Experimenting Linear Cryptanalysis

Baudoin Collard*, François-Xavier Standaert**

UCL Crypto Group, Microelectronics Laboratory, Université catholique de Louvain.
Place du Levant 3, B-1348, Louvain-la-Neuve, Belgium.

e-mails: baudoin.collard; fstandae@uclouvain.be

Introduction

Since the publication of linear cryptanalysis in the early 1990s, the precise understanding of the statistical properties involved in such attacks has proven to be a challenging and computationally intensive problem. As a consequence, a number of strategies have been developed, in order to design block ciphers secure against cryptanalysis, under reasonable assumptions. In this context, a good assessment of the hypotheses used for the evaluation of linear cryptanalysis and a careful measurement of the distance between actual constructions and theoretical expectations are of particular interest. In this chapter, we present a number of illustrative experiments that allow discussing these issues. Based on a concrete instance of block cipher with small block size, we first evaluate the distance between the so-called practical and provable security approaches for designing block ciphers. Then, we challenge the assumptions of key independence and key equivalence that are frequently used in linear cryptanalysis. Third, we put forward the difficulty of obtaining precise estimations of the distributions within a secure block cipher when the number of rounds increases. We also discuss the consequences of this observation for the key ranking strategies used in order to extract information from actual statistical biases. Finally, we provide systematic experiments of linear cryptanalysis using single and multiple approximations in order to confirm a number of intuitive views that can be found in former papers.

Summarizing, this chapter provides an experimental survey of the basic assumptions in linear cryptanalysis and its consequences for the design of modern block ciphers. It is structured as follows. Section 1 contains background information, including notations, definitions, related works, a specification of our target cipher and a description of Matsui's second algorithm for linear cryptanalysis. Section 2 contains an empirical evaluation of different assumptions in linear cryptanalysis. It discusses the linear hull effect, the pros and cons of the practical security approach, and the key independence and key equivalence hypotheses. Section 3 contains experiments on the test of key-dependent linear biases in different scenarios. Finally, Section 4 briefly browses though the consequences of our experiments and observations for more advanced statistical attacks.

* Work supported by the project Nanotic-Cosmos of the Walloon Region.
** Associate researcher of the Belgian Fund for Scientific Research (FNRS-F.R.S.).

1 Background

1.1 Notations & definitions

The following notations and definitions are standard in linear cryptanalysis. We borrow them from Liam Keliher's and Vincent Rijmen's PhD theses [29, 47].

Definition 1. *An iterated block cipher is an algorithm that transforms a plaintext block of a fixed size n into a ciphertext of identical size, under the influence of a key k, by a repeated application of an invertible transformation ρ, called the round transformation. Denoting the plaintext with x_0 and the ciphertext with x_R, the encryption operation can be written as:*

$$x_{r+1} = \rho_{k_r}(x_r), \qquad r = 1, 2, \ldots, R, \tag{1}$$

where the different k_r are the subkeys generated by a key scheduling algorithm.

For simplicity, we will consider n-bit keys and subkeys in the rest of the paper.

Definition 2. *Let $F : \{0,1\}^n \to \{0,1\}^n$ be a bijection and $\boldsymbol{a}$,$\boldsymbol{b}$ be two masks $\in \{0,1\}^n$. If $X \in \{0,1\}^n$ is a uniformly distributed random variable, then the linear approximation bias $LB(\boldsymbol{a}, \boldsymbol{b})$ is defined as:*

$$LB(\boldsymbol{a}, \boldsymbol{b}) = \Pr_X\{\boldsymbol{a} \bullet X = \boldsymbol{b} \bullet F(X)\} - \frac{1}{2}, \tag{2}$$

where $\bullet$ denotes the scalar product. If F is parametrized by a key K, we write $LB(\boldsymbol{a}, \boldsymbol{b}; K)$ and the expected linear bias $ELB(\boldsymbol{a}, \boldsymbol{b})$ is defined as:

$$ELB(\boldsymbol{a}, \boldsymbol{b}) = \mathop{\boldsymbol{E}}_K \left(LB(\boldsymbol{a}, \boldsymbol{b}; K)\right). \tag{3}$$

The linear bias can be computed for different transformations, e.g. a single S-box, a round function or a complete block cipher. Quite naturally, computing it precisely becomes computationally intensive as the transformation size increases.

Definition 3. *A one-round characteristic for the round i of an iterated block cipher is a pair of n-bit vectors $\langle \boldsymbol{a}_i, \boldsymbol{b}_i \rangle$ respectively corresponding to the input and output masks for this round. An R-round characteristic for rounds $1 \ldots R$ is an $(R+1)$-tuple of n-bit vectors $\Omega = \langle \boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_{R+1} \rangle$, where $\langle \boldsymbol{a}_i, \boldsymbol{a}_{i+1} \rangle$ respectively correspond to the input and output masks for the round i.*

Definition 4. *A Markov cipher is a block cipher in which the linear (and differential) biases of different rounds are independent of each other, assuming that uniformly random subkeys are used in the different rounds [34].*

Lemma 1 (Piling-up lemma). *Given a vector of independent subkeys $\tilde{K}$ and a Markov cipher, the linear characteristic bias of a characteristic Ω is defined as:*

$$LCB(\Omega, \tilde{K}) = 2^{R-1} \prod_{i=1}^{R} LB(\boldsymbol{a}^i, \boldsymbol{a}^{i+1}; k_i). \tag{4}$$

This lemma was introduced and proved in Matsui's Eurocrypt 1993 paper [36]. Note that in an iterated block cipher using a bitwise XOR key addition, the absolute value of $LCB(\Omega, \tilde{K})$ is independent of the subkey vector, but not its sign.

Definition 5. *Given input and output masks $\boldsymbol{a}$,$\boldsymbol{b}$, the linear hull $LH(\boldsymbol{a}, \boldsymbol{b})$ is the set of all R-round characteristics having $\boldsymbol{a}$ as input mask for round 1 and $\boldsymbol{b}$ as output mask for round R. The approximated linear hull $ALH(\boldsymbol{a}, \boldsymbol{b}, N_h)$ is the subset of N_h characteristics in $LH(\boldsymbol{a}, \boldsymbol{b})$ having the largest bias $LB(\boldsymbol{a}, \boldsymbol{b})$.*

Note that we slightly modified Nyberg's definition of linear hull [44], in order to take into account the possibility of using subsets of characteristics. Finally, the following definition is introduced to reflect the need of estimating the linear bias and expected linear bias from a subset of plaintexts (smaller than the codebook).

Definition 6. *Let $\mathcal{X}$ be the set of all plaintexts of a block cipher (aka the codebook) and $\mathcal{Y}_i \subset \mathcal{X}$, with $1 \leq i \leq N_s$, be different subsets of $\mathcal{X}$ containing N_p plaintexts chosen uniformly and independently. Then the linear bias sampled with N_s sets of N_p plaintexts $SLB(\boldsymbol{a}, \mathbf{b}, N_s, N_p)$ is defined as:*

$$SLB(\boldsymbol{a}, \mathbf{b}, N_s, N_p) = \sum_{i=1}^{N_s} \Pr\{i\} \cdot \left(\widehat{\Pr_{Y_i}}\{\boldsymbol{a} \bullet X = \boldsymbol{b} \bullet F(X)\} - \frac{1}{2} \right). \tag{5}$$

If F is parametrized by a key K, we write $SLB(\boldsymbol{a}, \boldsymbol{b}, N_s, N_p; K)$. If we additionally denote a uniformly selected subset of keys of cardinality N_k by $\mathcal{L} \subset \mathcal{K}$, the sampled expected linear bias $SELB(\boldsymbol{a}, \boldsymbol{b}, N_s, N_p, N_k)$ is defined as:

$$SELB(\boldsymbol{a}, \boldsymbol{b}, N_s, N_p, N_k) = \sum_{K \in \mathcal{L}} \Pr\{K\} \cdot SLB(\boldsymbol{a}, \boldsymbol{b}, N_s, N_p; K). \tag{6}$$

The SLB is usually referred to as the experimental bias in the literature.

1.2 Related works

Design approaches to prevent linear cryptanalysis

Worst case security. Let K be a block cipher master key. In the worst case, an adversary would perform linear cryptanalysis using masks $(\mathbf{a}, \mathbf{b})$ such that:

$$(\mathbf{a}, \mathbf{b}) = \underset{(\mathbf{x}, \mathbf{y})}{\operatorname{argmax}} \; LB(\mathbf{x}, \mathbf{y}; K). \tag{7}$$

Following [37], it directly yields the approximated data complexity of the attack[1]:

$$N \approx \frac{c}{\max \; LB(\mathbf{a}, \mathbf{b}; K)^2}, \tag{8}$$

where c is a small constant value. However, as will be detailed next, this worst case strategy cannot be directly exploited by actual adversaries. In practice, the

[1] More sophisticated approximations are discussed in [6, 24, 25, 51].

direct computation of Equation (7) is generally infeasible, both because of an unknown key and for computational reasons (estimating the bias of an n-bit permutation requires $n \cdot 2^{2n}$ operations, i.e. more than exhaustive key search).

The key equivalence hypothesis. As a consequence of the previous limitations, design strategies to prevent linear cryptanalysis usually start by assuming key equivalence, as introduced by Harpes et al. in [18]. That is, they assume that the linear bias will be close to its average value for the vast majority of keys:

$$LB(\mathbf{a}, \mathbf{b}; \tilde{K}) \approx ELB(\mathbf{a}, \mathbf{b}). \tag{9}$$

Practically secure block ciphers. Next to the key equivalence hypothesis, arguing about security against linear cryptanalysis requires to find ways to evaluate the linear biases in a computationally tractable manner. As shown by Lemma 1, a simple solution for this purpose is to use the concept of characteristic. For a Markov cipher and assuming independent round keys, the probability of an R-round characteristic can be computed as a product of 1-round characteristic probabilities. Hence, a designer can run an algorithm to search the characteristic Ω_{max} such that $LCB(\Omega_{max})$ is maximal and then assume:

$$ELB(\mathbf{a}, \mathbf{b}) \approx LCB(\Omega_{max}). \tag{10}$$

Lars Knusden calls a block cipher practically secure if the data complexity determined by this method is prohibitive [30]. Obviously, such an approach is only valid up to a certain extent and it may give rise to false intuitions. For example, increasing the number of rounds in a block cipher always reduces the linear characteristic bias, while the actual expected linear bias of a cipher cannot be decreased below a certain threshold, depending on its block size. Positively, the practical security approach is a simple way to estimate the number of rounds required for a block cipher to become hard to distinguish from a random permutation. It is also the basis of the *wide-trail strategy* [14], that has been successful for designing many modern block ciphers, most notably the AES Rijndael [15]. In the wide-trail strategy, one essentially ensures that each characteristic involves a large number of S-boxes in each of the block cipher rounds, and that these S-boxes do not have highly probable linear approximations.

Provable security. In contrast with the practical security approach, the theory of provable security against linear cryptanalysis attempts to compute expected linear biases, i.e. to consider all the characteristics in a linear hull rather than only the best one[2]. An example of design strategy based on such a theory was proposed by Matsui in [39], and gave rise to the design of the block cipher Misty [40]. A similar line of work was followed by Keliher et al. in [27] and applied to the AES Rijndael in [28]. The main benefit of provable security is to provide security guarantees that only rely on the key equivalence hypothesis, and assuming independent round keys. Its main limitation is the computational difficulty of finding tight bounds, when the number of block cipher rounds increases.

[2] In a very similar way, provable security against differential cryptanalysis considers the concept of differential rather than the one of differential characteristic [45].

Decorrelation theory. Vaudenay proposed an alternative solution for designing block ciphers with provable security against a large class of attacks in [52]. It essentially aims at preventing the use of the key equivalence hypothesis in an attack. For this purpose, decorrelation modules are used, that are key-dependent transformations making the linear bias (and differential probability) of a given approximation highly key-dependent, so that any attack that chooses a priori the input and output masks will fail. Examples of block ciphers based on the decorrelation theory include the cipher C [2] and the Krazy Feistel Cipher [3].

Experimental evaluations of assumptions and attacks

Following Matsui's experiments on the DES [37], various publications contain empirical evaluations of the linear cryptanalysis and its underlying assumptions. All these empirical works are tightly connected with our following analyzes. We list a few of them for illustration. First, Junod reported new results on the linear cryptanalysis of the DES in 2001 [23], together with a discussion of the attack's complexity. Rouvroy et al. performed similar experiments, exploiting the computational power of reconfigurable hardware [49]. Extensions of these works, considering the use of multiple linear approximations can be found, e.g. in [5, 12, 21]. Second, the assumption of independent round keys is discussed and tested experimentally by Knudsen and Mathiassen in [33]. They show that the key scheduling algorithms used in practical block cipher constructions has an impact on this assumption, but that the conclusions drawn when independent round keys are used should still reasonably hold with a good key schedule. Third, the relevance of the practical security approach is analyzed by Selçuk in [50], with experiments conducted against reduced versions of RC5 and Feistel ciphers. This paper also discusses the impossibility to accurately evaluate linear approximations in block ciphers by means of statistical sampling. A very similar approach was followed in [46], in which experiments are performed against small scale block ciphers, in order to review the practical security approach and the key equivalence hypothesis for various block sizes, S-boxes and diffusion layers. A counterpart of these experiments in the context of differential cryptanalysis can be found in [7]. Finally, Daemen and Rijmen investigated the statistical distributions of fixed-key linear probabilities, both theoretically and empirically, in [16]. Their results allow replacing the key equivalence hypothesis by a precise understanding of the fixed-key vs. average behavior in block ciphers.

1.3 Target cipher

The experiments in the next sections will be performed against the block cipher `SmallPresent`-[16], with 16-bit block size, described in [35]. `SmallPresent` is an iterated block cipher with R rounds, each of them made of a key addition layer, a substitution layer and a permutation layer. The key addition layer is a bitwise XOR, the substitution layer applies four 4-bit S-boxes to the state and the permutation layer is a simple wire crossing (or bit permutation). The key scheduling of `SmallPresent` shifts the key register, applies one S-box to the left-most four bits and adds a round constant to the right-most bits.

1.4 Matsui's second algorithm

Matsui's original description of the linear cryptanalysis comes with two algorithms that allow exploiting linear approximations in iterated block ciphers [36]. In this paper, we focus on the second one, that is intuitively pictured in Figure 1.

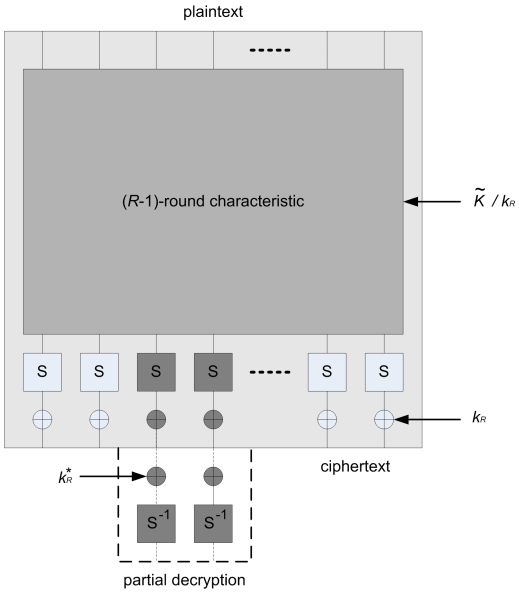

Fig. 1. Linear cryptanalysis with Matsui's M2 algorithm.

In this setting, an adversary exploits an $(R-1)$-round characteristic with good LCB. This characteristic ideally connects to a limited number of active S-boxes in the last block cipher round (e.g. the two dark grey boxes in Figure 1), so that by guessing a few key bits of k_r, it is possible to do a partial decryption of the ciphertext for these S-boxes[3]. As a result, the adversary is able to sample the linear bias of round $(R-1)$ under the various key hypotheses for k_r. We denote this sampled bias as $SLB(\mathbf{a}_1, \mathbf{a}_R; k_r)$. Eventually, the adversary selects the key candidate that maximizes some function of this sampled bias. Summarizing, a linear cryptanalysis using Matsui's M2 algorithm requires: (1) a way to generate good characteristics and (2) a way to test the key-dependent sampled biases.

1. Generation of characteristics. This step usually exploits a variety of heuristics, starting with the inspection of the non-linear components in block ciphers (e.g. S-boxes) and trying to extend partial approximations from one round to multiple rounds. For example, a branch-and-bound algorithm can be used for this purpose [38]. Intuitively, such an algorithm concatenates all possible 1-round approximations, and compares the resulting biases (estimated thanks to the piling-up

[3] The complexity of such a partial decryption for a n-bit key guess is in $\mathcal{O}(n \cdot 2^n)$ [11].

lemma) with a lower bound, in order to reject "bad" approximations. For certain ciphers, e.g. the DES, it allows to find the best characteristics. For other ones, e.g. SERPENT, the amount of possible approximations grows in such a way that maintaining a list of all possibly optimal characteristics becomes impossible as the number of rounds increases, because of memory constraints [10]. The precise description of these heuristics is out of the scope of this paper. Nevertheless, we mention that, as we investigate a 16-bit cipher, it was always possible to exhaustively find the best characteristics in the next sections.

2. Testing the key-dependent sampled biases. Again, various solutions are possible, of which the precise description is out of the scope of this paper. To keep it short, we will refer to two general types of approaches. The first one requires that a good estimation of the linear bias LB is available to the adversary, e.g. using the linear characteristic bias LCB. In this case, it is possible to apply a maximum likelihood approach. Under certain assumptions detailed in [25], this leads to the simple rule to select the key candidate that minimizes the Euclidean distance between the sampled and estimated bias values:

$$\hat{k_r} = \underset{k_r^*, \mathsf{P}(\Omega, \tilde{K})}{\operatorname{argmin}} \left(SLB(\mathbf{a}_1, \mathbf{a}_R; k_r) - LCB(\Omega, \tilde{K}) \right)^2, \tag{11}$$

where $\mathsf{P}(\Omega, \tilde{K})$ denotes the parity of the subkey bits used in the characteristic Ω. The second type of approach is used if a good estimation of the bias is not available. In this case, one can rely on different types of heuristics, of which a classical one is simply to select the key that maximizes the sampled linear bias:

$$\hat{k_r} = \underset{k_r^*}{\operatorname{argmax}} \left(SLB(\mathbf{a}_1, \mathbf{a}_R; k_r) \right). \tag{12}$$

The next sections provide different experiments in order to illustrate these two steps. Namely, Section 2 discusses the difference between the linear characteristics that can be efficiently generated by actual adversaries and the corresponding linear approximations. Then, Section 3 discusses the exploitation of these characteristics and their test with the previously described maximum likelihood and heuristic approaches, including experiments using multiple approximations.

2 Evaluation of characteristics and approximations

2.1 Computing linear hulls

The first experiments we performed relate to the notion of linear hull introduced by Nyberg in [44]. Thanks to the limited block size of our target cipher, it was possible to generate the entire linear hull for a given pair of input/output masks $(\mathbf{a}, \mathbf{b})$, with a branch-and-bound algorithm. In addition, it was also possible to set the lower bound of the branch-and-bound arbitrarily low, in order to generate all characteristics with non-zero LCB. As expected, the number of characteristics in the hull increased exponentially with the number of rounds.

In practice, we generated the linear hull for the 100 best characteristics of the cipher. The size of the linear hull was between 2 and 77 after three rounds, between 54 and 51,388 after four rounds and between 991 and 1,826,043 after five rounds. For more than five rounds, we could not generate the complete linear hull, because of limited computational resources. While far from being exhaustive, these experiments suggest that, for a fixed number of rounds, the number of characteristics in a linear hull varies considerably according to the choice of the input/output masks (in particular, the number of active S-boxes in these input/output masks has a significant impact in this respect).

Cheating with the full codebook. Considering small block ciphers allows to generate the full codebook. An interesting consequence of this possibility is that one can consider the block cipher with a fixed master key as a large 16-bit S-box. And for any pair of masks $(\mathbf{a}, \mathbf{b})$, we can then compute the exact linear bias $LB(\mathbf{a}, \mathbf{b}; K)$. There exists two equivalent ways allowing to perform this task. The first one is to compute the FFT of the S-box directly, e.g. using the technique described in [48], which is possible for any number of rounds. The second one is to exploit the knowledge of the linear hull, as long as it is available (i.e. for a reduced number of rounds), and to compute:

$$LB(\mathbf{a}, \mathbf{b}; K) = \sum_{\Omega \in LH(\mathbf{a}, \mathbf{b})} LCB(\Omega, \tilde{K}). \tag{13}$$

It is important to note that this equality does not straightforwardly lead to good approximations of the linear bias for the adversaries, as it only holds if the parity of all the subkey bits that are involved the characteristics (i.e. the bias signs) are taken into account. As the number of such parities increases with the size of the hull, it rapidly becomes intensive to guess by any practical adversary.

Also, Equation (13) is only tight if the linear bias is evaluated with the full codebook and linear hull. In this respect, an interesting experiment is to evaluate how the quality of the estimation for $LB(\mathbf{a}, \mathbf{b}; K)$ degrades when the number of plaintexts and characteristics in the hull decreases. For this purpose, we generated the complete linear hull for the linear approximation $(\mathbf{a}, \mathbf{b}) = (\texttt{0770}_{hex}, \texttt{0111}_{hex})$, for 5 rounds of `SmallPresent`-[16]. We found 916,841 characteristics corresponding to these masks. Their linear characteristic bias ranged from 2^{-9} for the best approximation to 2^{-24} for the 69,632 worst ones[4].

Figure 2 shows the impact of sampling with less plaintexts than the full codebook: every line in the picture represents one random key. We see that the distance between the exact linear bias $LB(\mathbf{a}, \mathbf{b}; K)$ and its sampled value $SLB(\mathbf{a}, \mathbf{b}, 1, N_p; K)$ only converges to zero when N_p equals 2^{16}. Figure 3 shows the impact of sampling with an approximated linear hull $ALH(\mathbf{a}, \mathbf{b}, N_h)$: every point in the picture represents a random key and its color scale indicates the size of the approximated hull N_h. We see that the approximation only converges towards the correct value of $LB(\mathbf{a}, \mathbf{b}; K)$ when N_h gets close to $916,841 \approx 2^{20}$.

[4] That is, the approximations with the smallest non-zero bias.

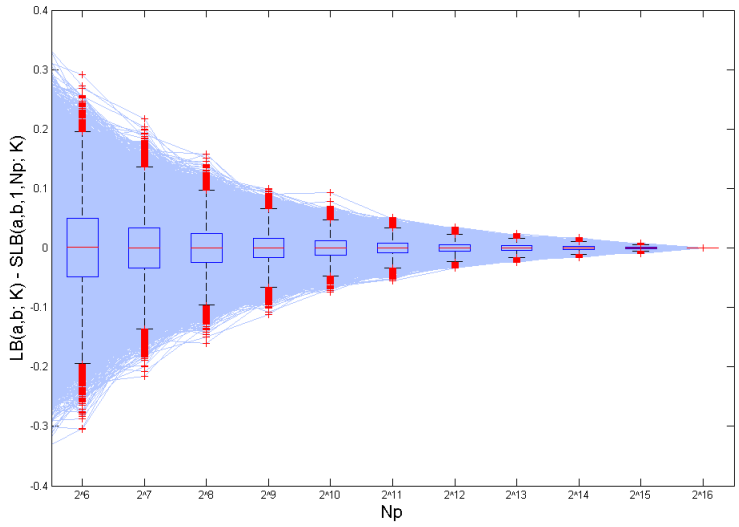

Fig. 2. Estimation of $LB(\mathbf{a}, \mathbf{b}; K)$ with the complete linear hull, in function of the number of plaintexts used in the sampling ($2^5 \leq N_p \leq 2^{16}$, $N_s = 1$), for different keys.

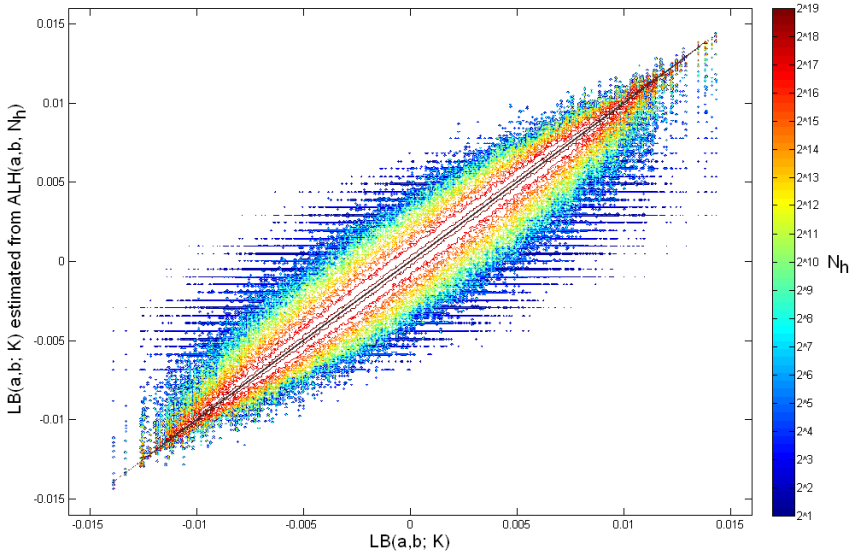

Fig. 3. Estimation of $LB(\mathbf{a}, \mathbf{b}; K)$ with the complete codebook, in function of the number of elements in the approximated linear hull ($2^1 \leq N_h \leq 2^{19}$), for different keys.

2.2 A note on Murphy's technical report

Sean Murphy recently posted a technical report about the existence of the so-called "linear hull effect" [42], first described by Nyberg in 1994. As this discussion closely relates to the experiments in this paper, we briefly comment on it. Summarizing, Murphy's observation relates to the following equality from [44]:

$$\mathop{\mathbf{E}}_{K}\left(LB(\mathbf{a},\mathbf{b};K)^2\right) = \sum_{\Omega \in LH(a,b)} \left(LCB(\Omega,\tilde{K})^2\right). \tag{14}$$

This equation can be illustrated with a simple example. Let us define a pair of masks $(\mathbf{a},\mathbf{b})$, such that the linear hull for a given cipher equals $LH(\mathbf{a},\mathbf{b}) = \{\Omega_1,\Omega_2\}$. Let us also assume that $LCB(\Omega_1,\tilde{K}) = \epsilon_1$ or $-\epsilon_1$, depending on the parity of $\tilde{K}$, and that $LCB(\Omega_2,\tilde{K}) = \epsilon_2$ or $-\epsilon_2$, depending on the parity of $\tilde{K}$. In this case, and assuming uniformly random keys, there will be four possible parities and the left part of Equation (14) can be simply evaluated as:

$$\begin{aligned}
\mathop{\mathbf{E}}_{K}\left(LB(\mathbf{a},\mathbf{b};K)^2\right) &= \frac{1}{4}\cdot\left((\epsilon_1+\epsilon_2)^2 + (\epsilon_1-\epsilon_2)^2 + (-\epsilon_1+\epsilon_2)^2 + (-\epsilon_1-\epsilon_2)^2\right), \\
&= \epsilon_1^2 + \epsilon_2^2, \\
&= \sum_{\Omega \in LH(a,b)} \left(LCB(\Omega,\tilde{K})^2\right).
\end{aligned}$$

Using this example as a case study, the issue raised by Murphy can be explained as follows. While Equation (14) is correct, it cannot be used to evaluate the average data complexity of a linear cryptanalysis. This is because the average data complexity is proportional to the average over the keys of the inverse of the squared linear bias. And this quantity is not equal to the inverse of the average over the keys of the squared linear bias. That is, in our example:

$$\frac{1}{\mathbf{E}_K\left(LB(\mathbf{a},\mathbf{b};K)^2\right)} = \frac{1}{\epsilon_1^2+\epsilon_2^2}, \tag{15}$$

$$\mathop{\mathbf{E}}_{K}\frac{1}{\left(LB(\mathbf{a},\mathbf{b};K)^2\right)} = \frac{\epsilon_1^2+\epsilon_2^2}{(\epsilon_1+\epsilon_2)^2\cdot(\epsilon_1-\epsilon_2)^2}. \tag{16}$$

As these two values are related by Jensen's inequality, Murphy correctly concludes that Equation (15) can only provide a lower bound for the data requirements of a linear cryptanalysis. The practical impact of this observation can easily be seen with the particular case in which $\epsilon_1 = -\epsilon_2$, leading to a zero value for the bias $LB(\mathbf{a},\mathbf{b};K)$. Clearly, such a situation is only interpreted correctly by Equation (16), resulting in an infinite data complexity. A more intuitive view of this result is that, by squaring the bias values, one looses the sign information that is crucial in combining the different approximations constructively.

We note that, while the previous observation is sound, it does not contradict the existence of a linear hull effect which, as will be discussed in the next section, relates to a trivial difference between a linear characteristic bias, estimated with

the piling-up lemma, and the linear bias of its corresponding masks. Also, it remains that Nyberg's relation in Equation (14) can be used to compute a lower bound for the data complexity of a linear cryptanalysis, in a less computationally intensive manner than with the direct computation of the linear biases.

2.3 Pros & cons of the practical security approach: the linear hull effect

In our following experiment, we ran a branch-and-bound algorithm in order to find the best characteristic Ω_{max}, for different number of block cipher rounds, and denoted the corresponding pairs of input and output masks as $\mathbf{a}_{max}, \mathbf{b}_{max}$. Then, we evaluated the following quantities in Figure 4:

$$\mathop{\mathbf{E}}_{K} \; LB(\mathbf{a}_{max}, \mathbf{b}_{max}; K),$$
$$\mathop{\mathbf{var}}_{K} \; LB(\mathbf{a}_{max}, \mathbf{b}_{max}; K).$$

In this figure, the continuously decreasing line represents the linear characteristic bias $LCB(\Omega_{max}, \tilde{K})$, estimated with the piling-up lemma. The bounded line represents the corresponding biases $LB(\mathbf{a}_{max}, \mathbf{b}_{max}; K)$, computed for different keys (dots representing the sample mean, crosses representing individual experiments with different keys). These results clearly allow to give an informal description of the linear hull effect. Namely, as soon as the number of characteristics in the hull $LH(\mathbf{a}_{max}, \mathbf{b}_{max})$ increases, the distance between $LCB(\Omega_{max}, \tilde{K})$ and $LB(\mathbf{a}_{max}, \mathbf{b}_{max}; K)$ also increases, resulting in an overestimation of the attack data complexity with $LCB(\Omega_{max}, \tilde{K})$. This clearly emphasizes the pros and

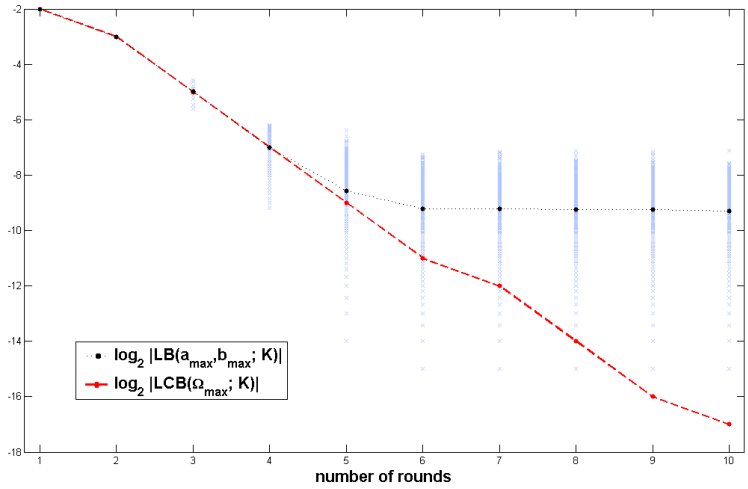

Fig. 4. Comparison between the linear characteristic bias $LCB(\Omega, \tilde{K})$ estimated with the piling-up lemma and the corresponding bias $LB(\mathbf{a}_{max}, \mathbf{b}_{max}; K)$, for different keys.

cons of the practical security approach. On the one hand, the relevance of single characteristics trivially vanishes as the number of rounds increases. This "linear hull effect" is in fact dominant as soon as the cipher is practically secure (according to Knudsen's definition). On the other hand, the pair of masks $(\mathbf{a}_{max}, \mathbf{b}_{max})$ generated with the branch-and-bound algorithm and exploited by the adversary no longer corresponds to the best approximation in this case. After a sufficient number of rounds, this best characteristic behaves like a random one: the mean of $LB(\mathbf{a}_{max}, \mathbf{b}_{max}; K)$ is smaller than $2^{-\frac{n}{2}}$ and its variance over the keys is large. It prevents the application of successful attacks exploiting these characteristics.

2.4 Best, worst and average cases

The previous section investigated the practical security approach and an adversary who can only find linear characteristics with a time complexity below the one of exhaustive key search, e.g. with a branch-and-bound algorithm. We now tackle the more theoretical situation in which the best linear approximations can be found, for each key. For this purpose, we again considered `SmallPresent`-[16] with various number of rounds and computed the maximum, minimum[5] and average values for $LB(\mathbf{a}, \mathbf{b}; K)$, averaged over the keys. That is, we computed:

$$\begin{aligned} &\mathop{\mathbf{E}}_{K} \max_{\mathbf{a},\mathbf{b}} LB(\mathbf{a}, \mathbf{b}; K), \\ &\mathop{\mathbf{E}}_{K} \min_{\mathbf{a},\mathbf{b}} LB(\mathbf{a}, \mathbf{b}; K), \\ &\mathop{\mathbf{E}}_{K} \mathop{\mathbf{E}}_{\mathbf{a},\mathbf{b}} LB(\mathbf{a}, \mathbf{b}; K), \end{aligned}$$

and the corresponding variances. The results are in Figure 5 and illustrate that:

- After a few first rounds for which the block cipher is practically insecure, the value of the maximum, minimum and average linear biases stabilize.
- Contrary to the case where the masks are fixed and determined by a branch-and-bound algorithm (as in the previous section), the variance of these quantities over the keys is small (and decreases with the block size [46]).
- Most importantly, the maximum bias value is between 2^{-6} and 2^{-7}.

This last point emphasizes that in theory, for a given cipher and key, a low data complexity (i.e. lower than exhaustive key search) linear cryptanalysis is always possible. However, this does not mean that the practical security approach is not good for designing ciphers. The goal of the practical security approach is not to prevent the existence of good linear approximations, but to make them hard to actually find and exploit, for computational reasons. This is because, for practically secure ciphers, the best approximations cannot be found anymore by chaining small approximations. Finding these best approximations requires (1) to enumerate the full linear hull and (2) to take the sign of each characteristic bias into account. In both cases, this implies computationally intensive tasks.

[5] That is, again, the approximations with the smallest non-zero bias.

Additionally, as far as Matsui's second algorithm is concerned, it is also necessary that these best approximations only imply a limited number of active S-boxes in the first/last rounds, so that an efficient key guessing can be performed.

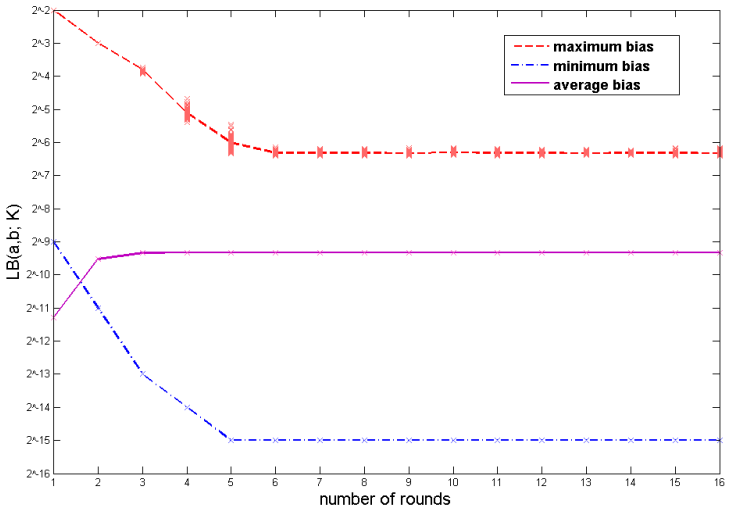

Fig. 5. Maximum, minimum and average values of the linear bias $LB(\mathbf{a}, \mathbf{b}; K)$.

2.5 Key independence and key equivalence hypotheses

Before moving to the empirical evaluation of Matsui's second algorithm, this section briefly discusses the key independence and key equivalence hypotheses. First, although we did not perform as intensive experiments as Knudsen and Mathiassen in [33], we mention that considering independent round keys or the key scheduling algorithm in [35] did not lead to significant differences in our previous experiments and conclusions. In general, it seems that the assumption of independent round keys is reasonably fulfilled by modern ciphers.

By contrast, our observations on the key equivalence hypothesis are contrasted. Clearly, Figure 4 shows that, given a fixed pair of masks $(\mathbf{a}, \mathbf{b})$, the key equivalence hypothesis of Equation (9) is not respected as soon as the linear hull effect increases. In other words, and independently of Murphy's observation, averaging the linear bias $LB(\mathbf{a}, \mathbf{b}; K)$ over the keys is not very significant in this case. On the other hand, Figure 5 shows that once a block cipher has a sufficient number of rounds for this linear hull effect to be dominant, the best linear approximations computed for each key have a minimum, maximum and average bias that does not strongly depend on the keys. Intuitively, once the number of rounds in the cipher is sufficient, changing the key is equivalent to adding one

round. In other words, the resulting cipher is hard to distinguish from a 16-bit random S-box, of which the maximum, minimum and average biases essentially depend on the block size. In order to confirm this intuition, we ran a last experiment in which we directly computed the distribution of the biases $LB(\mathbf{a}, \mathbf{b}; K)$ for independent round keys. The results in Figure 6 show that after a few rounds, the shape of this distribution does not change anymore, and is in fact hard to distinguish from the one of a random 16-bit S-box, as theoretically expected.

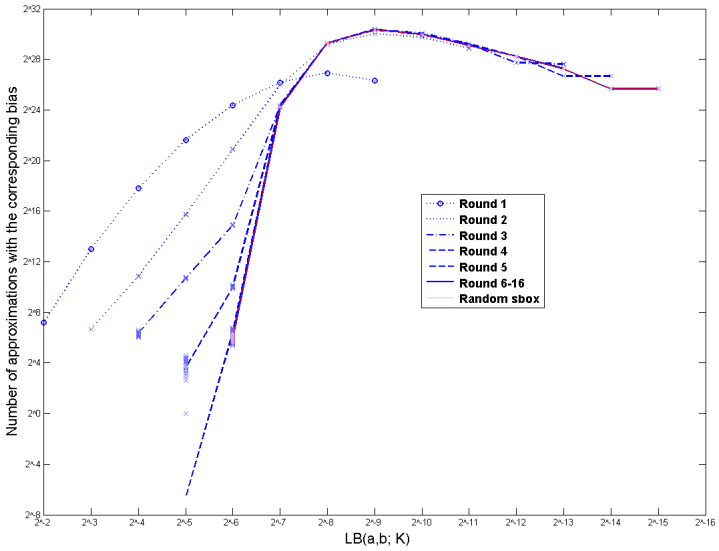

Fig. 6. Distributions of the linear biases for different number of rounds.

3 Test of key dependent sampled biases

The previous section discussed experiments related to the best selection of linear characteristics and their differences with linear approximations. In this section, we complement this evaluation by looking at the exploitation of these characteristics in a linear cryptanalysis. For this purpose, we will investigate how the two procedures to test sampled linear biases described in Section 1.4 perform with different linear characteristics and approximations. In addition, we will consider an extension of linear cryptanalysis using multiple linear approximations, which is a frequently considered solution to decrease the data complexity of the attack, first introduced in [26]. For this purpose, we will assume an adversary who has determined the set of m best characteristics of an $(R-1)$-round cipher. Then, we define the gain of a linear cryptanalysis attack as proposed in [5]:

Definition 7. *If an attack is used to recover an n-bit key and is expected to return the correct key after having checked on the average M candidates in the sorted list, then the gain of the attack, expressed in bits, is defined as:*

$$\gamma = -\log_2 \frac{2 \cdot M - 1}{2^n}. \tag{17}$$

For any set of approximation, we also use Biryukov et al.'s definition of capacity:

$$C(\mathbf{a}^{1:m}, \mathbf{b}^{1:m}; K) = 4 \cdot \sum_{i=1}^{m} LB(\mathbf{a}^i, \mathbf{b}^i; K)^2 \tag{18}$$

In [5], the data complexity of a linear cryptanalysis using multiple linear approximations is estimated with[6] $1/C(\mathbf{a}^{1:m}, \mathbf{b}^{1:m}; K)$. Finally, when considering multiple approximations, the test of key-dependent sampled biases is generalized as follows. First, the maximum likelihood key testing of Equation (11) becomes:

$$\tilde{k_r} = \underset{k_r^*}{\operatorname{argmin}} \sum_{i=1}^{m} \left(|SLB(\mathbf{a}_1^i, \mathbf{a}_R^i; k_r^*)| - |LCB(\Omega^i, \tilde{K}|) \right)^2. \tag{19}$$

It corresponds to a straightforward extension of Equation (11), where an absolute difference operator is added, in order to remove the need to guess the parities for multiple approximations (which becomes computationally hard as m increases). Next, the heuristic key testing procedure of Equation (12) becomes:

$$\tilde{k_r} = \underset{k_r^*}{\operatorname{argmax}} \left(\sum_{i=1}^{m} \left| SLB(\mathbf{a}_1^i, \mathbf{a}_R^i; k_r^*) \right| \right). \tag{20}$$

In practice, we performed systematic experiments against 4-round, 6-round and 8-round of SmallPresent, corresponding to contexts where the linear hull effect is negligible, starts to play a role and is eventually dominant. For each of these contexts, we computed the gain of different linear cryptanalysis attacks using multiple approximations, with $m = 2^i$ and $i \in [0; 13]$. For clarity purpose, the results of these experiments have been reported in Appendices A, B, C, D. They correspond to four different scenarios that we now detail.

Best characteristics generated with a branch-and-bound algorithm. This scenario, of which the results are given in Appendix A, Figure 9, is the most realistic one, as it corresponds to the only strategy applicable by an actual adversary who does not know the key and has limited computational power to find good characteristics. The figure confirms two important intuitions. First,

[6] As detailed in [43], this can lead to an overestimate of the attack gain, because of an argument similar to the one discussed in Section 2.2. However, as will be detailed next, the main limitation of this estimated data complexity is due to the unprecise knowledge of the linear biases in actual linear cryptanalyses.

the maximum likelihood approach only works as long as the linear characteristic bias $LCB(\Omega, \tilde{K})$ is a good approximation of the sampled linear bias $SLB(\mathbf{a}, \mathbf{b}, 1, N_p; K)$. But as soon as the linear hull effect appears (i.e. for 6 rounds in our experiments), it does not allow reaching high gains anymore. Second, the heuristic approach is successful for a few more rounds than the maximum likelihood one (i.e. up to 6 rounds in our experiments), as it does not require such a precise estimations of the bias. But when increasing the number of rounds, the best characteristics eventually behave as random approximations, as pointed out in Section 2.4, Figure 4. Hence, maximizing the sampled linear bias does not allow recovering secret information anymore in this case.

Best $(R-1)$-round approximations. First, let us mention again that this (and the next) scenario(s) do not correspond to realistic adversaries, as finding the best $(R-1)$-round approximations and having precise estimations of their biases is usually not possible for practically secure ciphers. However, as the previous experiments suggest that having a good estimation of the linear biases is critical for the success of a linear cryptanalysis, it is interesting to observe the behavior of the key testing procedures in an artificial context, where the adversary knows these best $(R-1)$ approximations and their exact bias, for each key. The results of our experiments in this case are plotted in Appendix B, Figure 10, and highlight the following facts. First, the gains of the two key ranking procedures is higher than when using characteristics and remains high even when increasing the number of rounds. Second, the impact of using multiple approximations is stronger when applying the heuristic key ranking. This is in fact directly related to the results of Section 2.1, Figure 2: for the sampled linear bias $SLB(\mathbf{a}, \mathbf{b}, 1, N_p; K)$ to be close to the actual bias $LB(\mathbf{a}, \mathbf{b}; K)$, we need to sample with a number of plaintexts N_p that is close to the full codebook. Hence, the maximum likelihood key ranking strategy only starts to extract information when N_p is large enough, while a heuristic key ranking is less sensitive to this need of good estimates for $LB(\mathbf{a}, \mathbf{b}; K)$. Note that, as discussed in [12], this last observation is particularly strong when using Matsui's second algorithm, because of the partial decryption process which maps the sampled biases at round R towards sampled biases at round $R-1$, with a non-linear S-box. It would be partially relaxed if Matsui's first algorithm was considered.

Random $(R-1)$-round approximations. The previous experiments suggest that the main parameter allowing a successful linear cryptanalysis with a maximum likelihood key ranking is the knowledge of a good estimated bias. In order to confirm this intuition, we launched another set of experiments, with randomly selected input and output masks, for which the exact value of $LB(\mathbf{a}, \mathbf{b}; K)$ was provided to the adversary. As illustrated in Appendix C, Figure 11, this context leads to gains for the maximum likelihood approach that are very close to the ones obtained with the best approximations in the previous paragraph. By contrast, the heuristic key ranking is totally ineffective in this case as the random masks selected do not correspond to maximum linear bias values.

$(R-1)$-round approximations with null bias. Eventually and for completeness, we launched a set of experiments with approximations having null bias. Figure 12 in Appendix D shows that this scenario can also lead to successful key recoveries. Interestingly, it can even work with a modified heuristic approach, provided that one selects the key candidate that minimizes the sampled linear bias (rather than the one maximizing it). Hence, attacks exploiting such approximations could actually be mounted against actual ciphers and constitute a counterpart to the use of impossible differentials in cryptanalysis [4]. They have been recently and independently analyzed by Bogdanov and Rijmen [9].

We conclude this section with two additional observations.

A note on the estimated data complexity of linear cryptanalysis with multiple approximations. As mentioned in footnote 4, the estimation of the gain in function of the data complexity in a linear cryptanalysis using multiple approximations is a difficult problem. One important issue is that most estimates published in the literature, e.g. [5], are based on a maximum likelihood key ranking procedure. But as witnessed by the previous experiments, such an approach is hardly applicable as soon as the linear hull effect in the cipher increases. Unfortunately, this is also the context in which exploiting multiple approximations would be useful in practice, in order to compensate the limited gain of a single approximation. In order to quantify the impact of multiple approximations in linear cryptanalysis, we additionally plotted the gain of attacks based on a heuristic key ranking, using the best approximations (as in Appendix B), in function of the product between the number of plaintexts used in the attack N_p and the capacity of the set of approximations $C(\mathbf{a}^{1:m}, \mathbf{b}^{1:m}; K)$, in Figure 7. If the data complexity was properly estimated with $1/C(\mathbf{a}^{1:m}, \mathbf{b}^{1:m}; K)$, the 14 curves in Figure 7 should be identical. For example, in cases where all the approximations have the same linear bias, it would imply that doubling the number of approximations in the attack would be equivalent to doubling the number of plaintexts, which is clearly not the case in our experiments. On the other hand, the figure also shows that we need approximately $2^5/c$ plaintexts to reach the maximum gain with the best approximation. And we need $2^{10}/c$ plaintexts to reach this gain with all the 2^{13} approximations. If the use of multiple approximations was completely ineffective, we would need $2^{18}/c$ plaintexts in this case (i.e. we would have a factor 2 between each of the curves in Figure 7). Summarizing, the improved data complexity of linear cryptanalysis attacks using multiple approximations falls between these two extremes. In this respect, we note again that these observations are mainly due to the use of Matsui's second algorithm with a heuristic key ranking procedure. But as discussed in [12], experiments performed with Matsui's first algorithm have a data complexity that is much closer to their estimated value with $1/C(\mathbf{a}^{1:m}, \mathbf{b}^{1:m}; K)$.

A note on the stationary area for block ciphers. Let us finally shortly comment on the notion of "stationary area" with respect to linear and differential cryptanalysis, discussed in [46]. Intuitively, a block cipher is in its stationary area if the distribution of its linear biases (and differential probabilities) do not vary

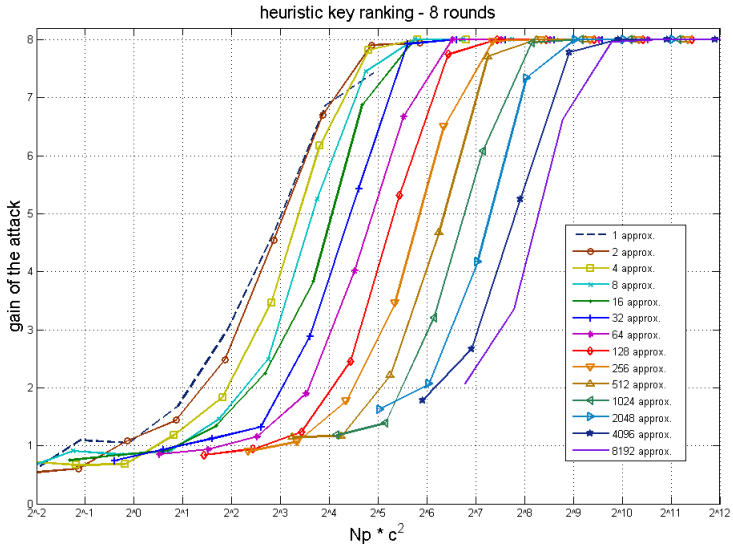

Fig. 7. Gain of linear cryptanalyses using multiple approximations with Matsui's second algorithm, in function of the number of texts N_p multiplied by the capacity $C(\mathbf{a}^{1:m}, \mathbf{b}^{1:m}; K)$, using the 1st to 8192th best approximations generated from the full codebook, for 8-round `SmallPresent` (maximum likelihood key ranking).

anymore with the number of rounds. This typically corresponds to the point where adding more rounds to the block cipher is not useful anymore from the point of view of statistical attacks. As pictured in Figure 6, SmallPresent-[16] becomes stationary after 6 rounds. This should imply that targeting any number of rounds larger than 6 with a given statistical attack should lead to similar gains. For illustration, we plotted in Figure 8 the gains of attacks against 8 and 19 rounds of SmallPresent-[16], with the best linear approximations. It clearly illustrates that attacks have very similar data complexities (experiments could also be launched with the best characteristics, with random approximations or approximations with zero bias and would lead to identical observations).

4 Beyond linear cryptanalysis

Many variations of linear cryptanalysis have been proposed in the literature. While they essentially rely on the same principles, they introduce tweaks that allow improving the effectiveness of the attack. In this section, we briefly survey some important results and their relations with our previous experiments[7].

[7] A number of these advanced strategies are analyzed carefully in the rest of this book.

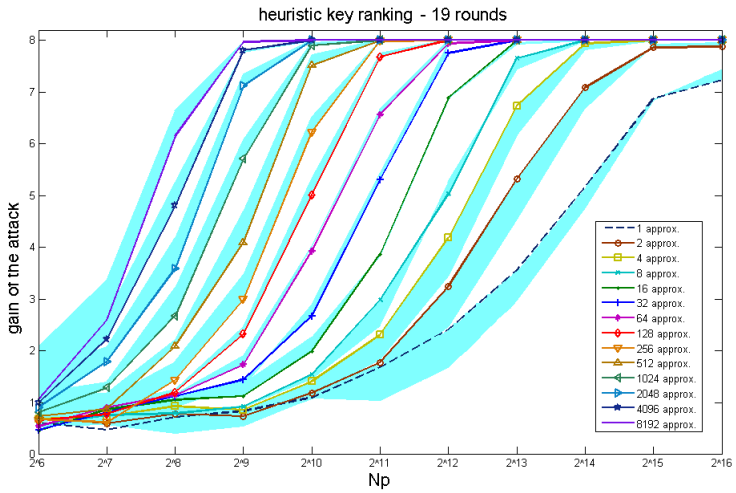

Fig. 8. Gain of linear cryptanalyses using multiple approximations with Matsui's 2nd. algorithm, in fct. of the number of texts N_p, using the 1st to 8192th best approx. generated from the full codebook, for 19-round `SmallPresent` (maximum likelihood key ranking). Shadowed curves show the difference with the results against 8 rounds).

Next to the use of non-linear approximations [31] or chosen plaintexts attacks [32], one of the most frequently investigated ways to extend linear cryptanalysis is to take advantage of non-uniform distributions in the plaintexts or ciphertexts (rather than simply biased linear approximations). This idea was first denoted as partitioning cryptanalysis [19] and directly leads to the question of how to find these non-uniform behaviors in a cipher. Hence, in practice, it usually relies on some specificities that a cryptanalysis may find, e.g. [17, 41]. Partitioning cryptanalysis has been analyzed by Baignères et al. in [1]. More recently, the idea of multidimensional cryptanalysis [20] and its application to linear cryptanalysis using multiple approximations [22] was based on very similar ideas. Quite naturally, these extensions suffer from the same limitations as linear cryptanalysis regarding the difficulty of obtaining precise estimates of the target distributions. In fact, the problem of estimating multidimensional distributions is generally more difficult than the one of estimating linear biases. Hence, these optimal attacks also need to be modified with heuristics when statistical hull effects appear. In this respect, we finally mention statistical saturation attacks as a typical example of such heuristics [13]. This attack exploits the non-uniform behavior of the diffusion layer in the block cipher PRESENT [8], in a chosen plaintext scenario. It can be viewed as a particular case of partitioning (or multidimensional) cryptanalysis in which one simply selects the key candidate that maximizes the distance with a uniform distribution over the m bits targeted in the attack, in order to avoid the need of precise estimated distributions in the key ranking.

5 Conclusion & open problems

This chapter mainly focused on the experimental review of a number of important assumptions used in linear cryptanalysis and its extensions. It highlights the difficulty of predicting the statistical behavior of a block cipher as its number of rounds increases, both for adversaries trying to exploit key-dependent biases, and for designers trying to accurately predict security bounds. As a consequence, our experiments confirm a tension between the practical and provable security approaches for designing block ciphers. They recall that security against linear cryptanalysis attacks is mainly due to the difficulty to find good approximations, and to their key dependency. On the positive side, this makes a case for the practical security approach, as such a key-dependency of the best approximations typically appears when no single characteristic can be used to predict the experimental biases (hence, when the data complexity estimated with single characteristics becomes prohibitive). On the negative side, it also limits our understanding of linear cryptanalysis, e.g. when determining the number of rounds needed for secure block ciphers. In this respect, one central scope for further research is to find efficient solutions for estimating the target distributions in statistical cryptanalyses. Also, and in view of the importance of key dependencies when studying the linear hull effect, it would be interesting to investigate whether there exist classes of keys for which a given approximation would allow successful attacks with high probability and large number of rounds. In other words, are there classes of keys that are weak for a given linear hull? From an adversarial point of view, the counterpart of this tension between theory and practice can be found in the key ranking procedures. Our experiments showed that in a number of contexts, attacks based on heuristics can perform better than maximum likelihood ones, because of imperfect bias estimations. Hence, finding the best heuristics to use in a given scenario is another interesting scope for further research. Let us finally mention that these questions have interesting concrete consequences. They typically relate to the extent to which one can trade data for time in statistical attacks (e.g. using multiple approximations). Hence, they relate to the question of the maximum key size for a fixed block cipher size.

References

1. T. Baignères, P. Junod, S. Vaudenay, *How Far Can We Go Beyond Linear Cryptanalysis?*, in the proceedings of ASIACRYPT 2004, LNCS, vol 3329, pp 432-450, Jeju Island, Korea, December 2004.
2. T. Baignères, M. Finiasz, *Dial C for Cipher*, in the proceedings of SAC 2006, LNCS, vol 4356, pp 76-95, Montreal, Canada, August 2006.
3. T. Baignères, M. Finiasz, *KFC - The Krazy Feistel Cipher*, in the proceedings of ASIACRYPT 2006, LNCS vol 4284, pp 380-395, Shanghai, China, December 2006.
4. E. Biham, A. Biryukov, A. Shamir, *Cryptanalysis of Skipjack Reduced to 31 Rounds Using Impossible Differentials*, in the proceedings of EUROCRYPT 1999, LNCS, vol 1592, pp 1223, Prague, Czech Republic, May 1999.

5. A. Biryukov, C. De Cannière, M. Quisquater, *On Multiple Linear Approximations*, in the proceedings of Crypto 2004, LNCS, vol 3152, pp 1-22, Santa Barbara, California, USA, August 2004.
6. C. Blondeau, B. Gérard, *On the Data complexity of Statistical Attacks Against Block Ciphers*, in the proceedings of the IMA International Conference on Cryptography and Coding 2009, full version available on the IACR ePrint archive: http://eprint.iacr.org/2009/064,
7. C. Blondeau, B. Gérard, *Links Between Theoretical and Effective Differential Probabilities: Experiments on PRESENT*, in the proceedings of the workshop in Tools for Cryptanalysis, pp 109-125, Edgham, UK, June 2010. full version available on the IACR ePrint archive: http://eprint.iacr.org/2010/261.
8. 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 the proceedings of CHES 2007, LNCS, vol 4727, pp 450-466, Vienna, Austia, September 2007.
9. A. Bogdanov, V. Rijmen, *Zero-Correlation Linear Cryptanalysis of Block Ciphers*, Cryptology ePrint Archive: Report 2011/123, http://eprint.iacr.org/2011/123.
10. B.Collard, F.X. Standaert, J.J. Quisquater, *Improved and Multiple Linear Cryptanalysis of Reduced Round Serpent*, in the proceedings of InsCrypt 2007, LNCS, vol 4990, pp 51-65, Xining, China, September 2007.
11. B. Collard, F.-X. Standaert, J.-J. Quisquater, *Improving the Time Complexity of Matsui's Linear Cryptanalysis*, in the proceedings of ICISC 2007, LNCS, vol 4817, pp 77-88, Seoul, Korea, November 2007.
12. B. Collard, F.-X. Standaert, J.-J. Quisquater, *Experiments on the Multiple Linear Cryptanalysis of Reduced Round Serpent*, in the proceedings of FSE 2008, LNCS, vol 5086, pp 382-397, Lausanne, Switzerland, February 2008.
13. B. Collard, F.-X. Standaert, *A Statistical Saturation Attack against the Block Cipher PRESENT*, in the proceedings of CT-RSA 2009, LNCS, vol 5473, pp 195-210, San Francisco, USA, April 2009.
14. J. Daemen, V. Rijmen, *The Wide Trail Design Strategy*, in the proceedings of the IMA International Conference on Cryptography and Coding 2001, LNCS, vol 2260, pp 222-238, Cirencester, UK, December 2001.
15. J. Daemen, V. Rijmen, *The Design of Rijndael: AES - The Advanced Encryption Standard*, Springer 2002.
16. J. Daemen, V. Rijmen, *Probability Distributions of Correlation and Differentials in Block Ciphers*, Journal of Mathematical Cryptology, vol 1, pp 12-35, 2007.
17. H. Gilbert, H. Handschuh, A. Joux, S. Vaudenay, *A Statistical Attack on RC6*, in the proceedings of FSE 2000, LNCS, vol 1978, pp 64-74, New York, USA, April 2000.
18. C. Harpes, G.G. Kramer, J.L. Massey, *A Generalization of Linear Cryptanalysis and the Applicability of Matsui's Piling-Up Lemma*, in the proceedings of Eurocrypt 1995, LNCS, vol 921, pp 24-38, Saint-Malo, France, May 1995.
19. C. Harpes, J.L. Massey, *Partitioning Cryptanalysis*, in the proceedings of FSE 1997, LNCS, vol 1267, pp 13-27, Haifa, Israel, January 1997.
20. M. Hermelin, J.Y. Cho, K. Nyberg, *Multidimensional Extension of Matsui's Algorithm 2*, in the proceedings of FSE 2009, LNCS, vol 5665, pp 209-227, Leuven, Belgium, February 2009.
21. M. Hermelin, J.Y. Cho, K. Nyberg, *Multidimensional Linear Cryptanalysis of Reduced Round Serpent*, in the proceedings of ACISP 2008, LNCS, vol 5107; pp 203-215, Wollollong, Australia, July 2008.

22. M. Hermelin, K. Nyberg, *Dependent Linear Approximations: The Algorithm of Biryukov and Others Revisited*, in the proceedings of CT-RSA 2010, LNCS, vol 5985, pp 318-333, San Franciscon California, USA, February 2010.
23. P. Junod, *On the Complexity of Matsui's Attack*, in the proceedings of SAC 2001, LNCS, vol 2259, pp 199-211, Toronto, Ontario, August 2001.
24. P. Junod, *On the Optimality of Linear, Differential, and Sequential Distinguishers*, in the proceedings of Eurocrypt 2003, LNCS, vol 2656, pp 17-32, Warsaw, Poland, May 2003.
25. P. Junod, S. Vaudenay, *Optimal Key Ranking Procedures in a Statistical Cryptanalysis* in the proceedings of FSE 2003, LNCS, vol 2887, pp 235-246, Lund, Sweden, February 2003.
26. B.S. Kaliski, M.J.B. Robshaw, *Linear Cryptanalysis using Multiple Approximations*, in the proceedings of CRYPTO 1994, LNCS, vol 839, pp. 26-39, Santa Barbara, California, USA, August 1994.
27. L. Keliher, H. Meijer, S.E. Tavares, *New Method for Upper Bounding the Maximum Average Linear Hull Probability for SPNs*, in the proceedings of Eurocrypt 2001, LNCS, vol 2045, pp 420-436, Innsbruck, Austria, May 2001.
28. L. Keliher, H. Meijer, S.E. Tavares, *Improving the Upper Bound on the Maximum Average Linear Hull Probability for Rijndael*, in the proceedings of SAC 2001, LNCS, vol 2259, pp 112-128, Toronto, Ontario, Canada, August 2001.
29. L. Keliher, *Linear Cryptanalysis of Substitution-Permutation Networks*, PhD Thesis, Queen's University, Kingston, Ontario, Canada, October 2003.
30. L.R. Knudsen, *Practically Secure Feistel Ciphers*, in the proceedings of FSE 1993, LNCS, vol 809, pp 211-221, Cambridge, UK, December 1993.
31. L.R. Knudsen, M.J.B. Robshaw, *Non-Linear Approximations in Linear Cryptoanalysis*, in the proceedings of EUROCRYPT 1996, LNCS, vol 1070, pp 224-236, Saragossa, Spain, May 1996.
32. L.R. Knudsen, J.E. Mathiassen, *A Chosen-Plaintext Linear Attack on DES*, in the proceedings of FSE 2000, LNCS, vol 1978, pp 262-272, New York, USA, April 2000.
33. L.R. Knudsen, J.E. Mathiassen, *On the Role of Key Schedules in Attacks on Iterated Ciphers*, in the proceedings of ESORICS 2004, LNCS, vol 3193, pp 322-334, Sophia Antipolis, France, September 2004.
34. X. Lai, J.L. Massey, *Markov Ciphers and Differentail Cryptanalysis*, in the proceedings of EUROCRYPT 1991, LNCS, vol 547, pp 17-38, Brighton, UK, April 1991.
35. G. Leander, *Small Scale Variants of the Block Cipher PRESENT*, Cryptology ePrint Archive, Report 2010/143, http://eprint.iacr.org/, 2010.
36. M. Matsui, *Linear cryptanalysis method for DES cipher*, in the proceedings of Eurocrypt 1993, LNCS, vol 765, pp 386-397, Lofthus, Norway, May 1993.
37. M. Matsui, *The First Experimental Cryptanalysis of the Data Encryption Standard*, in the proceedings of Crypto 1994, LNCS, vol 839, pp 1-11, Santa Barbara, California, USA, August 1994.
38. M. Matsui, *On Correlation Between the Order of S-boxes and the Strength of DES*, in the proceedings of Eurocrypt 1994, LNCS, vol 950, pp 366-375, Perugia, Italy, May 1994.
39. M. Matsui, *New Structure of Block Ciphers with Provable Security against Differential and Linear Cryptanalysis*, in the proceedings of FSE 1996, LNCS, vol 1039, pp 205-218, Cambridge, IK, February 1996.
40. M. Matsui, *New Block Encryption Algorithm MISTY*, in the proceedings of FSE 1997, LNCS, vol 1267, pp 54-68, Haifa, Israel, January 1997.

41. M. Minier, H. Gilbert, *Stochastic Cryptanalysis of Crypton*, in the proceedings of FSE 2000, LNCS, vol 1978, pp 121-133, New York, USA, April 2000.
42. S. Murphy, *The Effectiveness of the Linear Hull Effect*, Royal Holloway University of London, Technical Report RHUL-MA-2009-19, 2009.
43. S. Murphy, *Overestimates for the Gain of Multiple Linear Approximations*, Royal Holloway University of London, Technical Report RHUL-MA-2009-21, 2009.
44. K. Nyberg, *Linear Approximations of Block Ciphers*, in the proceedings of Eurocrypt 1994, LNCS, vol 950, pp 439-444, Perugia, Italy, May 1994.
45. K. Nyberg, L.R. Knudsen, *Provable Security against Differential Cryptanalysis*, Journal of Cryptology, vol 8, num 1, pp 27-37, 1995.
46. G. Piret, F.-X. Standaert, *Provable Security of Block Ciphers Against Linear Cryptanalysis - a Mission Impossible?*, Designs, Codes and Cryptography, vol 50, num 3, pp 325-338, March 2009.
47. V. Rijmen, *Cryptanalysis and Design of Iterated Block Ciphers*, PhD Thesis, Katholieke Universiteit Leuven, Belgium, October 1997.
48. T. Ritter, *Measuring boolean function nonlinearity by Walsh transform*, `http://www.ciphersbyritter.com/ARTS/MEASNONL.HTM`.
49. G. Rouvroy, F.-X. Standaert, J.-J. Quisquater, J.-D. Legat, *Efficient Uses of FPGAs for Implementations of DES and Its Experimental Linear Cryptanalysis*, IEEE Transactions on Computers, vol 52, num 4, pp 473-482, 2003.
50. A.A. Selçuk, *On Bias Estimation in Linear Cryptanalysis*, in the proceedings of Indocrypt 2000, LNCS, vol 1977, pp 52-66, Calcutta, India, December 2000.
51. A.A. Selçuk, *On Probability of Success in Linear and Differential Cryptanalysis*, Journal of Cryptology, vol 21, num 1, pp 131-147, 2008.
52. S. Vaudenay, *Decorrelation: A Theory for Block Cipher Security*, Journal of Cryptology, vol 16, num 4, pp 249-286, 2003.

A Best characteristics

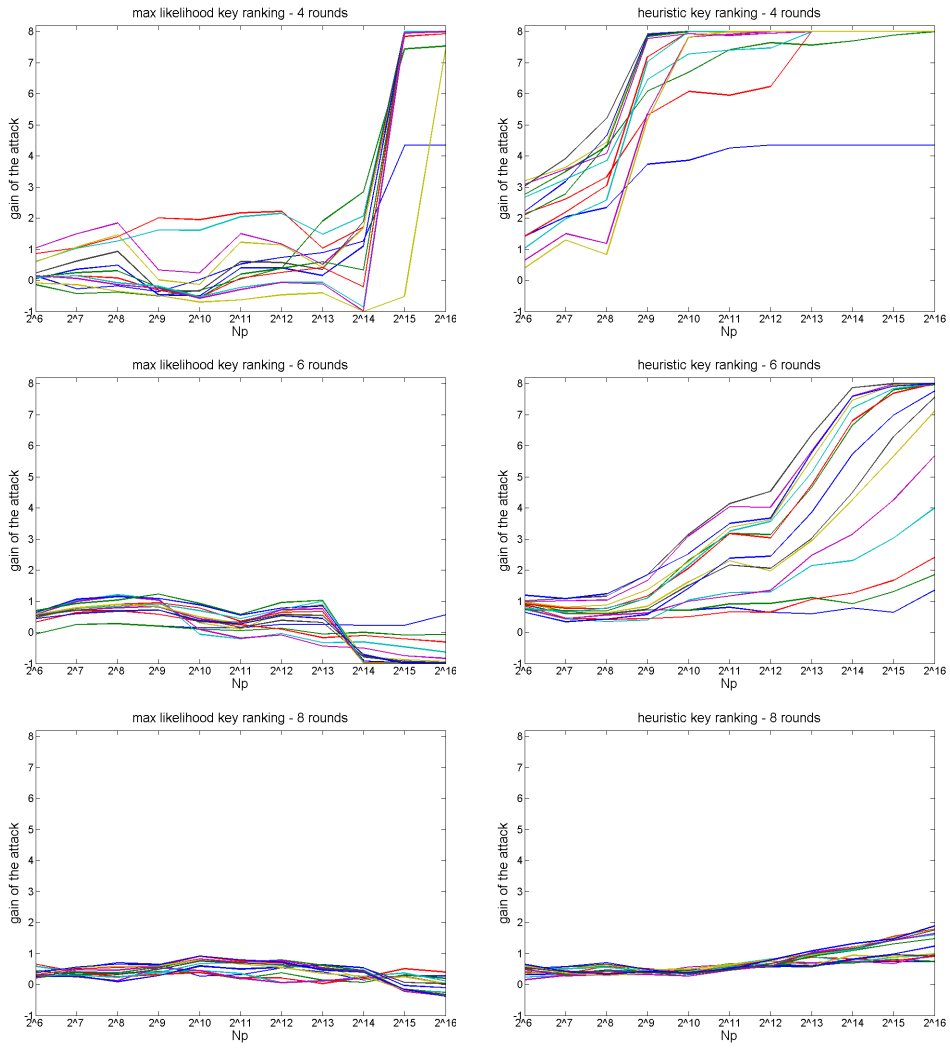

Fig. 9. Gain of linear cryptanalyses using multiple approximations with Matsui's second algorithm, in function of the number of texts N_p, using 1 to 8192 characteristics generated with a branch-and-bound algorithm, for 4-round, 6-round and 8-round `SmallPresent`. Left: maximum likelihood key ranking. Right: heuristic key ranking.

B Best approximations

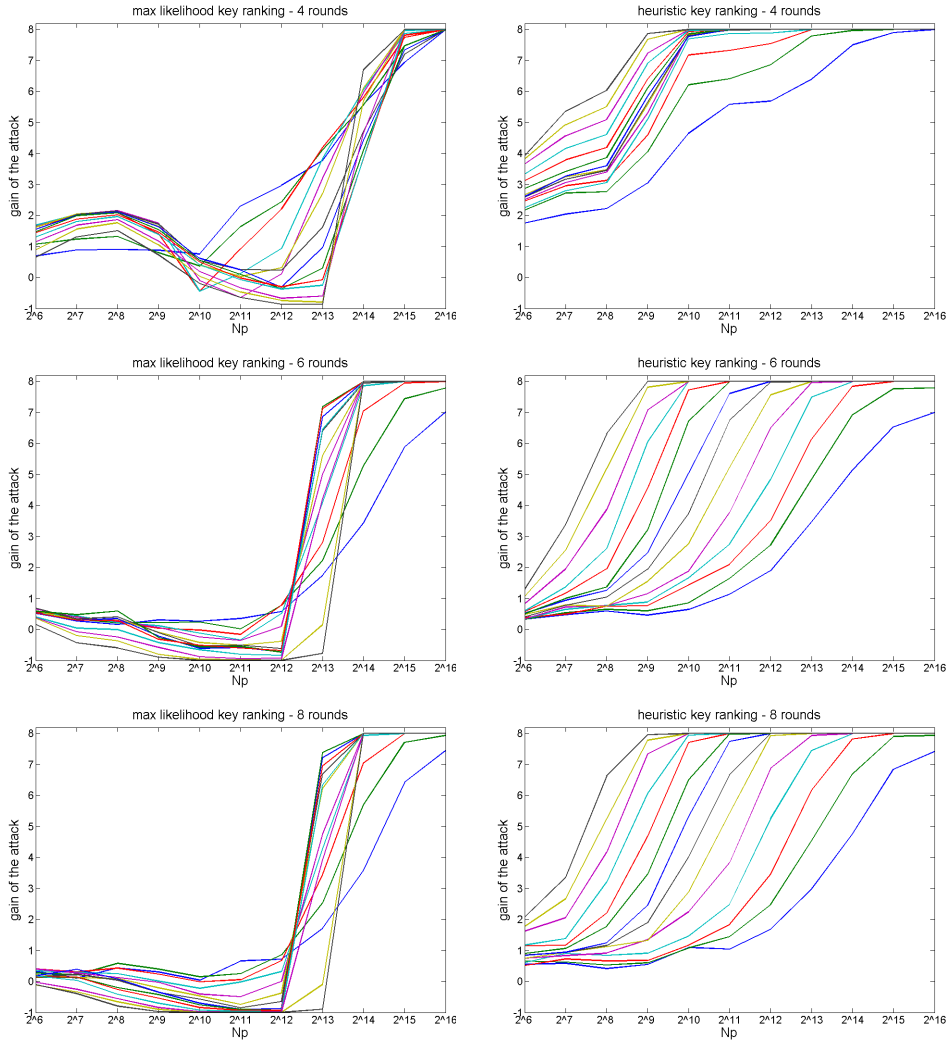

Fig. 10. Gain of linear cryptanalyses using multiple approximations with Matsui's second algorithm, in function of the number of texts N_p, using the 1st to 8192th best approximations generated from the full codebook, for 4-round, 6-round and 8-round `SmallPresent`. Left: maximum likelihood key ranking. Right: heuristic key ranking.

C Random approximations

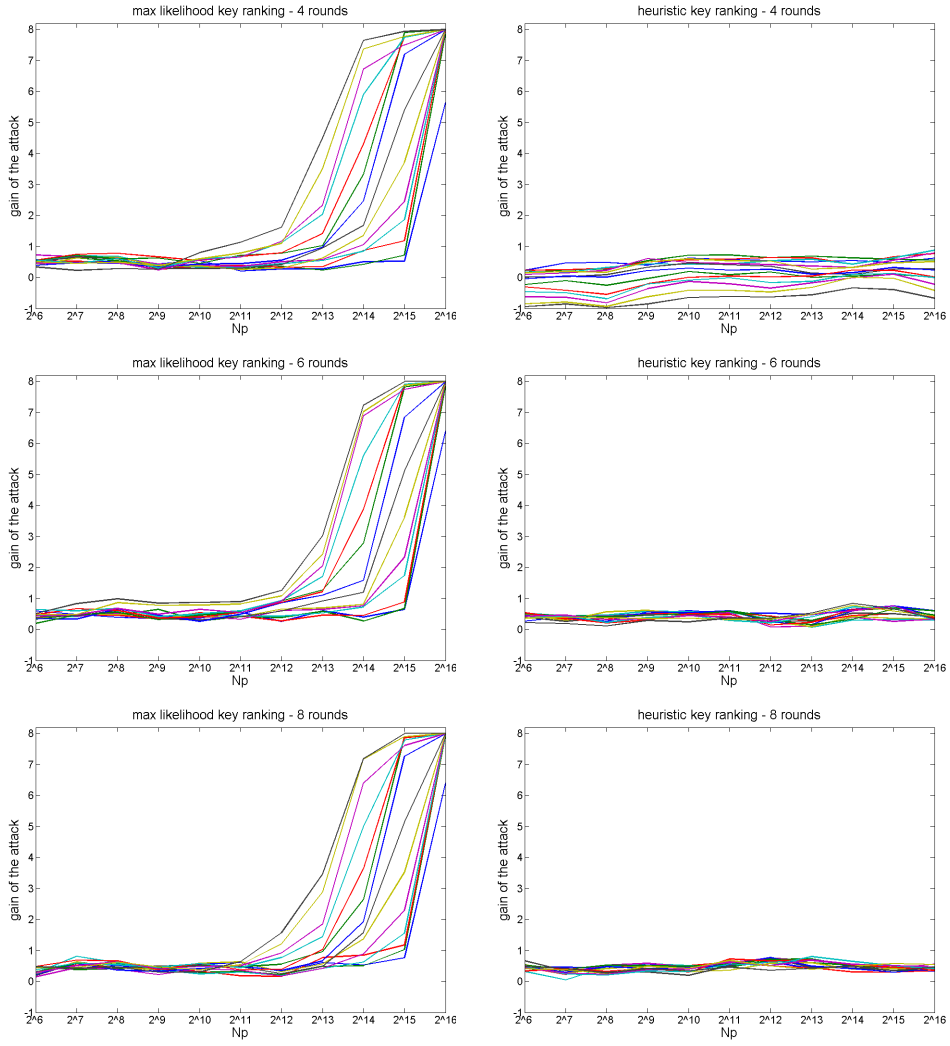

Fig. 11. Gain of linear cryptanalyses using multiple approximations with Matsui's second algorithm, in function of the number of texts N_p, using 1 to 8192 random approximations with biases estimated from the full codebook, for 4-, 6- and 8-round `SmallPresent`. Left: maximum likelihood key ranking. Right: heuristic key ranking.

D Approximations with null bias

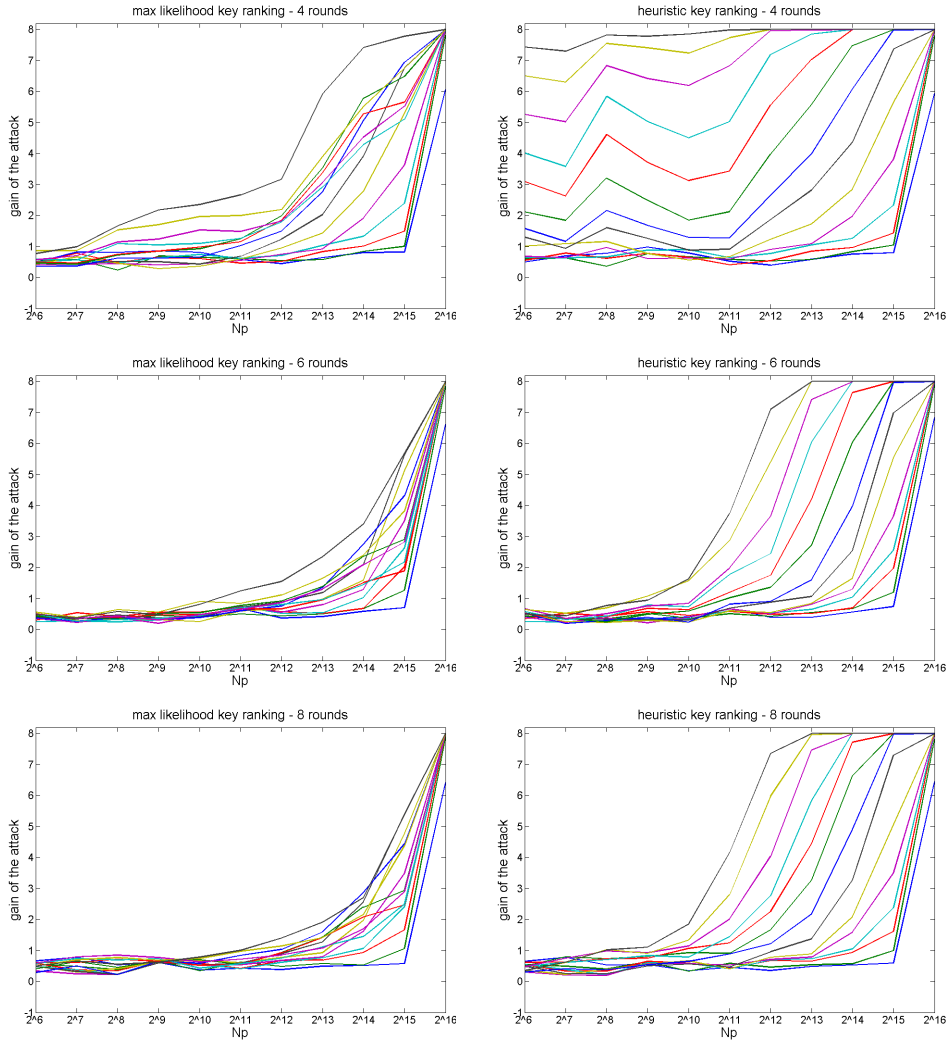

Fig. 12. Gain of linear cryptanalyses using multiple approximations with Matsui's second algorithm, in function of the number of texts N_p, using 1 to 8192 approximations with null bias estimated from the full codebook, for 4-round, 6-round and 8-round `SmallPresent`. Left: maximum likelihood key ranking. Right: heuristic key ranking.