The multidimensional differential-linear approximation, including the classical one as its special case, with nonzero input differences in $U^{\perp}$ and output masks in $W^{\perp}$, is clearly a truncated differential, and we have the following relationship between its bias and the truncated differential probability

$$P_{U^\perp,W^\perp} = \frac{|U^\perp| - 1}{|U^\perp|}\mathcal{E}_{U^\perp,W} + \frac{1}{|W|} + \frac{|W| - 1}{|W||U^\perp|}.$$

We can see that when the size of $U^\perp$ increases we can estimate

$$\mathcal{E}_{U^\perp,W} \approx P_{U^\perp,W^\perp} - \frac{1}{|W|}, \tag{19}$$

that is, the effect of the zero input difference can be ignored.

Since the goal of [13] was to establish a link between the multidimensional linear and the truncated differential attack, the used statistical model of the truncated differential attack has been computed asymptotically for a success probability of 50 %. However, a more accurate derivation of the number of input data pairs N_S^{TD} can be obtained from Proposition 3, since in the case where the estimate given by Eq. (19) is valid, the input data pairs N_S and N_S^{TD} of a multidimensional differential-linear and truncated attacks satisfy $N_S = N_S^{TD}$.

Corollary 4. *We assume that the input difference space $U^\perp$ is of reasonable size to justify the estimate given by Eq.* (19)*. Given a success probability P_S and an advantage a, the number of input data pairs N_S^{TD} required in a truncated differential attack with input differences in $U^\perp$, output differences in $W^\perp$, and probability $P_{U^\perp,W^\perp}$ close to* $1/|W|$*, is*

$$N_S^{TD} = \frac{[\Phi^{-1}(P_S) + \Phi^{-1}(1 - 2^{-a})]^2(1/|W| - 1/|W|^2)}{\left(P_{U^\perp,W^\perp} - 1/|W|\right)^2}. \tag{20}$$

From this result, assuming that the data are organized into structures in the same way in a multidimensional differential-linear and a truncated differential attack, we deduce that the respective data complexities N^{DL} and N^{TD} are equal.

A multidimensional differential-linear approximation can be described also as a multidimensional linear approximation. If only one input difference is used, then this link remains of theoretical nature, as it does not yield any useful multidimensional linear attack due to the huge number of input masks needed to perform the known-plaintext attack in practice. But as the size of the input difference space $U^\perp$ increases also this link becomes practical.

For the actual key-recovery attack algorithms and their time and memory complexities, we refer to [13]. We only remark here that the size of a structure used in the sampling of input data varies depending on the number rounds added in the encryption side of the key-recovery attack. The maximum size of the structure is equal to $|U^\perp|$ when no round is added before E_0, but is larger than this value, otherwise. This fact may affect the data complexities of the attacks and should be taken into account when the complexities of the known-plaintext and chosen plaintext types of attacks are compared.

7. Conclusion

In this paper, we studied and generalized differential-linear cryptanalysis. Starting from the observation that any differential-linear relation can be regarded as a truncated differential, we derive a general expression of its bias based on the link between differential probabilities and linear correlations provided by Chabaud and Vaudenay.

The exact expression given in Theorem 2 is valid under the sole assumption of round independence. As a side-note, we demonstrated that, in general and independent of the differential-linear context, linear- and differential round independence are equivalent properties or assumptions.

We also revisit previous studies and applications of differential-linear cryptanalysis, where the bias of the differential-linear approximation has often been estimated under some heuristic assumptions, implicitly or explicitly present in the derivations. Starting from the exact expression of the bias under the assumption of round independence of the parts of the cipher, we identify new additional assumptions for computing efficient estimates of it. Extensive experiments have been performed to test the validity of these assumptions. We furthermore extracted assumptions (cf Theorem 3 and Corollary 2) under which the previous estimates as derived by Lu [39] and Biham et al. [5] are valid. This is potentially of practical relevance, as it shows which properties of the cipher remain to be validated experimentally.

Although no new applications of differential-linear cryptanalysis are presented in this paper, the potential and generality of our sound framework is demonstrated by its ability to explain existing examples of differential-linear cryptanalysis. Our generalization of differential-linear cryptanalysis to multidimensional differential-linear cryptanalysis presented in Sect. 6 could be used in future work to improve differential-linear attacks in the same manner as multiple differentials and multidimensional liner attacks improved upon differential and linear attacks, respectively.

Finally, we encourage further research to obtain better estimates for the differential-linear bias. As can be seen in Figs. 2 and 3, while the theoretical estimate is always an underestimate as desired, the actual bias is significantly larger then the theoretical one. At least in the case of SmallPresent-[8], this would result in a significant overestimation of the data complexity of the corresponding attack.

Acknowledgements

The authors wish to thank the anonymous reviewers of this and earlier versions of the paper for insightful comments that helped us to improve our work.

References

[1] R. Anderson, E. Biham, L.R. Knudsen. Serpent: A Proposal for the Advanced Encryption Standard. NIST AES Proposal, 1998

[2] T. Baignères, *Quantitative Security of Block Ciphers: Designs and Cryptanalysis Tools*. PhD thesis, École polytechnique fédérale de Lausanne, 2008

[3] E. Biham, A. Biryukov, A. Shamir, Cryptanalysis of Skipjack Reduced to 31 Rounds Using Impossible Differentials, in J. Stern, editor, *EUROCRYPT*. LNCS, vol. 1592 (Springer, 1999), pp. 12–23

[4] E. Biham, O. Dunkelman, N. Keller, The Rectangle Attack - Rectangling the Serpent, in B. Pfitzmann, editor, *EUROCRYPT*. LNCS, vol. 2045 (Springer, 2001), pp. 340–357

[5] E. Biham, O. Dunkelman, N. Keller, Enhancing Differential-Linear Cryptanalysis, in Y. Zheng, editor, *ASIACRYPT*. LNCS, vol. 2501 (Springer, 2002), pp. 254–266

[6] E. Biham, O. Dunkelman, N. Keller, Differential-Linear Cryptanalysis of Serpent, in T. Johansson, editor, *FSE*. LNCS, vol. 2887 (Springer, 2003), pp. 9–21

[7] E. Biham, O. Dunkelman, N. Keller, New Combined Attacks on Block Ciphers, in H. Handschuh, H. Gilbert, editors, *FSE*. LNCS, vol. 3557 (Springer, 2005), pp. 126–144

[8] E. Biham, A. Shamir, Differential Cryptanalysis of DES-like Cryptosystems, in A. Menezes, S.A. Vanstone, editors, *CRYPTO*. LNCS, vol. 537 (Springer, 1990), pp. 2–21

[9] A. Biryukov, C. De Cannière, M. Quisquater, On Multiple Linear Approximations, in M.K. Franklin, editor, *CRYPTO*. LNCS, vol. 3152 (Springer, 2004), pp. 1–22

[10] C. Blondeau, Improbable differential from impossible differential: On the validity of the model, in G. Paul, S. Vaudenay, editors, *INDOCRYPT 2013*. LNCS, vol. 8250 (Springer, 2013), pp. 149–160

[11] C. Blondeau, G. Leander, K. Nyberg, Differential-Linear Cryptanalysis Revisited, in C. Cid, C. Rechberger, editors, *FSE 2014*. LNCS, vol. 8540 (Springer, 2015), pp 411–430

[12] C. Blondeau, K. Nyberg, New Links between Differential and Linear Cryptanalysis, in T. Johansson, P.Q. Nguyen, editors, *EUROCRYPT*. LNCS, vol. 7881 (Springer, 2013), pp. 388–404

[13] C. Blondeau, K. Nyberg, Links Between Truncated Differential and Multidimensional Linear Properties of Block Ciphers and Underlying Attack Complexities, in E. Oswald, P.Q. Nguyen, editors, *Eurocrypt 2014*, vol. 8441 (Springer-Verlag, 2014)

[14] A. Bogdanov, C. Boura, V. Rijmen, M. Wang, L. Wen, J. Zhao, Key difference invariant bias in block ciphers, in K. Sako, P. Sarkar, editors, *ASIACRYPT (1)*. LNCS, vol. 8269 (Springer, 2013), pp. 357–376

[15] A. Bogdanov, L.R. Knudsen, G. Leander, C. Paar, A. Poschmann, M.J.B. Robshaw, Y. Seurin, C. Vikkelsoe, PRESENT: An Ultra-Lightweight Block Cipher, in P. Paillier, I. Verbauwhede, editors, *CHES*. LNCS, vol. 4727 (Springer, 2007), pp. 450–466

[16] A. Bogdanov, V. Rijmen, Zero-Correlation Linear Cryptanalysis of Block Ciphers. *IACR Cryptol. ePrint Arch.***2011**, 123 (2011)

[17] A. Bogdanov, V. Rijmen, Linear hulls with correlation zero and linear cryptanalysis of block ciphers. *Des. Codes Cryptogr.***70**(3), 369–383 (2014)

[18] C. Carlet, Boolean functions for cryptography and error correcting, in Y. Crama, P.L. Hammer, editors, *Boolean Models and Methods in Mathematics, Computer Science, and Engineering* (Cambridge University Press, Oxford, 2010), pp. 257–397

[19] F. Chabaud, S. Vaudenay, Links Between Differential and Linear Cryptanalysis, in A. De Santis, editor, *EUROCRYPT*. LNCS, vol. 950 (Springer, 1994), pp. 356–365

[20] J.Y. Cho, Linear Cryptanalysis of Reduced-Round PRESENT, in J. Pieprzyk, editor, *CT-RSA*. LNCS, vol 5985 (Springer, 2010), pp. 302–317

[21] N. Courtois, CTC2 and fast algebraic attacks on block ciphers revisited. *IACR Cryptol. ePrint Arch.***2007**, 152 (2007)

[22] J. Daemen, V. Rijmen, *The Design of Rijndael: AES—The Advanced Encryption Standard*. Springer, 2002.

[23] C. Dobraunig, M. Eichlseder, F. Mendel, M. Schläffer, Cryptanalysis of Ascon, in K. Nyberg, editor, *CT-RSA 2015*. LNCS, vol. 9048 (Springer, 2015), pp. 371–387

[24] O. Dunkelman, S. Indesteege, N. Keller, A Differential-Linear Attack on 12-Round Serpent, in D.R. Chowdhury, V. Rijmen, A. Das, editors, *INDOCRYPT*. LNCS, vol. 5365 (Springer, 2008), pp. 308–321

[25] O. Dunkelman, N. Keller, Cryptanalysis of CTC2, in M. Fischlin, editor, *CT-RSA 2009*. LNCS, vol. 5473 (Springer, 2009), pp. 226–239

[26] C. Guo, H. Zhang, D. Lin, Estimating Differential-Linear Distinguishers and Applications to CTC2, in J. Lopez, Y. Wu, editors, *ISPEC 2015*. LNCS, vol. 9065 (Springer, 2015), pp. 220–234

[27] M. Hermelin, J.Y. Cho, K. Nyberg, Multidimensional Extension of Matsui's Algorithm 2, in O. Dunkelman, editor, *FSE*. LNCS, vol. 5665 (Springer, 2009), pp. 209–227

[28] T. Huang, I. Tjuawinata, H. Wu, Differential-Linear Cryptanalysis of ICEPOLE, in G. Leander, editor, *FSE 2015*. LNCS, vol. 9054 (Springer, 2015), pp. 243–263

[29] J. Kelsey, T. Kohno, B. Schneier, Amplified Boomerang Attacks Against Reduced-Round MARS and Serpent, in B. Schneier, editor, *FSE*. LNCS, vol. 1978 (Springer, 2000), pp. 75–93

[30] L.R. Knudsen, Truncated and Higher Order Differentials, in B. Preneel, editor, *FSE*. LNCS, vol. 1008 (Springer, 1994), pp. 196–211

[31] L.R. Knudsen, DEAL—A 128-bit Block-Cipher. NIST AES Proposal, 1998.

[32] X. Lai, Higher Order Derivatives and Differential Cryptanalysis, in R.E. Blahut, D.J. Costello, U. Maurer, T. Mittelholzer, editors, *Communications and Cryptography*. The Springer International Series in Engineering and Computer Science, vol. 276 (Springer US, 1994), pp. 227–233

[33] X. Lai, J.L. Massey, S. Murphy, Markov Ciphers and Differential Cryptanalysis, in D.W. Davies, editor, *EUROCRYPT*. LNCS, vol. 547 (Springer, 1991), pp. 17–38

[34] S.K. Langford, *Differential-Linear Cryptanalysis and Threshold Signatures*. Ph.D. Thesis, 1995.

[35] S.K. Langford, M.E. Hellman, Differential-Linear Cryptanalysis, in Y. Desmedt, editor, *CRYPTO*. LNCS, vol. 839 (Springer, 1994), pp. 17–25

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

[37] G. Leurent, Improved Differential-Linear Cryptanalysis of 7-round Chaskey with Partitioning, in M. Fischlin, J.-S. Coron, editors, *EUROCRYPT 2016*. LNCS, vol. 9665 (Springer, 2016), pp. 344–371

[38] Z. Liu, D. Gu, J. Zhang, W. Li, Differential-Multiple Linear Cryptanalysis, in F. Bao, M. Yung, D. Lin, J. Jing, editors, *Inscrypt*. LNCS, vol. 6151 (Springer, 2009), pp. 35–49

[39] J. Lu, A Methodology for Differential-Linear Cryptanalysis and Its Applications - (Extended Abstract), in A. Canteaut, editor, *FSE*. LNCS, vol. 7549 (Springer, 2012), pp. 69–89

[40] J. Lu, A methodology for differential-linear cryptanalysis and its applications. *Des. Codes Cryptogr.*, 77(1), 11–48 (2015)

[41] M. Matsui, Linear Cryptanalysis Method for DES Cipher, In T. Helleseth, editor, *EUROCRYPT*. LNCS, vol. 765 (Springer, 1993), pp. 386–397

[42] J. McLaughlin, J.A. Clark, Filtered nonlinear cryptanalysis of reduced-round Serpent, and the Wrong-Key Randomization Hypothesis. Cryptology ePrint Archive, Report 2013/089, 2013

[43] P. Morawiecki, K. Gaj, E. Homsirikamol, K. Matusiewicz, J. Pieprzyk, M. Rogawski, M. Srebrny, M. Wojcik, Icepole: High-speed, hardware-oriented authenticated encryption. Cryptology ePrint Archive, Report 2014/266, 2014. http://eprint.iacr.org/

[44] P.H. Nguyen, H. Wu, H. Wang, Improving the Algorithm 2 in Multidimensional Linear Cryptanalysis, in U. Parampalli, P. Hawkes, editors, *ACISP*. LNCS, vol. 6812 (Springer, 2011), pp. 61–74

[45] K. Nyberg, Linear Approximation of Block Ciphers. In A. De Santis, editor, *EUROCRYPT*. LNCS, vol. 950 (Springer, 1994), pp. 439–444

[46] A.A. Selçuk, On Probability of Success in Linear and Differential Cryptanalysis. *J. Cryptol.*, 21(1), 131–147 (2008)

[47] B. Sun, Z. Liu, V. Rijmen, R. Li, L. Ceng, Q. Wang, H. AlKhzaimi, C. Li, Links Among Impossible Differential, Integral and Zero-Correlation Linear Cryptanalysis, in R. Gennaro, M. Robshaw, editors, *CRYPTO, Part I*. LNCS, vol. 9215 (Springer, 2015), pp. 95–115

[48] D. Wagner, The Boomerang Attack, in L.R. Knudsen, editor, *FSE*. LNCS, vol. 1636 (Springer, 1999), pp. 156–170

[49] D. Wagner, Towards a Unifying View of Block Cipher Cryptanalysis, in B.K. Roy, W. Meier, editors, *FSE*. LNCS, vol. 3017 (Springer, 2004), pp. 16–33