On The Distribution of Linear Biases: Three Instructive Examples*

Mohamed Ahmed Abdelraheem[1], Martin Ågren[2], Peter Beelen[1], and Gregor Leander[1]

[1] Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark
{M.A.Abdelraheem,P.Beelen,G.Leander}@mat.dtu.dk
[2] Dept. of Electrical and Information Technology, Lund University, P.O. Box 118, SE-221 00 Lund, Sweden
martin.agren@eit.lth.se

Abstract. Despite the fact that we evidently have very good block ciphers at hand today, some fundamental questions on their security are still unsolved. One such fundamental problem is to precisely assess the security of a given block cipher with respect to linear cryptanalysis. In by far most of the cases we have to make (clearly wrong) assumptions, e.g., assume independent round-keys. Besides being unsatisfactory from a scientific perspective, the lack of fundamental understanding might have an impact on the performance of the ciphers we use. As we do not understand the security sufficiently enough, we often tend to embed a security margin – from an efficiency perspective nothing else than wasted performance. The aim of this paper is to stimulate research on these foundations of block ciphers. We do this by presenting three examples of ciphers that behave differently to what is normally assumed. Thus, on the one hand these examples serve as counter examples to common beliefs and on the other hand serve as a guideline for future work.

1 Introduction

IT Security plays an increasingly crucial role in everyday life and business. When talking on a mobile phone, when withdrawing money from an ATM or when buying goods over the internet, security plays a crucial role in both protecting the user and in maintaining public confidence in the system. Moreover, security techniques are often an enabler for innovative business models, e.g., iTunes and the Amazon Kindle require strong copyright protection mechanisms, or after-sale feature activation in modern cars. Virtually all modern security solutions are based on cryptographic primitives. Block ciphers are arguably the most widely used type of these primitives.

State-of-the-Art of Block Ciphers While great progress has been made in designing and analyzing block ciphers, fundamental aspects of these ciphers are still not understood.

* The full version of this paper is available at ePrint.

Besides being unsatisfactory from a scientific perspective, the lack of fundamental understanding might have consequence on the performance of the ciphers we use. As we do not understand the security sufficiently, we tend to embed a security margin in the ciphers we are using. From an efficiency perspective, a security margin is nothing else than wasted performance. While for some applications this might not be critical, for others it certainly is. In particular, when it comes to cryptography in the emerging field of pervasive computing, the computational resources of the devices in question are often highly constrained, and we can only allow close to zero overhead for a security margin.

Especially for the key-scheduling algorithm, the fundamental part of a cipher that is responsible for generating key-material from a master-key to be used at several places in the algorithm (see Figure 1 for an example), simplifying assumptions are standard. While these assumptions are strictly speaking wrong, the hope is that the behaviour of the real cipher does not differ significantly from the simplified model.

Linear Cryptanalysis One of the best known and most general applicable attacks on block ciphers is Matsui's linear attack [15]. Since its introduction many extensions and improvements have been made, and we mention a selection here. A more precise estimate for the success probability and the data complexity are given by Selçuk [21]. The effect of using more than one linear trail, referred to as linear hulls, has been introduced by Nyberg [17]; see also Daemen and Rijmen [8]. This has been used for example by Cho [6]. Multi-dimensional linear attacks have been studied by Hermelin, Cho, and Nyberg [10] as a way to further reduce the data complexity of the basic attack. We also like to mention the critics on the concept of linear hulls by Murphy [16]; see Leander [13] for a further discussion. One of the most recent developments is the idea to make use of unbiased linear approximations by Bogdanov et al. [4].

However, despite its discovery more than 15 years ago, and the many extensions introduced since then some very fundamental properties are not yet well understood.

In a nutshell, for claiming a cipher secure against linear attacks, one has to demonstrate that the cipher does not possess certain statistical irregularities. In almost all cases the best we can do is to bound the correlation of a single linear trail (see [20] for an exception), as this roughly corresponds to bounding the number of active Sboxes. Thus, using for example the well established wide-trail strategy used in AES [8], obtaining strong bounds on the correlation of a single trail is quite easy nowadays. However, when it comes to bounding the correlation of a linear approximation, or linear hull to emphasize its relation to many linear trails, no general convincing arguments are available. More precisely, the task of understanding the distribution of linear correlations over the keys is unsolved.

In order to be able to do so, it is in by far most of the cases necessary to assume that all (round) keys are independently and uniformly chosen or make the even stronger (and clearly wrong) assumption that distinct linear trails are independent.

While independent round-keys are hardly ever used in any real cipher, this assumption is on the one hand needed to make the analysis feasible and on the other hand often does not seem problematic as even with the keys not independently and uniformly chosen, most ciphers (experimentally) do not behave different from the expectation.

However, those experimental confirmations that the cipher behaves as assumed are inherently insufficient. For a 128 bit block cipher a single linear approximation that for a fraction of 2^{-30} of all keys has a correlation greater than 2^{-30} is something that, on the one hand, we clearly want to avoid but, on the other hand, we will never discover by experiments only.

Thus, it is important to really understand the distribution of bias, where the distribution is taken over all possible keys. Only by studying the entire distribution can weak keys possibly be identified (this has been pointed out previously, cf. for example [8]). Promising results along these lines include the work of Daemen and Rijmen [9] where the problem is clearly stated and attempts are made to give general statements. Unfortunately, as we will discuss below, one of the most general theorems is strictly speaking wrong.

Our Contribution The aim of this paper is to stimulate research on the foundations of block ciphers. We do this by presenting three examples of ciphers that behave differently to what is normally assumed. Thus, on the one hand these examples serve as counter examples to common beliefs and on the other hand serve as a guideline for future work. The value of our examples as guidelines for future work is specific for each example. The first example mainly limits the most general statements one can hope to prove, and in particular is a counter example to Theorem 22 in [9] where under rather natural conditions it was stated that the distribution of correlations is well approximated by a normal approximation and in particular one should expect many different possible values for the bias. The second example considers the influence of the key scheduling on the distribution of correlations. Here the variance (but not the shape) of the distribution significantly depends on the key-scheduling. For future work this suggests that highly non-linear key scheduling algorithms are superior to linear ones with respect to the distribution of correlations. Here highly non-linear has to be understood not as a vague criteria but in terms of minimizing the absolute values for all linear approximations. The last example is again related to key-scheduling but more so to symmetries in ciphers. We show a general equivalence of symmetries and linear approximations for weak-keys that exist for any number of rounds and illustrate this fact with an example.

The techniques used to analyze these examples are surprisingly diverse and of independent interest.

In order to facilitate the understanding of our examples without diving into too many details we give only an overview of the results of the first and the third example in Section 3. The details for the first example are postponed to Section 4 and for the third example to Section 5.

2 Notation and Preliminaries

Given an n bit function $F : \mathbb{F}_2^n \to \mathbb{F}_2^n$, a *linear approximation* is an equation of the form

$$\langle \alpha, x \rangle + \langle \beta, F(x) \rangle = 0,$$

where $\langle \cdot, \cdot \rangle$ denotes an inner product. The vector α is called the input mask and β is the output mask. The *bias* $\epsilon_F(\alpha, \beta) \in [-1/2, 1/2]$ of a linear approximation is defined as

$$\text{Prob}(\langle \alpha, x \rangle + \langle \beta, F(x) \rangle = 0) = \frac{1}{2} + \epsilon_F(\alpha, \beta),$$

where the probability is taken over all inputs x. The bias is the value of major importance for linear attacks. However, due to scaling reasons, it is much more convenient to work with the *correlation* $c_F(\alpha, \beta) \in [-1, 1]$ defined by

$$c_F(\alpha, \beta) = 2\epsilon_F(\alpha, \beta).$$

Another measure that we are going to use is the Fourier-transformation of F,

$$\widehat{F}(\alpha, \beta) = \sum_{x \in \mathbb{F}_2^n} (-1)^{\langle \beta, F(x) \rangle + \langle \alpha, x \rangle}.$$

Up to scaling $\widehat{F}(\alpha, \beta)$ is equivalent to the bias $\epsilon_F(\alpha, \beta)$ and the correlation $c_F(\alpha, \beta)$. More precisely,

$$\epsilon_F(\alpha, \beta) = \frac{c_F(\alpha, \beta)}{2} = \frac{\widehat{F}(\alpha, \beta)}{2^{n+1}}. \tag{1}$$

Linear Trails and Linear Hull Given a composite function F, i.e., $F_i : \mathbb{F}_2^n \to \mathbb{F}_2^n$ such that $F = F_r \circ \cdots \circ F_2 \circ F_1$, a *linear trail* θ is a collection of all intermediate masks

$$\theta = (\theta_0 = \alpha, \ldots, \theta_r = \beta)$$

and its correlation is defined by

$$C_\theta = \prod_i c_{F_i}(\theta_i, \theta_{i+1}).$$

It is well known, see e.g., [8], that the correlation of a linear approximation is the sum of all correlations of linear trails starting with the same mask α and ending with the same mask β, i.e.,

$$c_F(\alpha, \beta) = \sum_{\theta \mid \theta_0 = \alpha, \theta_r = \beta} C_\theta. \tag{2}$$

In this paper we are concerned with keyed permutations, more precisely with key-alternating ciphers as defined for example in [8, Section 2.4.2] and depicted

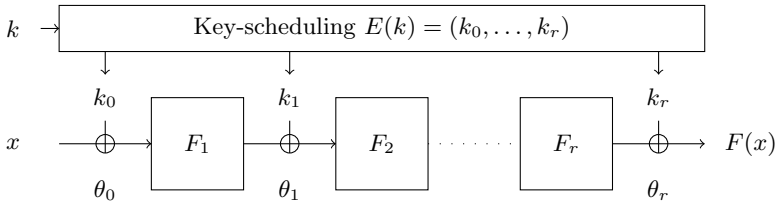


Fig. 1. A key-alternating cipher

in Fig. 1. An n bit key-alternating cipher with a k bit (master) key consists of round functions $F_i : \mathbb{F}_2^n \to \mathbb{F}_2^n$ and a key-scheduling algorithm $E : \mathbb{F}_2^k \to \mathbb{F}_2^{n(r+1)}$.

The dependence of the correlation of a linear trail is conceptually very simple, only the sign of the correlation depends on the key. More precisely,

$$C_\theta = (-1)^{\langle \theta, E(k) \rangle} \prod_i c_{F_i}(\theta_i, \theta_{i+1}).$$

Plugging this into Equation (2) leads to the following result.

$$c_F(\alpha, \beta) = \sum_{\theta \ \mid \ \theta_0 = \alpha, \theta_r = \beta} (-1)^{\langle \theta, E(k) \rangle + \sigma_\theta} |C_\theta| \tag{3}$$

It is exactly this equation that is often referred to as the *linear hull*: The correlation of a linear approximation is the key-dependent signed sum of the correlation of all trails.

In this work we are interested in the distribution over the keys of the linear correlation. Here one can think of each linear trail C_θ as a random variable with a fixed absolute value that with probability $1/2$ is positive and with probability $1/2$ is negative. In this setting the linear hull is the sum of those random variables.

While in general not a lot is known about this distribution, two important characteristics can be stated, assuming independent round-keys. First, as the average of a sum of random variables is the sum of the averages, the average bias is zero. Here, independent round-keys are used to ensure that each single trail has average zero. Moreover, it is easy to see that two distinct linear trails C_θ and C'_θ are pairwise independent. It follows (cf. Theorem 7.9.1 in [8]) that with independent round-keys the variance of the distribution, i.e., the average square correlation, is the sum of the squares of the correlations of all trails. We summarize this in the following proposition.

Proposition 1. *Assuming independent round-keys, i.e., $k = n(r+1)$ and $E(k) = k$, the average correlation is zero, i.e.,*

$$\mathbb{E}(C_\theta) = 0.$$

Moreover, the average square correlation is given by

$$\mathbb{E}(C_\theta^2) = \sum_i c_{F_i}(\theta_i, \theta_{i+1})^2.$$

Finally, we already note here an observation that we will make use of later.

Lemma 1. *If the key-scheduling $E : \mathbb{F}_2^k \to \mathbb{F}_2^{n(r+1)}$ is linear then two distinct linear trails C_θ and C'_θ are either independent or $C_\theta = \pm c \cdot C'_\theta$ for a constant c. More precisely C_θ and C'_θ are independent if and only if $E^*(\theta + \theta') \neq 0$ where E^* is the adjoint linear mapping.*

Proof. The lemma follows directly from the observation that

$$\langle \theta, E(k) \rangle + \langle \theta', E(k) \rangle = \langle E^*(\theta), k \rangle + \langle E^*(\theta'), k \rangle = \langle E^*(\theta + \theta'), k \rangle$$

and the general remark that a linear function $\ell(\cdot) = \langle a, \cdot \rangle$ is either balanced (if $a \neq 0$) or constant (if $a = 0$). Thus two trails are independent if and only if $E^*(\theta + \theta') \neq 0$. □

3 Our Results

In this section we briefly describe our examples, the results and their interpretation. As the first and the last example require more technical details for a full explanation, the exact analysis of those results are given in later sections.

3.1 Example I: The CUBE-Cipher

As a first example, we study the two round key-alternating cipher depicted in Fig. 2 with block size n, n odd. The round function $x \to x^3$ has to be understood as a mapping on the finite field $\mathbb{F}_{2^n}$ with 2^n elements (as n is odd this is a bijection). The key consists of three independent subkeys k_0, k_1, k_3 each of n bits. Obviously, and for various reasons [12, Section 8.4], this is an artificial example of a block cipher. However, for the purpose of this counterexample that does not matter – it is a counterexample anyway. Moreover, as this cipher clearly belongs to the class of key-alternating ciphers, general theorems on these have to either explicitly exclude this (and similar) examples or the statements have to cover this strange behavior as well.

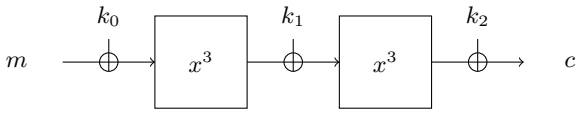


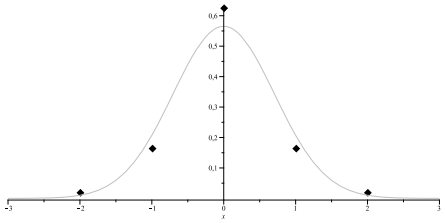
Fig. 2. The CUBE-cipher

As we will prove in Theorem 2, the number of trails is very large. Namely roughly 2^{n-2} trails with non-zero correlation exist. Furthermore, all trails have

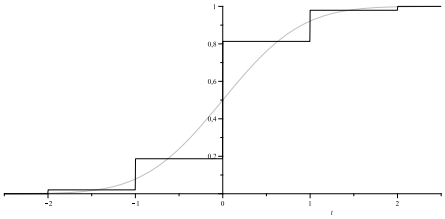
the same absolute correlation. This seems the ideal situation (we even have a parameter, n, that could go to infinity) for assuming that the distribution of correlations over the keys is very well approximated by a normal distribution, cf. Theorem 22 in [9].

Theorem 1 (Theorem 22 of [9]). *Given a key-alternating cipher with independent round-keys. If the number of linear trails with non-zero correlation is large and the square of the correlation of each linear trail is small compared to the variance of the distribution then the distribution is well approximated by a normal distribution.*

The intuition why a normal distribution should be a good approximation is that in this case the linear hull is the sum of a huge amount of pairwise(!) independent random variables. However, it turns out to be wrong. In particular, for any n, the correlation of the cipher actually takes only 5 different values. Thus, the roughly 2^{n-2} random variables are not independent at all. As an example of the real distribution compared to the assumed normal distribution we consider the case $n = 31$ (other cases behave very similarly). Figs. 3(a) and 3(b) show both distributions and make clear that the normal approximation is not a very good approximation and in particular does not get substantially better when n increases. In Section 4 we prove that in general only 5 values are obtained and we furthermore study the exact distribution of those 5 values.



(a) The (normalized) distribution of the CUBE-cipher vs the normal distribution



(b) The (normalized) cumulative probability distribution of the CUBE-cipher vs the normal distribution

Fig. 3. (Normalized) distributions of the CUBE-cipher vs the normal distribution

3.2 Example II: PRESENT with identical round-keys

Our second example is related to the block cipher PRESENT by Bogdanov et al., see [3] for details. As was previously shown, e.g., by Ohkuma [19], for an increasing number of rounds PRESENT exhibits many linear trails with only one active Sbox per round. Due to the design criteria of the Sbox, it follows that all those trails have the same linear bias. Besides, those trails are the ones with a maximal correlation (in absolute terms).

It was experimentally confirmed in [19] that the distribution of the correlation nicely follows a normal-distribution with mean zero and variance $2^{(-2r-1)^2}N$ where N is the number of the linear trails with only one active Sbox per rounds. Thus, experimentally, we can notice two facts: Firstly, for PRESENT different trails behave like independent random variables (in contrast to the CUBE-cipher) and secondly, the contribution of the non-optimal trails does not influence the distribution significantly.

Let us now come to a variant of PRESENT with identical round-keys[1] (and round-constants to avoid trivial slide attacks [1]). As it turns out this is an intriguing example of the influence of the key-scheduling on the distribution of the correlations. We started by performing experiments on a large number of random keys and observed that the total variance of the bias distribution for some linear approximations of PRESENT with identical round-keys is consistently bigger than that of standard PRESENT for any number of rounds ≥ 5. Fig. 4 shows the distribution of the linear correlation for the identical round-keys case vs. the original PRESENT key-scheduling for 17 rounds.

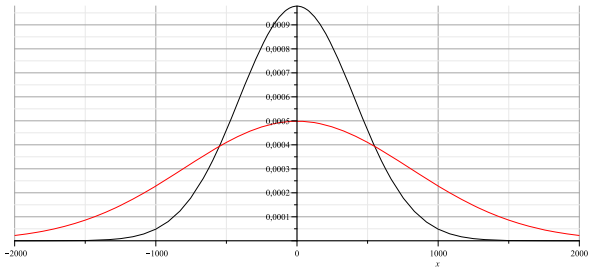


Fig. 4. The (normalized) probability distribution of the PRESENT-cipher with the usual vs the identical round-keys case

The difference is significant in the sense that more rounds of PRESENT with identical round-keys are vulnerable to linear attacks for a non-negligible fraction of keys. In other words, in this example it is indeed the choice of the key-scheduling that makes the cipher secure or insecure against linear cryptanalysis.

[1] Note that identical round-keys have been used before, see for example the block cipher NOEKEON [7]

To illustrate the difference consider a 20 round version. The fraction of keys with a squared bias larger than 2^{-55} is 33.7% in the case of identical round-keys but only 1.1% in the case of the standard PRESENT key-scheduling.

For computing the variance of a sum of random variables it is sufficient to study the pairwise covariance of the summands. Now, as mentioned above in Lemma 1 for a linear key-scheduling algorithm there are only two possibilities for the covariance. Either two trails are independent or identical up to a constant factor. In our particular case this constant is either 1 or -1 as all trails we consider have the same absolute correlation. Note furthermore that, following Lemma 1, two trails C_θ and C'_θ are identical (up to a constant ± 1) iff $E^*(\theta+\theta') = 0$ where E denotes the key-scheduling function. For identical round-keys, we have that $E(k) = (k, \ldots, k)$ and thus

$$E^*(\theta) = E^*(\theta_0, \ldots, \theta_r) = \sum_{i=0}^{r} \theta_i.$$

In other words two trails are identical iff the (xor) sums of all intermediate masks are identical. While in general, computing the number of trails is much more efficient than listing all trails, it is still feasible for $r \leq 20$ to compute the list of trails and sort this list according to the sum of the intermediate masks. Thus, for $r \leq 20$ we can relatively easily compute the expected variance for the PRESENT-variant with identical round-keys. Table 1 shows the expected variance (Var_2) of the bias distribution of the optimal linear approximation for a specific one bit input and output difference. For PRESENT with identical round-keys the expected variance is very close to the observed variance (ObsVar) sampled over 20000 random keys. Table 1 also shows the expected variance of the bias distribution of the optimal linear approximation of (standard) PRESENT (Var_1) along with the number of trails (N_1) with one active Sbox per round, and the number of trails (N_2) where the sign depends on the keybits, for number of rounds r, $15 \leq r \leq 20$.

Table 1. Analytical and experimental data on r-round reduced PRESENT (possibly with identical round-keys). N_1 is the number of all linear trials with one active Sbox. N_2 is the number of trails (among N_1) that don't behave the same (their absolute values are equal but the correlation sign is different and it changes according to the subkeys). Var_1 gives the expected variance of the bias of the optimal linear approximation of (standard) PRESENT, while Var_2 corresponds to PRESENT with identical round-keys. ObsVar is the experimentally observed variance sampled over 20000 random keys.

r	N_1	N_2	$\log_2(\mathrm{Var}_1)$	$\log_2(\mathrm{Var}_2)$	$\log_2(\mathrm{ObsVar})$
15	166375	12016	−44.66	−42.71	−42.71
16	435600	20039	−47.26	−45.15	−45.28
17	1140480	31799	−49.88	−47.63	−47.61
18	2985984	48223	−52.49	−50.03	−50.12
19	7817472	69528	−55.10	−52.50	−52.52
20	20466576	95125	−57.71	−54.88	−54.92

It is important to note that, while the CUBE-cipher is certainly an artificial design, PRESENT with identical round-keys is not. In this context we like to mention that it is precisely the behavior described here that resulted in the need to choose a different Sbox in the PRESENT-inspired sponge-based hash-function SPONGENT by Bogdanov et al. [2]. SPONGENT can be seen as a fixed key and large block size variant of PRESENT with identical round-keys.

3.3 Example III: PRINTCIPHER or Block Ciphers with Symmetries

It was already pointed out by Leander et al. [14] that for PRINTCIPHER-48 [11] by Knudsen et al. there exist strongly biased linear relations for any number of rounds. More precisely,

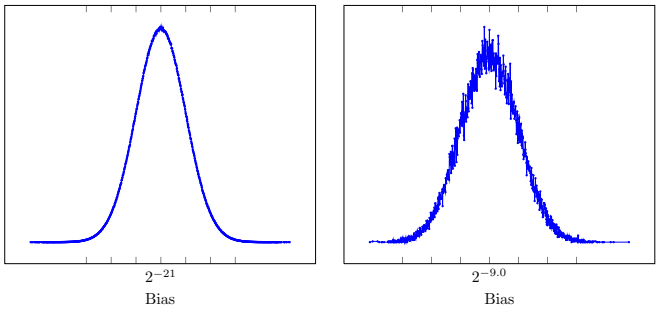
Proposition 2 (Corollary 2 in [14]). *For a fraction of* 2^{-28} *of all keys and for any round* $r \leq 48$ *there exists at least one linear approximation for* PRINTCIPHER-*48 with correlation at least* $2^{-16} - 2^{-32}$.

Here (cf. Section 5), we extend upon this analysis by showing an equivalence between a submatrix A of the correlation matrix that has an eigenvector with eigenvalue one and a round function that has an invariant subspace. This is crucial as this implies that this sub-matrix A, when taken to the r-th power, does not converge to the all-zero matrix. In particular in the case where there is a unique eigenvector with eigenvalue of norm 1, A^r converges to a non-zero constant. *This is equivalent to saying that trails with all intermediate masks determined by the invariant subspace cluster significantly for any number of rounds.*

Note that an invariant subspace in particular captures, as a special case, what is usually referred to as symmetries. Thus, besides PRINTCIPHER one could also imagine an identical-round-key variant of AES with round constants that do not destroy the symmetries introduced by the very structured and byte oriented linear layer of AES. This example reveals two interesting points. First, in such a situation of trail clustering, increasing the number of rounds *does not* help and secondly without the link to invariant subspaces it seems very hard to understand why certain trails should cluster even for an AES-like design that follows the wide-trail strategy. Moreover this clustering is not inherently limited to ciphers with weak mixing (e.g., PRINTCIPHER), but is rather a problem for all ciphers exhibiting symmetries.

Interestingly, in a restricted sense to be discussed in Section 5, the reciprocal statement holds as well. That is, if the cipher does not exhibit symmetries, then no sub-matrices (of a certain type) have eigenvectors (of a certain type) with eigenvalue 1. Thus by avoiding symmetries one also ensures that trail clustering for any number of rounds is highly unlikely.

Fig. 5 shows the difference of the distribution of the correlation for non-weak keys vs. the distribution for weak keys for a 24-bit version of PRINTCIPHER. Both graphs can be nicely approximated by normal distributions, however, the mean of the distribution for weak keys differ significantly from the origin.



(a) The non-weak xor keys (b) The weak xor keys

Fig. 5. The distribution of PRINTCIPHER-24 biases for a fixed permutation key. The experimentally observed means m are indicated. In both distributions, the standard deviation σ is approximately $2^{-13.0}$. Ticks have been placed at $m+k\sigma$, $k \in \{-3, \ldots, 3\}$.

4 Example I: The CUBE-Cipher

In this section we give a detailed analysis of the distribution of the correlations in the CUBE-cipher.[2]

We denote the block size by n, where in this example n has to be odd. First note that the initial and the final key-addition do not change the distribution. Thus, to simplify notation, we ignore them from now on. We therefore have to consider the function $F_k(x) = (x^3 + k)^3$. Moreover, to make the analysis easier, we focus on the input and output mask $1 \in \mathbb{F}_{2^n}$. That is, we are interested in the distribution of $\widehat{F}_k(1,1)$ for varying key k.

As a first step we show that the corresponding linear hull contains a very large number of trails with non-zero correlations. More precisely, the following holds (cf. the full version for the proof).

Theorem 2. *The number t of trails with non-zero correlation of the form*

$$1 \xrightarrow{x^3} \alpha \xrightarrow{x^3} 1$$

is

$$t = \frac{2^n + 1 - (a_1^n + a_2^n + a_3^n + a_4^n)}{4},$$

where a_i are the four (complex) roots of the polynomial $x^4 + x^3 + 2x + 4$. Furthermore, each trail has a correlation of $\pm 2^{1-n}$.

The next proposition shows that, despite the huge number of non-zero trails, only 5 values occur for the correlations.

Proposition 3. $\widehat{F}_k(1,1) \in \{0, \pm 2^{\frac{n+1}{2}}, \pm 2^{\frac{n+3}{2}}\}$

[2] Due to page limitations most of the proofs are given in the full version.

Proof. We denote by $\mu(x) = (-1)^{\mathrm{Tr}(x)}$, where $\mathrm{Tr}(x) = x + x^2 + x^4 + \ldots x^{2^{n-1}}$ is the trace mapping and note that $\mathrm{Tr}(xy)$ is the natural inner product on $\mathbb{F}_{2^n}$.

$$\begin{aligned}
\widehat{F}_k(1,1)^2 &= \sum_{x,y} \mu\left((x^3+k)^3 + (y^3+k)^3 + x + y\right) \\
&= \sum_{x,y} \mu\left(((x+y)^3+k)^3 + (y^3+k)^3 + x\right) \\
&= \sum_{x} \mu\left((x^3+k)^3 + x + k^3\right) \sum_{y} \mu\left(\left(x^{64} + (k^{16}+k^4)x^{16} + (k^8+k^2)x^4 + x\right) y^8\right) \\
&= 2^n \sum_{x \in M} \mu\left((x^3+k)^3 + x + k^3\right)
\end{aligned}$$

where

$$M = \{x \mid x^{64} + (k^{16}+k^4)x^{16} + (k^8+k^2)x^4 + x = 0\}.$$

Thus, we have to understand the kernel (and in particular its size) of the $\mathbb{F}_2$-linear mapping

$$P(x) = x^{64} + (k^{16}+k^4)x^{16} + (k^8+k^2)x^4 + x.$$

As a polynomial, P splits nicely into factors, i.e.,

$$P(x) = A_1(x^3)A_2(x^3)A_3(x^3)A_4(x^3)x,$$

with

$$\begin{aligned}
A_1(x) &= x^3 + k^2x + 1 \qquad & A_3(x) &= x^6 + x^4 + (k^4+k^2+1)x^2 + x + 1 \\
A_2(x) &= x^3 + (k^2+1)x + 1 \qquad & A_4(x) &= x^9 + x^3 + (k^8+k^2)x + 1.
\end{aligned}$$

For now, we show that $A_3(x)$ does not have any roots. Note that this is actually enough to prove the proposition, as this upper bounds the number of elements in M by $16 = 9+3+3+1$ and for general reasons we know that $|M| = 2^i$ with i odd. Thus $|M| \in \{2, 8\}$. Assume that $A_3(x) = 0$. Then

$$k^4 + k^2 + 1 = \frac{x^6 + x^4 + x + 1}{x^2} = x^4 + x^2 + \frac{1}{x} + \frac{1}{x^2}.$$

Applying the trace mapping to both sides implies $\mathrm{Tr}(1) = 0$.[3] A contradiction, as n is odd. □

This already demonstrates an unexpected behavior. The following theorem, proven in the full version of the paper, allows us to compute the exact distribution of the 5 occurring values for reasonably large n.

Theorem 3. *We have*

$$\#\{k \in \mathbb{F}_{2^n} \mid \widehat{F}_k(1,1)^2 = 2^{n+3}\} = \frac{1}{3}\#\{\beta \in \mathbb{F}_{2^n} \mid \beta \textit{ satisfies Eqns. (4)}\},$$

[3] Note that $\mathrm{Tr}(x) = \mathrm{Tr}(x^2)$ for all $x \in \mathbb{F}_{2^n}$.

where

$$\mathrm{Tr}\left(\left(\frac{1}{(\beta^2+\beta)}\right)^{1/9}\right) = \mathrm{Tr}\left(\left(\frac{\beta^3}{\beta^2+\beta}\right)^{1/9}\right) = \mathrm{Tr}\left(\left(\frac{(1+\beta)^3}{\beta^2+\beta}\right)^{1/9}\right) = 1. \quad (4)$$

The advantage of the above theorem is that it gives a fast way to compute the number A of $k \in \mathbb{F}_{2^n}$ such that $\widehat{F}_k(1,1) = 2^{n+3}$. Let us further denote by B the number of $k \in \mathbb{F}_{2^n}$ such that $\widehat{F}_k(1,1) = 2^{n+1}$. Then clearly

$$\sum_k \widehat{F}_k(1,1)^2 = A2^{n+3} + B2^{n+1}.$$

On the other hand, since knowing the number of trails with non-zero correlation together with their correlation values, corresponds to knowing the average square correlation, we have

$$\sum_k \widehat{F}_k(1,1)^2 = 2^{-n} \sum_\alpha \widehat{C}(1,\alpha)^2 \widehat{C}(\alpha,1)^2.$$

Using the above and Theorem 2 we see that

$$A2^{n+3} + B2^{n+1} = \sum_k \widehat{F}_k(1,1)^2 = 2^n(2^n + 1 - a_1^n - a_2^n - a_3^n - a_4^n),$$

where, as in Theorem 2, a_i are the four (complex) roots of the polynomial $x^4 + x^3 + 2x + 4$. Thus using Proposition 3 and the symmetry of the distribution (which can easily be proven in this case) one can now obtain the complete distribution for how many k, $F_k(1,1)$ obtains a particular value in $\{\pm 2^{(n+3)/2}, \pm 2^{(n+1)/2}, 0\}$ for reasonably large values of n. We give some examples below.

n	$-2^{(n+3)/2}$	$-2^{(n+1)/2}$	0	$2^{(n+1)/2}$	$2^{(n+3)/2}$
1	0	1	0	1	0
3	1	0	6	0	1
5	0	6	20	6	0
9	10	90	312	90	10
19	10868	87078	328396	87078	10868
31	44732008	357939982	1342139668	357939982	44732008

5 Example III: PRINTcipher or Block Ciphers with Symmetries

The results in this section, especially as summarized in Theorem 5, are quite general, applying to any block cipher (permutation) exhibiting these kinds of symmetries. For this reason, we do not describe PRINTcipher in detail here, but refer to the full version or [11] instead. We only note here that the round-key is the same in every round (there is a small round constant).

PRINTcipher is used for experiments in Section 3.3 and below. Smaller-state versions are not formally defined but are easy to extrapolate from [11]. We use the same round constants as in the first rounds of PRINTcipher-48. A PRINTcipher-key can be split into a permutation key and an xor key.

The Invariant Subspace Let us define a subspace $U \subset \mathbb{F}_2^n$, the orthogonal subspace $U^\perp = \{y : \langle x, y\rangle = 0, \forall x \in U^\perp\}$ and a constant $d \in \mathbb{F}_2^n$. Then, the invariant subspace property (cf. [14]) can be expressed as $F_i(U + d) = U + d$. In the case of PRINTCIPHER, the exact definitions of U, $U^\perp$, and d can be found in the full version of the paper. We only note that for PRINTCIPHER, $|U + d| = 2^{16}$, and the trails do not involve the round constants so the invariant subspace extends to the entire cipher F, regardless of the number of rounds. However, even if an invariant subspace only occurs for some rounds of the cipher, it can certainly be interesting in linear cryptanalysis as seen below.

Understanding the Large Correlations The correlation matrix (cf. [8]) $M_i = (c_{F_i}(\alpha, \beta))_{\alpha\beta}$ collects all correlation coefficients for a single round. We are interested in the submatrix $A = (a_{\alpha\beta})_{\alpha,\beta\in U^\perp}$ constructed through $a_{\alpha\beta} = c_{F_i}(\alpha, \beta)$ and its powers A^r. Thus A collects the correlations where input and output masks only involve the bits that govern the invariant subspace. In any correlation matrix, the first row and column are all-zero except for $c(0, 0) = 1$. We extract the sub-matrix $B = (a_{\alpha\beta})_{\alpha,\beta\in U^\perp\setminus\{0\}}$, since it will be slightly more convenient to use.

We should identify A_i with F_i, but the round constants do not affect A_i, so all A_i are equal. In particular $A_r A_{r-1} \dots A_1 = A^r$. Note how A^r describes the contribution to the linear hull from following trails with intermediate masks in $U^\perp$. More specifically, we can write Equation (3) as

$$c_F(\alpha, \beta) = \sum_{\substack{\theta \;\mid\; \theta_0=\alpha, \theta_r=\beta, \\ \theta_i\in U^\perp,\ \forall i}} (-1)^{\langle\theta, E(k)\rangle+\sigma_\theta}|C_\theta| + \sum_{\substack{\theta \;\mid\; \theta_0=\alpha, \theta_r=\beta, \\ \exists i:\ \theta_i\notin U^\perp}} (-1)^{\langle\theta, E(k)\rangle+\sigma_\theta}|C_\theta|,$$

where the first sum corresponds to element (α, β) of A^r. If elements of A^r have a large magnitude, then the corresponding elements of M^r have (at least) the same magnitude, unless the contributions from trails that go outside $U^\perp$ (essentially) cancel those from inside.

We now examine the asymptotic behaviour of A^r. Define $v = (v_\alpha)_{\alpha\in U^\perp}$ by $v_\alpha = (-1)^{\langle d,\alpha\rangle}$.

Lemma 2. *v^T is an eigenvector to A with eigenvalue 1, i.e., $Av^T = v^T$.*

We prove this lemma in the full version of the paper.

Now, in the case where there is no other (non-trivial) eigenvector with eigenvalue 1 the sequence A^r will converge (see the theorem below). This motivates the following definition.

Definition 1. *If the algebraic multiplicity of A's eigenvalue 1 is two and A has no other eigenvalue of absolute value 1, we say that A (or the corresponding cipher) has a* stable symmetry. *(The eigenvectors are that given in Lemma 2, and the vector $(1, 0, 0, \dots, 0)^T$.)*

For the following theorem, we use that A has eigenvalues with absolute value at most 1, the Schur decomposition of A and the relation between convergence of A^r and the spectrum of A [5].

Theorem 4. *If A has a stable symmetry then $B^r \to C = \frac{1}{2^{\dim(U)}-1} u^T u$, $r \to \infty$, $u = (v_\alpha)_{\alpha \in U^\perp \setminus \{0\}}$.*

If other contributions to $c_F(\alpha, \beta)$ are negligible, then all characteristics with non-zero $\alpha, \beta \in U^\perp$ have $c_F(\alpha, \beta) \approx \pm 2^{-\dim(U)}$ so bias $\epsilon_F(\alpha, \beta) \approx \pm 2^{-\dim(U)-1}$.

Equivalence Between Eigenvectors and Invariant Subspaces With the following theorem, which we prove in the full version, we establish a loose relation between symmetries in block ciphers and susceptibility to linear cryptanalysis. In the case of PRINTcipher, this was a negative result, but in case of block ciphers without symmetries, it is positive.

Theorem 5. *Consider an invertible vectorial Boolean function F, a subspace U, the orthogonal subspace $U^\perp$ and a vector d. Define $A = (a_{\alpha\beta})_{\alpha,\beta \in U^\perp}$ and $v = (v_\alpha)_{\alpha \in U^\perp}$, $v_\alpha = (-1)^{\langle d, \alpha \rangle}$. Then $Av^T = v^T$ if and only if $F(U + d) = U + d$.*

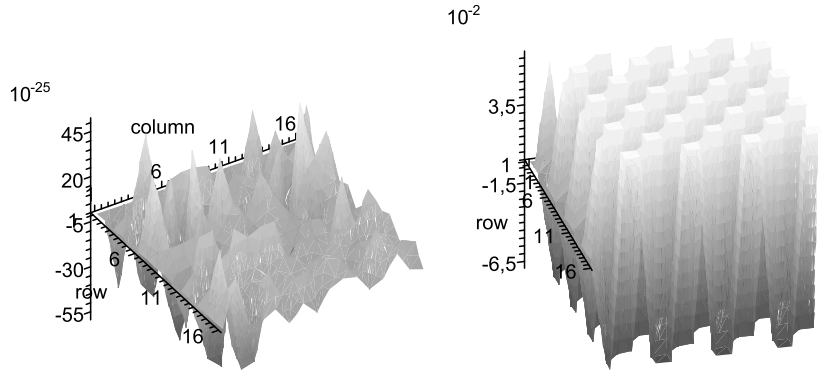
Experimental Results on PRINTcipher We have implemented PRINTcipher-48 for a key from the class of weak keys used as the main example in [14]. This allowed us to derive A and verify that $Av^T = v^T$. We could also derive the biases for 16 characteristics with $\alpha, \beta \in U^\perp$. All of them were close to $\pm 2^{-17}$ as suggested by the above analysis. This gives some circumstantial support to the idea that $B^{48} \approx C$, that PRINTcipher has a stable symmetry, and that this is the main contribution to $c_F(\alpha, \beta)$.

On PRINTcipher-12, we can derive B_{12} analogously to above. Here the stable symmetry can then be confirmed by deriving the eigenvalues numerically for all possible matrices B_{12}. Also, the convergence can be observed experimentally. Fig. 6(a) shows B_{12}^{100} for a non-weak key, while Fig. 6(b) corresponds to a weak key. The matrices clearly differ both in terms of magnitude and structure. Furthermore convergence is rather fast, B_{12}^{10} is already very close to the expected limit.

6 Conclusion and Future Work

We presented and analyzed three interesting examples of ciphers with a non-expected distribution of correlations. The first example mainly limits the most general statements one can hope to prove. General theorems on key-alternating ciphers have to deal with this strange behavior as well.

The second example considered the influence of the key scheduling on the distribution of correlations. For future work this suggests that highly non-linear key scheduling algorithms might be preferable (cf. also [18]). To see this consider the covariance between two different non-zero trails C_θ and $C_{\theta'}$ for $\theta = (\alpha, \theta_1, \ldots, \theta_{r-1}, \beta)$ and $\theta' = (\alpha, \theta'_1, \ldots, \theta'_{r-1}, \beta)$ in the case where the key-length equals the block length. Given $E(k) = (E_0(k), \ldots, E_r(k))$ we assume furthermore that all E_i are permutations and wlog $E_1(k) = k$. Denoting $\gamma = \theta_1 + \theta'_1$,



(a) The matrix B_{12}^{100} for a non-weak key (b) The matrix B_{12}^{100} for a weak key

Fig. 6. The matrix B_{12}^{100} for two different keys

$\delta = (\theta_2 + \theta_2', \ldots, \theta_{r-1} + \theta_{r-1}')$ and $E'(k) = (E_2(k), \ldots, E_{r-1}(k))$, in this setup the covariance is essentially determined by

$$\sum_k (-1)^{\langle \theta+\theta', E(k)\rangle} = \sum_k (-1)^{\langle \delta, E'(k)\rangle + \langle \gamma, k\rangle},$$

which is nothing else than the Fourier coefficient $\widehat{E'}(\gamma, \delta)$. Thus minimizing all covariances corresponds to minimizing the absolute value of $\widehat{E'}(\gamma, \delta)$ which in turn corresponds exactly to maximizing the nonlinearity of E'.

The last example is again related to key-scheduling but more so to symmetries in ciphers. We show a general equivalence of symmetries and linear approximations for weak keys that exist for any number of rounds. This is actually a positive result as it suggests that avoiding these symmetries makes clustering of trails unlikely. Future work is needed to either make this equivalence tighter or find examples of round-independent trail clustering that does not originate from symmetries.

We hope that this work stimulates further research on this fundamental topic.

Acknowledgments The second author was supported by the Swedish Foundation for Strategic Research (SSF) through its Strategic Center for High Speed Wireless Communication at Lund. The third and fourth authors gratefully acknowledge the support from the Danish National Research Foundation and the National Science Foundation of China (Grant No. 11061130539) for the Danish-Chinese Center for Applications of Algebraic Geometry in Coding Theory and Cryptography.

References

1. A. Biryukov and D. Wagner. Slide attacks. In L. R. Knudsen, editor, *FSE*, volume 1636 of *Lecture Notes in Computer Science*, pages 245–259. Springer, 1999.
2. A. Bogdanov, M. Knezevic, G. Leander, D. Toz, K. Varici, and I. Verbauwhede. SPONGENT: A lightweight hash function. In B. Preneel and T. Takagi, editors, *CHES*, volume 6917 of *Lecture Notes in Computer Science*, pages 312–325. Springer, 2011.
3. A. Bogdanov, L. R. Knudsen, G. Leander, C. Paar, A. Poschmann, M. J. B. Robshaw, Y. Seurin, and C. Vikkelsoe. PRESENT: An ultra-lightweight block cipher. In P. Paillier and I. Verbauwhede, editors, *CHES*, volume 4727 of *Lecture Notes in Computer Science*, pages 450–466. Springer, 2007.
4. A. Bogdanov and M. Wang. Zero correlation linear cryptanalysis with reduced data complexity. In *FSE*, 2012.
5. M. L. Buchanan and B. N. Parlett. The uniform convergence of matrix powers. *Numerische Mathematik*, 9(1):51–54, 1966.
6. J. Y. Cho. Linear cryptanalysis of reduced-round PRESENT. In J. Pieprzyk, editor, *CT-RSA*, volume 5985 of *Lecture Notes in Computer Science*, pages 302–317. Springer, 2010.
7. J. Daemen, M. Peeters, G. V. Assche, and V. Rijmen. Nessie proposal: NOEKEON. http://gro.noekeon.org/Noekeon-spec.pdf, 2000.
8. J. Daemen and V. Rijmen. *The design of Rijndael: AES - the Advanced Encryption Standard*. Springer, 2002.
9. J. Daemen and V. Rijmen. Probability distributions of correlation and differentials in block ciphers. *IACR Cryptology ePrint Archive*, 2005:212, 2005.
10. M. Hermelin, J. Y. Cho, and K. Nyberg. Multidimensional extension of Matsui's algorithm 2. In O. Dunkelman, editor, *FSE*, volume 5665 of *Lecture Notes in Computer Science*, pages 209–227. Springer, 2009.
11. L. R. Knudsen, G. Leander, A. Poschmann, and M. J. B. Robshaw. PRINTcipher: A block cipher for IC-printing. In S. Mangard and F.-X. Standaert, editors, *CHES*, volume 6225 of *Lecture Notes in Computer Science*, pages 16–32. Springer, 2010.
12. L. R. Knudsen and M. Robshaw. *The Block Cipher Companion*. Information security and cryptography. Springer, 2011.
13. G. Leander. On linear hulls, statistical saturation attacks, PRESENT and a cryptanalysis of PUFFIN. In K. G. Paterson, editor, *EUROCRYPT*, volume 6632 of *Lecture Notes in Computer Science*, pages 303–322. Springer, 2011.
14. G. Leander, M. A. Abdelraheem, H. AlKhzaimi, and E. Zenner. A cryptanalysis of PRINTcipher: The invariant subspace attack. In P. Rogaway, editor, *CRYPTO*, volume 6841 of *Lecture Notes in Computer Science*, pages 206–221. Springer, 2011.
15. M. Matsui. Linear cryptanalysis method for DES cipher. In T. Helleseth, editor, *EUROCRYPT*, volume 765 of *Lecture Notes in Computer Science*, pages 386–397. Springer, 1993.
16. S. Murphy. The effectiveness of the linear hull effect. *Technical report*, RHUL-MA-2009-19, 2009.
17. K. Nyberg. Linear approximation of block ciphers. In A. De Santis, editor, *EUROCRYPT*, volume 950 of *Lecture Notes in Computer Science*, pages 439–444. Springer, 1994.
18. K. Nyberg. Comments on key-scheduling, 2012. Personal communication.
19. K. Ohkuma. Weak keys of reduced-round PRESENT for linear cryptanalysis. In M. J. Jacobson Jr., V. Rijmen, and R. Safavi-Naini, editors, *Selected Areas in*

Cryptography, volume 5867 of *Lecture Notes in Computer Science*, pages 249–265. Springer, 2009.
20. S. Park, S. Sung, S. Lee, and J. Lim. Improving the upper bound on the maximum differential and the maximum linear hull probability for spn structures and aes. In *Fast Software Encryption*, pages 247–260. Springer, 2003.
21. A. A. Selçuk. On probability of success in linear and differential cryptanalysis. *J. Cryptology*, 21(1):131–147, 2008.