IACR Transactions on Symmetric Cryptology
ISSN 2519-173X, Vol. 2017, No. 3, pp. 169–191. DOI:10.13154/tosc.v2017.i3.169-191

Accurate Estimate of the Advantage of Impossible Differential Attacks

Céline Blondeau

Department of Computer Science, Aalto University, Helsinki, Finland
celine.blondeau@aalto.fi

Abstract. Impossible differential attacks, which are taking advantage of differentials that cannot occur, are powerful attacks for block cipher primitives. The power of such attacks is often measured in terms of the advantage — number of key-bits found during the key sieving phase — which determines the time complexity of the exhaustive key search phase. The statistical model used to compute this advantage has been introduced in the seminal work about the resistance of the DEAL cipher to impossible differential attacks. This model, which has not been modified since the end of the 1990s, is implicitly based on the Poisson approximation of the binomial distribution.
In this paper, we investigate this commonly used model and experimentally illustrate that random permutations do not follow it. Based on this observation, we propose more accurate estimates of the advantage of an impossible differential attack. The experiments illustrate the accuracy of the estimate derived from the multivariate hypergeometric distribution. The maximal advantage –using the full codebook– of an impossible differential attack is also derived.

Keywords: impossible differential · data complexity · time complexity · advantage · binomial distribution · multivariate distribution · multivariate hypergeometric distribution

1 Introduction

Impossible differential cryptanalysis has been introduced in the late 90's [Knu98] when analyzing the security of a new design, the DEAL cipher. The idea behind this attack is to take advantage of differentials which never appear for a given permutation. Since its introduction, the security of block ciphers [Knu98, BBS99a, BBS99b, SKU+00, BNS14a, Der16, Tod16] is analyzed with respect to this attack which is particularly powerful on word-oriented ciphers.

In the seminal papers [Knu98, BBS99a], an estimate of the data complexity of this attack, generalization of differential cryptanalysis, is presented. This estimate, which relies on the probability of not finding the targeted differentials for a random permutation, is based on an approximation of the binomial distribution. As for most statistical attacks, the online part of an impossible attack is divided into three phases, which are data generation, key sieving, and exhaustive key search. The time complexity of the attack is determined by the time needed to perform these three phases. Their relation has been summarized in [BNS14a]. In particular, the exhaustive key search time complexity is directly linked to the advantage which corresponds to the number of key-bits won during the key sieving phase.

Recently, Boura et al. [BNS14a, BNS14b, BLNPS17] proposed generic formulas for estimating the time complexity of an impossible differential attack. Using the relation between the key-bits involved in the attack, their analysis shows that the time complexity

Received: 2017-06-01, Accepted: 2017-08-01, Published: 2017-09-19

of some previous attacks was wrongly estimated. In the same papers, they provide a generic method to estimate the time complexity of the key sieving phase. The limit of this generic approach, which is accurate for many impossible differential attacks, is discussed in [DF16]. The concept of multiple impossible differential attacks is also introduced in [BNS14a, BLNPS17].

1.1 Motivation

The data complexity of a key recovery attack influences the time complexity of the three different phases. For the second phase, the time complexity is related to the number of pairs which are generated from the used messages and to the number of key candidates required to perform the partial encryption/decryption. The exhaustive key search time complexity is determined by the advantage we get from the used messages. From this point of view analyzing the relation between the advantage and the data complexity of an impossible differential attack is of great importance. The work presented in this paper focuses on this relation and completes the recent works [BNS14b, BNS14a, Der16] with respect to the time complexity of the key sieving phase.

The relation between data complexity and advantage of the attack is given by a statistical model. As seen recently for other statistical attacks [BT13, BN14, BN17, BTV16], providing a correct statistical model is of great importance for estimating the data and time complexity of the attack. For a better understanding of the model, experiments are usually performed. Impossible differential attacks are in a way different from other known statistical attacks, since by construction, the considered differentials are known to never appear for the given permutation. Meaning that for an attack using N plaintexts, only the ratio of permutations for which none of the targeted differentials is fulfilled is of interest.

In the impossible differential context, the classical statistical model is derived from the Poisson distribution approximating the binomial distribution. The preliminary experiments showed that this model does not always provide an accurate estimate of the advantage. By considering a model which allows repetition we show that the advantage of most impossible differential attacks was overestimated.

1.2 Contribution

Sampling over the plaintexts Classical statistical models for attacks in the differential context [BS90] take as sample a pair of messages fulfilling a particular difference. In this paper, we consider a sample as a single plaintext combined with its corresponding ciphertext. As explained in Section 3 to count the number of pairs which fulfill the differential we store in a multivariate vector some information on the encrypted ciphertexts.

Using this multivariate vector we explain how the problem of finding differentials is similar to the problem of finding collisions. While this problem is known as the birthday paradox problem, we show in this paper that this approach does not provide a good estimate of the advantage of an impossible differential attack.

Using the Multivariate Hypergeometric Distribution Impossible differential attacks are usually chosen plaintext attacks, and when implemented, the selected plaintexts are selected as distinct. However, the classical statistical model assumes that the involved random variables follow a binomial distribution. This is, in particular, the case for the random variables simulating the behavior of a differential for the wrong keys.

As shown recently in the linear context [BLNW12, BN17], when using distinct samples the theory should be handled with the hypergeometric distribution instead of the binomial distribution.

Based on the two previous remarks, we use in this paper the multivariate hypergeometric distribution to estimate the advantage of an impossible differential attack. The accuracy

of this new model is illustrated by experiments on random permutations. For educational matter, the theory developed in this paper is first explained for a distinguishing attack using only one structure in Theorem 1 and Corollary 1 and generalized to the case of multiple structures in Theorem 2. The performed experiments illustrate the accuracy of the new model.

If the impossible differential distinguisher is such that the number of involved input and output differentials is relatively small in comparison to the cipher size, the advantage estimate provided in Theorem 3 is not more difficult to compute than the previous estimate.

Using the Bi-Multivariate Hypergeometric Distribution The analysis based on the multivariate hypergeometric distribution assumes independence of the structures involved in the attack. As explained in this paper, this is not true in practice when the attack requires a large number of structures. This assumption can be removed if we instead use the bi-multivariate hypergeometric distribution to estimate the advantage of an impossible differential attack. This extremely accurate model, introduced in Section 4, is relatively difficult to handle in practice. However, we use it to derive the maximal advantage of an impossible differential attack and to illustrate the relative accuracy of the estimate derived using the multivariate hypergeometric distribution.

Impact on Existing Attacks After recalling the link between the data complexities of a distinguishing and a key recovery attack, we present the impact of our results on the time complexity of existing impossible differential attacks. In particular, we show that the maximal advantage of an impossible differential attack involving only one differential (as it is the case for the recent attacks on SIMON [BNS14a, DF16]), was optimistically computed and correct it from 1.4 to 0.72. The time complexity of the impossible differential attacks on SIMON is then closer to the time of the brute force attack. The case of multiple impossible differentials introduced in [BNS14a] is also developed in this paper.

Impact on Differential and Truncated Differential Attacks From the same theory, we could estimate the advantage of differential and truncated differential attacks. In this paper, we explain how we can derive the advantage of classical differential attacks when the parameters of the attack are such that the used threshold is small.

1.3 Outline

The outline of this paper is as follows. An introduction to impossible differential attacks is given in Section 2. In Section 3 we introduce the multivariate hypergeometric distribution and propose an accurate estimate of the advantage of an impossible differential attack. In Section 4 we develop on the bi-multivariate hypergeometric distribution and discuss the accuracy of the estimate provided in Section 3. Some remarks related to the known plaintext and known ciphertext impossible differential attacks as well as on attacks involving multiple truncated impossible differentials are provided in Section 5. The impact on key recovery attacks is developed in Section 6. In Section 7 we discuss how this result could impact the advantage of differential and truncated differential attacks. In Section 8 we conclude this paper.

2 Definitions and Related Work

2.1 Definitions

Given an integer d, we denote by $d!$ the factorial of d with the convention $0! = 1$. We denote by $\log_2(\cdot)$ the logarithm in base 2 and by $\ln(\cdot)$ the natural logarithm in base e and

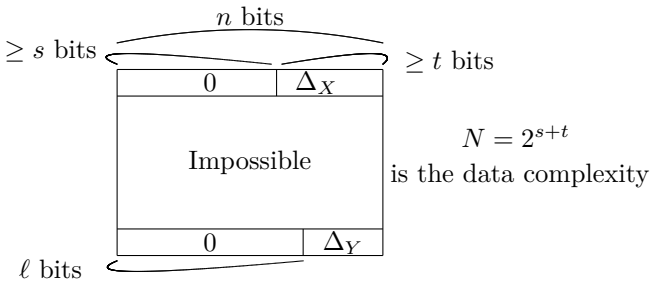

Figure 1: Notation used in this paper for an impossible differential distinguisher (Δ_X, Δ_Y)

by exp[·] its inverse.

We denote by E'_K a part of an n-bit block cipher. A truncated differential (Δ_X, Δ_Y) is said to be impossible if for all plaintext pairs $(x, x \oplus \delta_x)$ with $\delta_x \in \Delta_X \setminus \{0\}$, the ciphertext pairs $(y, y') = (E'_K(x), E'_K(x \oplus \delta_x))$ satisfy $\delta_y = y \oplus y' \notin \Delta_Y$.

In the first part of this paper, we assume that the sets, Δ_X and Δ_Y, of input and output differences are linear spaces. Otherwise, we can usually describe the impossible differential in terms of multiple truncated impossible differential and we refer to Subsection 3.4. We denote by $|\Delta_Y|$ the number of output differences (including the difference 0) and by ℓ the number of output bits classically equal to 0. More generally we have $\ell = n - \log_2(|\Delta_Y|)$. We take similar notation into use for the input difference space. In particular the linear space Δ_X of input differences includes the difference 0 and $|\Delta_X| \geq 2$. For instance, when only one input difference δ_x is in practice considered, Δ_X is the linear space $\{0, \delta_x\}$. A summary of the used notation is given in Figure 1.

Using N_S plaintext pairs, $(x, x \oplus \delta_x)$ with $\delta_x \in \Delta_X \setminus \{0\}$, we can distinguish E'_K from a random permutation if we find a ciphertext pair $(y, y \oplus \delta_y)$ such that $\delta_y \in \Delta_Y$. In this case we are certain that the oracle producing the ciphertexts is not E'_K.

The complexity of an impossible differential distinguisher is related to the number of plaintext pairs $(x, x \oplus \delta_x)$ necessary to find a pair $(y, y \oplus \delta_y)$ such that $\delta_y \in \Delta_Y$ for a random permutation. We denote by N_S the number of plaintext pairs used in an attack.

The data complexity N corresponds to the number of messages needed to generate the N_S plaintexts pairs. When $|\Delta_X| > 2$, taking advantage of the multiple input differences, this complexity can be smaller than the number of generated pairs. In general we define a structure as a subset of Δ_X of size $2 \leq W \leq |\Delta_X|$. We denote by $t = \log_2(W)$. While $2^t = W$ should be an integer value, t is not necessarily an integer number. Inside a structure we can generate $2^{t-1}(2^t - 1)$ pairs (x, x') satisfying $x \oplus x' \in \Delta_X \setminus \{0\}$.

For an attack, we might want to use multiple disjoint structures. We denote by 2^s the number of used structures. From $N = 2^{s+t} \leq 2^n$ plaintexts, we can then generate $N_S = 2^{s+t-1}(2^t - 1)$ plaintext pairs.

If the data complexity is larger than $|\Delta_X|$, to maximize the advantage, t is usually chosen equal to $\log_2(|\Delta_X|)$. When $|\Delta_X| = 2$, $t = 1$ and the number N_S of plaintext pairs used in the attack are generated from $N = 2N_S$ messages.

2.2 Related Work: Advantage of an Impossible Differential Distinguisher

In the impossible differential context, given N_S plaintext pairs, the false alarm error probability β corresponds to the ratio of random permutations for which the truncated differential (Δ_X, Δ_Y) is not fulfilled. Throughout this paper, we denote by $\beta_{\mathcal{X}}$ the false alarm error probability obtained under the assumption that the involved statistical

distribution is $\mathcal{X}$. We also provide estimates of the different $\beta_{\mathcal{X}}$ that we denote by $\tilde{\beta}_{\mathcal{X}}$. The advantage of the attack is simply $a = -\log_2(\beta)$. From this advantage, the time complexity of the exhaustive key search is 2^{k-a} encryptions where k is the key length. We denote by $a_{\mathcal{X}}$ the advantage associated to $\beta_{\mathcal{X}}$. Experimentally observed advantages will be denoted by $\hat{a}$.

In practice, in the online part of an impossible differential attack, we reject a key or decide that we are not dealing with the targeted cipher as soon as a pair fulfilling the truncated differential is found. For the theoretical analysis, in this paper, we use a scoring function T counting the number of occurrences of the truncated differential (Δ_X, Δ_Y) for a random permutation. In the context of impossible differential attacks, we focus on the probability that $T \geq 1$ or alternatively that $T = 0$.

In the seminal publication [Knu98], it is assumed that the random variable associated to T follows a binomial distribution with parameters $(N_S, p = 2^{-n}|\Delta_Y| = 2^{-\ell})$. For clarity, we denote this random variable by $T_{\mathcal{B}}$ and, by definition, the portion of random permutations for which the truncated differential is not fulfilled is

$$\beta_{\mathcal{B}} = P[\, T_{\mathcal{B}} = 0\,] = \binom{N_S}{0} p^0 (1-p)^{N_S} = (1-p)^{N_S}, \tag{1}$$

with $p = 2^{-\ell}$. Using the accurate approximation $(1-p)^{N_S} \approx \exp[-pN_S]$, the false alarm error probability is usually estimated to

$$\tilde{\beta}_{\mathcal{B}} = \exp[-pN_S]. \tag{2}$$

As recalled in [BNS14a], the relation between N_S and N depends if the data complexity is smaller than $|\Delta_X|$ or not. However, in general we have $N = \min(\sqrt{N_S}, N_S \cdot 2^{-|\Delta_X|+1})$. In this paper, we use the value s and t to compute the data complexity. We assume that $2^t \leq |\Delta_X|$ messages form a structure. Using 2^s structures with $s \leq n-t$, the data complexity is $N = 2^{s+t}$ from which, as explained in the previous section, $N_S = 2^{s+t-1}(2^t - 1)$ pairs can be generated. In this paper we do not use the classical approximation $(2^t - 1) \approx 2^t$. A discussion on the influence of this approximation on the advantage of an impossible differential attack when t is small is provided in Subsection 3.6 and Subsection 6.3.

2.3 Experiments and Motivation

Throughout this paper, we compare different estimates of β with some experimental ones obtained from 2^{26} 12-bit random permutations. As illustrated in the introduction of [BNS14a], it is usual to deal with sets Δ_X and Δ_Y of relatively small size therefore we usually assume that ℓ is large $(2^\ell = 2^n/|\Delta_Y|)$.

The experiments[1] of this paper are done as follow. From a PRNG we generate 2^{26} random permutations (independant of any classical cipher structure). For each permutation, we check if a truncated differential is fulfilled or not when using N distinct messages.

Before developing the new theory, we performed experiments on 12-bit random permutations with an impossible differential distinguisher satisfying $\ell = 9$ and $t = 5$. When the data complexity is $N = 2^{4+5}$ $(s = 4)$, using the theory, recalled in Subsection 2.2, we computed that the advantage should approximatively be of $\tilde{a}_{\mathcal{B}} = -\log_2(\tilde{\beta}_{\mathcal{B}}) = 22.36$ bits while experimentally derived advantages average to $\hat{a} = 20.03$ bits. For this example, the time complexity of the exhaustive key search is 4 times slower than expected.

In this paper we solve the question of providing a better estimate of this advantage. Further experiments on random permutations are provided in the different parts of this paper. Experiments illustrating the accuracy of the estimate on a key-recovery attack of a toy cipher are presented in Subsection 6.2.

[1] The experiments can be found at https://users.ics.aalto.fi/blondeau/exp_ID.zip

3 Using Multivariate Distributions

3.1 The Multivariate Vector

The previously derived statistical model assumes that for random permutations, the samples – pairs of messages with given difference – are binomially distributed. Meaning that it is assumed that the plaintext pairs are independent and that the sampling is done assuming replacement. In practice these assumptions are too strong.

In this section we explain how the assumption on the plaintext pairs can be relaxed to an assumption on the plaintexts themselves as it is done in the linear context. The question of replacement is also partially handled in this section. In particular we assume that there is no replacement of the messages drawn inside a given structure.

To count the number of occurrences of truncated differentials, the following approach is, in general, taken. We first encrypt all the plaintexts of a given structure, store the intermediate values and check for all the possible pairs if their ciphertext difference is in Δ_Y or equivalently equal to 0 on the ℓ targeted bits. In [BN14], to underline the link between multidimensional linear attacks and truncated differential attacks, it is suggested to count the number of occurrences of fulfilled truncated differences without comparing the pairs. For most classical parameters this method, which is given in the next lemma, is more time consuming than the classical one. In this paper, we use this approach to provide a theoretical analysis of the advantage of an impossible differential attack.

Lemma 1. *For plaintexts in a given structure (numbered j), subset of Δ_X, we can count the number of pairs fulfilling the truncated differential (Δ_X, Δ_Y) using the following approach:*

- *Define a table L of size 2^ℓ.*
- *Store in $L[i]$ the number of ciphertexts which have value i on these ℓ bits.*
- *The number of ciphertext pairs fulfilling the truncated differential (Δ_X, Δ_Y) is obtained by computing $S_j = \sum_{i=0}^{2^\ell-1} \left(L[i](L[i]-1)\right)/2$.*

When using multiple structures, the number T of pairs fulfilling the truncated differential is obtained by summing up the previous results $T = \sum_{j=0}^{2^s-1} S_j$.

It is easy to see that if only one ciphertext has a given value on these ℓ bits then $L[i](L[i]-1)/2 = 0$, meaning that no pairs have equal value on these ℓ bits.

In the impossible differential context, false alarms are brought by the random permutations for which, using N messages, we are not able to find a pair fulfilling the (truncated) differential. Using the previously defined approach, the problem of finding pairs is equivalent to the problem of finding collisions in the vector L.

If we draw more than 2^ℓ distinct messages, we necessarily have a collision in L, meaning that a truncated differential can only be impossible if $|\Delta_X| \leq 2^\ell$. Therefore, in this paper, we only consider sets Δ_X satisfying the previous relation.

In the next section, we analyze the advantage of an impossible differential attack using the multinomial and the multivariate hypergeometric distributions. These distributions which are generalizations of respectively the binomial and the hypergeometric distribution can be modeled as follow. Given an urn with balls of different colors, these distributions focus on the probability to obtain a certain repartition of the color after a certain number of draws. As for the two colors case, the multinomial distribution assumes replacement of the balls and the multivariate hypergeometric assumes a non-replacement of the balls.

The problem of finding a collision is usually referred as the birthday paradox. In Subsubsection 3.2.2 we provide details on this approach, in particular, we explain how the birthday bound is derived from the assumption that each $L[i]$ follows a binomial distribution. However, since a structure is a finite set, we explain in the same section

that this approach does not provide a good estimate of the advantage. We then prefer to first introduce the accurate model assuming that each $L[i]$ follows an hypergeometric distribution when the plaintexts inside a structure are distinct.

We first provide the explanation when considering only one structure and, in this case, we identify S_j as S.

3.2 Using Only one Structure

3.2.1 Theory in the Multivariate Hypergeometric Case

As explained in Subsection 2.2, to estimate the advantage of an impossible differential attack we should analyze the distribution of the random variable associated to the score value S. In this section, we consider that the vector L (defined in Lemma 1) follows a multivariate hypergeometric distribution and we denote the random variable associated to S by $S_{\mathcal{MH}}$. In the impossible differential context, only the probability of not finding a pair in Δ_Y is of interest. For a random cipher we have $P[\, S_{\mathcal{MH}} = 0\,] = P[\, \forall i \;\; L[i] \leq 1\,]$. In particular, if we take 2^t messages inside a structure we have $P[\; S_{\mathcal{MH}} = 0\;] = P[\,\text{exactly } 2^t \text{ values of } L \text{ are } 1\,]$.

In practical attacks plaintexts are used only once and therefore the hypergeometric distribution is the most accurate generalization of the Bernoulli distribution. When multiple outputs are involved in the impossible differential attacks, we consider the multivariate version of this distribution and we derive the following estimate of the false alarm error probability. As illustrated in Table 1, this estimate is accurate.

Theorem 1. *Given a truncated differential (Δ_X, Δ_Y) with $|\Delta_Y| = 2^{n-\ell}$. We assume that the random variable associated to the vector L described in Lemma 1 follows a multivariate hypergeometric distribution. Using $2^t \leq |\Delta_X|$ messages inside a structure, the false alarm error probability $\beta_{\mathcal{MH}}$ can be computed as follows*

$$\beta_{\mathcal{MH}} = P[\, S_{\mathcal{MH}} = 0\,] \quad = \quad \frac{\binom{2^\ell}{2^t}\binom{2^n/2^\ell}{1}^{2^t}}{\binom{2^n}{2^t}} = \frac{\binom{2^\ell}{2^t}}{\binom{2^n}{2^t}}(2^{n-\ell})^{2^t}. \tag{3}$$

Proof. For a random permutation we assume that each $L[i]$ has the same probability to be incremented. From the definition of the multivariate hypergeometric distribution the probability that $L = (1, \cdots, 1, 0, \cdots, 0)$ with the first 2^t values equal to 1 is $\binom{2^n/2^\ell}{1}^{2^t} / \binom{2^n}{2^t}$. As there are $\binom{2^\ell}{2^t}$ possible ways to have 2^t values of L equal to 1 we obtain the result. □

As for large numbers, the binomial coefficients are in general difficult to manipulate we approximate them. Using Stirling's approximation we obtain the following approximation $\tilde{\beta}_{\mathcal{MH}}$ of $\beta_{\mathcal{MH}}$.

Corollary 1. *We assume that the data complexity $N = 2^t > 4$ is less than or equal to the size of a structure. Using the multivariate hypergeometric distribution as in Theorem 1, we have the following approximation of the ratio of false positives*

$$\begin{aligned} P[\, S_{\mathcal{MH}} = 0\,] \quad \approx \quad & \sqrt{2^{\ell-n} \cdot \frac{2^n - 2^t}{2^\ell - 2^t}} \\ & \cdot \exp\left[(2^n - 2^t) \cdot \ln(1 - 2^{t-n}) + (2^t - 2^\ell) \cdot \ln(1 - 2^{t-\ell})\right]. \end{aligned} \tag{4}$$

Later we denote by $\tilde{\beta}_{\mathcal{MH}}$ this estimate.

Proof. This approximation is obtained thanks to Stirling's approximation of the factorial function. In particular we use the fact that given $x > 1$ we have $x! \approx \sqrt{2\pi x} x^x \exp[-x]$. From Equation 3 we have

$$\begin{aligned} P[\, S_{\mathcal{MH}} = 0\,] &= \frac{2^\ell!(2^n-2^t)!}{(2^\ell-2^t)!2^n!}(2^{n-\ell})^{2^t} \\ &\approx \exp[-2^\ell-(2^n-2^t)+(2^\ell-2^t)+2^n]\cdot\sqrt{\frac{2^\ell(2^n-2^t)}{(2^\ell-2^t)2^n}} \\ &\cdot \frac{(2^\ell)^{2^\ell}(2^n-2^t)^{2^n-2^t}}{(2^\ell-2^t)^{2^\ell-2^t}(2^n)^{2^n}}\frac{(2^n)^{2^t}}{(2^\ell)^{2^t}}. \end{aligned}$$

All the terms in the exponential cancel and from

$$\begin{aligned} \frac{(2^\ell)^{2^\ell}(2^n-2^t)^{2^n-2^t}}{(2^\ell-2^t)^{2^\ell-2^t}(2^n)^{2^n}}\frac{(2^n)^{2^t}}{(2^\ell)^{2^t}} &= (\frac{2^\ell-2^t}{2^\ell})^{2^t-2^\ell}(\frac{2^n-2^t}{2^n})^{2^n-2^t} \\ &= \exp\left[(2^n-2^t)\cdot\ln(1-2^{t-n})+(2^t-2^\ell)\cdot\ln(1-2^{t-\ell})\right] \end{aligned}$$

we obtain the result. □

3.2.2 Theory using the Multinomial Distribution

For many attacks $|\Delta_Y|$ is of same magnitude than $|\Delta_X|$ and $2^t \leq |\Delta_X|$ is much smaller than 2^ℓ. Therefore we could wonder if instead of using the multivariate hypergeometric distribution we could obtain a similar estimate of the advantage of an impossible differential attack using the classical multinomial distribution, which is the multivariate version of the binomial distribution. In that case we have the following estimate of the advantage of an impossible differential attack.

Lemma 2. *In the setting of Theorem 1, assuming now that the random variable associated to the vector L follows a multinomial distribution, we can compute the false alarm error probability $\beta_{\mathcal{MB}}$ as follows*

$$\beta_{\mathcal{MB}} = P[\, S_{\mathcal{MB}} = 0\,] = \frac{2^\ell!}{(2^\ell-2^t)!}(2^{-\ell})^{2^t}, \tag{5}$$

where $S_{\mathcal{MB}}$ is the random variable associated to S.

From Stirling's approximation we derive the following estimate

$$P[\, S_{\mathcal{MB}} = 0\,] \approx \sqrt{\frac{2^\ell}{2^\ell-2^t}}\cdot\exp[(2^t-2^\ell)\cdot\ln(1-2^{t-\ell})-2^t], \tag{6}$$

that we denote by $\tilde{\beta}_{\mathcal{MB}}$.

Proof. The proof of the first point is similar to the proof of Theorem 1 and is classically known as the birthday bound. From the definition of the multinomial distribution the probability that the first 2^t values of L are equal to 1 is $\frac{2^t!}{(1!)^{2^t}(0!)^{2^\ell-2^t}}(2^{-\ell})^{2^t}$. We conclude using the fact that we have $\binom{2^\ell}{2^t}$ interesting combinations. Equation 6 is derived using Stirling's approximation. □

Note that usually, in the birthday paradox context, another approximation of Equation 5 is used. This approximation is $\exp[-\frac{2^t(2^t-1)}{2^{\ell+1}}]$ and is similar to the approximation obtained from using the binomial distribution directly on the pairs (see Equation 2). The experiments given in the next section show that this approximation is not an accurate approximation of $\beta_{\mathcal{MB}}$ when ℓ is small. In the same section, presenting the results of our experiments, we illustrate that $\beta_{\mathcal{MB}}$ is not a good estimate of the false alarm error probability.

3.3 Experiments in Small Dimensions

As the previous formulas involve binomial coefficients, only theoretical computations in small dimensions are possible. In Table 1 we compare the averaged experimental advantage $\hat{a}$ obtained from different settings on these 12-bit random permutations with the theoretical results of the first part of this paper. Since 2^{26} random permutations are used only advantages smaller than 22 are considered.

The theory of this paper has been developed using a model in which pairs of plaintexts are not manipulated. As classical differential attacks are performed by comparing the difference between messages, the experiments have been performed in this respect and aims at verifying that the model is accurate in this context.

These experimental results illustrate that, in some cases, using the binomial distribution the theoretical advantage $a_{\mathcal{B}} = -\log_2(\beta_{\mathcal{B}})$ (and $a_{\mathcal{MB}} = -\log_2(\beta_{\mathcal{MB}})$ for the multinomial case) gives accurate estimate of $\hat{a}$. However, using the multivariate hypergeometric distribution the advantage estimate $a_{\mathcal{MH}} = -\log_2(\beta_{\mathcal{MH}})$ is always accurate. For the three distributions: binomial, multinomial and multivariate hypergeometric, the estimates $\tilde{a}_{\mathcal{B}}$, $\tilde{a}_{\mathcal{MB}}$, $\tilde{a}_{\mathcal{MH}}$ of $a_{\mathcal{B}}$, $a_{\mathcal{MB}}$, $a_{\mathcal{MH}}$ are accurate.

Table 1: Comparison between experimentally obtained advantages $\hat{a}$ and theoretical ones for 12-bit permutations. The data complexity is 2^t.

ℓ	t	$\hat{\mathbf{a}}$	$a_{\mathcal{B}}$	$\tilde{a}_{\mathcal{B}}$	$a_{\mathcal{MB}}$	$\tilde{a}_{\mathcal{MB}}$	$a_{\mathcal{MH}}$	$\tilde{a}_{\mathcal{MH}}$
6	3	**0.65**	0.64	0.63	0.66	0.66	**0.65**	0.65
6	4	**2.91**	2.73	2.71	2.95	2.95	**2.91**	2.91
6	5	**13.50**	11.27	11.18	13.67	13.67	**13.49**	13.49
8	4	**0.65**	0.68	0.68	0.69	0.69	**0.65**	0.65
8	5	**2.74**	2.80	2.80	2.92	2.92	**2.74**	2.74
8	6	**11.72**	11.38	11.36	12.44	12.44	**11.72**	11.72
10	5	**0.53**	0.70	0.70	0.71	0.71	**0.53**	0.53
10	6	**2.19**	2.84	2.84	2.90	2.90	**2.19**	2.19
10	7	**9.07**	11.46	11.45	11.96	11.96	**9.07**	9.07

Since these experiments clearly confirm that the multivariate hypergeometric approach is the most accurate one, we use this distribution in the next section to derive an estimate of the advantage when multiple structures are used.

3.4 Using Multiple Structures

In this section we assume that the attack requires more than one structure. In particular we denote by 2^s the number of used structures. To maximize the success of the attack we generally have that the data complexity $N = 2^s|\Delta_X|$ allows us to generate $N_S = 2^{s-1}|\Delta_X| \cdot (|\Delta_X| - 1)$ pairs. However, in this section we assume a more general setting where the data complexity is $N = 2^{s+t}$ with $2^t \leq |\Delta_X|$. For accuracy we do not use the usual approximation $(2^t - 1) \approx 2^t$.

We denote by T the scoring function recording the number of pairs which fulfill the truncated differential when multiple structures are used. By definition we have $T = \sum_{j=0}^{2^s-1} S_j$, where S_j is the score obtained for the j-th structure. In the impossible differential context, as recalled in Subsection 2.2, we are interested in the ratio of random permutations for which $T = 0$ when using N messages. Depending on the used statistical model, this ratio can be different.

The model presented in this section is based on the following assumption

Assumption 1. *The statistics obtained from the different structures are independent.*

The accuracy of this assumption is discussed in Section 4. Based on this assumption and on the model developed in Subsection 3.2 we obtain the following result.

Theorem 2. *Given 2^s $(s \leq m)$ structures containing 2^t messages with $t \leq \log_2(|\Delta_X|)$. Under Assumption 1, we assume that, for each structure involved in the attack, the random variables associated to the vectors L follow a multivariate hypergeometric distribution. The ratio $\beta_{\mathcal{MH}}$ of false positives is equal to*

$$\beta_{\mathcal{MH}} = P[\, T_{\mathcal{MH}} = 0\,] \quad = \quad P[\, S_{\mathcal{MH}} = 0\,]^{2^s}.$$

Using Stirling's approximation we obtain the following estimate $\tilde{\beta}_{\mathcal{MH}}$ of $\beta_{\mathcal{MH}}$.

$$\tilde{\beta}_{\mathcal{MH}} \approx \exp[(2^{s-1}+2^{n+s}-2^{s+t})\cdot\ln(1-2^{t-n})-(2^{s-1}+2^{\ell+s}-2^{s+t})\cdot\ln(1-2^{t-\ell})]. \quad (7)$$

Proof. From Assumption 1 and Corollary 1 we have

$$\begin{aligned}\tilde{\beta}_{\mathcal{MH}} \quad &\approx \quad \left(2^{\ell-n}\frac{(2^n-2^t)}{(2^\ell-2^t)}\right)^{2^{s-1}} \\ &\cdot \quad \exp[2^s\left((2^n-2^t)\cdot\ln(1-2^{t-n})+(2^t-2^\ell)\cdot\ln(1-2^{t-\ell})\right)].\end{aligned}$$

Using

$$\left(2^{\ell-n}\frac{(2^n-2^t)}{(2^\ell-2^t)}\right)^{2^{s-1}} = \exp\left[2^{s-1}\ln(\frac{1-2^{t-n}}{1-2^{t-\ell}})\right],$$

we derive the proof. □

3.5 Experiments

We present in Table 2 the results of our experiments for a varying number of structures. The difference between the experimental advantage $\hat{a}$ and the advantage $a_{\mathcal{B}}$ obtained from Equation 1 is particularly noticeable when the number of structures increases. However, in most cases, the advantage $a_{\mathcal{MH}}$ (and its approximation $\tilde{a}_{\mathcal{MH}}$) obtained from the use of the multivariate hypergeometric distribution (see Theorem 2) is relatively accurate. Due to the independence assumption a small difference between the theoretical and experimental advantage is sometimes noticeable when many structures are used. The accuracy of the formula, in this case, is discussed in Section 4.

The accuracy of the advantage estimate $\tilde{a}_{\mathcal{MH}}$ can also be observed in the same table.

3.6 Influence on a Concrete Distinguisher

The results presented in this paper have been experimentally verified on random permutations. However, it is interesting to see if the newly developed theory has a concrete impact on the advantage of a distinguishing attack on ciphers.

First, assuming that t is small in comparison to ℓ and n, we can derive an estimate of the false alarm error probability which is comparable to the historical estimate $\tilde{\beta}_{\mathcal{B}}$ recalled in Equation 1.

Theorem 3. *When t is small in comparison to n and ℓ and using the approximation $\ln(1-\frac{w}{2^\ell}) \approx -\frac{w}{2^\ell}$ for $1 \leq w \leq 2^t-1$ we have that $\beta_{\mathcal{MH}}$ is approximatively equal to*

$$\exp[-N_S(2^{-\ell}-2^{-n})],$$

where $N_S = 2^{s+t-1}(2^t-1)$. We denote this estimate by $\bar{\beta}_{\mathcal{MH}}$, and by

$$\bar{a}_{\mathcal{MH}} = \frac{2^{s+t-1}(2^t-1)}{\ln(2)}(2^{-\ell}-2^{-n})$$

the associated advantage.

Table 2: Comparison between experimentally obtained advantages $\hat{a}$ and theoretical ones for 12-bit permutations. Experiments obtained from 2^{s+t} messages.

ℓ	s	t	$s+t$	$\hat{a}$	$a_{\mathcal{B}}$	$\tilde{a}_{\mathcal{B}}$	$a_{\mathcal{MH}}$	$\tilde{a}_{\mathcal{MH}}$
7	0	3	3	**0.31**	0.32	0.32	**0.31**	0.31
7	2	3	5	**1.25**	1.27	1.26	**1.25**	1.25
7	4	3	7	**4.99**	5.07	5.05	**4.99**	4.99
7	6	3	9	**20.17**	20.28	20.20	19.97	19.97
7	0	5	5	**5.94**	5.61	5.59	**5.94**	5.94
7	2	5	7	**23.68**	22.45	22.36	23.76	23.76
9	0	3	3	**0.07**	0.08	0.08	**0.07**	0.07
9	2	3	5	**0.28**	0.32	0.32	**0.28**	0.28
9	4	3	7	**1.11**	1.26	1.26	**1.11**	1.11
9	6	3	9	**4.44**	5.05	5.05	**4.44**	4.44
9	8	3	11	**17.73**	20.22	20.20	17.77	17.77
9	0	5	5	**1.25**	1.40	1.40	**1.25**	1.25
9	2	5	7	**5.01**	5.60	5.59	**5.01**	5.01
9	4	5	9	**20.05**	22.38	22.36	20.03	20.03

Proof. From Theorem 2, for small t, we have

$$\begin{aligned}
\beta_{\mathcal{MH}} &= \left[\frac{\binom{2^\ell}{2^t}}{\binom{2^n}{2^t}}(2^{n-\ell})^{2^t}\right]^{2^s} \\
&= \left[\frac{\prod_{0\le w\le 2^t-1}(2^\ell-w)}{\prod_{0\le w\le 2^t-1}(2^n-w)}(2^{n-\ell})^{2^t}\right]^{2^s} = \left[\prod_{1\le w\le 2^t-1}\frac{2^\ell-w}{2^\ell}\frac{2^n}{2^n-w}\right]^{2^s} \\
&= \exp\left[2^s\left(\sum_{1\le w\le 2^t-1}\ln(1-w2^{-\ell})-\ln(1-w2^{-n})\right)\right] \\
&\approx \exp\left[2^s\left((-2^{-\ell}+2^{-n})\sum_{1\le w\le 2^t-1}w\right)\right] \\
&\approx \exp\left[2^s\left((2^{-n}-2^{-\ell})2^{t-1}(2^t-1)\right)\right] = \exp\left[N_S(2^{-n}-2^{-\ell})\right].
\end{aligned}$$

□

Note that this results is not true if the assumption that t is small in comparison to n and ℓ is not fulfilled. For instance, for 12-bit random permutations, setting $\ell = 6$, $t = 5$ and $s = 0$ we have $a_{\mathcal{MH}} = \tilde{a}_{\mathcal{MH}} = 13.49$ (see Table 1) while the estimate given in Theorem 3 is $\bar{a}_{\mathcal{MH}} = 11.01$.

We present in Table 3 our computational results for the impossible distinguisher on 14 rounds of LBlock of [WZ11]. This distinguisher is used in different key recovery attacks on 22 and 23 rounds of LBlock [BNS14a] . For this distinguisher, the parameters are $n = 64$, $\ell = 60$ and $|\Delta_X| = 2^4$. As $N > 2^4$, t is usually taken as equal to 4.

These computations illustrate that, for instance, using 2^{62} plaintexts an error of 2.5 bits was made using the binomial distribution. Note that these computational results can easily be computed from Equation 2 and Equation 7, even with a 64-bit cipher. The accuracy of the estimate provided in Theorem 3 is also illustrated in the table.

In the impossible differential attack on 23-rounds of LBlock [BNS14a] the time complexity is not dominated by the exhaustive key search and the results presented in Table 3 do not influence the total time complexity of the attack.

Table 3: The case of LBlock with $n = 64$, $\ell = 60$ and $t = 4$. Meaning that $N = 2^{s+4}$ and $N_S = 15N/2$.

$\log_2(N)$	$\tilde{a}_{\mathcal{MH}}$	$\bar{a}_{\mathcal{MH}}$	$\tilde{a}_{\mathcal{B}}$
58	2.54	2.54	2.69
59	5.07	5.07	5.38
60	10.14	10.14	10.77
61	20.29	20.29	21.54

$\log_2(N)$	$\tilde{a}_{\mathcal{MH}}$	$\bar{a}_{\mathcal{MH}}$	$\tilde{a}_{\mathcal{B}}$
62	40.58	40.58	43.07
63	81.15	81.15	86.15
64	162.30	162.30	172.30

4 Non-Independent Structures: When Close to the Full Codebook

In the previous sections, we assumed that for a random permutation we can consider independent structures. However, while this is true for random functions this assumption is wrong for permutations. From a theoretical point of view, this could have an impact on the advantage of the attack when the data complexity is close to the full codebook. In this section, we elaborate on this point. Taking into consideration the possible dependency between the structures can be handled with the bi-multivariate hypergeometric distribution. Generalizing the multivariate hypergeometric distribution, the bi-multivariate hypergeometric distribution is the distribution which focus on the repartition of draws of balls with two different properties such as color and size. In this section we use this distribution to give a bound on the maximal advantage of an impossible differential distinguisher.

Let us consider the following table

$a_{1,1}$	$a_{1,2}$	$\cdots$	$a_{1,2^\ell}$	r_1
$a_{2,1}$	$a_{2,2}$	$\cdots$	$a_{2,2^\ell}$	r_2
$\cdots$	$\cdots$	$\cdots$	$\cdots$	
$a_{2^s,1}$	$a_{2^s,2}$	$\cdots$	$a_{2^s,2^\ell}$	r_{2^s}
c_1	c_2	$\cdots$	c_{2^ℓ}	2^{s+t}

where for $1 \leq i \leq 2^s$ and $1 \leq j \leq 2^\ell$ we have $r_i = \sum_j a_{i,j}$ and $c_j = \sum_i a_{i,j}$. Each line of this table corresponds to the values of L given in the previous section for a given structure. For an attack using 2^s structures, we have 2^s lines. Using 2^t messages inside a structure, for each i, we have $r_i = 2^t$. As explained in Section 3, the truncated differential is not fulfilled if each $a_{i,j}$ takes only the value 0 or 1. Implying that, for all j, we have $c_j \leq 2^s$.

If we use the full codebook $n = s + t$, all outputs are equiprobable and all c_j's are equal to $2^{n-\ell}$. In practice, for a random permutation, using less than the full codebook we have $c_j \leq \min(2^s, 2^{n-\ell})$. The vector $(c_1, \cdots, c_{2^\ell})$ is denoted by C.

We chose to illustrate the new concepts, introduced in this section, with an impossible differential fulfilling the following properties $n = 4$, $\ell = 3$, $t = 1$, $s = 2$. For instance the following repartition is possible in the setting of Section 3 (independent structures) and impossible assuming that for all j, $c_j \leq \min(2^s, 2^{n-\ell}) \leq 2$.

	1	1	0	0	0	0	0	0
	1	0	0	1	0	0	0	0
	1	1	0	0	0	0	0	0
	1	0	0	0	0	1	0	0
$C:$	4	2	0	1	0	1	0	0

The probability to have a particular repartition can be expressed thanks to the bi-multivariate hypergeometric distribution.

Lemma 3. *Using the bi-multivariate hypergeometric distribution, the probability to obtain a given repartition of the bi-dimensional table after 2^{s+t} encryptions fulfilling $a_{i,j} \in \{0,1\}$, $r_i = 2^t$ and $C = (c_1, c_2, \cdots, c_{2^\ell})$ is*

$$P_C = \frac{(2^{n-\ell}!)^{2^\ell} \cdot (2^t!)^{2^s} \cdot (2^n - 2^{t+s})!}{2^n! \cdot \prod_j (\bar{c_j}!)},$$

with $\bar{c_j} = 2^{n-\ell} - c_j$.

Proof. Given the full codebook and a given repartition, after 2^{s+t} encryptions, we define, for all $1 \leq j \leq 2^\ell$, the quantity $\bar{c_j} = 2^{n-\ell} - c_j$ as the remaining possibilities for the j-th value after 2^{s+t} encryptions. The full repartition is then given by

$a_{1,1}$	$a_{1,2}$	$\cdots$	$a_{1,2^\ell}$
$a_{2,1}$	$a_{2,2}$	$\cdots$	$a_{2,2^\ell}$
$\cdots$	$\cdots$	$\cdots$	$\cdots$
$a_{2^s,1}$	$a_{2^s,2}$	$\cdots$	$a_{2^s,2^\ell}$
$\bar{c}_1$	$\bar{c}_2$	$\cdots$	$\bar{c}_{2^\ell}$

where $\sum_{i=1}^{2^s} r_i + \sum_{j=1}^{2^\ell} \bar{c}_j = \sum_{j=1}^{2^\ell} (c_j + \bar{c}_j) = 2^n$.

The result comes from the definition of the probability function of bi-multivariate hypergeometric distribution and is simplified by the fact that each $a_{i,j}$ are equal to 0 or 1 with factorial equal to 1. □

For instance, using 2^{2+1} messages ($s = 2$ and $t = 1$), we obtain the following repartitions

1	1	0	0	0	0	0	0
0	0	0	1	1	0	0	0
0	1	0	0	1	0	0	0
0	0	0	1	0	1	0	0

0	1	0	0	0	0	1	0
0	0	0	1	0	0	1	0
0	1	0	0	1	0	0	0
0	0	0	1	0	1	0	0

with same probability $P_C \approx 0.0000196$ for $n = 4$ (left: $C = (1, 2, 0, 2, 2, 1, 0, 0)$ and right $C = (0, 2, 0, 2, 1, 1, 2, 0)$).

To estimate the false alarm error probability of a given impossible differential we need the number of possible binary matrices for each vector $C = (c_1, c_2, \cdots, c_{2^\ell})$. We denote by M_C this number. From this approach, the false alarm error probability $\beta_{\mathcal{BMH}}$ is

$$\beta_{\mathcal{BMH}} = \sum_C M_C \cdot P_C. \tag{8}$$

As discussed in [GMW06], for the setting we are dealing with, only estimates of M_C can be provided.

Due to number of possible combinations for the vector C we can, in general, not use this approach to compute the advantage of an impossible differential attack. However, when using the full codebook the only possibility for the vector C is $C = (2^{n-\ell}, 2^{n-\ell}, \cdots, 2^{n-\ell})$. From this observation we derive the following result which take into consideration the non-independence of the samples when multiple structures are used in the attack.

Theorem 4. *The maximal advantage of an impossible differential distinguisher (Δ_X, Δ_Y) is*

$$a_{\max} = \frac{(2^{n-\ell} - 1)(2^t - 1)}{2\ln(2)} \left(1 + \mathcal{O}(2^{-min(n,\ell+t)})\right).$$

Proof. The maximal advantage is reached when all possible pairs are considered which is obtained from using the full codebook $N = 2^n = 2^s|\Delta_X|$ and setting $2^t = |\Delta_X|$. In that case we only have one possibility for the vector C which is $C = (2^{n-\ell}, 2^{n-\ell}, \cdots, 2^{n-\ell})$. For this vector, according to [GMW06], the number of matrices fulfilling the conditions given in this section is

$$M_c = \frac{2^n!}{(2^{n-\ell}!)^{2^\ell}(2^t!)^{2^s}} \exp[-b \cdot \left(1 + \mathcal{O}(2^{-min(n,\ell+t)})\right)],$$

with $b = \frac{(2^{n-\ell}-1)(2^t-1)}{2}$ and $s + t = n$. When $C = (2^{n-\ell}, 2^{n-\ell}, \cdots, 2^{n-\ell})$, from Lemma 3 we derive $P_C = \frac{(2^{n-\ell}!)^{2^\ell}(2^t!)^{2^s}}{2^n!}$ and $\beta_{\mathcal{BMH}} = \sum_C M_C \cdot P_C$ simplifies to $\beta_{\mathcal{BMH}} = \exp[-b\left(1 + \mathcal{O}(2^{-min(n,\ell+t)})\right)]$. The results is obtained from $a_{\max} = \log_2(\beta_{\mathcal{BMH}})$. □

Remark 1. This result gives some intuition on the validity of the assumption used in Theorem 3. In particular when t is relatively small in comparison to ℓ and n using $N_S = 2^{n-1}(2^t - 1)$ pairs, assuming independent structures as in Theorem 3, we have

$$\bar{\beta}_{\mathcal{MH}} = \exp[N_S(2^{-n} - 2^{-\ell})] = \exp[2^{n-1}(2^t - 1)(2^{-n} - 2^{-\ell})]$$

and $\bar{a}_{\mathcal{MH}} = \frac{(2^t - 1)(2^{n-\ell} - 1)}{2\ln(2)}$ corresponds to a_{max}.

If, as it is often the case in practice, the data complexity $N = 2^{s+t}$ is smaller than 2^ℓ then the relation between the structures is also minimal and the approximations provided in Subsection 3.4 should be valid. The experiments provided in Table 2 confirm this intuition. In these experiments, the small difference between $\hat{a}$ and $\bar{a}_{\mathcal{MH}}$ can be observed when $s + t \geq \ell$.

In the remainder of the paper we assume that t is small in comparison ℓ and n. In that case the advantage is given by Theorem 3. For simplicity, we denote by $\bar{a}$ the advantage $\bar{a}_{\mathcal{MH}}$.

5 Remarks

5.1 Encryption/Decryption

Depending on whether the attacker has access to the encryption oracle or the decryption oracle, chosen (or known) plaintext or chosen (or known) ciphertext impossible differential attacks are considered. In particular, if we have an impossible differential distinguisher (Δ_X, Δ_Y) on the encryption primitive, the differential (Δ_Y, Δ_X) is also impossible for the decryption function. Depending on the size of Δ_X and Δ_Y, it is usually assumed [BNS14a, BLNPS17] that, to minimize the complexity, chosen plaintext or chosen ciphertext impossible differential attacks should be considered.

For instance, taking the parameters $n = 32$, $\ell = 20$ and $t = 4$ as for the experimental attack of [BLNPS17] to reach an advantage of 16 bits, using Equation 2, the data complexity of a chosen plaintext attack is estimated to $2^{20.565}$ messages and of a chosen ciphertext to $2^{20.47}$ messages. Taking, as in [BLNPS17], the minimum of these two values give us a data complexity of $2^{20.47}$. In this section we use the newly derived theory to show that the data complexity is actually of $2^{20.565}$.

The following lemma state that when more than one structure is used, chosen plaintext and chosen ciphertext impossible differential distinguishers have equal advantages.

Lemma 4. *Given a data complexity of $N = 2^s|\Delta_X| = 2^{s+t} = 2^{s'}|\Delta_Y| = 2^{s'+t'}$ with $s > 0$ and $s' > 0$, the advantage of a chosen plaintext or a chosen ciphertext distinguishing attack computed from Theorem 3 is the same.*

Proof. For a chosen plaintext attack, we consider the encryption primitive and according to Theorem 3 we have $\bar{a}_{enc} = \frac{N}{2\ln(2)}(2^t - 1)(2^{-\ell} - 2^{-n})$.

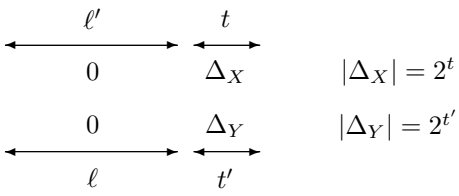

Figure 2: Notation encryption/decryption when (Δ_X, Δ_Y) is impossible. The notation ℓ and t are taken for an attack in the encryption side and ℓ' and t' in the decryption side.

Setting $t' = n - \ell$ and $\ell' = n - t$ as given in Figure 2 we obtain that

$$\begin{aligned} \bar{a}_{dec} &= \frac{N}{2\ln(2)}(2^{t'} - 1)(2^{-\ell'} - 2^{-n}) = \frac{N}{2\ln(2)}(2^{n-\ell} - 1)2^{-n}(2^t - 1) \\ &= \frac{N}{2\ln(2)}(2^{-\ell} - 2^{-n})(2^t - 1). \end{aligned}$$

□

Remark 2. Note that, using the historic estimate recalled in Equation 2, the difference between a chosen plaintext or a chosen ciphertext attack was noticeable for structures of small sizes. Using the notation introduced in the previous proof, the advantage of a chosen ciphertext attack derived from Equation 2 is $\frac{N}{2\ln(2)}(2^{t'} - 1)2^{-\ell'} = \frac{N}{2\ln(2)} \cdot 2^t(2^{-\ell} - 2^{-n})$ which is different from $\frac{N}{2\ln(2)} \cdot (2^t - 1)(2^{-\ell} - 2^{-n})$ when t is small.

Thanks to the previous result, the condition: "t small in comparison to ℓ and n", used in this paper, is equivalent to the condition: "$n - \ell$ small in comparison to $n - t$ and n".

5.2 Multiple Impossible Differentials

To improve the power of impossible differential attacks we can, as described in [BNS14a, BLNPS17], consider simultaneously multiple (truncated) impossible differentials. In this section we explain how to estimate the advantage of such multiple impossible differential attacks. We assume here that we have m_{in} sets of input differences and m_{out} sets of output differences. As we will see that the advantage only depends on the product of these values we define $m = m_{\text{in}}m_{\text{out}}$. For more details on the multiple impossible differential attacks, we refer to [BNS14a, BLNPS17].

Lemma 5. *When $m = m_{in}m_{out}$ truncated difference sets are used. Assuming that the data complexity $N = 2^{s+t}$ satisfies $m2^{s+t} < 2^\ell$ and that all structures have the same number of messages 2^t, the advantage of a multiple impossible distinguisher is*

$$a_{\mathcal{MH}} = -m \cdot s \cdot \log_2 \left(\frac{\binom{2^\ell}{2^t}}{\binom{2^n}{2^t}} (2^{n-\ell})^{2^t} \right). \tag{9}$$

If we assume also that $t \leq \log_2(|\Delta_X|)$ is small in comparison to ℓ and n as in Theorem 3, we have the following approximation

$$\bar{a} = m\frac{2^{s+t-1}(2^t - 1)}{\ln(2)}(2^{-\ell} - 2^{-n}). \tag{10}$$

Proof. In Section 4 we have seen that when the data complexity $N = 2^{s+t}$ is smaller than 2^ℓ then the influence of considering independant structures on the advantage is negligible.

Table 4: Experiments for $n = 12$, $t = 3$, $\ell = 9$, $s = 3$ with 2^{26} random functions. $\hat{a}$ denotes the experimental advantage, $a_{\mathcal{MH}}$ is taken from Equation 9, $\bar{a}$ is taken from Equation 10 and $\tilde{a}_{\mathcal{B}}$ is the advantage we obtain using the binomial distribution.

m_{in}	m_{out}	m	$\hat{a}$	$a_{\mathcal{MH}}$	$\bar{a}$	$\tilde{a}_{\mathcal{B}}$
1	2	2	**1.11**	**1.11**	1.11	1.26
2	1	2	**1.11**	**1.11**	1.11	1.26
1	3	3	**1.67**	**1.67**	1.66	1.89
2	3	6	**3.34**	**3.33**	3.31	3.79

Table 5: Experiments for $n = 12$, $t = 3$, $\ell = 8$, $s = 3$ with 2^{26} random functions. $\hat{a}$ denotes the experimental advantage, $a_{\mathcal{MH}}$ is taken from Equation 9, $\bar{a}$ is taken from Equation 10 and $\tilde{a}_{\mathcal{B}}$ is the advantage we obtain using the binomial distribution.

m_{in}	m_{out}	m	$\hat{a}$	$a_{\mathcal{MH}}$	$\bar{a}$	$\tilde{a}_{\mathcal{B}}$
1	1	1	**1.20**	**1.20**	1.18	1.26
1	2	2	**2.42**	**2.39**	2.37	2.52
1	3	3	**3.60**	**3.59**	3.55	3.79
2	2	4	**4.80**	**4.78**	4.73	5.04
2	3	6	**7.23**	7.18	7.10	7.57

Taking multiple inputs differences is, in that case, equivalent to using more structures. When taking multiple output differences, thanks to Lemma 4, the same reasoning can be applied. □

To confirm this result, experiments on 12-bit random permutations have been performed. The result of some of these experiments is provided in Table 4. Taking $2^{s+t} = 2^6$ plaintexts and $\ell = 9$, we illustrate that we have a relatively good estimate of the advantage of an impossible differential attack. Note that in that case $6 \cdot 2^{s+t} < 2^9$. The case where $\ell = 8$ is provided in Table 5. In that case the approximation $\bar{a}$ is less accurate since t is closer to ℓ. However, the estimate $a_{\mathcal{MH}}$ provided in Equation 9 is relatively accurate when $m2^{s+t} \leq 2^{\ell}$.

6 Key Recovery Attacks

6.1 Theory

Using the state-test technique [BNS14a, BLNPS17], data, time and memory complexities of the previously introduced key recovery impossible differential attacks have been improved. In this paper the focus is put on providing a better estimate of the exhaustive key search time complexity. To explain the relation between the data complexity of a distinguishing and a key recovery attack, we introduce the following notations which are summarized in Figure 3.

Given an impossible differential (Δ_X, Δ_Y) on r central rounds of an n-bit block cipher, to obtain an attack on $r_{\text{in}} + r + r_{\text{out}}$ rounds, we add r_{in} and r_{out} rounds before and after this distinguisher and define Δ_{in} and Δ_{out} as the sets of all possible input respectively output differences on the $r_{\text{in}} + r + r_{\text{out}}$ rounds.

Similarly, as for the distinguisher, we denote by $\ell' = n - \log_2(|\Delta_{\text{out}}|)$. We assume that the data complexity of the key recovery attack is $N^A = 2^{s'+t'}$ where $2^{t'} \leq |\Delta_{\text{in}}|$ corresponds to the structure size and $2^{s'}$ denotes the number of used structures.

In a key recovery attack, we are interested in finding the encryption key among the key candidates, possible round keys, in the outer rounds. To do so, the wrong-key

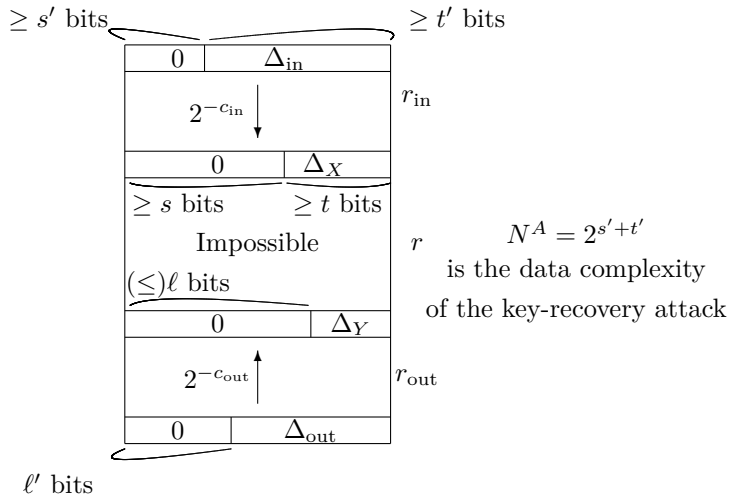

Figure 3: A block cipher with impossible differential distinguisher on the central rounds (Δ_X, Δ_Y).

randomization hypothesis[2] is usually assumed:

- For the correct key guess, the truncated differential (Δ_X, Δ_Y) on r rounds is never fulfilled.
- For the wrong key guesses, the truncated differential (Δ_X, Δ_Y) is not impossible and it is usually assumed that the obtained permutations behave like random permutations.

The number of pairs needed to perform the impossible differential attack is usually determined by the different sieving steps of the key recovery phase.

During the partial inversion phase, pairs which do not fulfill the targeted truncated differential path are discarded. As in [BNS14a, BLNPS17], we denote by $2^{-c_{\text{in}}}$ (resp. $2^{-c_{\text{out}}}$) the probability for a pair to be discarded during the partial encryption (resp. decryption) procedure.

Based on these notations, we can estimate the advantage of an impossible differential attack as follows.

Lemma 6. *Given $|\Delta_{\text{in}}|$, $|\Delta_{\text{out}}|$, c_{in}, c_{out} as defined in this section and assuming the setting of Theorem 3. The advantage of an impossible differential key recovery attack using $N^A = 2^{s'+t'}$ plaintexts is approximatively*

$$2^{s'+t'-1}(2^{t'}-1)\frac{(|\Delta_{\text{in}}|2^{-c_{\text{in}}}-1)}{(|\Delta_{\text{in}}|-1)}(2^{-\ell'-c_{\text{out}}}-2^{-n})/\ln(2).$$

If $N^A \geq |\Delta_{\text{in}}|$, this value is maximized for $t' = \log_2(|\Delta_{\text{in}}|)$. Using m truncated impossible differentials, the advantage is multiplied by m if $m2^{s+t} < 2^{\ell}$ and all structures have size $2^{t'}$.

Proof. In this proof we consider the partial encryption and decryption separately.

- Starting with $2^{s'} \geq 1$ structures of size $2^{t'} = |\Delta_{\text{in}}|$, after partial encryption, we have $2^s = 2^{s'+c_{\text{in}}}$ structures of size $|\Delta_{\text{in}}|2^{-c_{\text{in}}}$. It might happen that $|\Delta_{\text{in}}|2^{-c_{\text{in}}}$ is smaller than $|\Delta_X|$. In that case the data complexity should be computed in accordance with the size of the structure being $2^t = |\Delta_{\text{in}}|2^{-c_{\text{in}}}$.

[2]While the wrong key-randomization hypothesis has been revisited in the linear context, in this paper we consider only the classical one. The question of adapting the results, assuming the revisited wrong key randomization hypothesis remains open.

- When less than one structure is used for the attack ($2^{t'} \leq |\Delta_{\text{in}}|$) then the sieve $2^{-c_{\text{in}}}$ influence the number of available pairs by a factor $(|\Delta_{\text{in}}|2^{-c_{\text{in}}} - 1)/(|\Delta_{\text{in}}| - 1)$.
- From $2^{\ell'} = \frac{|\Delta_{\text{out}}|}{2^n}$, after partial decryption, we have $\ell' + c_{\text{out}}$ bits equal to zero. In practice if $\ell' + c_{\text{out}} < n - \log_2(|\Delta_Y|$ then $\ell \leq n - \log_2(|\Delta_Y|)$ and the analysis should be done with $\ell = \ell' + c_{\text{out}}$.

□

6.2 Experiments

Usual key-recovery attacks relies on the wrong-key-randomization hypothesis. To the best of our knowledge very few experiments have been performed to check the validity of this assumption. In this section we present the results of our experiments on a 16-bit cipher similar to CLEFIA (4 branches of 4 bits each) with independant rounds key. For the performed key recovery attacks, 3 and 4 rounds have been partially inverted. The setting and results of our experiments are given in Table 6. These experiments illustrate that the model presented in this paper is more accurate than the previous model. In particular, with respect to the possible relation between the partially inverted rounds, these results show that the theory is accurate when enough round keys are targeted in the attack.

Table 6: Key-recovery attack on a 16-bit Feistel with 4 branches. Taking an impossible differential distinguisher with $|\Delta_X| = |\Delta_Y| = 2^4$. $\hat{a}$ denotes the experimental advantage, $\bar{a}$ is derived from Lemma 6
Left: 1 round before the distinguisher, 2 rounds after. The settings are $\log_2(|\Delta_{\text{in}}|) = 8$, $\log_2(|\Delta_{\text{out}}|) = 12$, $c_{\text{in}} = 4$, $c_{\text{out}} = 8$.
Right: 2 rounds before the distinguisher, 2 rounds after. The settings are $\log_2(|\Delta_{\text{in}}|) = 12$, $\log_2(|\Delta_{\text{out}}|) = 12$, $c_{\text{in}} = 8$, $c_{\text{out}} = 8$.

$\log(N)$	s'	t'	$\hat{a}$	$\bar{a}$	$\tilde{a}_{\mathcal{B}}$
10	2	8	2.53	2.54	2.71
11	3	8	5.01	5.07	5.41
12	4	8	9.81	10.14	10.82

$\log(N)$	s'	t'	$\hat{a}$	$\bar{a}$	$\tilde{a}_{\mathcal{B}}$
10	0	10	0.51	0.51	0.68
11	0	11	2.53	2.54	2.70
12	0	12	10.14	10.14	10.82

6.3 Impossible Differential Involving Only One Differential

In [BNS14a, DF16] impossible differential attacks on 19 and 20 rounds of SIMON32/64 are presented. The full codebook is taken into consideration for these attacks where the dominant factor of the time complexity is the exhaustive key search. Using only one impossible differential the advantage of the attack is evaluated using the binomial distribution (see [BNS14a, DF16]) to 1.41 bits.

In the case of one differential, we show in the proof of the following lemma that all approaches provide a similar advantage. In particular the advantage of any impossible differential attack using only one differential does not exceed 0.73 bits.

Lemma 7. *Given a n-bit block cipher. Using only one impossible differential, the maximal advantage is of $\frac{1}{2\ln(2)}\big(1 + \mathcal{O}(2^{-n})\big) \approx 0.72$ bits. With a key of length k bits, the time complexity of the exhaustive key search can not be smaller than approximatively $2^{k-0.72}$.*

Proof. As explained in Section 2, when only one differential is involved, the parameters are $\ell = n - 1$ and $t = 1$. The maximum advantage is reached when using the full codebook. From Theorem 4, we then have $a_{max} = 1/(2\ln(2))\big(1 + \mathcal{O}(2^{-n})\big) \approx 0.72$.

□

Table 7: Experimental advantage of a single impossible differential attack when $N = 2^n$ ($N_S = 2^{n-1}$). Illustrating that the advantage a satisfies $a \approx 0.72$.

n	$\hat{a}$
8	0.7164
10	0.7224

n	$\hat{a}$
12	0.7194
14	0.71932

This advantage has also been experimentally verified for random permutations of different sizes. The results are summarized in Table 7.

Based on this result we can claim that the impossible differential attack on 20 rounds of SIMON32/64 [DF16] has time complexity $2^{63.3}$ instead of $2^{62.6}$ when the full codebook is used. This remark holds also for the impossible differential attacks on other versions of SIMON presented in [BNS14a].

6.4 Application to LBlock

In the key recovery attack on 23 rounds of LBlock presented in [BLNPS17] the parameters are: $\Delta_{in} = 48$, $\Delta_{out} = 32$ $c_{\text{in}} + c_{\text{out}} = 72$ and $m = 2^6$. In [BLNPS17], the proposed attack has a data complexity of $2^{55.5}$. The time complexity has been computed for an advantage of 30.6 bits. With the newly developed theory, we can state that for this data complexity the advantage is only of 28.69 bits. However, the overall time complexity of the attack (2^{72} encryptions) is not modified since the time complexity is dominated by the key sieving phase and not by the exhaustive key search.

6.5 Application to CRYPTON

In [BLNPS17] the impossible differential attack on 7 rounds of CRYPTON has been improved. The parameters of the attack are equivalent to the following ones: $\Delta_{in} = 32$, $\Delta_{out} = 64$, $c_{in} = 24$, $c_{out} = 48 + 14.36$, $m_{in} = 4 \cdot 4 = 16$ and $m_{out} = 4 \cdot 6 \cdot 6 = 144$. Based on these parameters, the advantage of the attack was previously estimated to 148.44 bits for a data complexity of $2^{114.9}$ known plaintexts. The new analysis provided in this paper allow us to state that for this data complexity the advantage of this attack is close to 145.45 bits if we consider that it exists a clever way to generate the 4 input structures (see details in [BLNPS17]).

7 A note on Differential and Truncated Differential Attacks

For the (truncated) differential attacks [Knu94, MSAK99, SKU+00] with $p_* = P[\, E_K(x) \oplus E_K(x \oplus \delta) \in \Delta_Y | \delta \in \Delta_X \,]$ significantly larger than the uniform probability, the data complexity is usually estimated in the same way as it is done for a classical differential attack [BS90, BGT11]. For instance, in [BS90], the authors use the Poisson approximation of the binomial distribution to estimate both the advantage and the success probability of the attack.

For such attacks, a threshold $\theta \geq 1$ is usually fixed and a permutation is accepted as potentially correct if more than θ pairs fulfilling the (truncated) differential are found. The ratio of such random permutations defines the false alarm error probability.

When the threshold is small, as it is the case for instance for the truncated differential distinguisher on 7 rounds of E2 (the threshold is fixed to 1) [MSAK99] or for the differential attack on the DES cipher [BS90], we could extend our method to estimate the advantage of a (truncated) differential attack.

7.0.1 When the threshold is fixed to 1

Lemma 8. *Given a truncated differential* (Δ_X, Δ_Y) *with probability* p_* *much larger than* $p = 2^{-n}|\Delta_Y|$*. For a threshold* $\Theta = 1$*, the false alarm error probability is*

$$\beta_T = P[\, T_T \geq 1 \,] = 1 - P[\, T_{\mathcal{MH}} = 0 \,],$$

with $P[\, T_{\mathcal{MH}} = 0 \,]$ *as in Theorem 1.*

For larger thresholds, the model developed in this paper is harder to implement. As an illustration, we derive an expression of the advantage when the threshold fixed to 2.

7.0.2 When the threshold is fixed to 2

To derive the false alarm error probability of a (truncated) differential distinguisher when the threshold θ is fixed to 2, we first need to estimate the probability to obtain one pair using only one structure. Using the previously defined notation this probability is denoted by $P[\, S_{\mathcal{MH}} = 1 \,]$.

To get one pair, the vector L of size 2^ℓ defined in Section 3 should have one value equals to 2 and $(2^t - 2)$ values equal to 1. From the multivariate hypergeometric distribution when we only have one structure we then get that

$$\begin{aligned} P[\, S_{\mathcal{MH}} = 1 \,] &= \binom{2^\ell}{1}\binom{2^\ell - 1}{2^t - 2}\frac{\binom{2^{n-\ell}}{2}\binom{2^{n-\ell}}{1}^{2^t-2}}{\binom{2^n}{2^t}} \\ &= 2^{\ell-1}(2^{n-\ell} - 1)(2^{n-\ell})^{2^t-1}\frac{(2^\ell - 1)!2^t!(2^n - 2^t)!}{(2^t - 2)!(2^\ell - 2^t + 1)!2^n!}. \end{aligned}$$

When the threshold is fixed to 2, given 2^{s+t} messages grouped in 2^s structures and assuming the independence of the structures, the false alarm error probability of the truncated differential (Δ_X, Δ_Y) is

$$P[\, T_{\mathcal{MH}} \geq 2 \,] = 1 - 2^s \cdot P[\, S_{\mathcal{MH}} = 1 \,] \cdot P[\, S_{\mathcal{MH}} = 0 \,]^{2^s-1} - P[\, S_{\mathcal{MH}} = 0 \,]^{2^s}.$$

Note that in a large part of this paper we have shown that, by considering the binomial distribution, the advantage of an impossible differential was overestimated. This phenomenon is converse for classical differential or truncated differential attacks. As illustrated in Lemma 8, for the case where the threshold is fixed it 1, the classically used model (based on the binomial distribution) underestimates the advantage of a differential or truncated differential attack. Meaning that in practice, for the same data complexity and the same threshold, the time complexity of the exhaustive key search is smaller than previously computed. However research remains to be done regarding how we could apply this theory to estimate the success probability of a (truncated) differential attack.

8 Conclusion

Using a multivariate approach, we provided a better estimate of the advantage of an impossible differential attack. When the number of involved differentials is small in comparison to the cipher size we show that given N_S, the total number of pairs used for the attack, the advantage of an impossible differential attack should be computed as $\frac{N_S}{\ln(2)}\left(2^{-\ell} - 2^{-n}\right)$ instead of $\frac{N_S}{\ln(2)}\left(2^{-\ell}\right)$ as previously estimated. The impact of the newly developed theory is illustrated by practical examples. In particular, we treated the case of impossible differential attacks involving only one differential and show that the new advantage estimate influences the time complexity of the attack. More generally this study

shows that, for an impossible differential attack, the time complexity of the exhaustive key search has always been overestimated. We believe that this theory could also have an impact on the complexity of other statistical attacks as illustrated in the last part of this paper.

Acknowledgement

I would like to thank María Naya-Plasencia, Kaisa Nyberg and Patrick Derbez as well as the reviewers for their comments on an earlier version of this paper.

References

[BBS99a] Eli Biham, Alex Biryukov, and Adi Shamir. Cryptanalysis of Skipjack reduced to 31 rounds using impossible differentials. In Jacques Stern, editor, *Advances in Cryptology - EUROCRYPT '99, International Conference on the Theory and Application of Cryptographic Techniques, Prague, Czech Republic, May 2-6, 1999, Proceeding*, volume 1592 of *Lecture Notes in Computer Science*, pages 12–23. Springer, 1999.

[BBS99b] Eli Biham, Alex Biryukov, and Adi Shamir. Miss in the middle attacks on IDEA and Khufu. In Lars R. Knudsen, editor, *Fast Software Encryption, 6th International Workshop, FSE '99, Rome, Italy, March 24-26, 1999, Proceedings*, volume 1636 of *Lecture Notes in Computer Science*, pages 124–138. Springer, 1999.

[BGT11] Céline Blondeau, Benoît Gérard, and Jean-Pierre Tillich. Accurate estimates of the data complexity and success probability for various cryptanalyses. *Des. Codes Cryptography*, 59(1-3):3–34, 2011.

[BLNPS17] Christina Boura, Virginie Lallemand, María Naya-Plasencia, and Valentin Suder. Making the impossible possible. *Journal of Cryptology*, pages 1–33, 2017.

[BLNW12] Andrey Bogdanov, Gregor Leander, Kaisa Nyberg, and Meiqin Wang. Integral and multidimensional linear distinguishers with correlation zero. In Xiaoyun Wang and Kazue Sako, editors, *Advances in Cryptology - ASIACRYPT 2012 - 18th International Conference on the Theory and Application of Cryptology and Information Security, Beijing, China, December 2-6, 2012. Proceedings*, volume 7658 of *Lecture Notes in Computer Science*, pages 244–261. Springer, 2012.

[BN14] Céline Blondeau and Kaisa Nyberg. Links between truncated differential and multidimensional linear properties of block ciphers and underlying attack complexities. In Phong Q. Nguyen and Elisabeth Oswald, editors, *Advances in Cryptology - EUROCRYPT 2014 - 33rd Annual International Conference on the Theory and Applications of Cryptographic Techniques, Copenhagen, Denmark, May 11-15, 2014. Proceedings*, volume 8441 of *Lecture Notes in Computer Science*, pages 165–182. Springer, 2014.

[BN17] Céline Blondeau and Kaisa Nyberg. Joint data and key distribution of simple, multiple, and multidimensional linear cryptanalysis test statistic and its impact to data complexity. *Des. Codes Cryptography*, 82(1-2):319–349, 2017.

[BNS14a] Christina Boura, María Naya-Plasencia, and Valentin Suder. Scrutinizing and improving impossible differential attacks: Applications to CLEFIA, Camellia, LBlock and Simon. In Palash Sarkar and Tetsu Iwata, editors, *Advances in Cryptology - ASIACRYPT 2014 - 20th International Conference on the Theory and Application of Cryptology and Information Security, Kaoshiung, Taiwan, R.O.C., December 7-11, 2014. Proceedings, Part I*, volume 8873 of *Lecture Notes in Computer Science*, pages 179–199. Springer, 2014.

[BNS14b] Christina Boura, María Naya-Plasencia, and Valentin Suder. Scrutinizing and improving impossible differential attacks: Applications to CLEFIA, Camellia, LBlock and Simon (full version). *IACR Cryptology ePrint Archive*, 2014:699, 2014.

[BS90] Eli Biham and Adi Shamir. Differential cryptanalysis of DES-like cryptosystems. In Alfred Menezes and Scott A. Vanstone, editors, *Advances in Cryptology - CRYPTO '90, 10th Annual International Cryptology Conference, Santa Barbara, California, USA, August 11-15, 1990, Proceedings*, volume 537 of *Lecture Notes in Computer Science*, pages 2–21. Springer, 1990.

[BT13] Andrey Bogdanov and Elmar Tischhauser. On the wrong key randomisation and key equivalence hypotheses in Matsui's Algorithm 2. In Shiho Moriai, editor, *FSE 2013*, volume 8424 of *Lecture Notes in Computer Science*, pages 19–38. Springer, 2013.

[BTV16] Andrey Bogdanov, Elmar Tischhauser, and Philip S. Vejre. Multivariate Linear Cryptanalysis: The Past and Future of PRESENT. *IACR Cryptology ePrint Archive*, 2016:667, June 2016.

[Der16] Patrick Derbez. Note on impossible differential attacks. In Thomas Peyrin, editor, *Fast Software Encryption - 23rd International Conference, FSE 2016, Bochum, Germany, March 20-23, 2016, Revised Selected Papers*, volume 9783 of *Lecture Notes in Computer Science*, pages 416–427. Springer, 2016.

[DF16] Patrick Derbez and Pierre-Alain Fouque. Automatic search of meet-in-the-middle and impossible differential attacks. In Matthew Robshaw and Jonathan Katz, editors, *Advances in Cryptology - CRYPTO 2016 - 36th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 14-18, 2016, Proceedings, Part II*, volume 9815 of *Lecture Notes in Computer Science*, pages 157–184. Springer, 2016.

[GMW06] Catherine Greenhill, Brendan D. McKay, and Xiaoji Wang. Asymptotic enumeration of sparse 0-1 matrices with irregular row and column sums. *Journal of Combinatorial Theory, Series A*, 113(2):291 – 324, 2006.

[Knu94] Lars R. Knudsen. Truncated and higher order differentials. In Bart Preneel, editor, *Fast Software Encryption: Second International Workshop. Leuven, Belgium, 14-16 December 1994, Proceedings*, volume 1008 of *Lecture Notes in Computer Science*, pages 196–211. Springer, 1994.

[Knu98] Lars Knudsen. DEAL - A 128-bit Block Cipher, 1998. Technical report no. 151. Department of Informatics, University of Bergen, Norway.

[MSAK99] Shiho Moriai, Makoto Sugita, Kazumaro Aoki, and Masayuki Kanda. Security of E2 against truncated differential cryptanalysis. In Howard M. Heys and Carlisle M. Adams, editors, *Selected Areas in Cryptography, 6th Annual International Workshop, SAC'99, Kingston, Ontario, Canada, August 9-10,*

1999, Proceedings, volume 1758 of *Lecture Notes in Computer Science*, pages 106–117. Springer, 1999.

[SKU+00] Makoto Sugita, Kazukuni Kobara, Kazuhiro Uehara, Shuji Kubota, and Hideki Imai. Relationships among differential, truncated differential, impossible differential cryptanalyses against word-oriented block ciphers like RIJNDAEL, E2. In *AES Candidate Conference*, pages 242–254, 2000.

[Tod16] Yosuke Todo. Impossible differential attack against 14-round *Piccolo*-80 without relying on full code book. *IEICE Transactions*, 99-A(1):154–157, 2016.

[WZ11] Wenling Wu and Lei Zhang. Lblock: A lightweight block cipher. In Javier Lopez and Gene Tsudik, editors, *Applied Cryptography and Network Security - 9th International Conference, ACNS 2011, Nerja, Spain, June 7-10, 2011. Proceedings*, volume 6715 of *Lecture Notes in Computer Science*, pages 327–344, 2011.