Links Between Theoretical and Effective Differential Probabilities: Experiments on PRESENT

C. Blondeau and B. Gérard

INRIA project-team SECRET, France
{celine.blondeau, benoit.gerard}@inria.fr

Abstract. Recent iterated ciphers have been designed to be resistant to differential cryptanalysis. This implies that cryptanalysts have to deal with differentials having so small probabilities that, for a fixed key, the whole codebook may not be sufficient to detect it. The question is then, do these theoretically computed small probabilities have any sense? We propose here a deep study of differential and differential trail probabilities supported by experimental results obtained on a reduced version of PRESENT.
Keywords : differential cryptanalysis, differential probability, iterated block cipher, PRESENT.

1 Introduction

Differential cryptanalysis has first been applied to the *Data Encryption Standard* (DES) in the early 90's by E. Biham and A. Shamir [BS91,BS93]. Since then, many ciphers have been cryptanalyzed using differential cryptanalysis or one of the large family of variants (truncated differential [Knu94], higher order differential [Knu94], impossible differential [BBS99], ...).

The basic differential cryptanalysis is based on a differential over r rounds of the cipher.

Definition 1. *A r-rounds differential characteristic*
A r-rounds differential characteristic *is a couple* $(\delta_0, \delta_r) \in \mathbb{F}_2^m \times \mathbb{F}_2^m$. *The* probability of a differential (δ_0, δ_r) *is*

$$p_* \stackrel{\text{def}}{=} \Pr_{\mathbf{X}}\left[F_K^r(X) \oplus F_K^r(X \oplus \delta_0) = \delta_r\right],$$

where m is the input/output size of the cipher and F the round function.

Then, r rounds of the cipher can be distinguished from a random permutation using this differential. To break $r+1$ rounds of the block cipher, we look for differentials on r rounds. Then, for all the possible subkeys for the last round, the attacker does a partial decryption of the ciphertext pairs and count the number of time δ_r appears. For a wrong candidate, the probability that δ_r appears is around 2^{-m} and for the correct subkey, this probability is around $p_* + 2^{-m}$. It is widely accepted that the number of pairs needed to distinguish those two probabilities is of order p_*^{-1} if the so called *signal-to-noise ratio* is large enough $\left(S_N = \dfrac{p_*}{2^{-m}}\right)$.

Recent ciphers, the *Advanded Encryption Standard* (AES) for instance, have been designed to be resistant to the basic differential cryptanalysis. Nevertheless, when a new cipher is proposed, cryptanalysts try to mount the best possible linear and differential attacks. In the case of PRESENT[BKL+07], the cipher we used for the experiments, the actual best published attack is the one of Wang [Wan08]. But actually, there is lacks in the data complexity estimate of this attack. First of all, an erroneous way of counting the number of subkey hits leads to a too optimistic signal-to-noise ratio of 17.63 and the use of the formula of Selçuk [Sel08] for the success probability is not tight in the case of differential cryptanalysis. We are not going to discuss this here because it is out of the scope of this paper and it does not challenge the validity of the attack that still has a very high success rate when taking these remarks into account. The second point is that the probabilities p_* of the differentials may be underestimated because they are obtained taking into account only one way of going from the input difference to the output difference. This is discussed in Section 4.

Contributions of this work.
In this work, we first present the cipher we used for experiments Section 2. Then, in Section 3, we focus on differential trails that is sets of intermediate differences taken by a pair that matches a differential. The classical way of estimating a trail probability relies on some hypotheses that are not true. Nevertheless, experiments show that this theoretical probability makes sense as an average of the probability over the keys. In Section 4 we recall that a differential probability is the sum of the corresponding trail probabilities. Then, we present some experiments about the key dependency of this differential probability and give a way of understanding them. Finally, we conclude in Section 5 and sum-up the results as well as the problematics left as open questions.

2 PRESENT

To confirm our conjectures, we made experiments on a lightweight cipher presented in 2007 at *CHES* conference: PRESENT [BKL+07]. The structure of PRESENT is a *Substitution Permutation Network* (SPN).

2.1 Reduced version of PRESENT: SMALLPRESENT-[s]

For the experiments to be meaningful, we need to be able to exhaustively compute the ciphertexts corresponding to all possible plaintexts for all possible keys. That is the reason why we chose to work on a reduced version of PRESENT named SMALLPRESENT-[s] [Lea10]. The family of SMALLPRESENT-[s] has been designed to be used for such experiments. The value of s indicate the number of Sboxes per round. The Sboxes of PRESENT are all the same and are defined in $\mathbb{F}_2^4$. This transformation is described in Table 1. The size of the message is then $4s$. In this article we make experiments on SMALLPRESENT-[4] that is the version with 4 Sboxes. The full version of PRESENT has 16 Sboxes. One round of SMALLPRESENT-[4] and PRESENT are respectively depicted in Figure 8, Figure 9 (Appendix A).

2.2 Different key schedules for SMALLPRESENT-[4]

The problem with the reduced cipher presented in [Lea10] is the key schedule. Actually, in the whole PRESENT, most of the bits of a subkey are directly used in the subkey of the next round. Since for SMALLPRESENT-[s], the number of key bits is always 80 but the state size is only $4 \cdot s$, this is not true anymore for a small s. We decide to introduce two additional key schedules for our experiments.

1. *Same key*: The cipher has a master key that has the same size as the state and each subkey is equal to this master key. Therefore SMALLPRESENT-[4] is parameterized by a 16 bits master key.
2. *80-bits*: This key schedule is the one used in the full version of PRESENT and proposed in [Lea10].
3. *20-bits*: a homemade key schedule used with SMALLPRESENT-[4] similar to the one of the full version.
   The master key is represented as $K = k_{19}k_{18}\dots k_0$. At round i the 16-bits round key $K_i = k_{19}k_{18}\dots k_4$ consists in the 16 left-most bits of the current content of register K. After extracting the round key K_i, the key register is updated as follows:
   (a) $[k_{19}k_{18}\dots k_1k_0] = [k_6k_5\dots k_8k_7]$
   (b) $[k_{19}k_{18}k_{17}k_{16}] = S[k_{19}k_{18}k_{17}k_{16}]$
   (c) $[k_7k_6k_5k_4k_3] = [k_7k_6k_5k_4k_3] \oplus \textit{roundcounter}$
   The key is rotated by 13 bit positions to the left, the left most four bits are passed through the PRESENT Sbox, and the *roundcounter* value is exclusive-ored with bits $k_8k_7k_6k_5k_4$. We keep the 5-bits counter version. But we only study less than 7 rounds of SMALLPRESENT-[4] so the counter can be represented in 3 bits.

3 Differential trail probability

3.1 Notation

Let us denote by K the master key. The round subkeys derived from K are denoted by $K_1, K_2, \dots, K_r$. Let $F_{K_i} : \mathbb{F}_2^m \mapsto \mathbb{F}_2^m$ be a round function of a block cipher. We will denote by F_K^r the application of r rounds of the block cipher.

$$F_K^r = F_{K_r} \circ F_{K_{r-1}} \circ \cdots \circ F_{K_1}.$$

In this section, we focus on differentials that have been defined in Definition 1. In general, there is not one but many ways to go from the input difference to the output difference of a differential characteristic. Such a way is called *differential trail*.

Definition 2. *Differential trail*
A differential trail *of a cipher is a* $(r+1)$*-tuple* $(\beta_0, \beta_1, \cdots, \beta_r) \in (\mathbb{F}_2^m)^{r+1}$ *of intermediate differences at each round. The* probability of a differential trail $\beta = (\beta_0, \beta_1, \cdots, \beta_r)$ *is defined by*

$$p_\beta \stackrel{\text{def}}{=} \Pr_{\mathbf{X}}\left[\forall i \;\; F_K^i(X) \oplus F_K^i(X \oplus \beta_0) = \beta_i\right].$$

Computing the exact value of a trail probability is not possible for real ciphers since it need to encipher the whole codebook for all possible keys. The classical way of estimating a trail probability is to chain trails on 1 round. This approach is based on a formalism introduced by Lai, Massey and Murphy [LMM91].

Assumption
The sequence of differences at each round output forms a Markov chain.

Assuming this, the theoretical probability of a trail $\beta = (\beta_0, \beta_1, \cdots, \beta_r)$ is computed as follow

$$p_\beta^t = \prod_{i=1}^{r} \Pr_{\mathbf{X}} \left[F_{K_i}(X) \oplus F_{K_i}(X \oplus \beta_{i-1}) = \beta_i \right].$$

But actually, this assumption is not true (at least for SPN ciphers). We give a counter example in the following subsection and focus on the key dependency of the probability of a trail.

3.2 Key dependency of a trail

The main difficulty in the pre-computation phase is the search of good differential trails. The computed value of the probability of a differential trail relies on the widely used aforementioned assumption. Thus, in some case the theoretical value of the probability of a differential trail does not match with reality. For one round it is easy to see that the theoretical value corresponds to the real one. But for more than one round it seems legitimate to wonder what does this theoretical probability mean.
As we see below, the probability of a differential trail can be influenced by the choice of the master key used to encipher samples. Thus, we override the previous definition of a differential trial probability with the following one.

Definition 3. *Differential trail probability*
The probability of a differential trail $\beta = (\beta_0, \beta_1, \cdots, \beta_r)$ *is defined by*

$$p_\beta \stackrel{\text{def}}{=} \Pr_{\mathbf{X},\mathbf{K}} \left[\forall i \;\; F_K^i(X) \oplus F_K^i(X \oplus \beta_0) = \beta_i \right].$$

We now take into account the choice of the master key in the variable we use. For a r-rounds differential trail $\beta = (\beta_0, \beta_1, \cdots, \beta_r)$, let us define

$$\begin{aligned} T_K &\stackrel{\text{def}}{=} \frac{1}{2} \#\{X \in \mathbb{F}_2^m | F_K^i(X) \oplus F_K^i(X + \beta_0) = \beta_i \quad \forall 1 \leq i \leq r\}, \\ T[j] &\stackrel{\text{def}}{=} \#\{K | T_K = j\}. \end{aligned} \tag{1}$$

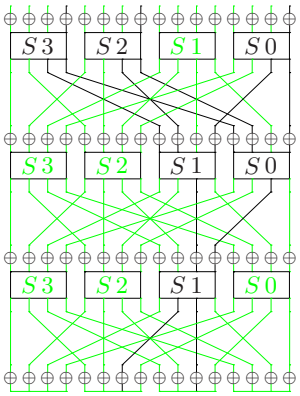
Let n_k be the number of bits of the master key. The real value of the trail probability is

$$p_\beta = 2^{-m-1} \sum_{K \in \mathbb{F}_2^{n_k}} T_K = 2^{-m-1-n_k} \sum_j T[j] \cdot j.$$

To motivate these new notation, we give an example where the key dependency is obvious.

Example of a trails with experimental probability not equal to the theoretical one

We illustrate this phenomena by a differential trail over 3 rounds on SMALLPRESENT-[4]: $\beta = (\texttt{0x1101}, \texttt{0xdd}, \texttt{0x30}, \texttt{0x220})$ (see Figure 1).



j	0	8	16
$T[j]$	131072	524288	393216

Fig. 1. Trail $\beta = (\texttt{0x1101}, \texttt{0xdd}, \texttt{0x30}, \texttt{0x220})$ and the corresponding $T[j]$'s

First, we are going to compute the theoretical probability of this trail. Assuming that the rounds of SMALLPRESENT-[4] are independent we can easily compute the theoretical value of this differential trail.

- *Round 1* 3 S-boxes with input difference `0x1` and output difference `0x3`.
- *Round 2* 2 S-boxes with input difference `0xd` and output difference `0x2`.
- *Round 3* 1 S-box with input difference `0x3` and output difference `0x6`.

We have 6-Sboxes with transition probability 2^{-2} therefore $p_\beta^t = 2^{-12}$. This means that the number of plaintext x such that $(X, X \oplus \texttt{0x1101})$ follows this trail for a fixed key should be $2^{16-1} \cdot 2^{-12} = 2^3$. We made experiments to check this assumption. For a fixed key we computed the number of plaintexts that follows this trail. In Figure 1 are given the values taken by $T[j]$ for all keys in $\mathbb{F}_2^{20}$ using the 20-bit key schedule. We can see that there are three kinds of key leading to three different values of $T[j]$. We found many of such trails on 3 rounds that split the space of keys.

Experiments on this trail show that for a fixed key, the theoretical probability of a differential trail do not always match with the real value of this trail probability. Experiments also show that for some keys the trail can be impossible. This can be of real significance because such phenomenon is also existing on 3 rounds of PRESENT. That means that, maybe, some differential trails used in a differential cryptanalysis may not have the expected probability for most of the keys.

Averaging the probabilities over the keys, we see that the effective probability of this differential trail is $2^{-11.6}$ (the theoretical one is 2^{-12}). This difference between real

and theoretical probabilities may weaken (or strengthen) some attacks. Does a lot of trails have such a difference between their theoretical value and the average probability? In the next subsection we will show that, most of the times, the theoretical value of a differential trail is close to the effective one (averaged over the keys).

3.3 Theoretical probability and average probability of a trail over the keys

To understand this phenomenon, we made some experiments on PRESENT. We observed that the theoretical probability is likely to be the average of the trail probability over all the possible keys. We make experiments on SMALLPRESENT-[4] with different key schedules. Let us recall that the theoretical probability of the differential trail β is denoted by p_β^t and the effective one (averaged over the keys) is denoted by p_β. In Figure 2, Figure 3 and Figure 4, we have computed the difference between $\log(p_\beta^t)$ and $\log(p_\beta)$ for 500 random trails.

- In *Figure 2* we assume that the round subkeys are derived from the 20-bits key schedule. We average the probabilities over the whole set of 2^{20} keys to obtain the value p_β.
- In *Figure 3* we assume that all the round subkeys are the same. We average the probabilities over the whole set of 2^{16} keys to obtain the value p_β.
- In *Figure 4* we assume that the round subkeys are derived from the 80-bits key schedule. Since we cannot average probabilities over the 2^{80} possible master keys, the computed value is obtained by an average over 2^{20} keys.

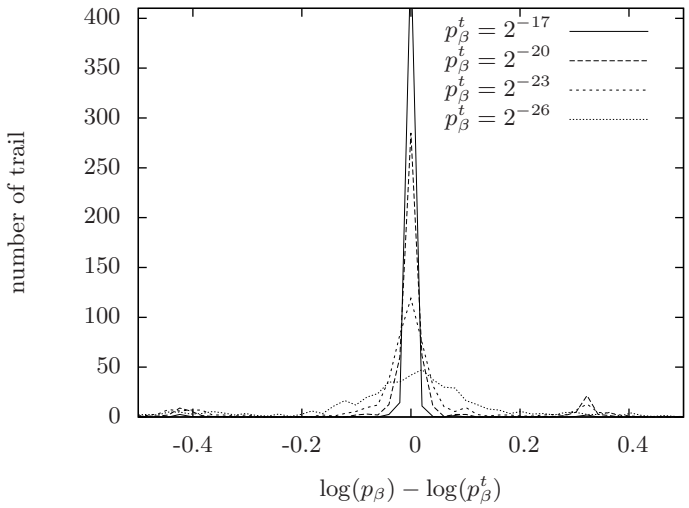


Fig. 2. Number of trails as a function of $\log(p_\beta) - \log(p_\beta^t)$ with a sample of 500 trails on 5 rounds and the 20-bits key schedule

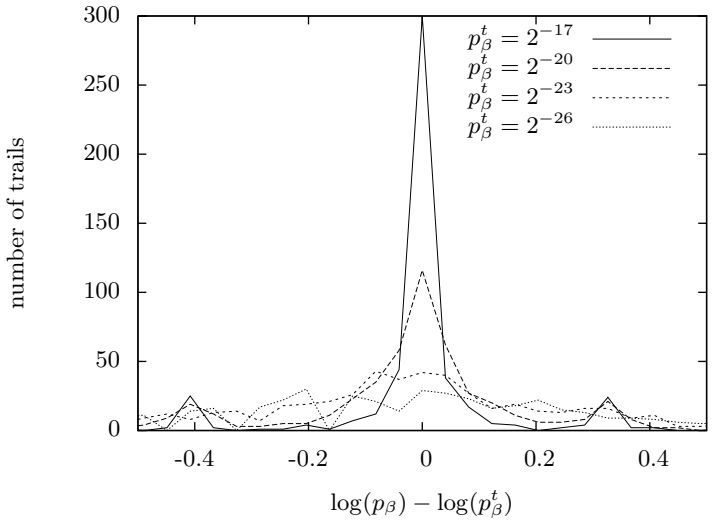


Fig. 3. Number of trails as a function of $\log(p_\beta) - \log(p^t_\beta)$ with a sample of 500 trails on 5 rounds and the same subkey for all rounds

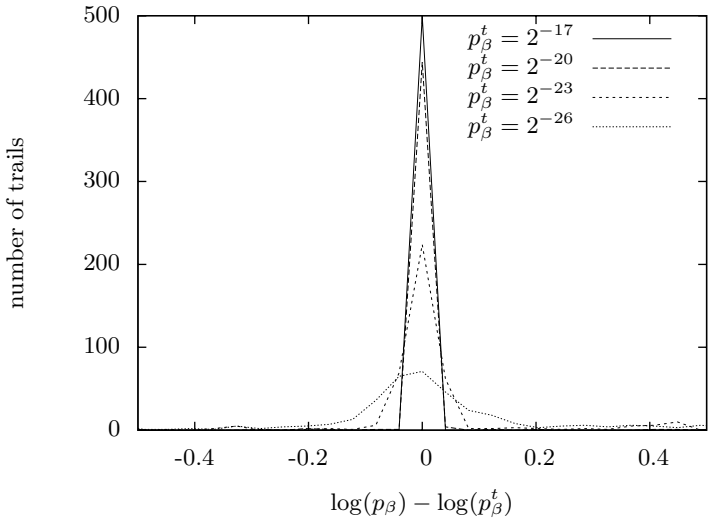


Fig. 4. Number of trails as a function of $\log(p_\beta) - \log(p^t_\beta)$ with a sample of 500 trails on 5 rounds and 2^{20} keys using the 80-bits key schedule

Remark:

We can see that the phenomenon is not the same depending on the key schedule. Indeed, in Figure 3 when the same key is taken over the 5 rounds the dependency of the round key is more important than in Figure 4 where the 80-bits key schedule is used implying that all the key bits are used only once (on average). This remark has motivate

the 20-bits key schedule we use in the following experiments that seems to be the most appropriate.

Experiments show that the average proportion of pairs satisfying a differential trail is close to the theoretical probability. We can observe that this behavior is getting worse as the probability is decreasing.
Nevertheless, it seems to be some symmetry what leads to the idea that taking enough trails will correct this and give better results.

4 Differential probability

4.1 Differential probability and trail probabilities

The first thing to say here is that the probability of a differential is the sum of the probabilities of the corresponding differential trails.

Lemma 1. *Let (δ_0, δ_r) be a r-rounds differential characteristic. Then the probability p_* of this differential is*

$$p_* = \sum_{\beta=(\delta_0,\beta_1,\ldots,\beta_{r-1},\delta_r)} p_\beta.$$

Proof. This is essentially due to the fact that a pair that matches a trail cannot match any other (they are disjoint events) and thus $\Pr[\cup_i A_i] = \sum_i \Pr[A_i]$.

For a large number of rounds, it is impossible to compute the probability of a differential (δ_0, δ_r) because there is too much differential trails that go from δ_0 to δ_r. Actually, in differential cryptanalysis, one uses a lower bound on the probability of the differential (δ_0, δ_r) by considering the sum of the likeliest trail probabilities.

In Section 3, we saw that the effective trail probability may not match with the theoretical one. Nevertheless, it seems to be some symmetry what leads to the idea that the sum of theoretical trail probabilities may give a good estimate of a differential probability (averaging over the key).

We made some experiments on 5 rounds of SMALLPRESENT-[4] with the 20 bits key schedule to see how many trails are required to get a good estimate of a differential probability. We computed the sum of the theoretical probabilities of many trails corresponding to the same differential characteristic. Since the cipher is small we also have computed the effective value of the differential by averaging over all plaintexts and all keys. In Figure 5 we have plotted the difference between both values for 20 differential characteristics. We can see that taking many trails give a better estimation of the differential probability.

Looking at the results in Figure 5 we can wonder whether it is possible to determine the number of trails to consider for estimating a differential probability. In this example we see that taking 2^7 trails seems to be sufficient but when we look at the whole cipher it is obviously not enough (see the following paragraph).

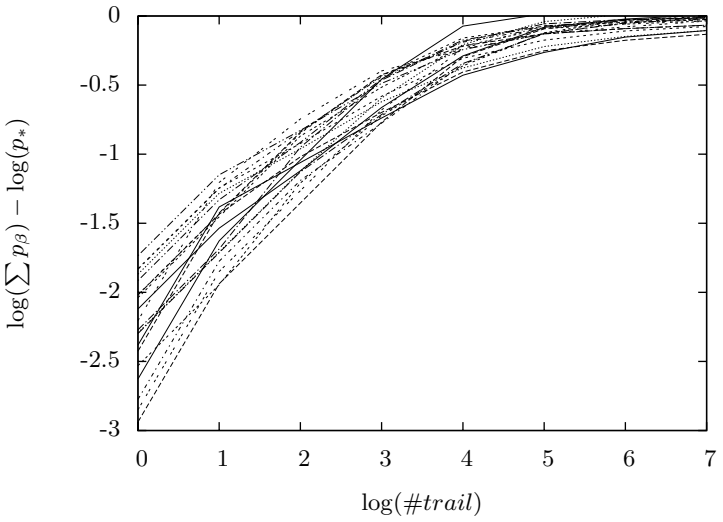


Fig. 5. Convergence of the sum of trails probabilities to the real value of the differential characteristic

Remark on Wang's paper [Wan08]
In [Wan08], the target is the whole PRESENT (64 bits). The differential used is

$$(d_0, d_{14}) = (\texttt{0x000000070000700}, \texttt{0x0000000900000009})$$

on 14 rounds obtained by iterating 3 times a differential trail on 4 rounds and by adding one more round at the beginning and at the end.

We observe some things about this trail.

- This trail on 4 rounds is not one of the best since it has a probability equal to 2^{-18} and we found a lot of differential trails with probability 2^{-12}. Nevertheless, it is the best iterative differential on 4 rounds.
- There exists lots of differential trails on 14 rounds with probability 2^{-62} and 2^{-62} is the best value of the probability of a differential trail over 14 rounds. Therefore the trail taken by Wang is one of the best.
- We have theoretically computed all differential trails with input difference d_0, output difference d_{14} and probability greater than 2^{-73}. We have sum-up the results in Figure 6. By taking the sum of the probabilities of the 2^k best trails, we observe that the probability of the differential characteristic (d_0, d_{14}) is greater than $2^{-57.53}$.

4.2 Key dependency of a differential probability

We now consider a differential (δ_0, δ_r) that is to be used in a differential cryptanalysis. The attacker will get some samples enciphered with a fixed master key. Depending on this key, the real probability of the differential will be smaller/equal/larger than the theoretically computed value.

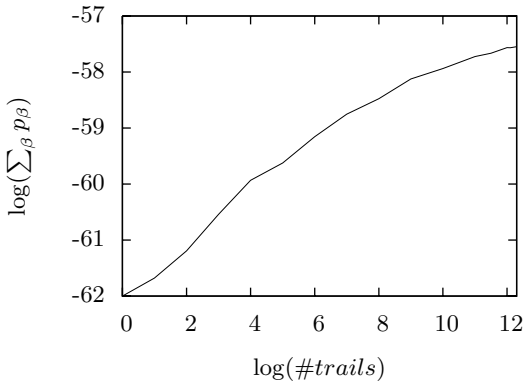


Fig. 6. Probability of the differential characteristic presented in [Wan08]as a function of the logarithm of the number of trails

For a fixed key K, let us denote by D_K the number of pairs of plaintexts with input difference δ_0 that lead to an output difference δ_r. Since we do not want to count a pair twice, we introduce a $\frac{1}{2}$ coefficient.

$$D_K \stackrel{\text{def}}{=} \frac{1}{2}\#\{X|F_K^r(X) + F_K^r(X+\delta_0) = \delta_r\}.$$

We are going to study the distribution of D_K's.

$$D[j] \stackrel{\text{def}}{=} \#\{K|D_K = j\}.$$

Previous experiments shown that the theoretical value of a differential probability (δ_0, δ_r) obtained by summing the theoretical probabilities of the corresponding trails is a good estimate of the proportion of pairs matching this differential averaged over the keys. Then, an intuitive way of looking at this problem is the following.

We look at the set of couples $(X, K) \in \mathbb{F}_2^m \times \mathbb{F}_2^{n_k}$. The proportion of pairs (X, K) such that

$$F_K^r(X) \oplus F_K^r(X \oplus \delta_0) = \delta_r \tag{2}$$

is p_* that we estimate summing the theoretical probabilities of the trails. Then, fixing a key can be seen as taking simultaneously and at random 2^m pairs. In this setting, the number of pairs fulfilling (2) is a random variable that follows an hyper-geometric distribution with parameters $(2^m, p_*, 2^{m+n_k})$. Since for cryptographic applications, $m = O(n_k) \gg 1$, this distribution can be approximated by a binomial law of parameters $(2^m, p_*)$.

We made some experiments on 5 rounds of SMALLPRESENT-[4] to check this. Using the 20-bits key schedule we computed the repartition of the D_K's. In Figure 7, we see that the D_K's seems to follow a binomial distribution as the intuition we got suggested.

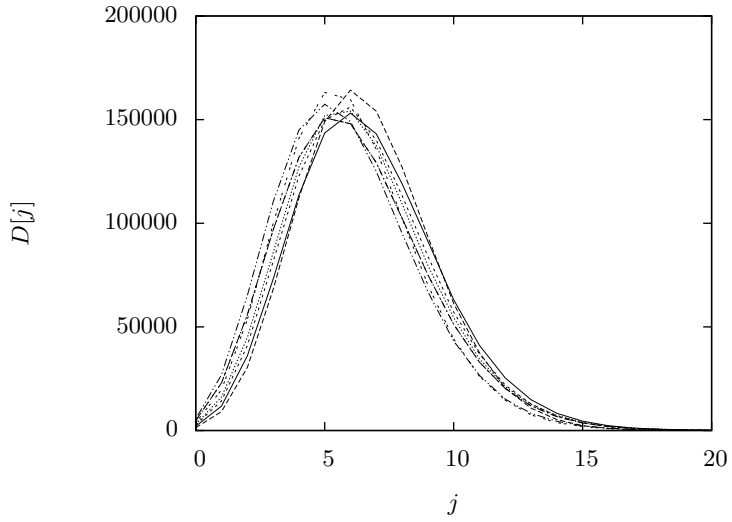


Fig. 7. Distribution of $D[j]$'s for 8 differentials over 5 rounds of SMALLPRESENT-[4]

This observation should be taken into account when computing the success probability of an attack. Let us denote by p_*^t the theoretical probability of the differential characteristic used in a cryptanalysis. We recall that n_k is the number of key bits. For $K \in \mathbb{F}_2^{n_k}$, the effective probability of the differential is $\frac{D_K}{2^{m-1}}$ where D_K is a random variable that follows a binomial distribution of parameters $(2^{m-1}, p_*^t)$. If we denote by $P_S(p_*)$ the success probability of a differential cryptanalysis using a differential with **effective** probability p_* (see [Sel08,BGT10]). Then the success probability of a differential cryptanalysis using a differential with **theoretical** probability p_*^t is

$$P_S = \sum_{i=0}^{2^{m-1}} P_S\left(\frac{i}{2^{m-1}}\right) \cdot \left[(p_*^t)^i (1-p_*^t)^{2^{m-1}-i} \binom{2^{m-1}}{i}\right]. \tag{3}$$

5 Conclusion

We have presented lots of experiments on differential cryptanalysis. The main teaching of this work is that claimed complexities of differential cryptanalyses on recent ciphers may be under/over-estimated.

The first point is the fact that estimating a differential probability with the probability of its main trail is really not suitable. To illustrate the first point, we estimated the probability of a differential used in [Wan08] to $2^{-57.53}$ while the author only takes into account the best trail and provides an estimate of 2^{-62}.

The second point is the key dependency of a differential probability. The theoretical probability of a differential makes sense if we average effective probabilities over the keys. But, an attacker uses a differential to attack a cipher parameterized by a fixed key. Then, the fact that the differential probability depends on the key has to be taken

into account. In Section 4, we propose a way to understand this dependency based on the hyper-geometric distribution and give a new formula for the success probability of a differential cryptanalysis.

This work give some elements for understanding differential cryptanalysis but it still remains some open questions. The two main problematics that seems to be of great interest are the following.

- The theoretical probability of a trail seems to be less meaningful as this probability decreases. How far does this theoretical value make sense?
- How can we get a good estimate of a differential probability without finding all the corresponding differential trails?

References

[BBS99] E. Biham, A. Biryukov, and A. Shamir. Cryptanalysis of Skipjack Reduced to 31 Rounds Using Impossible Differentials. In *EUROCRYPT '99*, volume 1592 of *LNCS*, pages 12–23, 1999.

[BGT10] C. Blondeau, B. Gérard, and J.-P. Tillich. Accurate Estimates of the Data Complexity and Success Probability for Various Cryptanalyses . *DCC special issue on Coding and Cryptography*, 2010. To appear.

[BKL+07] A. Bogdanov, L. R. Knudsen, G. Leander, C. Paar, A. Poschmann, M. J. B. Robshaw, Y. Seurin, and C. Vikkelsoe. PRESENT: An Ultra-Lightweight Block Cipher. In *CHES '07*, volume 4727 of *LNCS*, pages 450–466. SV, 2007.

[BS91] E. Biham and A. Shamir. Differential Cryptanalysis of DES-like Cryptosystems. *Journal of Cryptology*, 4(1):3–72, 1991.

[BS93] E. Biham and A. Shamir. Differential Cryptanalysis of the Full 16-round DES. In *CRYPTO'92*, volume 740 of *LNCS*, pages 487–496. Springer–Verlag, 1993.

[Knu94] L. R. Knudsen. Truncated and Higher Order Differentials. In *FSE '94*, volume 1008 of *LNCS*, pages 196–211. Springer–Verlag, 1994.

[Lea10] Gregor Leander. Small Scale Variants Of The Block Cipher PRESENT. Cryptology ePrint Archive, Report 2010/143, 2010. http://eprint.iacr.org/.

[LMM91] X. Lai, J. L. Massey, and S. Murphy. Markov Ciphers and Differential Cryptanalysis. In *EUROCRYPT '91*, volume 547, pages 17–38, 1991.

[Sel08] A. A. Selçuk. On Probability of Success in Linear and Differential Cryptanalysis. *Journal of Cryptology*, 21(1):131–147, 2008.

[Wan08] Meiqin Wang. Differential Cryptanalysis of Reduced-Round PRESENT. In *AFRICACRYPT '08*, volume 5023 of *LNCS*, pages 40–49. SV, 2008.

A Characteristics of PRESENT

x	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
S(x)	C	5	6	B	9	0	A	D	3	E	F	8	4	7	1	2

Table 1. The S-box of PRESENT

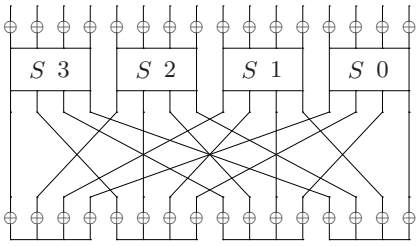


Fig. 8. 1 round of SMALLPRESENT-[4]

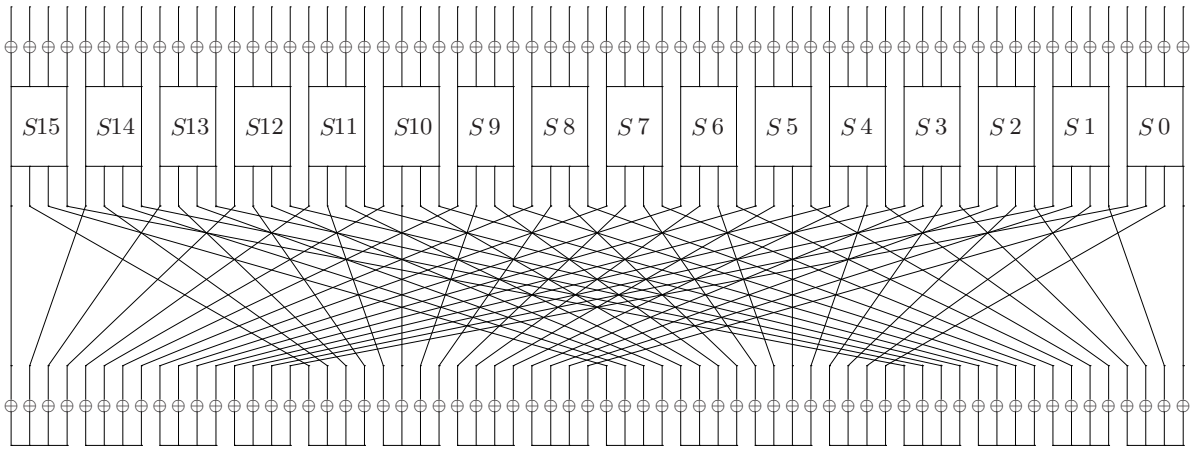


Fig. 9. 1 round of PRESENT