Rigorous Analysis of Truncated Differentials for 5-round AES

Lorenzo Grassi and Christian Rechberger

IAIK, Graz University of Technology, Austria
firstname.lastname@iaik.tugraz.at

Abstract. Since the development of cryptanalysis of AES and AES-like constructions in the late 1990s, the set of inputs (or a subset of it) which differ only in one diagonal has special importance. It appears in various (truncated) differential, integral, and impossible differential attacks, among others.

In this paper we present new techniques to analyze this special set of inputs, and report on new properties. In cryptanalysis, statements about the probability distribution of output differences are of interest. Until recently such statements were only possible for up to 4 rounds of AES (many results since two decades), and the only property described for 5 rounds is the multiple-of-8 property (Eurocrypt 2017). On the other hand, our understanding of this property is far from complete: e.g. does this property influence the average number of output pairs that lie in a particular subspace (i.e. the mean) and/or other probabilistic parameters?

Here we answer these questions by considering more generally the probability distribution of the number of different pairs of corresponding ciphertexts that lie in certain subspaces after 5 rounds. The variance of such a distribution is shown to be higher than for a random permutation, which immediately follows from the Eurocrypt 2017 result. *Surprisingly*, also the mean of the distribution is significantly different from random, something which cannot be explained by the multiple-of-8 property. To show this, a new approach is developed. For a rigorous proof of it, we need an APN-like assumption on the S-Box which closely resembles the AES-Sbox.

Keywords: AES · Truncated-Differential Cryptanalysis · Distinguisher

Contents

1 Introduction 2

2 Preliminary 5
2.1 Brief Description of AES 5
2.2 Properties of an S-Box 6

3 “Multiple-of-8” Property for 5-round AES 7

4 Probability Distribution for 5-round AES 9
4.1 Initial Considerations 10
4.2 About the “Uniform Distribution of Solutions of eq. (1)” 12
4.3 Comparison between the Prob. Distribution of 5-round AES and of a Random Permutation 13

5 Proof of Theorem 2 – Average Number of Collisions for 5-round AES 14
5.1 Reduction to the Middle Round 14
5.1.1 Idea of the Proof 14
5.2 A “Simpler” Case: 2^{16} Texts with Two Equal Generating Variables 15
5.3 Generic Case: 2^{32} Texts 17
5.3.1 First Case 18
5.3.2 Second Case 18
5.3.3 Third Case 19
5.3.4 Total Number of Different - not null - Common Solutions 19
5.4 Remarks 20

6 Proof of Theorem 2 – Variance – and of Lemma 1 20
6.1 Variance of the Prob. Distribution for 5-round AES 21
6.2 Proof of Lemma 1 21

7 Relation among Multiple-of-8, Mean and Variance 22

8 Practical Results on AES 23
8.1 5-round AES defined over $(\mathbb{F}_{2^n})^{4\times 4}$ 24
8.2 Practical Results on 4-bit AES 24

9 Truncated Differential Distinguishers for 5-round AES 27
9.1 Truncated Differential Distinguisher based on the Variance 28
9.2 Useful Approximation for the Prob. Distribution of 5-round AES 28
9.3 Truncated Differential Distinguisher based on the Mean 31

10 Property of the S-Box and 5-round Truncated Distinguisher based on the Mean 32
10.1 Preliminary Considerations and Practical Results 33
10.2 Observations and (possible) Explanation 33
10.3 MixColumns Dependence 35

11 Open Problems 35

A Subspace Trails Cryptanalysis for AES 40

B Number of Collisions – Random Permutation 41

C Truncated Differential on 5-round AES: MDS MixColumns Matrix 42

D Computational Cost of Algorithm 1 **44**
D.1 Details of the Secret-Key Distinguisher of Sect. 9.3 45

E Probability distribution for 5-round AES over $(\mathbb{F}_{2^n})^{4\times 4}$ **45**
E.1 Average Number of Collisions 46
E.2 Variance 50
E.3 Observation on the Variance – Comparison with a Random Permutation . . 51

F Key-Recovery Attacks on 5-round AES **52**
F.1 Generic Strategy 53
F.1.1 Wrong-Key Randomization Hypothesis 54
F.1.2 Implementation Strategy 54
F.1.3 Practical Tests on small-scale AES 55
F.2 Multiple-of-n Key-Recovery Attack 55
F.3 Truncated Diff. Attack based on the *Mean* 56
F.4 Truncated Diff. Attack based on the *Variance* 57

G Details of used S-Box **59**

1 Introduction

Since its conception by Biham and Shamir [BS91] in their effort to break the Data Encryption Standard (DES), differential cryptanalysis has been successfully applied in many cases such that any modern cipher is expected to have strong security arguments against this attack. Differential attacks exploit the fact that pairs of plaintexts with certain differences yield other differences in the corresponding ciphertexts with a non-uniform probability distribution. The methodology of differential cryptanalysis has been extended several times with a number of attack vectors, most importantly truncated differentials [Lai94, Knu95] – where only part of the difference between pairs of texts is considered, impossible differentials [BBS99] – where differences with zero-probability are exploited, and higher-order differentials [Lai94, Knu95].

AES (Advanced Encryption Standard) [DR02] is probably the most used and studied block cipher. Any cryptanalytic improvement on this cipher should thus be a good indicator of the novelty and quality of a new cryptanalytic technique. AES with its wide-trail strategy was designed to withstand differential and linear cryptanalysis, so pure versions of these techniques have limited applications in attacks. As is state of the art, truncated differential distinguishers which are *independent* of the secret key – that exploit differences which hold with probability different from 0 – can be set up for at most 3-round AES, while impossible differential ones which are *independent* of the secret key can be set up for at most 4-round AES.

Rigorous Analysis of the Probability Distribution of 5-round AES

Recently at Eurocrypt 2017, a new property which is *independent* of the secret key has been found for 5-round AES [GRR17a]. By appropriate choices of a number of input pairs, it is possible to make sure that the number of times that the difference of the resulting output pairs lie in a particular subspace[1] $\mathcal{ID}$ is always a multiple of 8. Such a distinguisher – re-called in detail in Sect. 3 – has then been exploited in e.g. [Gra18] and [BDK+18] to set up new competitive distinguishers and key-recovery attacks on round-reduced AES. Meanwhile, at Asiacrypt 2017, Rønjom, Bardeh and Helleseth [RBH17] presented new secret-key distinguishers for 3- to 6-round AES, which are based on the "yoyo-game", which means they require adaptively chosen ciphertexts in addition to chosen plaintexts.

Several open questions arise from the result provided in [GRR17a]: does this property influence e.g. the average number of output pairs that lie in a particular subspace (i.e. the mean)? Are other parameters (e.g. the variance, the skewness, ...) affected by the multiple-of-8 property?

Systematization of Knowledge: Diagonal Set of Plaintexts & Probability Distribution after 5-round AES. In this paper, given a diagonal set of plaintexts (that is, a set of plaintexts with one active diagonal), we consider the probability distribution of the corresponding number of pairs of ciphertexts that are equal in one fixed anti-diagonal after 5-round AES (without the final MixColumns operation) – equivalently, that belong to the same coset of a particular subspace $\mathcal{ID}$ – denoted in the following as the "number of collisions".

While a lot is known about the properties of a diagonal set of plaintexts for up to 4-round AES, a complete analysis for 5 or more rounds AES is still missing. E.g. given a diagonal set of plaintexts and the corresponding ciphertexts after 4 rounds, it is well known that the XOR-sum of the ciphertexts is equal to zero – see integral cryptanalysis [DKR97], or that each pair of ciphertexts cannot be equal in any of the four anti-diagonals

[1] A pair of texts has a certain difference if and only if the texts belong to the same coset of a particular subspace $\mathcal{X}$.

Table 1: *Expected properties of a diagonal set after 5-round encryption.* Given a set of 2^{32} chosen plaintexts all equal in three diagonals (that is, a diagonal set), we consider the *distribution* of the number of different pairs of ciphertexts that are equal in one anti-diagonal (equivalently, that lie in a particular subspace $\mathcal{ID}_I$ for $I \subseteq \{0,1,2,3\}$ fixed with $|I| = 3$ – as defined in Def. 5). *Theoretical expected values for mean and variance* of these distributions are given in this table for 5-round AES and for a random permutation. Practical results on AES are close and are discussed in Sect. 8.

	Random Permutation	**5-round AES**
Mean (Theorem 2)	$2\,147\,483\,647.5 \approx 2^{31}$	$2\,147\,484\,685.6 \approx 2^{31} + 2^{10}$
Variance (Theorem 2)	$2\,147\,483\,647 \approx 2^{31}$	$76\,842\,293\,834.905 \approx 2^{36.161}$
Multiple-of-8 [GRR17a]		✓

(as shown by Biham and Keller in [BK01]). For the first time, here we perform and propose a *precise theoretical differential analysis* of such distribution after 5-round AES (with an APN-like assumption on the S-Box which closely resembles the AES-Sbox), supported by practical implementations and verification. A numerical summary is given in Table 1.

Our Contribution: Probability Distribution for 5-round AES

Given 2^{32} chosen plaintexts with an active diagonal, we present a complete – and practical verified – analysis about the theoretical probability distribution of the number of pairs of corresponding ciphertexts that are equal in one fixed anti-diagonal after 5-round AES (without the final MixColumns operation). Compared to the case in which the ciphertexts are generated by permutation drawn at random (or for simplicity, a pseudo-random permutation), our results can be summarized as following:

Mean of 5-round AES. Firstly, by appropriate choice of 2^{32} plaintexts in a diagonal space $\mathcal{D}$, we prove for the first time that *the number of times that the resulting output pairs are equal in one fixed anti-diagonal* (equivalently, the number of times that the difference of the resulting output pairs lie in a particular subspace $\mathcal{ID}$) *is (a little) bigger for 5-round AES than for a random permutation, independently of the secret key.* A complete proof of this result – under an "APN-like" assumption on the S-Box which closely resembles the AES S-Box – can be found in Sect. 5.

An important technical contribution of this result is the new and original way in which such numbers are derived. To the best of our knowledge, such an approach to compute the probabilities exploited by our distinguisher *is new in the literature* and it is general enough to be applied to any AES like-cipher, providing new possible future results about truncated differential distinguishers.

Variance of 5-round AES. As a second contribution, *we theoretically compute the variance of the probability distribution just defined, and we show that it is higher (by a factor of approximately 36) for 5-round AES than for a random permutation.* As showed in Sect. 6.1, such property – whose proof is based on the "multiple-of-8" result [GRR17a] proposed at Eurocrypt 2017 – is independent of the secret key.

Probability Distribution of 5-round AES. By combining the multiple-of-8 property presented in [GRR17a], mixture differential cryptanalysis [Gra18] and the results just mentioned about the mean and the variance, in Sect. 4 we are able – for the first time – *to precisely formulate the probability distribution for 5-round AES.* In particular, we show that – given 2^{32} plaintexts with one active diagonal – the probability distribution of the number of different pairs of ciphertexts which are equal in one fixed anti-diagonal after 5-round AES (without the final MixColumns operation) is well described *by a sum of independent binomial distributions* $\mathcal{B}(n, p)$, that is

$$\text{5-AES} = 2^3 \times \mathcal{B}(n_3, p_3) + 2^{10} \times \mathcal{B}(n_{10}, p_{10}) + 2^{17} \times \mathcal{B}(n_{17}, p_{17})$$

where the exact values of n_3, n_{10}, n_{17} and p_3, p_{10}, p_{17} are provided in the following.

Relations between Multiple-of-8, Mean and Variance Properties

What is the relation between the multiple-of-8 property and the results just given about the mean and the variance of the probability distribution of 5-round AES? As explained in details in Sect. 7, *there is no "obvious relation" between the multiple-of-8 property and the result on the average number of collisions*, while the multiple-of-8 property and the variance result are strictly related.

In particular, the multiple-of-8 property and the result on the average number of collisions are derived from (completely) different considerations and argumentation. As we are going to show:

- the fact that the number of collisions is always a multiple of 8 for AES does *not* imply that such number is on average higher/equal/lower for AES;
- the fact that the number of collisions is on average higher for AES does *not* imply that it is a multiple-of-8.

New Truncated Differential Distinguishers for 5-round AES

As a concrete example of possible (practical) application of these results, in Sect. 9 we formulate

- the *first truncated diff. distinguisher for 5-round AES based on the mean*
- the *first truncated diff. distinguisher for 5-round AES based on the variance*

which are both *independent of the secret-key.* To the best of our knowledge, this is the *first time* that *a differential distinguisher exploits the variance parameter* – instead of the mean value (usually used in the literature) – to distinguish a cipher from a random permutation. More details about their data and computational costs can be found in the following.

Before going on, it is important to remark a crucial difference between our distinguishers and the others currently present in the literature for up to 4-round AES. Truncated differential distinguishers in the literature consider the probability that, *given random pairs of plaintexts*, the corresponding pairs of ciphertexts are equal or not in certain anti-diagonal(s). In order to set up our results up to 5-round AES, the price that has to be paid is less freedom in the input set. That is, in our case one must consider a particular set of input plaintexts – that is, an *entire coset* of a particular diagonal subspace rather than random pairs of texts – to appreciate a difference in the probability of the previous event. More details are given in the following.

Open Problems as Future Work

As already mentioned, the result regarding the average number of pairs of ciphertexts that are equal in one fixed anti-diagonal is theoretically computed under an "APN-like" assumption on the S-Box, which closely resembles the AES S-Box. Even if this assumption is restrictive, it resembles criteria used to design an S-Box which is strong against differential cryptanalysis. Hence, many ciphers in the literature are built using S-Boxes which (are close to) satisfy it. Nevertheless, in Sect. 10 we start an analysis regarding the influence of the S-Box on the average number of pairs of ciphertexts that satisfy the previous requirement. Due to the results of our practical tests on small-scale AES-like ciphers, we conjecture a possible link between the average number of collisions and some particular properties of the S-Box (described in the following). The problem to better understand and (if possible) formally prove such a link is open for future investigation.

Besides the question just mentioned, other open problems that arise from our work are listed in Sect. 11, and they mainly regard *(1st)* the possibility to set up secret-key distinguishers based on the skewness and *(2nd)* the possibility to set up a truncated diff. distinguisher on 6-round AES by extending the strategy presented in this paper for 5-round.

2 Preliminary

2.1 Brief Description of AES

The Advanced Encryption Standard [DR02] is a *Substitution-Permutation network* that supports key size of 128, 192 and 256 bits. The 128-bit plaintext initializes the internal state as a 4×4 matrix of bytes as values in the finite field $\mathbb{F}_{256}$, defined using the irreducible polynomial $X^8 + X^4 + X^3 + X + 1$. Depending on the version of AES, N_r rounds are applied to the state: $N_r = 10$ for AES-128, $N_r = 12$ for AES-192 and $N_r = 14$ for AES-256. An AES round applies four operations to the state matrix:

- *SubBytes* (S-Box) - applying the same 8-bit to 8-bit invertible S-Box 16 times in parallel on each byte of the state (provides non-linearity in the cipher);
- *ShiftRows* (SR) - cyclic shift of each row to the left;
- *MixColumns* (MC) - multiplication of each column by a constant 4×4 invertible matrix (MC and SR provide diffusion in the cipher);
- *AddRoundKey* (ARK) - XORing the state with a 128-bit subkey.

One round of AES can be described as $R(x) = K \oplus MC \circ SR \circ$ S-Box(x). In the first round an additional AddRoundKey operation (using a whitening key) is applied, and in the last round the MixColumns operation is omitted.

The Notation Used in the Paper. Let x denote a plaintext, a ciphertext, an intermediate state or a key. Then $x_{i,j}$ with $i, j \in \{0, ..., 3\}$ denotes the byte in the row i and in the column j. We denote by R one round[2] of AES, while we denote r rounds of AES by R^r (where we use the notation R^r_f in the case in which the last MixColumns operation is omitted).

We also define the diagonal and the anti-diagonal of a text as following. The i-th *diagonal* of a 4×4 matrix A is defined as the elements that lie on row r and column c such that $r - c = i \bmod 4$. The i-th *anti-diagonal* of a 4×4 matrix A is defined as the elements that lie on row r and column c such that $r + c = i \bmod 4$.

Finally, in the paper we often use the term "partial collision" (or "*collision*") when two texts are equal in certain bytes. We recall that this is equivalent to the fact that they belong to the same coset of a given subspace $\mathcal{X}$. For this reason, we sometimes use the subspace notation defined in [GRR17b] (briefly recalled in App. A) in order to present our results.

[2] Sometimes we use the notation R_K instead of R to highlight the round key K.

2.2 Properties of an S-Box

Given a bijective S-Box function on $\mathbb{F}_{2^n}$, let $\Delta_I, \Delta_O \in \mathbb{F}_{2^n}$. As we are going to show, some results of this paper depend on the probability distribution of the number of solutions of the following equation

$$\text{S-Box}(x \oplus \Delta_I) \oplus \text{S-Box}(x) = \Delta_O. \tag{1}$$

For this reason, we first recall some properties of the S-Box function.

In the following, we limit ourselves to consider the cases $\Delta_I \neq 0$ and $\Delta_O \neq 0$ – if $\Delta_O = 0$, the equation admits solution if and only if $\Delta_I = 0$ (the S-Box is bijective). Moreover, we denote by n_{Δ_I,Δ_O} the number of solutions x of the previous equation.

Mean Value. *Independently of the details of the S-Box*, the *mean value*[3] (or the average value) of n_{Δ_I,Δ_O} is equal to

$$\mathbb{E}[n_{\Delta_I,\Delta_O}] = \frac{2^n}{2^n - 1} \simeq 1 + 2^{-n} + 2^{-2n} + \ldots \tag{2}$$

(where $\mathbb{E}[n_{\Delta_I,\Delta_O}] \simeq 1 + 2^{-7.995}$ for the case $n = 8$). Indeed, observe that for each x and for each $\Delta_I \neq 0$ there exists $\Delta_O \neq 0$ (since S-Box is bijective) that satisfies eq. (1). Since there are 2^n different x and $2^n - 1$ different values of Δ_I and Δ_O, the average number of solutions is $\frac{2^n \cdot (2^n-1)}{(2^n-1)^2} = \frac{2^n}{(2^n-1)}$ independently of the details of the (bijective) S-Box.

Variance. In the following, we denote by $Var(n_{\Delta_I,\Delta_O})$ the *variance*[4] of n_{Δ_I,Δ_O}. This quantity *depends on the details of the S-Box*, in particular on the distribution of n_{Δ_I,Δ_O} with respect to Δ_I and Δ_O. For the AES S-Box case, for each $\Delta_I \neq 0$ there are 128 values of $\Delta_O \neq 0$ for which equation (1) has no solution, 126 values of $\Delta_O \neq 0$ for which equation (1) has 2 solutions ($\hat{x}$ is a solution iff $\hat{x} \oplus \Delta_I$ is a solution) and finally 1 value of $\Delta_O \neq 0$ for which equation (1) has 4 solutions. The variance for the AES S-Box is so equal to $Var_{AES}(n_{\Delta_I,\Delta_O}) = 2^2 \cdot \frac{126}{255} + 4^2 \cdot \frac{1}{255} - \left(\frac{256}{255}\right)^2 = \frac{67\,064}{65\,025}$.

Maximum Differential Probability. The *Maximum Differential Probability* DP_{max} of an S-Box is defined as

$$DP_{max} = \max_{\Delta_I \neq 0, \Delta_O} \frac{n_{\Delta_I,\Delta_O}}{2^n}. \tag{3}$$

Since all entries of the differential distribution table are even, DP_{max} is always bigger than or equal to 2^{-n+1} (i.e. $DP_{max} \geq 2^{-n+1}$). *Permutations with* $DP_{max} = 2^{-n+1}$ *are called Almost Perfect Nonlinear (APN).*

"Homogeneous" S-Box. Finally, given $\Delta_I \neq 0$ (resp. $\Delta_O \neq 0$), consider the probability distribution of n_{Δ_I,Δ_O} w.r.t. $\Delta_O \neq 0$ (resp. $\Delta_I \neq 0$). In the follow-up, we say that the S-Box is (differential) "*homogeneous*" if such distribution is independent of Δ_I (resp. Δ_O).

Just to give some concrete examples and w.r.t. this definition[5], the AES S-Box is differential "homogeneous" since for each $\Delta_I \neq 0$ (fixed), $Prob(n_{\Delta_I,\Delta_O} = 2) = \frac{126}{255}$ and $Prob(n_{\Delta_I,\Delta_O} = 4) = \frac{1}{255}$. The PRINCE S-Box (recalled in App. G) is instead not differential "homogeneous", since $Prob(n_{\Delta_I,\Delta_O} = 4)$ depends on $\Delta_I \neq 0$: e.g. $Prob(n_{\Delta_I,\Delta_O} = 4) = 0$ if $\Delta_I = 0\text{x}F$ (i.e. $n_{0\text{x}F,\Delta_O} \neq 4$ for all Δ_O) while $Prob(n_{\Delta_I,\Delta_O} = 4) = \frac{2}{15}$ if $\Delta_I = 0\text{x}A$ (two values of Δ_O satisfy $n_{0\text{x}A,\Delta_O} = 4$).

[3] In the case of a discrete probability distribution of a random variable X, the mean $E[X] \equiv \mu$ is defined as $\mu = \sum x \cdot P(x)$, i.e. the sum over every possible value x weighted by the probability of that value $P(x)$.

[4] In the case of a discrete probability distribution of a random variable X, the variance $Var(X) \equiv \sigma^2$ is defined as $\sigma^2 = E[(X - E[X])^2 = E[X^2] - E[X]^2$.

[5] We remark that this property is different from the "uniform" one introduced e.g. in [Nyb93]. In particular, an S-Box is differentially δ-uniform if the number of solutions n_{Δ_I,Δ_O} of (1) satisfies $n_{\Delta_I,\Delta_O} \leq \delta$ for each $\Delta_I \neq 0$ and $\Delta_O \neq 0$.

3 “Multiple-of-8” Property for 5-round AES

As already recalled in the introduction, the first secret-key distinguisher independent of the secret-key for 5-round AES – called “multiple-of-8” property [GRR17a] – has been presented at Eurocrypt 2017. In there, authors show that given a set of plaintexts with $1 \leq d \leq 3$ active diagonal(s), the corresponding ciphertexts after 5-round AES satisfy a particular property: the number of different pairs of ciphertexts which are equal in $1 \leq a \leq 3$ fixed anti-diagonal(s) (assuming the final MixColumns operation has been omitted) is always a multiple of 8 independently of the secret key.

Definition 1. Given two different texts $t^1, t^2 \in \mathbb{F}_{2^b}^{4\times 4}$, we say that $t^1 \leq t^2$ if $t^1 = t^2$ or if there exists $i, j \in \{0,1,2,3\}$ s.t. (1st) $t^1_{k,l} = t^2_{k,l}$ for all $k, l \in \{0,1,2,3\}$ with $k+4\cdot l < i+4\cdot j$ and (2nd) $t^1_{i,j} < t^2_{i,j}$. Moreover, we say that $t^1 < t^2$ if $t^1 \leq t^2$ (w.r.t. the previous definition) and $t^1 \neq t^2$.

Theorem 1 ([GRR17a])**.** *Given $2^{32\cdot d}$ plaintexts with $1 \leq d \leq 3$ active diagonals (equivalently, in the same coset of a diagonal subspace $\mathcal{D}_I$ for a certain $I \subseteq \{0,1,2,3\}$ with $|I| = d$), consider corresponding ciphertexts after 5 rounds, that is (p^i, c^i) for $i = 0, ..., 2^{32\cdot|I|} - 1$ where $c^i = R^5(p^i)$. Assuming the final MixColumns operation is omitted, the number n of different pairs of ciphertexts (c^i, c^j) for $c^i < c^j$ that are equal in $1 \leq a \leq 3$ anti-diagonals (equivalently, that belong to the same coset of a subspace $\mathcal{ID}_J$ for a certain $J \subseteq \{0,1,2,3\}$ with $|J| = 4 - a$)*

$$n := |\{(p^i, c^i), (p^j, c^j) \,|\, \forall p^i, p^j \in \mathcal{D}_I \oplus a,\; p^i < p^j \text{ and } c^i \oplus c^j \in \mathcal{ID}_J\}|. \tag{4}$$

is always a multiple of 8, independently of the secret key, of the details of the S-Box and of the MixColumns matrix.

We refer to [GRR17a, Gra18] for a detailed proof (see also [GRR17b] and App. A for a *formal definition* of the subspaces $\mathcal{D}_I, \mathcal{C}_I, \mathcal{ID}_I, \mathcal{M}_I \subseteq \mathbb{F}_{2^8}^{4\times 4}$), and we limit ourselves here to recall and highlight the main concepts that are useful for the follow-up. For completeness, we mention that a re-visitation of such proof has been recently proposed in [BCC19], where authors show that the above property is an immediate consequences of an equivalence relation on the input pairs, under which the difference at the output of the round function is invariant. In there, authors also prove that the branch number of the linear layer does not influence the validity of the multiple-of-8 property.

Idea of the Proof. First of all, a set of plaintexts with d active diagonals (i.e. a coset of $\mathcal{D}_I$) is always mapped into a set of texts with d active columns (i.e. a coset of $\mathcal{C}_I$) after one round, and in a coset of a mixed space $\mathcal{M}_I$ after two rounds, e.g.:

$$\underbrace{\begin{bmatrix} A & C & C & C \\ C & A & C & C \\ C & C & A & C \\ C & C & C & A \end{bmatrix}}_{\equiv \mathcal{D}_I \oplus \delta \text{ (diagonal subspace)}} \xrightarrow[\text{prob. 1}]{R(\cdot)} \underbrace{\begin{bmatrix} A & C & C & C \\ A & C & C & C \\ A & C & C & C \\ A & C & C & C \end{bmatrix}}_{\equiv \mathcal{C}_I \oplus \gamma \text{ (column subspace)}} \xrightarrow[\text{prob. 1}]{R(\cdot)} \underbrace{MC \times \begin{bmatrix} A & C & C & C \\ C & C & C & A \\ C & C & A & C \\ C & A & C & C \end{bmatrix}}_{\equiv MC(\mathcal{ID}_I) \oplus \omega \equiv \mathcal{M}_I \oplus \omega \text{ (mixed subspace)}}$$

where A and C denote respectively an active byte and a constant one. In a more compact way (see [GRR17b] and App. A for details):

$$R^2(\mathcal{D}_I \oplus \delta) = R(\mathcal{C}_I \oplus \gamma) = MC(\mathcal{ID}_I) \oplus \omega \equiv \mathcal{M}_I \oplus \omega,$$

which implies

$$\mathcal{D}_I \oplus \delta \xrightarrow[\text{prob. 1}]{R(\cdot)} \mathcal{C}_I \oplus \gamma \xrightarrow{R^2(\cdot)} \mathcal{D}_J \oplus \delta' \xrightarrow[\text{prob. 1}]{R_f^2(\cdot)} \mathcal{ID}_J \oplus \omega'.$$

Due to the 1-/2-round truncated diff. with prob. 1, one can focus only on the two central rounds $\mathcal{C}_I \oplus \gamma \to \mathcal{D}_J \oplus \delta'$ and prove a result equivalent to Theorem 1: given $2^{32\cdot d}$ plaintexts with $1 \le d \le 3$ active *column(s)*, the idea is to prove that the number of different pairs of texts which are equal in $1 \le a \le 3$ *diagonal(s) after 2 rounds* (equivalently, that are in the same coset of a diagonal subspace $\mathcal{D}_J$ with $|J| = 4 - a$) is always a multiple of 8.

To do this, the idea is to show that, given p^1 and p^2 in $\mathcal{C}_I \oplus \gamma$ as before, they are equal in $1 \le a \le 3$ diagonal(s) after 2 rounds *if and only if* there exist other *related*[6] pair of plaintexts s^1 and s^2 in $\mathcal{C}_I \oplus \gamma$ that are equal in (the same) a diagonal(s) after 2 rounds, that is:

$$R^2(p^1) \oplus R^2(p^2) \in \mathcal{D}_J \qquad \text{iff} \qquad R^2(s^1) \oplus R^2(s^2) \in \mathcal{D}_J. \tag{5}$$

Using the "super-Sbox" notation

$$\textit{super-Sbox}(\cdot) = \text{S-Box} \circ ARK \circ MC \circ \text{S-Box}(\cdot)$$

and since 2-round AES can be rewritten as

$$R^2(\cdot) = ARK \circ MC \circ SR \circ \textit{super-Sbox} \circ SR(\cdot),$$

note that the result (5) is immediately verified if one proves that $R^2(p^1) \oplus R^2(p^2) = R^2(s^1) \oplus R^2(s^2)$, or equivalently

$$\textit{super-Sbox}(\tilde{p}^1) \oplus \textit{super-Sbox}(\tilde{p}^2) = \textit{super-Sbox}(s^1) \oplus \textit{super-Sbox}(s^2).$$

For simplicity, we limit ourselves to give all the details for the case $d = |I| = 1$ – the proof for the other cases is analogous. Consider two elements $\tilde{p}^1 = SR(p^1)$ and $\tilde{p}^2 = SR(p^2)$ that differ in one anti-diagonal (equivalently, p^1 and p^2 differ in one column), i.e. that are in the same coset of $SR(\mathcal{C}_I) \oplus a \equiv \mathcal{ID}_I \oplus a$ for $a \in \mathcal{ID}_I^{\perp}$. Since $\tilde{p}^1$ and $\tilde{p}^2$ differ in only one anti-diagonal, there exist $x_0^j, x_1^j, x_2^j, x_3^j \in \mathbb{F}_{2^8}$ for $j = 1, 2$ such that:

$$\tilde{p}^1 = a \oplus \begin{bmatrix} x_0^1 & 0 & 0 & 0 \\ 0 & 0 & 0 & x_1^1 \\ 0 & 0 & x_2^1 & 0 \\ 0 & x_3^1 & 0 & 0 \end{bmatrix}, \qquad \tilde{p}^2 = a \oplus \begin{bmatrix} x_0^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & x_1^2 \\ 0 & 0 & x_2^2 & 0 \\ 0 & x_3^2 & 0 & 0 \end{bmatrix}.$$

For the follow-up, we say that $\tilde{p}^j$ is "generated" by the variables $(x_0^j, x_1^j, x_2^j, x_3^j)$, that is

$$\tilde{p}^j \equiv (x_0^j, x_1^j, x_2^j, x_3^j) = a \oplus \bigoplus_{i=0}^{3} x_i^j \cdot e_{i,-i \bmod 4}.$$

Let $\mathscr{S}_{\tilde{p}^1,\tilde{p}^2}$ be the set of pairs of texts in $\mathcal{ID}_I \oplus a$ obtained by swapping the generating variables of $\tilde{p}^1$ and $\tilde{p}^2$. More formally, let $\Phi \subseteq \{0, 1, 2, 3\}$ s.t.

$$\forall i \in \Phi : \; x_i^1 = x_i^2 \qquad \text{and} \qquad \forall i \notin \Phi : \; x_i^1 \neq x_i^2.$$

Given $\tilde{p}^1$ and $\tilde{p}^2$ as before, the set $\mathscr{S}_{\tilde{p}^1,\tilde{p}^2}$ contains all $2^{8\cdot|\Phi|+(3-|\Phi|)} = 2^{3+7\cdot|\Phi|}$ pairs of texts (s_1, s_2) for all $\Psi \subseteq \{0, 1, 2, 3\}$ and for all $\alpha_0, ..., \alpha_{|\Phi|} \in \mathbb{F}_{2^8}$ s.t.

$$\begin{aligned} s^1 &= a \oplus \bigoplus_{i \in \{0,1,2,3\} \setminus \Phi} \left\{ \left[x_i^1 \cdot \delta_i(\Psi) \oplus x_i^2 \cdot (1 - \delta_i(\Psi)) \right] \cdot e_{i,-i \bmod 4} \right\} \oplus \bigoplus_{j \in \Phi} \alpha_j \cdot e_{j,-j \bmod 4} \\ s^2 &= a \oplus \bigoplus_{i \in \{0,1,2,3\} \setminus \Phi} \left\{ \left[x_i^2 \cdot \delta_i(\Psi) \oplus x_i^1 \cdot (1 - \delta_i(\Psi)) \right] \cdot e_{i,-i \bmod 4} \right\} \oplus \bigoplus_{j \in \Phi} \alpha_j \cdot e_{j,-j \bmod 4} \end{aligned} \tag{6}$$

[6] As shown in [Gra18], the pairs of texts (p^1, p^2) and (s^1, s^2) are not independent, in the sense that they share the same generating variables.

where $\delta_x(A)$ is the Dirac measure defined as

$$\delta_x(A) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}$$

As shown in details in [GRR17a, Gra18], *since each column of $\tilde{p}^1$ and $\tilde{p}^2$ depends on different and independent variables, since the super-Sbox works independently on each column and since the XOR-sum is commutative*, it follows that

$$\forall s^1, s^2 \in \mathscr{S}_{\tilde{p}^1,\tilde{p}^2} : \qquad \textit{super-Sbox}(\tilde{p}^1) \oplus \textit{super-Sbox}(\tilde{p}^2) = \textit{super-Sbox}(s^1) \oplus \textit{super-Sbox}(s^2).$$

Finally, since the cardinality of each set $S_{\tilde{p}^1,\tilde{p}^2}$ is always a multiple of 8, it follows that the number of collisions must be a multiple of 8.

Open Problems: Starting Point of Our Work

As already mentioned in the introduction, several open questions arise from the result provided in [GRR17a]. In particular, given a set of $2^{32\cdot d}$ plaintexts with $1 \le d \le 3$ active diagonal(s) (eq., in the same coset of $\mathcal{D}_I$ for $I \subseteq \{0,1,2,3\}$ with $d = |I|$), consider the probability distribution of the number of pairs of ciphertexts which are equal in $1 \le a \le 3$ fixed anti-diagonal(s) (assuming the final MixColumns operation has been omitted):

- is it possible to say something about the mean, the variance and the skewness of this distribution?
- *does the multiple-of-8 property influence e.g. the average number of output pairs that lie in a particular subspace (i.e. the mean)? Are other parameters (e.g. the variance, the skewness, ...) affected by the multiple-of-8 property?*

In the following, we (partially) answer these questions.

4 Probability Distribution for 5-round AES

Given a set of 2^{32} plaintexts with one active diagonal, the probability distribution of the number of pairs of ciphertexts which are equal in one fixed anti-diagonal (assuming the final MixColumns operation is omitted) after 5-round AES is given in the following Theorem.

Theorem 2. *Consider an AES-like cipher that works with texts in $\mathbb{F}_{2^8}^{4\times 4}$, such that the MixColumns matrix is an MDS matrix[7] and s.t. the solutions of eq. (1) are uniformly distributed for each input/output difference $\Delta_I \neq 0$ and $\Delta_O \neq 0$.*

Consider 2^{32} plaintexts p^i for $i = 0, 1, ..., 2^{32} - 1$ with one active diagonal (equivalently, in a coset of a diagonal subspace $\mathcal{D}_i$ for $i \in \{0,1,2,3\}$), and the corresponding ciphertexts after 5 rounds without the final MixColumns operation, that is $c^i = R^5(p^i)$. The probability distribution of the number of different pairs of ciphertexts (c^i, c^j) with $c^i < c^j$ that are equal in 1 anti-diagonal (equivalently, that belong to the same coset of $\mathcal{ID}_J$ for $J \subseteq \{0,1,2,3\}$ fixed with $|J| = 3$) – denoted in the following by 5-AES – is described by

$$\textit{5-AES} = 2^3 \times X_3 + 2^{10} \times X_{10} + 2^{17} \times X_{17} \tag{7}$$

where

$$\forall i = 3, 10, 17 : \qquad X_i \sim \mathcal{B}(n_i, p_i) \tag{8}$$

[7] A matrix $M \in \mathbb{F}_{2^b}^{n\times n}$ is called *Maximum Distance Separable* (MDS) matrix iff it has branch number $B(M)$ equal to $B(M) = n+1$. The branch number is defined as $B(M) = \min_{0\neq x\in\mathbb{F}_{2^b}^n}\{wt(x) + wt(M(x))\}$ where wt is the hamming weight. Similarly, a $n \times n$ matrix is "*almost MDS*" if its branch number is n.

are binomial distributions s.t.

$$\begin{aligned} n_3 &= 2^{28} \cdot (2^8-1)^4 & p_3 &= 2^{-32} + 2^{-53.983};\\ n_{10} &= 2^{23} \cdot (2^8-1)^3 & p_{10} &= 2^{-32} - 2^{-45.989};\\ n_{17} &= 3 \cdot 2^{15} \cdot (2^8-1)^2 & p_{17} &= 2^{-32} + 2^{-37.986}. \end{aligned}$$

In particular, the probability distribution of the number of different pairs of ciphertexts that are equal in one fixed anti-diagonal has mean value $\mu = 2\,147\,484\,685.6$ *and standard deviation* $\sigma = 277\,204.426$.

For completeness, we mention that the same result holds also in the decryption direction (that is, using chosen ciphertexts instead of chosen plaintexts).

Lemma 1. *Consider an AES-like cipher that works with texts in* $\mathbb{F}_{2^8}^{4\times 4}$ *and for which the assumptions of Theorem 2 hold.*

Consider 2^{32} *plaintexts* p^i *for* $i = 0, 1, ..., 2^{32}-1$ *with one active diagonal (equivalently, in a coset of a diagonal space* $\mathcal{D}_i$ *for* $i \in \{0,1,2,3\}$*), and the corresponding ciphertexts after 5 rounds without the final MixColumns operation, that is* $c^i = R^5(p^i)$. *The probability to have* $n \in \mathbb{N}$ *different pairs of ciphertexts* (c^i, c^j) *with* $c^i < c^j$ *that are equal in one fixed anti-diagonal (equivalently, that belong to the same coset of* $\mathcal{ID}_J$ *for* $J \subseteq \{0,1,2,3\}$ *fixed with* $|J| = 3$*) is given by:*

$$Prob(n) = \begin{cases} 0 & \text{if } n \text{ mod } 8 \neq 0 \\ \sum_{(k_3,k_{10},k_{17})\in K_n} \left[\prod_{i\in\{3,10,17\}} \underbrace{\binom{n_i}{k_i} \cdot (p_i)^{k_i} \cdot (1-p_i)^{n_i-k_i}}_{\sim\mathcal{B}(n_i,p_i)} \right] & \text{otherwise} \end{cases}$$

where

$$K_n = \left\{ (k_3, k_{10}, k_{17}) \in \mathbb{N}\times\mathbb{N}\times\mathbb{N} \,\middle|\, 0 \le k_i \le n_i \;\text{ and }\; 2^3 \cdot k_3 + 2^{10} \cdot k_{10} + 2^{17} \cdot k_{17} = n \right\}$$

and where n_i *and* p_i *for* $i = 3, 10, 17$ *are given in Theorem 2.*

Note that $Prob(n > [2^3 \cdot n_3 + 2^{10} \cdot n_{10} + 2^{17} \cdot n_{17}]) = 0$.

4.1 Initial Considerations

In order to prove the results of Theorem 2 and Lemma 1, we start by formally computing the values n_i for $i = 3, 10, 17$ and by proving the result given in (7) - (8). In the next sections, we formally compute the probabilities p_i for $i = 3, 10, 17$, the value of the mean and the variance, and the probability given in Lemma 1.

Multiple-of-8 Property & Mixture Differential Cryptanalysis. First of all, given 2^{32} plaintexts with one active diagonal, *the corresponding pairs of ciphertexts are not independent/unrelated.* Due to the multiple-of-8 property [GRR17a] and of mixture differential cryptanalysis [Gra18], *these pairs of texts can be divided in sets* $\mathscr{S}$ *defined as in (6)* s.t.

- $2^3 \times n_3$ sets have cardinality 8 – each one of these sets contains pair of plaintexts for which the generating variables are all different after 1-round encryption;
- $2^{10} \times n_{10}$ sets have cardinality 2^{10} – each one of these sets contains pair of plaintexts for which one out of the four generating variables is equal (and three are different) after 1-round encryption;

- $2^{17} \times n_{17}$ sets have cardinality 2^{17} – each one of these sets contains pair of plaintexts for which two out of the four generating variables are equal (and two are different) after 1-round encryption;
- $2^{24} \times n_{24}$ sets have cardinality 2^{24} – each one of these sets contains pair of plaintexts for which three out of the four generating variables are equal (and one is different) after 1-round encryption.

The values of $n_3, n_{10}, n_{17}, n_{24}$ are computed in details in the next paragraph. All these sets $\mathscr{S}$ just given have particular properties, namely:

1. the two ciphertexts of pairs in the same set are either all equal or all different in $1 \le a \le 3$ anti-diagonals (equivalently, either belong or not belong to the same coset of $\mathcal{ID}_J$ where $J \subseteq \{0, 1, 2, 3\}$ where $a = 4 - |J|$);
2. pairs of texts of different sets are independent (in the sense that pair of texts of different sets do not satisfy the property just given for the case of pairs of texts that belong to the same set $\mathscr{S}$).

In other words, given a set of pairs just defined, it is *not* possible that some pairs of ciphertexts in such a set are equal in $1 \le a \le 3$ anti-diagonals (i.e. belong to the same coset of $\mathcal{ID}_J$) after 5 rounds, while other pairs of ciphertexts in the same set are different in those a anti-diagonals. We refer to [Gra18] for more details about this fact (exploited to set up the Mixture Differential distinguisher).

About the Values of n_3, n_{10}, n_{17}. Given a set of 2^{32} chosen texts with one active column[8], let's first compute *the number of different pairs of texts with v equal generating variables for* $0 \le v \le 3$. Note that given such a set of chosen texts, it is possible to construct $\binom{2^{32}}{2} = 2^{31} \cdot (2^{32} - 1) \simeq 2^{63}$ different pairs. Among them, the number of pairs of texts with $0 \le v \le 3$ equal generating variables (and $4 - v$ different generating variables) *after one round* is given by

$$\binom{4}{v} \cdot 2^{31} \cdot (2^8 - 1)^{4-v}. \tag{9}$$

Indeed, note that if v variables are equal for the two texts of the given pair, then these variables can take $(2^8)^v$ different values. For each one of the remaining $4 - v$ variables, the variables must be different for the two texts. Thus, these $4 - v$ variables can take exactly $\left[2^8 \cdot (2^8 - 1)\right]^{4-v}/2$ different values. The result follows immediately since there are $\binom{4}{v}$ different combinations of v variables.

Consider now separately the sets of pairs of texts with "no equal generating variables" (namely, $v = 0$), the set of of pairs of texts with "one equal (and three different) generating variable(s)" (namely, $v = 1$) and finally the set of of pairs of texts with "two equal (and two different) generating variable" (namely, $v = 2$). It follows that the number $2^{7 \cdot v + 3} \times n_{7 \cdot v + 3}$ of sets $\mathscr{S}$ of pairs of texts with v equal generating variables for $v = 0, 1, 2, 3$ satisfies

$$2^{7 \cdot v + 3} \times n_{7 \cdot v + 3} = \underbrace{\binom{4}{v} \cdot 2^{31} \cdot (2^8 - 1)^{4-v}}_{\equiv (9)}$$

which implies

$$n_{7 \cdot v + 3} = \frac{1}{2^{7 \cdot v + 3}} \times \binom{4}{v} \cdot 2^{31} \cdot (2^8 - 1)^{4-v}. \tag{10}$$

The values of n_3, n_{10}, n_{17} – which correspond respectively to $v = 0, 1$ and 2 – are simple obtained exploiting the previous formula.

[8] Remember that one active diagonal is mapped to one active column after 1-round AES encryption.

About Binomial Distributions $X_i \sim \mathcal{B}(n_i, p_i)$ for $i = 3, 10, 17$. Before going on, we remark that if three out of the four generating variables of the plaintexts are equal after 1-round encryption, then the corresponding ciphertexts cannot be equal in any anti-diagonal (due to e.g. the impossible diff. trail on 4-round AES given in Theorem 5 – App. A).

Due to the previous facts, it follows that the probability of the event "$n = 8 \cdot n'$ pairs of ciphertexts equal in one fixed anti-diagonal" for $n' \in \mathbb{N}$ – equivalently, "$n = 8 \cdot n'$ collisions" in a coset of $\mathcal{ID}_J$ for $J \subseteq \{0, 1, 2, 3\}$ with $|J| = 3$ – corresponds to the sum of the probabilities to have "$2^3 \cdot k_3$ collisions in the first set *and* $2^{10} \cdot k_{10}$ collisions in the second set *and* $2^{17} \cdot k_{17}$ collisions in the third set" *for each* k_3, k_{10}, k_{17} *such that* $2^3 \cdot k_3 + 2^{10} \cdot k_{10} + 2^{17} \cdot k_{17} = n$.

Moreover, note that each one of these (independent) events is well characterized by a *binomial distribution*[9]. By definition, a binomial distribution with parameters n and p is the discrete probability distribution of the number of successes in a sequence of n independent yes/no experiments, each of which yields success with probability p. In our case, given n pairs of texts, each one of them satisfies or not the above property/requirement with the same probability p.

Due to all these initial considerations, it follows that the distribution 5-AES of the number of collisions for the AES case is given by

$$\text{5-AES} = 2^3 \times X_3 + 2^{10} \times X_{10} + 2^{17} \times X_{17}$$

where $X_i \sim \mathcal{B}(n_i, p_i)$ for $i = 3, 10, 17$ are binomial distributions.

4.2 About the "Uniform Distribution of Solutions of eq. (1)"

Before going on, we discuss the assumptions of the Theorem, focusing on the one related to the properties/details of the S-Box. The fact that "the solutions of eq. (1) are uniformly distributed for each $\Delta_I \neq 0$ and $\Delta_O \neq 0$" basically corresponds of an S-Box that satisfies the following properties: *(1st) it is "homogeneous"* (see Sect. 2.2 for a detailed definition) *and (2nd) its $Var(n_{\Delta_I,\Delta_O})$ is as "low" as possible. This is close to being true if the S-Box is APN, or if the SBox is "close" to be APN.* Note that even if the variance $Var(n_{\Delta_I,\Delta_O})$ is related "in some sense" to DP_{max}, *S-Boxes with equal DP_{max} can have very different variance.* Moreover, the variance of an S-Box S_1 can be bigger than the corresponding variance of an S-Box S_2 even if DP_{max} of S_1 is lower than DP_{max} of S_2 (see Table 3 in Sect. 10 for concrete examples).

Although much is known for (bijective) APN permutations in odd dimension, currently only little is known for the case of even dimension and what is known relies heavily on computer checking. In particular, there is no APN permutation of dimension 4 [LP07], while there is at least one APN permutation, up to equivalence, of dimension 6 – called the Dillon's permutation [BDMW10]. The question of finding an APN bijective (n, n)-function for even $n \geq 8$ is still open.

As a result, in the case of dimensions equal to a power of 2 (e.g. $\mathbb{F}_{2^4}$ or $\mathbb{F}_{2^8}$), *the only (known) S-Box that (approximately) matches the assumptions of the Theorem in dimensions 4 or 8 is the one generated by the multiplicative-inverse permutation unless affine equivalence relations*[10], as for example the AES S-Box, which is not APN but differentially 4-uniform [Nyb91] (e.g. note that the variance of the AES S-Box is 67 064/65 025 vs 64 004/65 025 of an APN S-Box). As we are going to show, our practical results on small-scale AES (for

[9] We remember that the mean μ and the variance σ^2 of a binomial distribution $\mathcal{B}(n, p)$ are respectively given by $\mu = n \cdot p$ and $\sigma^2 = n \cdot p \cdot (1 - p)$.

[10] Variance, homogeneous differential property and DP_{max} of an S-Box $\mathcal{S}$ remain unchanged if affine transformations are applied in the domain or co-domain of $\mathcal{S}$. Thus, consider two S-Boxes $\mathcal{S}, \mathcal{S}' : \mathbb{F}_2^n \to \mathbb{F}_2^n$. Let $A, B \in \mathbb{F}_2^{n \times n}$ be two invertible $n \times n$ matrices and $a, b \in \mathbb{F}_2^n$. $\mathcal{S}$ and $\mathcal{S}'$ are *affine equivalent* iff $\mathcal{S}'(x) = B \cdot [\mathcal{S}(A \cdot x + a)] + b \; \forall x \in \mathbb{F}_2^n$.

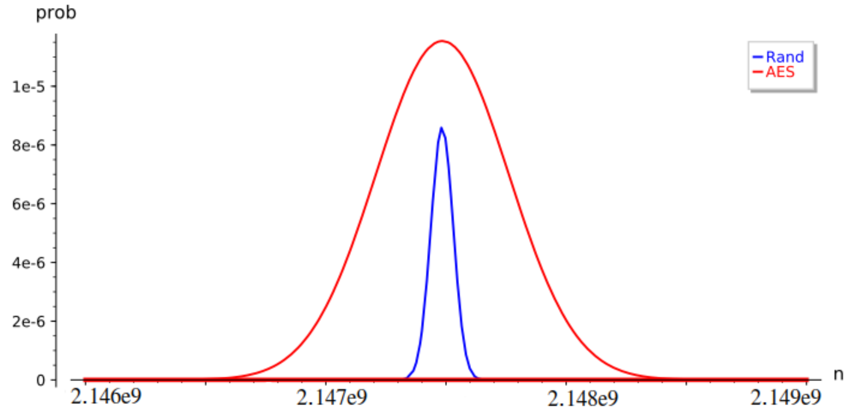

Figure 1: Comparison between the theoretical probability distribution of the number of collisions between 5-round AES (*approximated* – only here – by a normal distribution) and a random permutation.

which the S-Box has the same property as the full-size AES one) are very close to the one predicted by the previous Theorem.

We remark that even if the assumptions on the S-Box of Theorem 2 are restrictive, they match criteria used to design an S-Box which is strong against differential cryptanalysis. As a result, many ciphers in the literature are built using S-Boxes which (are close to) satisfy the assumptions of Theorem 2.

Finally, we emphasize that if the S-Box does not satisfy the required properties related to the assumption of the Theorem, then the number of collisions can be different from the one previously given. To be more concrete, in Sect. 10 we provide several practical examples of the dependency of the number of collisions for small-scale AES-like ciphers w.r.t. the properties of the S-Box, and we provide theoretical argumentation to explain the influence of the S-Box. In the case in which the assumption about the S-Box is not fulfilled, it turns out that also the details of the MixColumns matrix can influence the average number of collisions.

4.3 Comparison between the Prob. Distribution of 5-round AES and of a Random Permutation

The previous results regarding the probability distribution for 5-round AES are not only of theoretically interest. As we are going to show, they can also be exploited in order to set up new truncated differential distinguishers for 5-round AES, which are independent of the secret-key. Thus, consider 2^{32} plaintexts in the same coset of a diagonal subspace $\mathcal{D}_i$, and the corresponding (cipher)texts generated by a random permutation $\Pi(\cdot)$ (or by an ideal cipher). *What is the probability distribution of the number of different pairs of (cipher)texts generated by a random permutation (or by an ideal cipher)* $\Pi(\cdot)$ *which are equal in one fixed anti-diagonal (assuming the final MixColumns operation is omitted)?*

Proposition 1. *Consider* 2^{32} *plaintexts* p^i *for* $i = 0, 1, ..., 2^{32} - 1$ *with one active diagonal (equivalently, a coset of a diagonal space* $\mathcal{D}_i$ *for* $i \in \{0, 1, 2, 3\}$*), and the corresponding (cipher)texts generated by a random permutation* Π*, that is* $c^i = \Pi(p^i)$*. The probability distribution of the number of different pairs of ciphertexts* (c^i, c^j) *with* $c^i < c^j$ *that are equal in one anti-diagonal (equivalently, belong to the same coset of* $\mathcal{ID}_J$ *for* $J \subseteq \{0, 1, 2, 3\}$

fixed with $|J| = 3$) is well approximated by a binomial distribution $\mathcal{B}(n,p)$, where

$$n = \binom{2^{32}}{2} = 2^{31} \cdot (2^{32} - 1) \qquad \textit{and} \qquad p = 2^{-32}.$$

The average number of collisions of such distribution is equal to $2^{31} - 0.5 = 2\,147\,483\,647.5$, while its variance is equal to $2\,147\,483\,647 \simeq 2^{31}$.

The main differences between the two distributions are the following:

- independently of the secret key, the average number of pairs of ciphertexts which are equal in one fixed anti-diagonal (equivalently, the number of collisions in $\mathcal{ID}_J$ for $|J| = 3$ fixed) is a (little) bigger for 5-round AES than for a random permutation (approximately $1\,038.1$ more collisions);

- independently of the secret key, the variance of the probability distribution of the number of collisions is a (much) bigger for 5-round AES than for a random permutation (approximately of a factor 36).

To highlight this difference, Fig. 1 proposes a comparison between the probability distribution of the number of collisions for the AES case – approximated here for simplicity by a normal distribution – in red (where the probability to have $n \neq 8 \cdot n'$ collisions – i.e., where n is not a multiple of 8 – is zero) and of the random case in blue.

5 Proof of Theorem 2 – Average Number of Collisions for 5-round AES

In this section, we formally compute the probabilities p_3, p_{10}, p_{17} given in Theorem 2, and the average number of collisions for 5-round AES.

5.1 Reduction to the Middle Round

In order to prove the probability p_3, p_{10} and p_{17} given in Theorem 2, the idea is to prove an equivalent result on a single round. Using the truncated differential for 2-round AES recalled in Sect. 3, note that

$$\mathcal{D}_i \oplus \delta \xrightarrow[\text{prob. } 1]{R^2(\cdot)} \mathcal{M}_i \oplus \omega \xrightarrow{R(\cdot)} \mathcal{D}_J \oplus \delta' \xrightarrow[\text{prob. } 1]{R_f^2(\cdot)} \mathcal{ID}_J \oplus \omega' \tag{11}$$

where $\mathcal{D}_i$ is a set of texts with 1 active diagonal, and where $\mathcal{ID}_J = SR(\mathcal{C}_J)$ is a set of texts with $|J|$ active anti-diagonals.

Working on the middle round, we are going to prove an equivalent result by *(1st)* considering 2^{32} plaintexts p^j for $j = 0, 1, ..., 2^{32} - 1$ in a coset of a mixed space $\mathcal{M}_i$, that is $\mathcal{M}_i \oplus \omega$ for $i \in \{0, 1, 2, 3\}$ and $\omega \in \mathcal{M}_i^{\perp}$ and the corresponding ciphertexts after 1 round, that is $c^j = R(p^j)$, and by *(2nd)* considering the probability distribution of different pairs of ciphertexts (c^l, c^j) with $c^l < c^j$ that are equal in one fixed diagonal (that is, that belong to the same coset of $\mathcal{D}_J$ for $J \subseteq \{0, 1, 2, 3\}$ fixed with $|J| = 3$).

5.1.1 Idea of the Proof

For simplicity, we limit ourselves to consider plaintexts in the same coset of $\mathcal{M}_0$ and to count the number of texts which are equal in the first diagonal after one round (the other cases are analogous). By definition of $\mathcal{M}_0$ (see [GRR17b] and App. A for details), if $p^1, p^2 \in \mathcal{M}_0 \oplus \omega$ there exist $x^i, y^i, z^i, w^i \in \mathbb{F}_{2^8}$ for $i = 1, 2$ such that:

$$p^i = \omega \oplus \begin{bmatrix} 2 \cdot x^i & y^i & z^i & 3 \cdot w^i \\ x^i & y^i & 3 \cdot z^i & 2 \cdot w^i \\ x^i & 3 \cdot y^i & 2 \cdot z^i & w^i \\ 3 \cdot x^i & 2 \cdot y^i & z^i & w^i \end{bmatrix}$$

where $2 \equiv$ 0x02 and $3 \equiv$ 0x03. In the following, we say that p^1 is "generated" by the generating variables (x^1, y^1, z^1, w^1) and that p^2 is "generated" by the generating variables (x^2, y^2, z^2, w^2) – as before, we denote it by $p^i \equiv (x^i, y^i, z^i, w^i)$ for $i = 1, 2$.

In the following, we consider separately the following cases

- 3 variables are equal, e.g. $x^1 \neq x^2$ and $y^1 = y^2$, $z^1 = z^2$, $w^1 = w^2$;
- 2 variables are equal, e.g. $x^1 \neq x^2$,$y^1 \neq y^2$ and $z^1 = z^2$, $w^1 = w^2$;
- 1 variable is equal, e.g. $x^1 \neq x^2$, $y^1 \neq y^2$, $z^1 \neq z^2$ and $w^1 = w^2$;
- all variables are different, e.g. $x^1 \neq x^2$, $y^1 \neq y^2$, $z^1 \neq z^2$, $w^1 \neq w^2$.

As already mentioned before, if 3 variables are equal (e.g. $y^1 = y^2$, $z^1 = z^2$ and $w^1 = w^2$), then $p^1 \oplus p^2 \in (\mathcal{M}_0 \cap \mathcal{C}_j) \subseteq \mathcal{C}_j$ for a certain $j \in \{0, 1, 2, 3\}$, and $R(p^1) \oplus R(p^2) \in \mathcal{M}_j$. Due to e.g. the impossible differential trail given in Theorem 5 (see App. A), it follows that $R(p^1)$ and $R(p^2)$ can not be equal in any diagonal (equivalently, $R(p^1) \oplus R(p^2) \notin \mathcal{D}_J$ for each J with $|J| \leq 3$). Thus, in the following we limit ourselves to consider the case in which at least 2 generating variables are different.

Remark. In the following, we start by considering a subset of 2^{16} texts with only 2 active bytes. This case is much simpler to analyze than the generic one, and it allows ourselves to highlight the crucial points of the proof. The same approach (opportunely modified) is then used to study the case of 2^{32} texts in the same coset of $\mathcal{M}_0$.

5.2 A "Simpler" Case: 2^{16} Texts with Two Equal Generating Variables

As a first case, we consider 2^{16} texts in which (at least) two generating variables are equal, e.g. $z^1 = z^2$ and $w^1 = w^2$. Given two texts p^1 generated by $(x^1, y^1, 0, 0)$ and p^2 generated by $(x^2, y^2, 0, 0)$, they are equal in the first diagonal after one round if and only if the following four equations are satisfied

$$\begin{aligned}
(R(p^1) \oplus R(p^2))_{0,0} &= 2 \cdot (\text{S-Box}(2 \cdot x^1 \oplus a_{0,0}) \oplus \text{S-Box}(2 \cdot x^2 \oplus a_{0,0})) \oplus \\
&\oplus 3 \cdot (\text{S-Box}(y^1 \oplus a_{1,1}) \oplus \text{S-Box}(y^2 \oplus a_{1,1})) = 0, \\
(R(p^1) \oplus R(p^2))_{1,1} &= \text{S-Box}(3 \cdot x^1 \oplus a_{3,0}) \oplus \text{S-Box}(3 \cdot x^2 \oplus a_{3,0}) \oplus \\
&\oplus \text{S-Box}(y^1 \oplus a_{0,1}) \oplus \text{S-Box}(y^2 \oplus a_{0,1}) = 0, \\
(R(p^1) \oplus R(p^2))_{2,2} &= 2 \cdot (\text{S-Box}(x^1 \oplus a_{2,0}) \oplus \text{S-Box}(x^2 \oplus a_{2,0})) \oplus \\
&\oplus 3 \cdot (\text{S-Box}(2 \cdot y^1 \oplus a_{3,1}) \oplus \text{S-Box}(2 \cdot y^2 \oplus a_{3,1})) = 0, \\
(R(p^1) \oplus R(p^2))_{3,3} &= \text{S-Box}(x^1 \oplus a_{1,0}) \oplus \text{S-Box}(x^2 \oplus a_{1,0}) \oplus \\
&\oplus \text{S-Box}(3 \cdot y^1 \oplus a_{2,1}) \oplus \text{S-Box}(3 \cdot y^2 \oplus a_{2,1}) = 0.
\end{aligned}$$

where $a_{\cdot,\cdot}$ depends on the initial key and on the constant ω that defines the coset. Equivalently, four equations of the form

$$\begin{aligned}
&A \cdot \big[\text{S-Box}(B \cdot x^1 \oplus a) \oplus \text{S-Box}(B \cdot x^2 \oplus a)\big] \oplus \\
&\oplus C \cdot \big[\text{S-Box}(D \cdot y^1 \oplus c) \oplus \text{S-Box}(D \cdot y^2 \oplus c)\big] = 0
\end{aligned} \tag{12}$$

must be satisfied, where A, B, C, D depend only on the MixColumns matrix, while a, c depend on the secret key and on the initial constant that defines the coset.

Number of Solutions of Each Equation. Consider one of these four equations. By simple observation, equation (12) is satisfied if and only if[11] the following system of equations is satisfied

$$\begin{aligned}\text{S-Box}(\hat{x} \oplus \Delta_I) \oplus \text{S-Box}(\hat{x}) &= \Delta_O \\ \text{S-Box}(\hat{y} \oplus \Delta'_I) \oplus \text{S-Box}(\hat{y}) &= \Delta'_O \\ \Delta'_O &= C^{-1} \cdot A \cdot \Delta_O\end{aligned} \tag{13}$$

for each value of Δ_O, where $\hat{x} = B \cdot x^1 \oplus a$, $\Delta_I = B \cdot (x^1 \oplus x^2)$, $\hat{y} = D \cdot y^1 \oplus c$ and $\Delta'_I = D \cdot (y^1 \oplus y^2)$.

What is the number of different (not null) solutions $[(x^1, y^1), (x^2, y^2)]$ *of eq. (12)?* Given $\Delta_O \neq 0$, each one of the first two equations of (13) admits 256 different solutions $(\hat{x}, \Delta_I)$ (resp. $(\hat{y}, \Delta'_I)$) – note that for each value of $\hat{x} \in \mathbb{F}_{2^8}$, there exists $\Delta_I \neq 0$ that satisfies the first equation (similar for $\hat{y}$ and Δ'_I). It follows that the number of different (not null) solutions $[(x^1, y^1), (x^2, y^2)]$ of eq. (12) - considering all the 255 possible values of Δ_O - is exactly equal to

$$\frac{1}{2} \cdot 255 \cdot \left(\frac{256}{255} \cdot 255\right)^2 = 255 \cdot 2^{15}$$

independently of the details of the S-Box. The factor 1/2 is due to the fact that we consider only different solutions, that is two solutions of the form $(p^1 \equiv (x^1, y^1), p^2 \equiv (x^2, y^2))$ and $(p^2 \equiv (x^1, y^1), p^1 \equiv (x^2, y^2))$ are considered equivalent. In other words, a solution $[(x^1, y^1), (x^2, y^2)]$ is considered to be valid if $x^2 \neq x^1$ and $y^1 < y^2$.

Probability of Common Solutions. Knowing the number of solutions of one eq. (12), *what is the number of common - different (not null) - solutions* $[(x^1, y^1), (x^2, y^2)]$ *of 4 equations of the form (12)?* We have just seen that each equation of the form (12) has exactly $255 \cdot 2^{15}$ different (not null) solutions $[(x^1, y^1), (x^2, y^2)]$. The probability that two equations admit the same solution (i.e. that $[(x^1, y^1), (x^2, y^2)]$ – solution of one equation – is equal to $[(\hat{x}^1, \hat{y}^1), (\hat{x}^2, \hat{y}^2)]$ – solution of another equation) is

$$(256 \cdot 255)^{-1} \cdot (255 \cdot 128)^{-1} = 255^{-2} \cdot 2^{-15}. \tag{14}$$

To explain this probability, the first term $(256 \cdot 255)^{-1}$ is due to the fact that $x^1 = \hat{x}^1$ with probability 256^{-1} while $x^2 = \hat{x}^2$ with probability 255^{-1}, since by assumption x^2 (resp. $\hat{x}^2$) cannot be equal to x^1 (resp. $\hat{x}^1$). The second term $(128 \cdot 255)^{-1}$ is due to the assumption on the second variable, that is $y^1 < y^2$. To explain it[12], note that the possible number of pairs (y^1, y^2) with $y^1 < y^2$ is $\sum_{i=0}^{255} i = \frac{255 \cdot (255+1)}{2} = 255 \cdot 128$. It follows that y^1 and y^2 are equal to $\hat{y}^1$ and $\hat{y}^2$ with prob. $(128 \cdot 255)^{-1}$.

Total Number of Different – not null – Common Solutions. In conclusion, the average number of common - different (not null) - solutions $[(x^1, y^1), (x^2, y^2)]$ of 4 equations of the form (12) is given by

$$(255 \cdot 2^{15})^4 \cdot (255^{-2} \cdot 2^{-15})^3 = \frac{2^{15}}{255^2} \simeq 0.503929258 \simeq 2^{-1} + 2^{-7.992}$$

[11] Observe that the equality $\Delta'_O = C^{-1} \cdot A \cdot \Delta_O$ is well defined, since no coefficient of an MDS matrix $M \in \mathbb{F}_{2^b}^{4\times 4}$ is equal to zero. Indeed, by definition, *a matrix M is MDS if and only if all square sub-matrices of M are of full rank.*

[12] As examples, if $y^1 = 0\text{x}0$ then y^2 can take 255 different values (all values except 0), if $y^1 = 0\text{x}1$ then y^2 can take 254 different values (all values except 0x0, 0x1) and so on - if $y^1 = d$ with $0 \leq d \leq 255$ then y^2 can take $255 - d$ different values.

For comparison, given plaintexts in the same coset of $\mathcal{M}_0 \cap \mathcal{C}_{0,1}$ and the corresponding ciphertexts generated by a random permutation, the average number of pairs of ciphertexts that belong to the same coset of $\mathcal{D}_J$ is approximately given by

$$\binom{2^{16}}{2} \cdot (2^{-8})^4 = \frac{2^{16}-1}{2^{17}} \simeq 0.499992371 \simeq 2^{-1} - 2^{-17}$$

Remark on Prob. (14). We highlight that *probability (14) (strongly) depends on the assumptions that*

- the solutions of eq. (1) – so the numbers n_{Δ_I,Δ_O} – are uniformly distributed for each $\Delta_I \neq 0$ and $\Delta_O \neq 0$;
- there is "*no (obvious/non-trivial) relations*" between the solutions of the studied system of four equations of the form (12); in other words, the four equations (12) must be independent/unrelated, in the sense that the solution of one equation must *not* be a solution of another one with probability different (bigger/smaller) than (14).

Let's focus here on this second requirement. A relation among solutions of different equations *can* arise if some relations hold between the coefficients A, B, C, D of different equations of the form (12). Since these are exactly the coefficients of the MixColumns matrix and since such matrix is MDS, no linear relation among the rows/columns of any submatrix exists. More details about this are given in the following – see also App. C.

5.3 Generic Case: 2^{32} Texts

Using the same strategy just presented, we now consider the case of 2^{32} texts in the same coset of $\mathcal{M}_0$. As before, two texts p^1, p^2 are equal in one diagonal after one round if and only if the following four equations

$$\begin{aligned} A \cdot [\text{S-Box}(B \cdot x \oplus b) \oplus \text{S-Box}(B \cdot x' \oplus b)] \oplus \\ \oplus C \cdot [\text{S-Box}(D \cdot y \oplus d) \oplus \text{S-Box}(D \cdot y' \oplus d)] \oplus \\ \oplus E \cdot [\text{S-Box}(F \cdot z \oplus f) \oplus \text{S-Box}(F \cdot z' \oplus f)] \oplus \\ \oplus G \cdot [\text{S-Box}(H \cdot w \oplus h) \oplus \text{S-Box}(H \cdot w' \oplus h)] = 0 \end{aligned} \tag{15}$$

are satisfied, where A, B, C, D, E, F, G, H depend only on the MixColumns matrix, while b, d, f, h depend on the secret key and on the constant that defined the initial coset. As before, each one of these equations is equivalent to a system of equations like (13), that is:

$$\begin{aligned} &\text{S-Box}(x \oplus \Delta_I) \oplus \text{S-Box}(x) = \Delta_O && \text{S-Box}(y \oplus \Delta'_I) \oplus \text{S-Box}(y) = \Delta'_O \\ &\text{S-Box}(z \oplus \Delta''_I) \oplus \text{S-Box}(z) = \Delta''_O && \text{S-Box}(w \oplus \Delta'''_I) \oplus \text{S-Box}(w) = \Delta'''_O \end{aligned}$$

together with one of the following conditions

1. $\Delta'''_O = \Delta''_O = 0$ and $\Delta'_O = C^{-1} \cdot A \cdot \Delta_O \neq 0$ or analogous (6 possibilities);
2. $\Delta'''_O = 0$ and $\Delta_O, \Delta'_O, \Delta''_O \neq 0$ and $\Delta''_O = E^{-1} \cdot (A \cdot \Delta_O \oplus C \cdot \Delta'_O)$ or analogous, for a total of 4 possibilities;
3. $\Delta_O, \Delta'_O, \Delta''_O, \Delta'''_O \neq 0$ and $\Delta'''_O = G^{-1} \cdot (A \cdot \Delta_O \oplus C \cdot \Delta'_O \oplus E \cdot \Delta''_O)$.

5.3.1 First Case

Note that the first case ($\Delta_O^{'''} = \Delta_O^{''} = 0$) is analogous to the case in which two generating variables are equal. In order to compute the number of collisions for this case is sufficient to re-use the previous computation.

In the case $\Delta_O^{'''} = \Delta_O^{''} = 0$ and $\Delta_O^{'} = C^{-1} \cdot A \cdot \Delta_O \neq 0$, the only possible solutions of the third and fourth equations are of the form $(z, \Delta_I^{''} = 0)$ and $(w, \Delta_I^{'''} = 0)$ for each possible value of z and w. Using the same computation as before, the average number of (not null) common solutions for this case is

$$\binom{4}{2} \cdot 256^2 \cdot \frac{2^{15}}{255^2} = \frac{2^{32}}{21\,675} \simeq 198\,153.047. \tag{16}$$

About Probability p_{17}. This number can be used in order to compute the probability p_{17} – given in Theorem 2 – that pairs of texts with two equal (and two different) generating variables are equal in one diagonal after one round. Indeed, by definition of probability:

$$p_{17} = \frac{1}{2^{17} \times n_{17}} \cdot \frac{2^{32}}{21\,675} = 2^{-32} + 2^{-37.98588}, \tag{17}$$

where $2^{17} \times n_{17}$ is the *total* number of pairs of texts with two equal (and two different) generating variables.

5.3.2 Second Case

Consider now the case $\Delta_O^{'''} = 0$ and $\Delta_O, \Delta_O^{'}, \Delta_O^{''} \neq 0$ (i.e. $\Delta_I, \Delta_I^{'}, \Delta_I^{''} \neq 0$). First of all, note that $\Delta_O \neq 0$ can take 255 different values, while $\Delta_O^{'} \neq 0$ can take only 254 different values. Indeed, it must be different from 0 and from $C^{-1} \cdot A \cdot \Delta_O$ (if $\Delta_O^{'} = C^{-1} \cdot A \cdot \Delta_O$, then $\Delta_O^{''} = 0$ which is excluded by assumption).

Using the same argumentation given before, for each equation (15) the number of different solutions $[(x^1, y^1, z^1, w^1), (x^2, y^2, z^2, w^2)]$ – where $z^1 < z^2$ and where $w^1 = w^2$ – is given by

$$\binom{4}{1} \cdot 256 \cdot \left[\frac{1}{2} \cdot 255 \cdot 254 \cdot \left(255 \cdot \frac{256}{255} \right)^3 \right] = 32\,385 \cdot 2^{34},$$

where the initial factor $\binom{4}{1} \cdot 256$ is due to the condition $w^1 = w^2$ and on the fact that there are 4 analogous cases (namely, $x^1 = x^2$ or $y^1 = y^2$ or $z^1 = z^2$). Similar to before, the probability that two equations of the form (15) – where $w^1 = w^2$ – have a common solution is given by

$$(256 \cdot 255)^{-2} \cdot (128 \cdot 255)^{-1} = 2^{-23} \cdot 255^{-3}$$

under the assumption (1st) of uniform distribution of the solutions n_{Δ_I, Δ_O} of eq. (1) and *(2nd)* that there is "no (obvious/non-trivial) relation" between the solutions of the studied system of four equations of the form (15). It follows that for this second case we expect on average

$$(32\,385 \cdot 2^{34})^4 \cdot (2^{-23} \cdot 255^{-3})^3 = \frac{127^4 \cdot 2^{37}}{255^5} \simeq 33\,160\,710.047 \tag{18}$$

different - not null - common solutions for the 4 equations of the form (15).

About Probability p_{10}. This number can be used in order to compute the probability p_{10} – given in Theorem 2 – that pairs of texts with one equal (and three different) generating variable(s) are equal in one diagonal after one round. Indeed, by definition of probability:

$$p_{10} = \frac{1}{2^{10} \times n_{10}} \cdot \frac{127^4 \cdot 2^{37}}{255^5} = 2^{-32} - 2^{-45.98874}, \tag{19}$$

where $2^{10} \times n_{10}$ is the *total* number of pairs of texts with one equal (and three different) generating variables.

5.3.3 Third Case

We finally consider the case $\Delta_O, \Delta'_O, \Delta''_O, \Delta'''_O \neq 0$. By simple computation, the number of different values that satisfy

$$\Delta'''_O = G^{-1} \cdot (A \cdot \Delta_O \oplus C \cdot \Delta'_O \oplus E \cdot \Delta''_O).$$

is given by $255^3 - (255 \cdot 254) = 16\,516\,605$. Indeed, the total number of $\Delta_O, \Delta'_O, \Delta''_O \neq 0$ is 255^3, while $255 \cdot 254$ is the total number of values $\Delta_O, \Delta'_O, \Delta''_O \neq 0$ for which Δ'''_O is equal to zero (which is not possible since $\Delta'''_O \neq 0$ by assumption). In more detail, *firstly* observe that for each value of Δ_O there is a value of Δ'_O that satisfies $A \cdot \Delta_O = C \cdot \Delta'_O$. For this pair of values $(\Delta_O, \Delta'_O = C^{-1} \cdot A \cdot \Delta_O)$, the previous equation – which reduces to $\Delta'''_O = G^{-1} \cdot E \cdot \Delta''_O$ is always different from zero, since $\Delta''_O \neq 0$. *Secondly*, for each one of the $255 \cdot 254$ values of the pair $(\Delta_O, \Delta'_O \neq C^{-1} \cdot A \cdot \Delta_O)$, there is only one value of Δ''_O such that the previous equation is equal to zero.

As a result, the total number of different solutions $[(x^1, y^1, z^1, w^1), (x^2, y^2, z^2, w^2)]$ with $w^1 < w^2$ of each equation corresponding to (15) is

$$\frac{1}{2} \cdot 16\,516\,605 \cdot \left(255 \cdot \frac{256}{255}\right)^4 = 16\,516\,605 \cdot 2^{31}.$$

Since the probability that two solutions $[(x^1, y^1, z^1, w^1), (x^2, y^2, z^2, w^2)]$ and $[(\hat{x}^1, \hat{y}^1, \hat{z}^1, \hat{w}^1), (\hat{x}^2, \hat{y}^2, \hat{z}^2, \hat{w}^2)]$ are equal is $(255 \cdot 256)^{-3} \cdot (255 \cdot 128)^{-1} = 255^{-4} \cdot 2^{-31}$ - *under the assumptions (1st)* of uniform distribution of the solutions of eq. (1) and *(2nd)* that there is “no (obvious/non-trivial) relation” between the solutions of the studied system of four equations of the form (15), the average number of (non null) common solutions (with no equal generating variables) is

$$\left(16\,516\,605 \cdot 2^{31}\right)^4 \cdot (255^{-4} \cdot 2^{-31})^3 = \frac{64\,771^4 \cdot 2^{31}}{255^8} \simeq 2\,114\,125\,822.5 \qquad (20)$$

About Probability p_3. This number can be used in order to compute the probability p_3 – given in Theorem 2 – that pair of texts with no equal generating variable are equal in one diagonal after one round. Indeed, by definition of probability:

$$p_3 = \frac{1}{2^3 \times n_3} \cdot \frac{64\,771^4 \cdot 2^{31}}{255^8} = 2^{-32} + 2^{-53.98306}, \qquad (21)$$

where $2^{17} \times n_{17}$ is the *total* number of pairs of texts with no equal generating variable.

5.3.4 Total Number of Different - not null - Common Solutions

By simple computation, given plaintexts in the same coset of $\mathcal{M}_0$, the number of different pairs of ciphertexts that are equal in one fixed diagonal after 1-round (equivalently, the number of collisions in $\mathcal{D}_J$ for $|J| = 3$) is

$$2\,114\,125\,822.5 + 33\,160\,710.047 + 198\,153.047 \simeq 2\,147\,484\,685.594 \simeq 2^{31} + 2^{10.02}$$

For comparison, if the ciphertexts are randomly generated, the number of different pairs of ciphertexts that are equal in one fixed diagonal (equivalently, the number of collisions in $\mathcal{D}_J$ for $|J| = 3$) is

$$\binom{2^{32}}{2} \cdot 2^{-32} \simeq 2\,147\,483\,647.5 = 2^{31} - 2^{-1}$$

In other words, on average there are

$$2\,147\,484\,685.594 - 2\,147\,483\,647.5 \simeq 1\,038.094$$

more collisions for 5-round AES than for a random permutation.

Finally, since the number of possible pairs of texts is $2^{31} \cdot (2^{32} - 1)$, the probability for the AES case that a couple of ciphertexts (c^1, c^2) satisfies $c^1 \oplus c^2 \in \mathcal{D}_J$ for $|J| = 3$ fixed is equal to

$$p_{AES} \simeq \frac{2\,147\,484\,685.594}{2^{31} \cdot (2^{32} - 1)} \simeq 2^{-32} + 2^{-52.9803}$$

versus 2^{-32} for the random case.

5.4 Remarks

On the MDS Requirement for the MixColumns Matrix. As already mentioned, the assumption that the MixColumns matrix is MDS is *crucial* when computing the number of solutions of a system of 4 equations of the form (12) or (15) – remember that the coefficients $A, B, C, ...$ are the coefficients of the MixColumns matrix.

To give more evidence, assume by contradiction that the matrix is not MDS, and focus on a system of equations e.g. of the form (12). First of all, if some coefficients of the MixColumns matrix are equal to zero, then some of such equations become trivial. E.g. if $C = 0$ then an equation of the form (12) is satisfied by $x^1 = x^2$ and by $y^1 \neq y^2$. In such a case, a problem arises since the arguments given for the case (12) hold only under the assumption $x^1 \neq x^2$ and $y^1 \neq y^2$. If the case $x^1 = x^2$ is admitted, the previous proof must be modified accordingly: e.g. the number of solutions of an equation of the form (12) for $C = 0$ becomes $255 \cdot 2^{16}$, which differs from $255^2 \cdot 2^8$ given in the text, and the previous result is in general not true anymore.

What happens if the MixColumns matrix is not MDS and if all its coefficients are not null? Consider a system of four equations of the form (12) or (15). Since the matrix is not MDS, then there exists a $r \times r$ submatrix (for $2 \leq r \leq 3$) whose determinant is equal to zero (this means that there exists a linear relation between the rows/columns of this matrix). This fact can have effects on the probability that the studied system of four equations of the form (15) has a common solution(s). In particular, *it can happen that the solutions of different equations of this system are not unrelated/independent, which is a crucial assumption exploited in the previous proof in order to compute the probability that different equations admit the same solution(s).* To better understand this fact, we show a concrete example in App. C.

Random Permutation and Probability 2^{-32}. Given 2^{32} texts generated by a random permutation, one can construct 2^{63} different pairs which are *not independent.* For example, consider a pair of texts (t^1, t^2). Given another text t^3, if $t^1 \oplus t^3 \in \mathcal{ID}_J$ – equivalently, if t^1 and t^2 are equal in $4 - |J|$ anti-diagonal(s) – and $t^2 \oplus t^3 \in \mathcal{ID}_J$ for a particular subspace $\mathcal{ID}_J$, then (t^1, t^2) belong to the same coset of $\mathcal{ID}_J$ with prob. 1 (by definition of subspace). Thus, one may think that the probability that (t^1, t^2) are in the same coset of $\mathcal{ID}_J$ is different than $2^{-32\cdot(4-|J|)}$. In App. B, we prove that *even if the pairs are not independent, the probability that each pair (t^1, t^2) satisfies the property to belong to the same coset of $\mathcal{ID}_J$ is exactly* $2^{-32\cdot(4-|J|)}$ (that is, 2^{-32} if $|J| = 3$ for J fixed).

6 Proof of Theorem 2 – Variance – and of Lemma 1

Using the result just given, we finally compute the variance of the probability distribution for 5-round AES given in Theorem 2, and the probability given in Lemma 1. To do this, we exploit the fact that the probability distribution of 5-round AES is well described by

$$\text{5-AES} = 2^3 \cdot X_3 + 2^{10} \cdot X_{10} + 2^{17} \cdot X_{17}$$

where X_i are binomial distributions. We recall that the pairs of texts with no equal generating variable are represented by $2^3 \cdot X_3$, the pairs of texts with 1 equal generating variable (and 3 different) are represented by $2^{10} \cdot X_{10}$ and finally the pairs of texts with 2 equal generating variables (and 2 different) are represented by $2^{17} \cdot X_{17}$.

6.1 Variance of the Prob. Distribution for 5-round AES

Properties of the Variance. First of all, remember that the cases X_3, X_{10} and X_{17} are independent. In other words, the behavior of a pair of texts with v equal generating variables is independent of another pair with $\hat{v}$ equal generating variables where $\hat{v} \neq v$. One property of the variance is that, given x independent variables $X_1, ..., X_x$, the variance of $Y = X_1 + ... + X_x$ is given by $Var(Y) = Var(X_1) + ... + Var(X_x)$. It follows that the total variance of the probability distribution for 5-round AES case is given by

$$\begin{aligned} Var(\text{5-AES}) &= Var(2^3 \cdot X_3) + Var(2^{10} \cdot X_{10}) + Var(2^{17} \cdot X_{17}) = \\ &= 2^6 \cdot Var(X_3) + 2^{20} \cdot Var(X_{10}) + 2^{34} \cdot Var(X_{17}), \end{aligned}$$

where $Var(\alpha \cdot X) = \alpha^2 \cdot Var(X)$ for each constant α.

About $Var(X_3), Var(X_{10})$ and $Var(X_{17})$. As seen before, X_3, X_{10} and X_{17} are well described by binomial distribution $\mathcal{B}(n_{7\cdot v+3}, p_{7\cdot v+3})$ for $v = 0, 1, 2$, where the values of $n_{7\cdot v+3}$ and $p_{7\cdot v+3}$ are given before. Since the variance of a binomial distribution $X_{7\cdot v+3} \sim \mathcal{B}(n_{7\cdot v+3}, p_{7\cdot v+3})$ is given by $Var(X_{7\cdot v+3}) = n_{7\cdot v+3} \times p_{7\cdot v+3} \times (1 - p_{7\cdot v+3})$, it follows that

$$\begin{aligned} Var(X_3) &= n_3 \cdot p_3 \cdot (1 - p_3) \approx 264\,265\,727.751 \\ Var(X_{10}) &= n_{10} \cdot p_{10} \cdot (1 - p_{10}) \approx 32\,383.506 \\ Var(X_{17}) &= n_{17} \cdot p_{17} \cdot (1 - p_{17}) \approx 1.51179 \end{aligned}$$

We remark one more time that:

- $2^{7\cdot v+3} \times X_{7\cdot v+3}$ represents the probability distribution of the number of pairs of ciphertexts that are equal in one anti-diagonal *given all the pairs of plaintexts* with "v equal (and $4 - v$ different) generating variable(s)";
- $X_{7\cdot v+3}$ represents the probability distribution of the number of pairs of ciphertexts that are equal in one anti-diagonal *given only the pairs of plaintexts* with "v equal (and $4 - v$ different) generating variable(s)" *that are independent.*

Final Result. By combining all previous results, it follows that

$$Var(\text{5-AES}) = 2^6 \times \underbrace{264\,265\,727.751}_{\simeq Var(X_3)} + 2^{20} \times \underbrace{32\,383.506}_{\simeq Var(X_{10})} + 2^{34} \times \underbrace{1.51179}_{\simeq Var(X_{17})} \simeq 2^{36.16118}.$$

6.2 Proof of Lemma 1

Consider 2^{32} plaintexts p^i for $i = 0, 1, ..., 2^{32} - 1$ with one active diagonal (equivalently, in a coset of a diagonal space $\mathcal{D}_i$), and the corresponding ciphertexts after 5 rounds without the final MixColumns operation, that is $c^i = R^5(p^i)$. As last thing, we formally compute the probability to have $n \in \mathbb{N}$ different pairs of ciphertexts (c^i, c^j) with $c^i < c^j$ that are equal in one fixed anti-diagonal (i.e. that belong to the same coset of $\mathcal{ID}_J$ for $J \subseteq \{0,1,2,3\}$ fixed with $|J| = 3$), namely n collisions.

Given $n \in \mathbb{N}$, note that $Prob(\text{5-AES} = n) = 0$ if $n \neq 8 \cdot n'$ is not a multiple of 8 (due to the multiple-of-8 property). Thus, assume $n = 8 \cdot n'$ for $n' \in \mathbb{N}$:

$$Prob\,(\text{5-AES} = n) := Prob\left([2^3 \cdot X_3 + 2^{10} \cdot X_{10} + 2^{17} \cdot X_{17}] = n\right) =$$
$$= \sum_{n_3, n_{10}, n_{17}} Prob([2^3 \cdot X_3] = n_3) \times Prob([2^{10} \cdot X_{10}] = n_{10}) \times Prob([2^{17} \cdot X_{17}] = n_{17}) \times$$
$$\times\, Prob\left([2^3 \cdot X_3 + 2^{10} \cdot X_{10} + 2^{17} \cdot X_{17}] = n \,\middle|\, 2^3 \cdot X_3 = n_3, 2^{10} \cdot X_{10} = n_{10}, 2^{17} \cdot X_{17} = n_{17}\right)$$

where remember that the distributions X_3, X_{10} and X_{17} are independent.

Since $Prob([2^i \cdot X_i] = n_i) = 0$ if there is no $k_i \in \mathbb{N}$ s.t. $n_i \neq 2^i \cdot k_i$ for $i \in \{3, 10, 17\}$, and since

$$Prob\begin{pmatrix} [2^3 \cdot X_3 + 2^{10} \cdot X_{10} + 2^{17} \cdot X_{17}] = n \ \textit{assuming that} \\ 2^3 \cdot X_3 = n_3,\ 2^{10} \cdot X_{10} = n_{10},\ 2^{17} \cdot X_{17} = n_{17} \end{pmatrix} = \begin{cases} 1 & \text{if } n_3 + n_{10} + n_{17} = n \\ 0 & \text{otherwise} \end{cases}$$

it follows that $Prob\,(\text{5-AES} = n)$ is equal to

$$\sum_{k_3, k_{10}, k_{17} \in K_n} Prob(X_3 = k_3) \times Prob(X_{10} = k_{10}) \times Prob(X_{17} = k_{17})$$

where

$$K_n = \left\{(k_3, k_{10}, k_{17}) \in \mathbb{N} \times \mathbb{N} \times \mathbb{N} \,\middle|\, 0 \leq k_i \leq n_i \text{ and } 2^3 \cdot k_3 + 2^{10} \cdot k_{10} + 2^{17} \cdot k_{17} = n\right\}.$$

The probability given in Lemma 1 is finally obtained by the fact that X_i are binomial distributions:

$$Prob(X_i = x) = \binom{n_i}{x} \cdot (p_i)^x \cdot (1 - p_i)^{n_i - x},$$

where n_i and p_i for $i = 3, 10, 17$ are given in Theorem 2.

7 Relation among Multiple-of-8, Mean and Variance

Here we briefly discuss the relations among the multiple-of-8 property, the fact that the average number of collisions – the mean in the following – is bigger for AES than for a random permutation and the fact that the variance of the number of collisions is (much) higher for AES than for a random permutation. As briefly mentioned at the beginning, *there is no "obvious relation" between the multiple-of-8 property and the result on the mean*, while the multiple-of-8 property and the result on the variance are strictly related.

Relation between Multiple-of-8 Property and the Mean

As we just said, the multiple-of-8 property and the result on the mean are unrelated:

- the fact that the number of collisions is always a multiple of 8 for AES does *not* imply that such number is on average bigger/equal/smaller for AES;

- the fact that the number of collisions is on average higher for AES does *not* imply that it is a multiple-of-8.

To give concrete examples, we practically computed the average number of collisions for 4-bit AES *when the AES S-Box is replaced by the S-Box of other ciphers* – see Table 3 in Sect. 10. *In all cases, the number of collisions always satisfies the multiple of 8 property. Instead, depending on the S-Box details, it's possible that the number of collisions is bigger or smaller (or potentially equal) for AES than for a random permutation* – more details in the following. This supports the arguments that these two properties are independent.

Moreover, we emphasize that:

Multiple-of-8: as proved in [GRR17a] and recalled in Sect. 3, the multiple-of-8 property holds since given two texts in p^1, p^2 in a coset of a column space (namely, $\mathcal{C}_i \oplus \gamma$ for $i \in \{0, 1, 2, 3\}$), *there exist other pairs of texts $s^1, s^2 \in \mathcal{C}_i \oplus \gamma$ defined by different combinations of the generating variables of p^1 and p^2* satisfy the following equivalence

$$R^2(p^1) \oplus R^2(p^2) = R^2(s^1) \oplus R^2(s^2).$$

This result is independent of the details of the S-Box and of the details of the MixColumns matrix (as shown in [BCC19]), *and it is completely deterministic* (everything is deterministic, i.e. probability plays no role).

Mean: in this case, given 2^{32} texts in the same coset of a mixed subspace $\mathcal{M}_i$ for $i \in \{0, 1, 2, 3\}$, one considers the average number of pairs of texts that are equal in one diagonal (equivalently, that belong to the same coset of a diagonal subspace $\mathcal{D}_J$ for $J \subseteq \{0, 1, 2, 3\}$ with $|J| = 3$) after 1-round AES encryption. By contrast to the multiple-of-8 property, *the result regarding the average number of collisions is due to (precise) probabilistic considerations, under particular assumptions on the S-Box.* To the best of our knowledge, *this is the first time that a similar approach is used in the literature.*

Relation between Multiple-of-8 Property and the Variance

Conversely, as we have just seen, the multiple-of-8 property and the variance are strictly related. Roughly speaking, due to the multiple-of-8 property (i.e. due the fact that the pairs of texts are not independent), the probability distribution of the number of collisions Y can be rewritten as $Y = \alpha \times X$ for a constant $\alpha > 1$, where X is the probability distribution of the number of collisions for the *independent/unrelated* pairs of texts (see Theorem 2 for details). Since

$$Var(Y) = Var(\alpha \times X) = \alpha^2 \times Var(X),$$

it turns out that the variance for 5-round AES is higher than the corresponding variance of a random permutation (note instead that the mean value does not have this property, since $\mathbb{E}[Y] = \alpha \times \mathbb{E}[X]$).

8 Practical Results on AES

We have practically verified the mean and the variance for 5-round AES given above – see Theorem 2 – using a C/C++ implementation[13]. In particular, we have verified the mean value on a small-scale AES as proposed in [CMR05], and the variance value both on full-size and on the small-scale AES. We limit ourselves to recall that the AES small-scale S-Box is defined in the same way as the full-size one and that it has the same properties as the full-size one, with the only exception that each word is composed of 8 bits for full-size AES and of 4 bits for the small-scale one. We emphasize that our verification on the

[13] The source codes of the distinguishers are available at https://github.com/Krypto-iaik/TruncatedDiff5roundAES

small-scale variant of AES is strong evidence for it to hold for the full-size AES, since the strategy used to theoretically compute such probabilities is independent of the fact that each word of AES is of 4 or 8 bits.

8.1 5-round AES defined over $(\mathbb{F}_{2^n})^{4\times 4}$

First of all, we propose a generic result about the average number of collisions for 5-round AES defined over $\mathbb{F}_{2^n}^{4\times 4}$.

Theorem 3. *Consider an AES-like cipher that works with texts in $\mathbb{F}_{2^n}^{4\times 4}$, such that the MixColumns matrix is an MDS matrix and s.t. the solutions of eq. (1) are uniformly distributed for each input/output difference $\Delta_I \neq 0$ and $\Delta_O \neq 0$.*

Consider 2^{4n} plaintexts p^i for $i = 0, 1, ..., 2^{4n} - 1$ with one active diagonal (equivalently, in a coset of a diagonal space $\mathcal{D}_i$ for $i \in \{0, 1, 2, 3\}$), and the corresponding ciphertexts after 5 rounds without the final MixColumns operation, that is $c^i = R^5(p^i)$. The probability distribution 5-AES of the number of different pairs of ciphertexts (c^i, c^j) with $c^i < c^j$ that are equal in one anti-diagonal (equivalently, that belong to the same coset of $\mathcal{ID}_J$ for $J \subseteq \{0, 1, 2, 3\}$ fixed with $|J| = 3$) is described by

$$5\text{-}AES = 2^3 \times X_3 + 2^{n+2} \times X_{n+2} + 2^{2n+1} \times X_{2n+1}$$

where

$$\forall i = 3, n+2, 3n+1 : \qquad X_i \sim \mathcal{B}(n_i, p_i)$$

are binomial distributions s.t.

$$n_3 = 2^{4n-4} \cdot (2^n - 1)^4 \qquad p_3 = \frac{(2^{2n} - 3 \cdot 2^n + 3)^4}{(2^n - 1)^{12}};$$

$$n_{n+2} = 2^{3n-1} \cdot (2^n - 1)^3 \qquad p_{n+2} = \frac{(2^{n-1} - 1)^4 \cdot 2^4}{(2^n - 1)^8};$$

$$n_{2n+1} = 3 \cdot 2^{2n-1} \cdot (2^n - 1)^2 \qquad p_{2n+1} = \frac{1}{(2^n - 1)^4}.$$

The average number of different pairs of ciphertexts (c^i, c^j) with $c^i \leq c^j$ for $i \neq j$ that belong to the same coset of $\mathcal{ID}_J$ for $J \subseteq \{0, 1, 2, 3\}$ fixed with $|J| = 3$ is equal to

$$\frac{2^{4n-1} \cdot (2^{2n} - 3 \cdot 2^n + 3)^4}{(2^n - 1)^8} + \frac{(2^{n-1} - 1)^4 \cdot 2^{4n+5}}{(2^n - 1)^5} + 3 \cdot \frac{2^{4n}}{(2^n - 1)^2}, \tag{22}$$

while the variance of such distribution is given by

$$\frac{2^{4n+2} \cdot (2^{2n} - 3 \cdot 2^n + 3)^4}{(2^n - 1)^8} + \frac{(2^{n-1} - 1)^4 \cdot 2^{5n+7}}{(2^n - 1)^5} + \frac{3 \cdot 2^{6n+1}}{(2^n - 1)^2} \tag{23}$$

The proof of this result is similar to the one just given for the case $\mathbb{F}_{2^8}^{4\times 4}$, and it can be found in App. E.

Before going on, we emphasize that the ratio between the variance of 5-round AES and the one of a random permutation is (almost) constant and equal to 36 *for each size n s.t.* $n \geq 8$, independently of the secret-key, of the details of the S-Box and of the MixColumns matrix. A formal proof of this fact can be found in App. E.3 (see also Fig. 5).

8.2 Practical Results on 4-bit AES

Theoretical Results. To compare the practical values with the theoretical ones, we first re-propose the main results of Theorem 2 for the case of small-scale AES.

Lemma 2. *Consider an AES-like cipher that works with texts in $\mathbb{F}_{2^4}^{4\times 4}$ and for which the assumptions of Theorem 2 hold.*

Consider 2^{16} plaintexts p^i for $i = 0, 1, ..., 2^{16} - 1$ with one active diagonal (equivalently, in a coset of a diagonal space $\mathcal{D}_i$ for $i \in \{0, 1, 2, 3\}$), and the corresponding ciphertexts after 5 rounds without the final MixColumns operation, that is $c^i = R^5(p^i)$. The probability distribution of the number of different pairs of ciphertexts (c^i, c^j) with $c^i \leq c^j$ for $i \neq j$ that are equal in one anti-diagonal (equivalently, that belong to the same coset of $\mathcal{ID}_J$ for $J \subseteq \{0, 1, 2, 3\}$ fixed with $|J| = 3$) has mean value $\mu = 32\,847.124$ and variance $\sigma^2 = 982\,466.615$ (equivalently, standard deviation $\sigma = 991.195$).

For comparison, if the ciphertexts are generated by a random permutation, the probability distribution of the number of collisions is well approximated by a normal distribution with mean value $\mu = 32\,767.5$ and variance $\sigma^2 = 32\,767$ (equivalently, standard deviation $\sigma = 181.017$).

Practical Results. In order to test our results, we took the variance over 320 initial cosets for full-size AES, while we took the average number of collisions and the variance over respectively $125\,000 \simeq 2^{17}$ and over 100 initial cosets for the small-scale one. The variance results for full-size AES[14] are given in the following

$$\sigma_T^2 = 76\,842\,293\,834.905 \simeq 2^{36.161} \qquad \sigma_P^2 = 73\,288\,132\,411.36 \simeq 2^{36.093}$$

where the subscript $\cdot_T$ denotes the theoretical value and the subscript $\cdot_P$ the practical one.

Our practical results for small-scale AES regarding the mean - denoted by μ - are

$$\mu_{AES}^T = 32\,847.124 \qquad \mu_{rand}^T = 32\,767.5$$
$$\mu_{AES}^P = 32\,848.57 \qquad \mu_{rand}^P = 32\,768.2$$

while our practical results for small-scale AES regarding the variance - denoted by σ^2 - are

$$\sigma_{AES}^T = 1036.58 \qquad \sigma_{rand}^T = 181.02$$
$$\sigma_{AES}^P = 1027.93 \qquad \sigma_{rand}^P = 182.42$$

where – as before – the superscript $\cdot^T$ and $\cdot^P$ denote resp. the theoretical and the practical values.

Fig. 2 highlights the difference between the *practical* probability distribution of the number of collisions for small-scale AES and for a random permutation.

Remark – Mean and Mode. *About Fig. 2, it is important not to confuse the mean and the mode.* In particular, consider a random variable X with a finite number of outcomes $x_0, x_1, ..., x_n$ occurring with probabilities $p_0, p_1, ..., p_n$ respectively (where $\sum_i p_i = 1$):

mean: the mean – or *expected value* – of such a random variable X is defined as $\mu = \mathbb{E}[X] = \sum_{i=0}^n p_i \times x_i$;

mode: the mode of a set of data values is the outcome $x_i \in X$ (if exists) for $0 \leq i \leq n$ that appears most often, that is for which the corresponding probability p_i satisfies $p_i > p_j$ for all $j \in \{0, ..., n\} \setminus \{i\}$.

In our case, consider the probability distribution for 5-round AES: the mean of such distribution is approximately equal to $\mu_{AES}^P = 32\,848.57$, while the mode is approximately equal to 32 560. For the random case, the mean and the mode are approximately equal (the distribution is approximately symmetric).

[14] We remark that one would need more than one year of computation on our cluster to test the distinguisher based on the mean with its $\approx 2^{16}$ initial cosets.

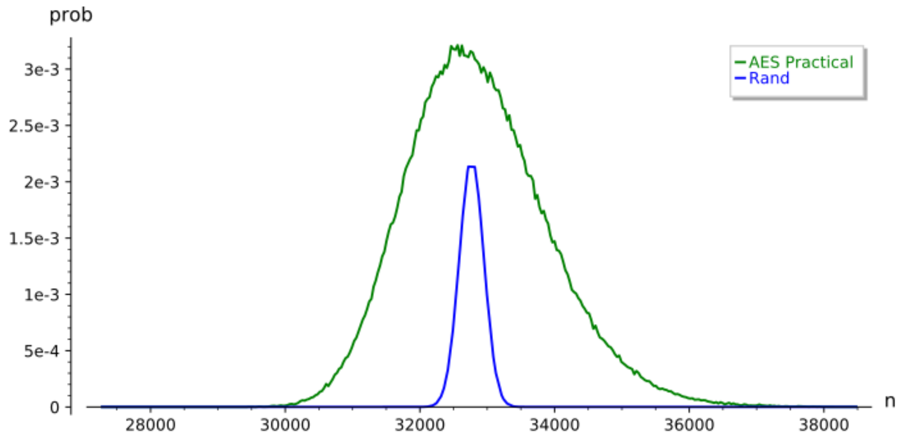

Figure 2: Comparison between the *practical* probabilistic distributions of the number of collisions of small-scale 5-round AES and of a random permutation.

A Distinguisher based on the Skewness? It is important to have in mind that *for skewed* (i.e. asymmetric) *distributions, the mean is not necessarily the same as the most likely value, i.e. the mode.* In particular, the mean and the mode coincide only in the case in which the skewness is equal to zero, which is the case of e.g. a normal distribution (which is always symmetric).

The skewness is a parameter that measures the asymmetry of the probability distribution of a real-valued random variable about its mean. The skewness value can be positive or negative, or undefined. In particular, referring to Figure[15] 3, the skew is negative if the left tail is longer (i.e. the mass of the distribution is concentrated on the right of the figure), while it is positive if the right tail is longer (i.e. the mass of the distribution is concentrated on the left of the figure).

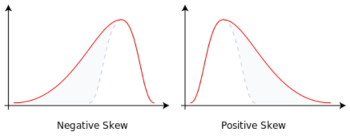

Figure 3: Examples of negative and positive skew.

The skewness of a random variable X is the third standardized moment γ, defined as:

$$\gamma = \mathbb{E}\left[\left(\frac{X-\mu}{\sigma}\right)^3\right]$$

where $\mathbb{E}[\cdot]$ is the mean value operator, $\mu \equiv \mathbb{E}[X]$ the mean value and $\sigma^2 \equiv Var(X)$ the variance[16]. For the particular case of a binomial distribution $\mathcal{B}(n, p)$, the skewness is given by

$$\gamma = \frac{1 - 2 \cdot p}{\sqrt{n \cdot p \cdot (1-p)}}, \tag{24}$$

which is close to zero if $p \approx 1/2$ or if $n \cdot p \gg 1$.

[15] Figure re-printed from Wikipedia https://en.wikipedia.org/wiki/Skewness

[16] For a sample of n values, an estimator z for the skewness is given by $z = \left\{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{X})^3\right\} / \left\{\frac{1}{n}\sum_{i=1}^{n}[(x_i - \bar{X})^2])^{3/2}\right\}$ where $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$.

Table 2: *Secret-Key Distinguishers for 5-round AES.* The complexity is measured in minimum number of chosen plaintexts/ciphertexts (CP/CC) or/and adaptive chosen plaintexts/ciphertexts (ACP/ACC) which are needed to distinguish 5-round AES from a random permutation with prob. bigger than 95%. Time complexity is measured in equivalent encryptions (E) or memory accesses (M) or XOR operations (XOR) - using the common approximation 20 M $\approx$ 1-round E. Our distinguishers are in bold.

Property	Data (CP/CC)	Cost	Ref.
Yoyo	2^{12} CP + $2^{25.8}$ ACC	$2^{24.8}$ XOR	[RBH17]
Multiple-of-8	2^{32} CP	$2^{35.6}$ M $\approx 2^{29}$ E	[GRR17a]
Variance Diff.	$\mathbf{2^{34}}$ **CP**	$\mathbf{2^{37.6}}$ **M** $\approx \mathbf{2^{31}}$ **E**	**Sect. 9.1**
Truncated Diff.	$\mathbf{2^{48.96}}$ **CP**	$\mathbf{2^{52.6}}$ **M** $\approx \mathbf{2^{46}}$ **E**	**Sect. 9.3**
Prob. Mixture Diff.	2^{52}	$2^{71.5}$ M $\approx 2^{64.9}$ E	[Gra17]

The probability distribution for 5-round AES is *not* Symmetric. Interestingly, it is possible to observe an *asymmetry in the (small-scale) 5-round AES distribution.*

By Fig. 2 and Fig. 4, it is possible to observe that small-scale 5-round AES distribution has positive skew, while the skew of the random distribution is approximately equal to zero. We practically computed these values both for full-size AES and for small-scale one using 2^9 initial cosets, and we got the following results:

$$\gamma^{AES} \simeq 0.43786 \qquad\qquad \gamma^{AES}_{\text{small-scale}} \simeq 0.4687$$

while – as expected – we got that the skew of a random permutation is close to 0 (hence, the probability distribution of a random permutation is well described by a normal one).

It follows that also the skewness can be used to set up a distinguisher. We leave the open problem to theoretically compute these numbers, both for small-scale AES and full-size AES, and to set up a corresponding distinguisher.

9 Truncated Differential Distinguishers for 5-round AES

In this section, we present new truncated differential distinguishers for 5-round AES.

As already recalled in the introduction, differential attacks [BS91] exploit the fact that pairs of plaintexts with certain differences yield other differences in the corresponding ciphertexts with a non-uniform probability distribution. A variant of this attack/distinguisher is the truncated differential one [Knu95], in which the attacker can predict only part of the difference between pairs of texts. Using the subspace terminology, given pairs of plaintexts that belong to the same coset of a subspace $\mathcal{X}$, one consider the probability that the corresponding ciphertexts belong to the same coset of a subspace $\mathcal{Y}$ to set up an attack – see [BLN17] for details. Another type of differential trail is the impossible differential one, where one exploits the fact that pairs of plaintexts that belong to the same coset of a subspace $\mathcal{X}$ cannot belong to the same coset of $\mathcal{Y}$ after a certain number of rounds. For the AES case, truncated differential distinguisher and impossible differential one – which are independent of the key – can be set up for up to 3 and for up to 4 rounds respectively of AES. In both cases, the subspaces $\mathcal{X}$ and $\mathcal{Y}$ correspond respectively to $\mathcal{D}_I$ and $\mathcal{ID}_J$, as shown in detail in [GRR17b].

The truncated differential distinguishers that we are going to present work in a similar way. The main difference with other differential distinguishers in the literature is the fact that our distinguishers work if and only if one considers entire cosets of a particular space $\mathcal{X}$ and not random pairs of texts. Moreover, for the first time in the literature we are able to present a truncated differential distinguisher based on the variance.

9.1 Truncated Differential Distinguisher based on the Variance

Since *the variance of the AES case is different from the one of the random case independently of the secret-key*, we can exploit this fact e.g. to set up a new secret-key distinguisher for 5-round AES.

The idea is very simple. Given n different cosets of a diagonal space $\mathcal{D}_i$ (that is, sets of texts with one active diagonal), one counts the number of different pairs of ciphertexts that are equal in one fixed anti-diagonal (equivalently, that belong to the same coset of $\mathcal{ID}_J$ for $J \subseteq \{0, 1, 2, 3\}$ with $|J| = 3$). Then, one computes the variance: by previous result, the highest one corresponds to the AES case.

We practically tested this distinguisher on a small-scale AES. Since the ratio between the variances for full-size 5-round AES and for a random permutation Π is similar to the same ratio for the case of small-scale AES, that is

$$\frac{Var(\text{AES}^{\text{8-bit}})}{Var(\Pi^{\text{8-bit}})} \simeq \frac{2^{36.161}}{2^{31}} \approx 35.8 \quad versus \quad \frac{Var(\text{AES}^{\text{4-bit}})}{Var(\Pi^{\text{4-bit}})} \simeq \frac{2^{20.035}}{2^{15}} \approx 32.8,$$

we conjecture that the results obtained for the small-scale AES are applicable as well to full-size AES.

By practical tests[17] (the following probability of success have been computed over 2 500 tests) on small-scale AES:

- a single initial coset of $\mathcal{D}_i$ allows to distinguish small-scale AES from a random permutation with prob. 98%
- two initial cosets of $\mathcal{D}_i$ allows to distinguish small-scale AES from a random permutation with prob. 99.9%

where note that for each initial coset of $\mathcal{D}_i$, it is possible to compute the number of collisions with respect to four different anti-diagonals, or equivalently four different subspaces $\mathcal{M}_J$. Moreover, we emphasize that *the goal of this distinguisher is not to compute the exact value of the variance for the two cases, but just to distinguish them.* In other words, the distinguisher works if the variance for AES is bigger than the one of a random permutation, even if it does not return the exact value of the two variances. Thus, due to the big gap between the two cases, 2 initial cosets of $\mathcal{D}_i$ are sufficient for this goal (even if they are not sufficient – in general – to compute the exact value of the two variances).

As a result, one can distinguish the two cases using $n \geq 2$ initial cosets. Due to the relation between small-scale AES and full-size AES previously discussed, we conjecture that the same number of initial cosets is sufficient to distinguish (full-size) AES from a random permutation (using this distinguisher based on the variance). However, just to have more confidence, we choose an arbitrary value of 4 initial cosets in order to set up the distinguisher, for a data cost of $4 \cdot 2^{32} = 2^{34}$ chosen plaintexts distributed in 4 initial cosets of $\mathcal{D}_i$. The computational cost is well approximated by the cost to compute the number of collisions. Using Algorithm 1 - described in details in App. D, the cost is well approximated by 4 (initial cosets) $\times 4$ (anti-diagonals) $\times 3 \cdot 2^{32}$ (table look-ups) $\simeq 2^{37.6}$ table look-ups, that is approximately 2^{31} five-round encryptions.

9.2 Useful Approximation for the Prob. Distribution of 5-round AES

In order to propose a truncated differential distinguisher based on the mean, we first need an approximation of the probability distribution for 5-round AES given in Theorem 2 – Lemma 1. The approximation given in the following turns out to be (very) useful in

[17] Given a set of $n \gg 1$ equally likely values, an *unbiased* estimator for the variance is given by $Var(X) = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{X})^2$ where $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} x_i$.

```
Data: $2^{32}$ (plaintext, ciphertext) pairs $(p^i, c^i)$ for $i = 0, ..., 2^{32} - 1$ with one active
      diagonal (equivalently, in a coset of a diagonal subspace $\mathcal{D}_I$ with $|I| = 1$).
Result: Number $n$ of pairs of ciphertexts which are equal in the $j$-th anti-diagonal
        (equivalently, that belong in the same coset of $\mathcal{ID}_J$ where
        $J \equiv \{0,1,2,3\} \setminus j$)
Let $A[0, ..., 2^{32} - 1]$ an array initialized to zero;
for $i$ from $0$ to $2^{32} - 1$ do
    $x \leftarrow \sum_{k=0}^{3} c^i_{k,j-k} \cdot 256^k$;    // $c^i_{k,j-k}$ denotes the byte of $c^i$ in row $k$ and
      column $j - k \mod 4$
    $A[x] \leftarrow A[x] + 1$; // $A[x]$ denotes the value stored in the $x$-th address
      of the array $A$
end
$n \leftarrow \sum_{i=0}^{2^{32}-1} A[i] \cdot (A[i] - 1)/2$;    // $n \equiv$ Number of Collisions
return $n$.
```

Algorithm 1: *Secret-Key Distinguisher for 5 Rounds of AES.* It count the number of pairs of ciphertexts which are equal in the j-th anti-diagonal (equivalently, the number of collisions in the same coset of $\mathcal{ID}_J$ (where $J \equiv \{0,1,2,3\} \setminus j$).

applications where the skewness does *not* play a crucial role, that is in applications which are (almost) independent of the bias in the skew.

As given in Theorem 2, the probability distribution for 5-round AES is well described by

$$\text{5-AES} = 2^3 \times \mathcal{B}(n_3, p_3) + 2^{10} \times \mathcal{B}(n_{10}, p_{10}) + 2^{17} \times \mathcal{B}(n_{17}, p_{17})$$

where $\mathcal{B}(n,p)$ are binomial distributions. Since $n_3, n_{10}, n_{17} \gg 1$ and $p_3, p_{10}, p_{17} \ll 1$, a first possibility would be to approximate each binomial distribution by a Poisson one, i.e. $\mathcal{B}(n,p) \approx \mathcal{P}(\lambda)$ where $\lambda = n \cdot p$. On the other hand, given the probability distribution $2^3 \cdot \mathcal{P}(n_3 \cdot p_3) + 2^{10} \times \mathcal{P}(n_{10} \cdot p_{10}) + 2^{17} \times \mathcal{P}(n_{17} \cdot p_{17})$, it seems hard to derive a closed "simple" formula[18] which describes the probability to have n collisions in the ciphertexts. The same occurs using e.g. a Gamma distribution.

Another possibility would be to approximate the binomial distributions using the corresponding normal ones. The De Moivre-Laplace Theorem claims that the normal distribution is a good approximation of the binomial one *if* the skewness of the binomial distribution – given in (24) – is close to zero. In our case, $\mathcal{B}(n_3, p_3)$ and $\mathcal{B}(n_{10}, p_{10})$ can be well approximated by a normal distribution, since their skewness are close to zero[19]. Unfortunately, this is not the case of X_{17}:

$$skew(X_3) \approx 2^{-14} \qquad skew(X_{10}) \approx 2^{-7.5} \qquad skew(X_{17}) \approx 0.813 \approx 2^{-0.3}.$$

On the other hand, the number of pairs represented by X_{17} (that is, the pairs of texts with two equal generating variables) is very small compared to the number of all possible pairs of texts, precisely $\frac{3 \cdot 2^{15} \cdot (2^8-1)^2}{2^{31} \cdot (2^{32}-1)} \approx 2^{-30.4} \approx (7.1 \cdot 10^{-8})$ %. For this reason – and with the only goal to set up the truncated diff. distinguisher in the following, we make use of this approximation.

Finally, we exploit the fact that the sum of normally distributed random variables is also normally distributed, that is if $X \sim N(\mu_X, \sigma_X^2)$ and $Y \sim N(\mu_Y, \sigma_Y^2)$, then $Z = X + Y \sim N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$, in order to get an approximation for the probability distribution of 5-round AES.

[18] While it is well known that $\mathcal{P}(\lambda_1) + \mathcal{P}(\lambda_2) = \mathcal{P}(\lambda_1 + \lambda_2)$, to the best of our knowledge there is no closed formula for the case $a_1 \cdot \mathcal{P}(\lambda_1) + a_2 \cdot \mathcal{P}(\lambda_2)$ for $a_1 \neq a_2$.

[19] Note that $skew(\alpha \cdot X) = sign(\alpha) \cdot skew(X)$ where $sign(\alpha) = -1$ if $\alpha < 0$, and 1 otherwise.

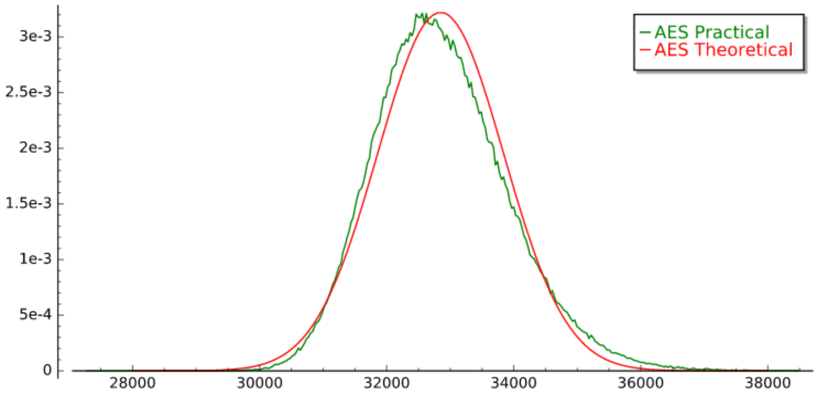

Figure 4: Comparison between the probability distribution of the number of collisions between theoretical small-scale 5-round AES (approximated by a normal distribution) and a practical one.

Corollary 1. *Consider an AES-like cipher that works with texts in $\mathbb{F}_{2^8}^{4\times 4}$ and for which the assumptions of Theorem 2 hold.*

Consider 2^{32} plaintexts p^i for $i = 0, 1, ..., 2^{32} - 1$ with one active diagonal (equivalently, in a coset of a diagonal space $\mathcal{D}_i$ for $i \in \{0, 1, 2, 3\}$), the corresponding ciphertexts after 5 rounds without the final MixColumns operation, that is $c^i = R^5(p^i)$. The probability distribution of the number of different pairs of ciphertexts (c^i, c^j) with $c^i < c^j$ that are equal in one anti-diagonal (equivalently, that belong to the same coset of $\mathcal{ID}_J$ for $J \subseteq \{0, 1, 2, 3\}$ fixed with $|J| = 3$) is approximated by a normal distribution $\mathcal{N}(\mu, \sigma^2)$, where the mean value is equal to $\mu = 2\,147\,484\,685.6 = 2^{32} + 1\,037.6$ and standard deviation is equal to $\sigma = 277\,204.426$.

Roughly speaking, the distribution of the number of collisions for the AES case is approximated by $8 \times X$, where X is a normal distribution with mean value and variance as given in Corollary 1. In more details, the (discrete) probability to have $n \in \mathbb{N}$ collisions is given by:

$$Prob(n \mid \mu, \sigma^2) = \begin{cases} 0 & \text{if } n \bmod 8 \neq 0 \\ \underbrace{\frac{8}{\sqrt{2 \cdot \pi \cdot \sigma^2}} \cdot e^{-\frac{(n-\mu)^2}{2 \cdot \sigma^2}}}_{\sim 8 \times N(\mu, \sigma^2)} & \text{otherwise} \end{cases}$$

A comparison between the real probability distribution for small-scale 5-round AES and the (theoretical) one approximated by a normal distribution is proposed in Fig. 4. As expected, while e.g. the variance of the two distributions are equal, while the main difference is the skewness, since the skewness of a normal distribution is zero, while the one of the probability distribution for 5-round AES is approximately 0.4 (as already pointed out before, do not to confuse the mean with the mode – see Sect. 8 for details).

Finally, a brief explanation about the factor 8 in the probability $Prob(n \mid \mu, \sigma^2)$. Let $Prob(n)$ be the probability - just defined - to have n collisions for 5-round AES. Since $Prob(n \neq 8 \cdot n') = 0$ (i.e. the probability to have n collisions is zero if n is not a multiple of 8), we highlight that *the factor 8 guarantees that the total probability is equal to 1:*

$$\sum_n Prob(n) = \sum_{n = 8 \cdot n'} \frac{8}{\sqrt{2 \cdot \pi \cdot \sigma^2}} \cdot e^{-\frac{(n-\mu)^2}{2 \cdot \sigma^2}} = \sum_{n'} \frac{1}{\sqrt{2 \cdot \pi \cdot (\sigma/8)^2}} \cdot e^{-\frac{(n' - (\mu/8))^2}{2 \cdot (\sigma/8)^2}} = 1.$$

9.3 Truncated Differential Distinguisher based on the Mean

Using previous results, we are now able to present and set up a new distinguisher based on truncated differential trail for 5-round AES, which exploits the fact that the average number of pairs of ciphertexts which are equal in one anti-diagonal (equivalently, that belong in the same coset of $\mathcal{ID}_J$ for $|J| = 3$) is a little bigger for AES than for a random permutation.

As discussed in Sect. 6.1, the number of collisions for 5-round AES and for the random permutation can be described by normal distributions. Moreover, to derive concrete numbers for our distinguisher, we can simply consider the difference of the two distributions, which is again a normal distribution. That is, given $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$, then $X - Y \sim \mathcal{N}(\mu, \sigma^2) = \mathcal{N}(\mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2)$. Indeed, note that to distinguish the two cases, it is sufficient to guarantee that the average number of pairs that satisfy the required property for the random case is smaller than for AES. As a result, the mean μ and the variance σ^2 of the difference between the AES and the random distributions are

$$\begin{aligned}
&\mu = |\mu_{AES} - \mu_{rand}| = n \cdot |p_{AES} - p_{rand}| \\
&\sigma^2 = \sigma_{rand}^2 + \sigma_{AES}^2 = n \cdot \big[p_{rand} \cdot (1 - p_{rand}) + 35.593 \cdot p_{AES} \cdot (1 - p_{AES})\big]
\end{aligned}$$

Since the probability density of the normal distribution is $f(x \mid \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, it follows that

$$prob = \int_{-\infty}^{0} \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}} \mathrm{d}x = \int_{-\infty}^{-\mu/\sigma} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2}} \mathrm{d}x = \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{-\mu}{\sigma\sqrt{2}}\right)\right],$$

where erf(x) is the error function, defined as the probability of a random variable with normal distribution of mean 0 and variance 1/2 falling in the range $[-x, x]$. We emphasize that the integral is computed in the range $(-\infty, 0]$ since we are interested only in the case in which the average number of pairs with the required property in the random case is smaller than in the AES case.

In order to have a probability of success bigger than *prob*, n has to satisfy

$$n > \frac{2 \cdot [p_{rand} \cdot (1 - p_{rand}) + 35.593 \cdot p_{AES} \cdot (1 - p_{AES})]}{(p_{rand} - p_{AES})^2} \cdot \Big[\operatorname{erfinv}\big(2 \cdot prob - 1\big)\Big]^2,$$

where erfinv(x) is the inverse error function. For the case $p_{rand}, p_{AES} \ll 1$, a good approximation of n is given by[20]

$$n > \frac{73.186 \cdot \max(p_{rand},\, p_{AES})}{(p_{rand} - p_{AES})^2} \cdot \Big[\operatorname{erfinv}\big(2 \cdot prob - 1\big)\Big]^2. \tag{25}$$

It follows that in order to have a probability of success bigger than 95%, the number of pairs must satisfy $n \geq 2^{78.374}$ and since $p_{rand} \approx p_{AES} \approx 2^{-30}$ and $|p_{rand} - p_{AES}| \approx 2^{-50.98}$. Given 2^{32} chosen plaintexts with one active diagonal (equivalently, a coset of a diagonal space $\mathcal{D}_i$), it is possible to construct approximately 2^{63} different pairs of texts, which means that the distinguisher requires $2^{15.374}$ different cosets for a data cost of $2^{47.374}$ chosen plaintexts.

[20] Observe: $p_{rand} \cdot (1 - p_{rand}) + 35.593 \cdot p_{AES} \cdot (1 - p_{AES}) < p_{rand} + 35.593 \cdot p_{AES} < 36.593 \cdot \max(p_{rand}, p_{AES})$.

Practical Results on small-scale AES. Since the previous result has been obtained under the assumption that the distribution of AES is well approximated by a normal distribution, we practically tested the probability of success of such distinguisher on a small-scale AES. Using the same computation as before, it turns out that for small-scale AES – where $\mu_{AES^{\text{4-bit}}} = n \cdot p_{AES^{\text{4-bit}}}$ and $\sigma^2_{AES^{\text{4-bit}}} = 29.983 \cdot n \cdot p_{AES^{\text{4-bit}}} \cdot (1 - p_{AES^{\text{4-bit}}})$ – one needs

$$n > \frac{59.965 \cdot \max(p_{rand^{\text{4-bit}}}, p_{AES^{\text{4-bit}}})}{(p_{rand^{\text{4-bit}}} - p_{AES^{\text{4-bit}}})^2} \cdot \left[\text{erfinv}\big(2 \cdot prob - 1\big)\right]^2.$$

different pairs of texts to set up the distinguisher with probability *prob*. In order to have a probability of success higher than 95% and since $p_{rand^{\text{4-bit}}} \approx p_{AES^{\text{4-bit}}} \approx 2^{-14}$ and $|p_{rand^{\text{4-bit}}} - p_{AES^{\text{4-bit}}}| \approx 2^{-22.68485}$, it follows that the number of pairs must satisfy $n \geq 2^{37.48}$. Since each coset of $\mathcal{D}_i$ contains 2^{16} different texts and approximately 2^{31} different pairs, this means that the distinguisher requires $2^{6.48} \simeq 90$ cosets for a data cost of $2^{22.48}$ chosen plaintexts.

By practical tests on small-scale AES:

- 90 initial cosets of $\mathcal{D}_i$ allows to distinguish small-scale AES from a random permutation with prob. 92% (close to 95% used before);
- 180 initial cosets of $\mathcal{D}_i$ allows to distinguish small-scale AES from a random permutation with prob. 98.5%;
- 270 initial cosets of $\mathcal{D}_i$ allows to distinguish small-scale AES from a random permutation with prob. 99.9%;

where the previous probability of success have been computed over 2 500 tests. The fact that the probability of success is a little smaller than expected is well justified by the approximation of the probability distribution for the AES case. Due to these results, due to the similarity between small-scale AES and AES (e.g. the value of the skewness is similar for these two cases – see Sect. 8) and just to have more confidence, we choose an arbitrary value of $3 \cdot 2^{15.375} = 2^{16.96}$ initial cosets in order to set up the distinguisher for AES, for a data cost of $2^{16.96} \cdot 2^{32} = 2^{48.96}$ chosen plaintexts distributed in $2^{16.96}$ initial cosets of $\mathcal{D}_j$.

The Computational Cost. We have just seen that $2^{48.96}$ chosen plaintexts (i.e. $2^{16.96}$ cosets of a diagonal subspace $\mathcal{D}_i$ for $i \in \{0, 1, 2, 3\}$) are sufficient to distinguish a random permutation from 5 rounds of AES, simply counting the number of pairs of ciphertexts that belong to the same coset of $\mathcal{ID}_J$ for $|J| = 3$ and using the fact that this number is bigger for AES. Here we give an estimation of the computational cost of the distinguisher, which is (approximately) given by the cost to count the number of collisions. Using Algorithm 1, the total computational cost can be well approximated by $2^{52.6}$ table look-ups, or equivalently 2^{46} five-round encryptions of AES (using the approximation[21] 20 table look-ups ≈ 1 round of AES). All details can be found in App. D.1.

10 Property of the S-Box and 5-round Truncated Distinguisher based on the Mean

In this paper, we have presented a new truncated property for 5-round AES-like ciphers in the case in which "the solutions of equation (1) are uniformly distributed for each

[21] Even if this approximation (largely used in the literature) is not formally correct – the size of the table of an S-Box look-up is smaller than the size of the table used for our proposed distinguisher, it allows to give a comparison between our distinguishers and the others currently present in the literature.

input/output difference $\Delta_I \neq 0$ and $\Delta_O \neq 0$", which is close to being true if the S-Box is APN, or if the SBox is "close" to be APN. Even if no S-Box (completely) satisfies this assumption in $\mathbb{F}_{2^4}$ or $\mathbb{F}_{2^8}$, *the theoretically results of Theorem 2 are matched by the practical results obtained for the AES S-Box, which approximately satisfies the assumptions of such Theorem* (as discussed in Sect. 4.2). Thus, natural questions arise: *What happens when the AES S-Box is changed with an S-Box that does not satisfy (at all) the assumptions of Theorem 2? Is it possible to naturally extend our results to any general case?*

We have studied this problem working on small-scale AES, and by practical results the answer to the second question seems to be negative. In other words, our theory does not extend naturally to generic S-Box, but it should be modified depending on the particular properties/details of the S-Box function.

10.1 Preliminary Considerations and Practical Results

Roughly speaking, the results/proof provided in Sect. 5 work under the assumption that – for each non-null Δ_I and Δ_O – the variance of the probability distribution of the number of solutions x of the equation S-Box$(x \oplus \Delta_I)\oplus$ S-Box$(x) = \Delta_O$ is close to zero. Obviously, this cannot be the case. *Since the variance of such distribution depends on the details of the S-Box, we expect that our theoretical results match the practical ones when one works with an S-Box that minimizes such variance.* As already explained in Sect. 4.2, even if this (almost) happens for an APN S-Box, note that the variance of such distribution for the AES S-Box is very close to the variance of such distribution for an APN S-Box ("Variance APN = 64004/65025" versus "Variance AES = 67064/65025").

So, *what happens in the case of an S-Box that do not satisfy the previous assumption?* Here we start an analysis in order to better understand which properties of the S-Box play a crucial role when computing the average number of collisions for 5-round AES. In more details, we did several practical tests by counting the average number of collisions in the case in which the AES S-Box is replaced with other S-Box permutations present in the literature - PRINCE [BCG+12], MIDORI [BBI+15], KLEIN [GNL12], PRESENT [BKL+07], RECTANGLE [ZBL+15], NOEKEON [DPAR00] and PRIDE [ADK+14] - and with some "toy" S-Boxes. For our tests, given 2^{16} plaintexts with one active diagonal (equivalently, in the same coset of $\mathcal{D}_i$ for $i \in \{0, 1, 2, 3\}$), we counted the average number of pairs of ciphertexts that are equal in one fixed anti-diagonal (equivalently, that belong in the same coset of $\mathcal{ID}_J$ for J fixed with $|J| = 3$), and we computed the mean. The obtained results are listed in Table 3, where we also highlight some properties of the used S-Box (definitions and differential spectrum of the used S-Boxes are given in App. G) and the difference between the number of collisions found by experiments and the theoretical number 32 847.124 under the assumptions of Theorem 2 (while the average number of collisions for a random permutation is 32 767.5). For each AES-like cipher, we used 125 000 $\simeq 2^{17}$ different initial cosets (values given in the table are the average ones), where new (random) keys are generated for each test.

We emphasize that, while all these AES-like ciphers satisfy the multiple-of-8 property, for some of them the average number of collisions is bigger than for the case of a random permutation (e.g. AES S-Box), while for others it is smaller (e.g. Toy-12 S-Box). This supports again the *independence* of the multiple-of-8 property from the fact that the average number of collisions is bigger for 5-round AES.

10.2 Observations and (possible) Explanation

As expected, *the (absolute) difference between the found number of collisions and the theoretical one seems to increase when the variance (of the S-Box) increases, while it seems to be independent of the maximum differential probability DP_{max}.* Moreover, the difference between the theoretical number of collisions (given under the assumptions of Theorem 2,

Table 3: In the following table, we provide the results of our practical tests about the number of different pairs of ciphertexts that belong to the same coset of $\mathcal{M}_J$ for J fixed with $|J| = 3$ when *the AES S-Box is replaced by the S-Box of other ciphers.* Together with the number of collisions, we provide the most relevant properties of the S-Box ("Var" denotes the variance of the probability distribution that describes the number of solutions of the eq. S-Box$(x \oplus \Delta_I) \oplus$ S-Box$(x) = \Delta_O$ – see also App. G for details) and the "Diff." (= Difference) between the practical and the theoretical number (= 32 847.124) of collisions - under the assumptions of Theorem 2.

AES-like Cipher	**Numb. Collisions**	**Diff.**	$2^4 \times$ **DP**$_{\mathbf{max}}$	**Var**	**Homogeneous**
AES S-Box	32 848.6	+1.5	4	344/225	✓
KLEIN S-Box	32 849.8	+2.7	4	344/225	
MIDORI SB_1 S-Box	32 843.0	−4.10	4	344/225	
PRINCE S-Box	32 852.7	+5.5	4	344/225	
Toy-6 S-Box	32 840.1	−7.1	6	392/225	
RECTANGLE S-Box	32 861.2	+14.0	4	416/225	
NOEKEON S-Box	32 878.7	+31.6	4	416/225	
MIDORI SB_0 S-Box	32 882.8	+35.7	4	416/225	
PRESENT S-Box	32 886.3	+39.2	4	416/225	
PRIDE S-Box	32 806.6	−40.5	4	416/225	
Toy-8 S-Box	32 815.7	−31.5	8	464/225	
Toy-10 S-Box	32 919.0	+71.9	10	864/225	
Toy-12 S-Box	32 684.1	−163.0	12	896/225	

i.e. that the number of solutions n_{Δ_I,Δ_O} of equation (1) are uniform distributed) and the practical one is minimum when the S-Box almost satisfies the assumption of Theorem 2 (e.g. the AES S-Box).

To explain these results, we must refer to the proof of Theorem 2 given in Sect. 5. The idea is to consider a system of 4 equations of the generic form (15), and to look for common solutions. In the case in which the solutions (in particular, the number of solutions n_{Δ_I,Δ_O}) of equation (1) are uniformly distributed, the probability that a possible solution satisfies all the 4 equations of the system is well approximated by $(255^{-4} \cdot 2^{-31})^3$, as explained in the proof of Sect. 5. This allows to (theoretical) predict the average number of common solutions, and so of collisions. Instead, in the case in which the solutions (in particular, the number of solutions n_{Δ_I,Δ_O}) of equation (1) are not uniform distributed (e.g. if the variance of the S-Box is not "small"), the probability to have a common solution is in general different from the one just given. As a result, the number of solutions of a system of equations like (15) can be bigger or smaller w.r.t. the one given in Theorem 2 (and the difference can be also non-negligible). It follows that the number of collisions is influenced by the details of the S-Box (as expected). *As future work, an open problem is to theoretically prove this conjecture about the link between the average number of collisions and the variance of the S-Box, and to theoretically derive the numbers given in Table 3.*

What about the distinguisher based on the variance (Sect. 6.1)? To compute the value of the variance, we have exploited the "multiple-of-8" property [GRR17a], the properties of the Variance (if X is a random variable and a a scalar, then $Var(\alpha \cdot X) = \alpha^2 \cdot Var(X)$) and the probability p_{AES} that – given a pair of plaintexts in a coset of $\mathcal{D}_i$ – two ciphertexts are equal in one fixed anti-diagonal (eq. belong to the same coset of $\mathcal{ID}_J$) after 5 rounds. This probability p_{AES} proposed in Sect. 4.2 depends on the details of the S-Box, as we have just seen. It follows that also the value of the variance depends on it. On the other hand, we found by practical tests that *the value of the variance changes much less than the corresponding value of the mean* when the S-Box changes. In general, the value of the variance is "almost" independent of the details of the S-Box. Moreover, since the variance for an AES-like cipher is much bigger than the one of a random permutation, the proposed distinguisher works even if the value of the variance is (a little) different than the one given in Theorem 2.

10.3 MixColumns Dependence

Until now, we have focused only on the details of the S-Box. *How does the average number of collisions depend on the details of the MixColumns matrix?*

MDS Matrix: "Good" vs "Bad" S-Box. We start by focusing on the case in which the MixColumns matrix is MDS, and then we briefly discuss the other cases. *If the S-Box satisfies the assumptions of Theorem 2, then the average number of collisions is (almost) independent of the MixColumns matrix details. Instead, if the S-Box does not satisfy the previous requirement, this number depends also on the details of the MixColumns matrix.* In particular, in this last case the solutions (and the corresponding number n_{Δ_I,Δ_O}) of equation (1) are not uniform distributed with respect to $\Delta_I \neq 0$ and $\Delta_O \neq 0$, and so the number of solutions of a system of 4 equations of the generic form (15) depends both on the details of the S-Box and of the linear layer. Indeed, remember that a system of equations of the generic form (15) depends on the coefficients of the MixColumns matrix, and so also the fact that a common solution exists.

To give a practical example, consider the (circulant) MixColumns matrix defined as $MC = circ(0x01, 0x03, 0x02, 0x02)$, that is the AES MixColumns matrix where 0x01 is replaced by 0x02 and vice-versa. We got that the number of collisions in the case of AES S-Box is 32 850.32, while in the case of PRESENT S-Box is 32 872.95. Thus, a difference in the MixColumns matrix implies almost no difference for the AES S-Box case (on average, there are +1.75 collisions for this new MDS matrix), while a higher difference occurs for the PRESENT S-Box case (on average, there are −13.37 collisions for this new MDS matrix). Again, this is due to the fact that the probability that a system of 4 equations of the generic form (15) admits a common solution depends both on the details of the S-Box and of the linear layer, in the case in which the S-Box is not "good" (w.r.t. assumptions of Theorem 2). Similar results can be obtained using different MDS MixColumns matrices.

Non-MDS Matrix. Finally, if the AES MixColumns matrix is replaced by an "almost MDS" one (which does not satisfy the assumptions of Theorem 2), then the number of collisions can be different with respect to the one predicted by Theorem 2 also in the case of "good" S-Box. As example, using the Midori matrix $MC_{Midori} = circ(0x00, 0x01, 0x01, 0x01)$ and the AES S-Box, the number of collisions after 5 rounds is on average 31 883.27 (instead of a theoretical number of 32 847.124). The same occurs also using a MixColumns matrix which is not MDS and for which all coefficients are different than zero. E.g. using the matrix $circ(0x02, 0x01, 0x01, 0x01)$ and the AES S-Box, the number of collisions after 5 rounds is on average 33 377.93 (instead of a theoretical number of 32 847.124).

11 Open Problems

To conclude, we collect and list the main open problems that arise from this paper.

Distinguisher based on the Mean: "APN-like" Assumption and Other Problems

- In order to theoretically describe the 5-round AES truncated diff. distinguisher based on the mean, we need particular assumptions on the S-Box, which are related to the fact that the number of solutions x of the equation $\text{S-Box}(x \oplus \Delta_I) \oplus \text{S-Box}(x) = \Delta_O$ are uniformly distributed for each non-null input/output difference Δ_I, Δ_O. Given the probability distribution of the number of solutions of such equation, a relation exists between the fact that its variance is “small” and that the solutions are uniform distributed.

On the other hand, consider the practical results on small-scale AES proposed in Sect. 10. The just given property on the variance of such distribution seems *not* to be sufficient to describe such results. E.g. focusing on the average number of collisions for different “S-Boxes with the same variance” – Table 3, it turns out that the average number of collisions is not (approximately) equal for all the “S-Boxes with the same variance”, as we should expect.

An open problem is to understand *which parameters/properties of the S-Box really influence the average number of collisions for 5-round AES (equivalently, play a crucial role on the fact that the solutions of the equation (1) are uniformly distributed for each non-zero input/output difference). Is it possible to theoretically compute a confidence interval* $[M-m, M+m]$ *– which depends on the these parameters/properties of the S-Box – such that all average numbers of collisions given in Table 3 fall into such interval with high probability?*

- The 5-round AES truncated differential distinguisher based on the mean seems to depend also on the details of the MixColumns matrix. It could be interesting to better understand this fact: *which details of the MixColumns matrix influence* – especially in the case of “bad” S-Boxes – *the average number of collisions for 5-round AES? Is it possible to better estimate the average number of collisions taking into account these details? Is it possible to theoretically compute the average number of collisions – equivalently, to reformulate the proof given in Sect. 5 – using a “weaker” assumption than the MDS one? Is it possible to theoretically predict the average number of collisions when working with a matrix*[22] *which is not MDS?*
- Let’s focus again on the 5-round AES truncated differential distinguisher based on the mean, in particular on its proof. In order to theoretically compute the average number of collisions, we focus and work only on the middle round of

$$\mathcal{D}_I \oplus \delta \xrightarrow[\text{prob. } 1]{R^2(\cdot)} \mathcal{M}_I \oplus \omega \xrightarrow{R(\cdot)} \mathcal{D}_J \oplus \delta' \xrightarrow[\text{prob. } 1]{R_f^2(\cdot)} \mathcal{ID}_J \oplus \omega'.$$

  In particular, given two plaintexts $p^1, p^2 \in \mathcal{M}_I \oplus \omega$, the fact that they belong to the same coset of $\mathcal{D}_J$ after one round – that is, $R(p^1) \oplus R(p^2) \in \mathcal{D}_J$ – can be re-written as a system of equations that involves the S-Box operation of the form

$$\bigoplus_{\{i,j\}\in\mathcal{I}} A_{i,j} \times \left[\text{S-Box}(B_{i,j} \cdot p^1_{i,j} \oplus C_{i,j}) \oplus \text{S-Box}(B_{i,j} \cdot p^2_{i,j} \oplus C_{i,j})\right] = 0$$

  for a set of index $\mathcal{I}$ and for some constants A, B, C (which depend only on the MixColumns matrix and on the secret key). A similar analysis can be performed for 6-round AES, by replacing the S-Box operation with the “super-Sbox” notation, where

$$\textit{super-Sbox}(\cdot) = \text{S-Box} \circ ARK \circ MC \circ \text{S-Box}(\cdot).$$

  Indeed, note that *(1st)* 6-round AES can be re-written as

$$\mathcal{D}_I \oplus \delta \xrightarrow[\text{prob. } 1]{R^2(\cdot)} \mathcal{M}_I \oplus \omega \xrightarrow{R^2(\cdot)} \mathcal{D}_J \oplus \delta' \xrightarrow[\text{prob. } 1]{R_f^2(\cdot)} \mathcal{ID}_J \oplus \omega',$$

  and *(2nd)* given two plaintexts $p^1, p^2 \in \mathcal{M}_I \oplus \omega$, the fact that they belong to the same coset of $\mathcal{D}_J$ after two rounds - that is, $R^2(p^1) \oplus R^2(p^2) \in \mathcal{D}_J$ - can be re-written

[22] Note that the MixColumns matrix of many AES-like lightweight ciphers in the literature is not MDS.

as a system of equations (similar to the one just given) that involves the super-Sbox. An open problem is to modify the proof given in Sect. 5 in order to get a similar result about the average number of collisions for 6-round AES. *Is it possible to set up a truncated differential distinguisher for 6-round AES which is independent of the secret key?*

Distinguisher based on the Variance and on the Skew

- In Sect. 9.1, we propose the first truncated differential distinguisher based on the variance which is (much) more competitive - both for the computational and data costs - than the corresponding one based on the mean. However, an open problem is *to formally study* the probability of success of such distinguisher. Is it possible to set up similar distinguishers for other ciphers?
- As highlighted in Sect. 8, the skewness of the probability distribution of 5-round AES is different from the one of a random permutation. In particular, the bias in the skewness seems to be stronger than the bias in the mean. As a result, also this parameter could be potentially used to set up distinguishers and/or key-recovery attacks. The problem *to theoretically compute the value of the skewness* of such probabilistic distributions is open for future research: does it depend on the details of the S-Box?

Acknowledgements. Authors thank Matthias Niel for implementing on C/C++ the truncated diff. distinguishers on 6- and 8-bit AES presented in this paper. Authors thank Anne Canteaut for pointing out the property about the ratio between the variance of 5-round AES and of a random permutation described in App. E.3. Finally, authors also thank anonymous reviewers for their valuable comments and suggestions. L. Grassi has been supported by IOV42.

References

[ADK+14] Martin R. Albrecht, Benedikt Driessen, Elif Bilge Kavun, Gregor Leander, Christof Paar, and Tolga Yalçın. Block Ciphers - Focus on the Linear Layer (feat. PRIDE). In *Advances in Cryptology - CRYPTO 2014*, volume 8616 of *LNCS*, pages 57–76, 2014.

[BBI+15] Subhadeep Banik, Andrey Bogdanov, Takanori Isobe, Kyoji Shibutani, Harunaga Hiwatari, Toru Akishita, and Francesco Regazzoni. Midori: A Block Cipher for Low Energy. In *Advances in Cryptology - ASIACRYPT 2015*, volume 9453 of *LNCS*, pages 411–436, 2015.

[BBS99] Eli Biham, Alex Biryukov, and Adi Shamir. Cryptanalysis of Skipjack Reduced to 31 Rounds Using Impossible Differentials. In *Advances in Cryptology - EUROCRYPT 1999*, volume 1592 of *LNCS*, pages 12–23, 1999.

[BCC19] Christina Boura, Anne Canteaut, and Daniel Coggia. A General Proof Framework for Recent AES Distinguishers. *IACR Trans. Symmetric Cryptol.*, 2019(1):170–191, 2019.

[BCG+12] Julia Borghoff, Anne Canteaut, Tim Güneysu, Elif Bilge Kavun, Miroslav Knezevic, Lars R. Knudsen, Gregor Leander, Ventzislav Nikov, Christof Paar, Christian Rechberger, Peter Rombouts, Søren S. Thomsen, and Tolga Yalçın. PRINCE – A Low-Latency Block Cipher for Pervasive Computing

Applications. In *Advances in Cryptology - ASIACRYPT 2012*, volume 7658 of *LNCS*, pages 208–225, 2012.

[BDK+18] Achiya Bar-On, Orr Dunkelman, Nathan Keller, Eyal Ronen, and Adi Shamir. Improved Key Recovery Attacks on Reduced-Round AES with Practical Data and Memory Complexities. In *Advances in Cryptology - CRYPTO 2018*, volume 10992 of *LNCS*, pages 185–212, 2018.

[BDMW10] K. A. Browning, J. F. Dillon, M. T. McQuistan, and A. J. Wolfe. An APN permutation in dimension six. *IEEE Transactions on Information Theory*, 518:33–42, 2010.

[BK01] Eli Biham and Nathan Keller. Cryptanalysis of Reduced Variants of Rijndael, 2001. unpublished, http://csrc.nist.gov/archive/aes/round2/conf3/papers/35-ebiham.pdf.

[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 *Cryptographic Hardware and Embedded Systems - CHES 2007*, volume 4727 of *LNCS*, pages 450–466, 2007.

[BLN17] Céline Blondeau, Gregor Leander, and Kaisa Nyberg. Differential-Linear Cryptanalysis Revisited. *Journal of Cryptology*, 30(3):859–888, 2017.

[BS91] Eli Biham and Adi Shamir. Differential cryptanalysis of DES-like cryptosystems. *Journal of Cryptology*, 4(1):3–72, Jan 1991.

[CMR05] Carlos Cid, Sean Murphy, and Matthew J. B. Robshaw. Small Scale Variants of the AES. In *Fast Software Encryption - FSE 2005*, volume 9054 of *LNCS*, pages 145–162, 2005.

[Der13] Patrick Derbez. Meet-in-the-middle attacks on AES. PhD thesis, Ecole Normale Supérieure de Paris, 2013. https://tel.archives-ouvertes.fr/tel-00918146.

[DKR97] Joan Daemen, Lars R. Knudsen, and Vincent Rijmen. The Block Cipher Square. In *Fast Software Encryption - FSE 1997*, volume 1267 of *LNCS*, pages 149–165, 1997.

[DPAR00] Joan Daemen, Michaël Peeters, Gilles Van Assche, and Vincent Rijmen. Nessie proposal: the block cipher Noekeon. Nessie submission, 2000. http://gro.noekeon.org/.

[DR02] Joan Daemen and Vincent Rijmen. *The Design of Rijndael: AES - The Advanced Encryption Standard*. Information Security and Cryptography. Springer, 2002.

[GNL12] Zheng Gong, Svetla Nikova, and Yee Wei Law. KLEIN: A New Family of Lightweight Block Ciphers. In *RFID 2011*, volume 7055 of *LNCS*, pages 1–18, 2012.

[Gra17] Lorenzo Grassi. Mixture Differential Cryptanalysis and Structural Truncated Differential Attacks on round-reduced AES. Cryptology ePrint Archive, Report 2017/832, 2017. https://eprint.iacr.org/2017/832.

[Gra18] Lorenzo Grassi. Mixture Differential Cryptanalysis: a New Approach to Distinguishers and Attacks on round-reduced AES. *IACR Trans. Symmetric Cryptol.*, 2018(2):133–160, 2018.

[GRR17a] Lorenzo Grassi, Christian Rechberger, and Sondre Rønjom. A New Structural-Differential Property of 5-Round AES. In *Advances in Cryptology - EUROCRYPT 2017*, volume 10211 of *LNCS*, pages 289–317, 2017.

[GRR17b] Lorenzo Grassi, Christian Rechberger, and Sondre Rønjom. Subspace Trail Cryptanalysis and its Applications to AES. *IACR Transactions on Symmetric Cryptology*, 2016(2):192–225, 2017.

[Knu95] Lars R. Knudsen. Truncated and higher order differentials. In *Fast Software Encryption - FSE 1994*, volume 1008 of *LNCS*, pages 196–211, 1995.

[LAAZ11] Gregor Leander, Mohamed Ahmed Abdelraheem, Hoda AlKhzaimi, and Erik Zenner. A Cryptanalysis of PRINTcipher: The Invariant Subspace Attack. In *Advances in Cryptology - CRYPTO 2011*, volume 6841 of *LNCS*, pages 206–221, 2011.

[Lai94] Xuejia Lai. *Higher Order Derivatives and Differential Cryptanalysis*, pages 227–233. Springer US, 1994.

[LP07] G. Leander and A. Poschmann. On the Classification of 4 Bit S-Boxes. In *Arithmetic of Finite Fields - WAIFI 2007*, volume 4547 of *LNCS*, pages 159–176, 2007.

[Nyb91] Kaisa Nyberg. Perfect nonlinear S-boxes. In *Advances in Cryptology - EUROCRYPT 1991*, volume 547 of *LNCS*, pages 378–386, 1991.

[Nyb93] Kaisa Nyberg. Differentially Uniform Mappings for Cryptography. In *Advances in Cryptology – EUROCRYPT 1993*, volume 765 of *LNCS*, pages 55–64, 1993.

[RBH17] Sondre Rønjom, Navid Ghaedi Bardeh, and Tor Helleseth. Yoyo Tricks with AES. In *Advances in Cryptology - ASIACRYPT 2017*, volume 10624 of *LNCS*, pages 217–243, 2017.

[Tie16] Tyge Tiessen. Polytopic Cryptanalysis. In *Advances in Cryptology - EUROCRYPT 2016*, volume 9665 of *LNCS*, pages 214–239, 2016.

[Tun12] Michael Tunstall. Improved “Partial Sums"-based Square Attack on AES. In *International Conference on Security and Cryptography - SECRYPT 2012*, volume 4817 of *LNCS*, pages 25–34, 2012.

[ZBL$^+$15] WenTao Zhang, ZhenZhen Bao, DongDai Lin, Vincent Rijmen, BoHan Yang, and Ingrid Verbauwhede. RECTANGLE: a bit-slice lightweight block cipher suitable for multiple platforms. *Science China Information Sciences*, 58(12):1–15, Dec 2015.

A Subspace Trails Cryptanalysis for AES

In this section, we recall few details about the subspace trails of AES presented in [GRR17b]. Let F denote a round function in an iterative block cipher and let $V \oplus a$ denote a coset of a vector space V. Then if $F(V \oplus a) = V \oplus a$ we say that $V \oplus a$ is an *invariant coset* [LAAZ11] of the subspace V for the function F. As shown in [GRR17b], this concept can be generalized to *trails of subspaces*.

Definition 2. Let F denote a round function in an iterative block cipher and let $(V_1, V_2, ..., V_{r+1})$ denote a set of $r+1$ subspaces with $\dim(V_i) \leq \dim(V_{i+1})$. If for each $i = 1, ..., r$ and for each a_i, there exists a_{i+1} s.t. $F(V_i \oplus a_i) \subseteq V_{i+1} \oplus a_{i+1}$, then $(V_1, V_2, ..., V_{r+1})$ is *subspace trail* of length r for the function F.

This means that if F^t denotes the application of t rounds with fixed keys, then $F^t(V_1 \oplus a_1) \subseteq V_{t+1} \oplus a_{t+1}$.

Subspace Trails for AES. In the following, we only work with vectors and vector spaces over $\mathbb{F}_{2^8}^{4\times 4}$, and we denote by $\{e_{0,0}, ..., e_{3,3}\}$ the unit vectors of $\mathbb{F}_{2^8}^{4\times 4}$ (e.g. $e_{i,j}$ has a single 1 in row i and column j).

Definition 3. The *column spaces* $\mathcal{C}_i$ are defined as $\mathcal{C}_i = \langle e_{0,i}, e_{1,i}, e_{2,i}, e_{3,i}\rangle$.

For instance, $\mathcal{C}_0$ corresponds to the symbolic matrix

$$\mathcal{C}_0 = \left\{ \begin{bmatrix} x_1 & 0 & 0 & 0 \\ x_2 & 0 & 0 & 0 \\ x_3 & 0 & 0 & 0 \\ x_4 & 0 & 0 & 0 \end{bmatrix} \middle| \forall x_1, x_2, x_3, x_4 \in \mathbb{F}_{2^8} \right\} \equiv \begin{bmatrix} x_1 & 0 & 0 & 0 \\ x_2 & 0 & 0 & 0 \\ x_3 & 0 & 0 & 0 \\ x_4 & 0 & 0 & 0 \end{bmatrix}.$$

Definition 4. The *diagonal spaces* $\mathcal{D}_i$ are defined as $\mathcal{D}_i = SR^{-1}(\mathcal{C}_i)$. Similarly, the *inverse-diagonal spaces* $\mathcal{ID}_i$ are defined as $\mathcal{ID}_i = SR(\mathcal{C}_i)$.

For instance, $\mathcal{D}_0$ and $\mathcal{ID}_0$ correspond to symbolic matrix

$$\mathcal{D}_0 \equiv \begin{bmatrix} x_1 & 0 & 0 & 0 \\ 0 & x_2 & 0 & 0 \\ 0 & 0 & x_3 & 0 \\ 0 & 0 & 0 & x_4 \end{bmatrix}, \qquad \mathcal{ID}_0 \equiv \begin{bmatrix} x_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & x_2 \\ 0 & 0 & x_3 & 0 \\ 0 & x_4 & 0 & 0 \end{bmatrix}$$

for all $x_1, x_2, x_3, x_4 \in \mathbb{F}_{2^8}$.

Definition 5. The *i-th mixed spaces* $\mathcal{M}_i$ are defined as $\mathcal{M}_i = MC(\mathcal{ID}_i)$.

For instance, $\mathcal{M}_0$ corresponds to symbolic matrix

$$\mathcal{M}_0 \equiv \begin{bmatrix} \text{0x02} \cdot x_1 & x_4 & x_3 & \text{0x03} \cdot x_2 \\ x_1 & x_4 & \text{0x03} \cdot x_3 & \text{0x02} \cdot x_2 \\ x_1 & \text{0x03} \cdot x_4 & \text{0x02} \cdot x_3 & x_2 \\ \text{0x03} \cdot x_1 & \text{0x02} \cdot x_4 & x_3 & x_2 \end{bmatrix}$$

for all $x_1, x_2, x_3, x_4 \in \mathbb{F}_{2^8}$.

Definition 6. For $I \subseteq \{0, 1, 2, 3\}$, let $\mathcal{C}_I$, $\mathcal{D}_I$, $\mathcal{ID}_I$ and $\mathcal{M}_I$ be defined as

$$\mathcal{C}_I = \bigoplus_{i\in I} \mathcal{C}_i, \qquad \mathcal{D}_I = \bigoplus_{i\in I} \mathcal{D}_i, \qquad \mathcal{ID}_I = \bigoplus_{i\in I} \mathcal{ID}_i, \qquad \mathcal{M}_I = \bigoplus_{i\in I} \mathcal{M}_i.$$

As shown in detail in [GRR17b], for any coset $\mathcal{D}_I \oplus a$ there exists unique $b \in \mathcal{C}_I^\perp$ such that $R(\mathcal{D}_I \oplus a) = \mathcal{C}_I \oplus b$. Similarly, for any coset $\mathcal{C}_I \oplus a$ there exists unique $b \in \mathcal{M}_I^\perp$ such that $R(\mathcal{C}_I \oplus a) = \mathcal{M}_I \oplus b$.

Theorem 4 ([GRR17b])**.** *For each $I \subseteq \{0, 1, 2, 3\}$ and for each $a \in \mathcal{D}_I^\perp$, there exists one and only one $b \in \mathcal{M}_I^\perp$ s.t. $R^2(\mathcal{D}_I \oplus a) = \mathcal{M}_I \oplus b$.*

We refer to [GRR17b] for a complete proof of this theorem, and we limit ourselves to emphasize that b depends on the initial coset of $\mathcal{D}_I$ defined by a and on the secret key k.

Observe that if X is a subspace, $X \oplus a$ is a coset of X and x and y are two elements of the (same) coset $X \oplus a$, then $x \oplus y \in X$. It follows that:

Lemma 3. *For all x, y and for all $I \subseteq \{0, 1, 2, 3\}$:*

$$Prob(R^2(x) \oplus R^2(y) \in \mathcal{M}_I \,|\, x \oplus y \in \mathcal{D}_I) = 1.$$

We finally recall that for each $I, J \subseteq \{0, 1, 2, 3\}$ then $\mathcal{M}_I \cap \mathcal{D}_J = \{0\}$ if and only if $|I| + |J| \leq 4$, as demonstrated in [GRR17b]. It follows that:

Theorem 5 ([GRR17b])**.** *Let $I, J \subseteq \{0, 1, 2, 3\}$ such that $|I| + |J| \leq 4$. For all x, y:*

$$Prob(R^4(x) \oplus R^4(y) \in \mathcal{M}_I \,|\, x \oplus y \in \mathcal{D}_J, \quad x \neq y) = 0.$$

B Number of Collisions – Random Permutation

Consider 2^{32} plaintexts in the same coset of a diagonal space $\mathcal{D}_i$. In Sect. 5, we approximately compute the number of different pairs of ciphertexts *generated by a random permutation* that belong to the same coset of $\mathcal{M}_J$ after 5-round for $|J| = 3$ (or in the same coset of $\mathcal{ID}_J$ if the MixColumns matrix is omitted). This number is approximately given by

$$\binom{2^{32}}{2} \cdot 2^{-32} \simeq 2\,147\,483\,647.5 = 2^{31} - 2^{-1}.$$

Here we show that this is a good approximation.

In the previous computation, we assume that all the pairs are independent. However, this is not the case. Indeed, consider three texts, that is t^1, t^2 and t^3, and the corresponding three couples, that is $(t^1, t^2), (t^1, t^3)$ and (t^2, t^3). Three possible events can happen:

- if $t^1 \oplus t^2 \in \mathcal{M}_J$ and $t^1 \oplus t^3 \in \mathcal{M}_J$, then $t^2 \oplus t^3 \in \mathcal{M}_J$ with probability 1 (since $\mathcal{M}_J$ is a subspace);
- if $t^1 \oplus t^2 \in \mathcal{M}_J$ and $t^1 \oplus t^3 \notin \mathcal{M}_J$ (or vice-versa), then $t^2 \oplus t^3 \notin \mathcal{M}_J$ with probability 1 (since $\mathcal{M}_J$ is a subspace);
- if $t^1 \oplus t^2 \notin \mathcal{M}_J$ and $t^1 \oplus t^3 \notin \mathcal{M}_J$, then both the events $t^2 \oplus t^3 \in \mathcal{M}_J$ and $t^2 \oplus t^3 \notin \mathcal{M}_J$ are possible; in particular, $t^2 \oplus t^3 \in \mathcal{M}_J$ with *approximately* prob. $2^{-32 \cdot (4-|J|)}$.

On the other hand, *what is the probability that a pair of texts (p, q) satisfy $p \oplus q \in \mathcal{M}_J$? In the following, we prove that such probability is equal to $2^{-32 \cdot (|J|-4)}$.*

To answer the previous question, first of all, it is important to focus on the previous last event and to theoretically compute a better approximation of this probability. For our goal, we focus on the case $|J| = 3$ with J fixed. We are going to show that the last probability is well approximated by $2^{-32} \cdot (1 - 2^{-32})^{-1}$. Since $t^1 \oplus t^2 \notin \mathcal{M}_J$, it follows that the difference on the J-th anti-diagonal is different from (0,0,0,0), i.e. they can take only one of $2^{32} - 1$ possible values different from (0,0,0,0). Similar consideration holds for $t^1 \oplus t^3 \notin \mathcal{M}_J$. Since $t^2 \oplus t^3 = (t^1 \oplus t^2) \oplus (t^1 \oplus t^3)$, it follows that the difference of the

J-th anti-diagonal of $t^2 \oplus t^3$ is equal to zero if the difference of the J-th anti-diagonal of $t^1 \oplus t^2$ is equal to the difference of the J-th anti-diagonal of $t^1 \oplus t^3$. Since this happens with probability $(2^{32}-1)^{-1}$, it follows that the probability that $t^1 \oplus t^3 \in \mathcal{M}_J$ is

$$(2^{32}-1)^{-1} = 2^{-32} \cdot (1-2^{-32})^{-1} \approx 2^{-32} + 2^{-64} - 2^{-96} + ...$$

To have more confidence about this fact, note that:

- $t^1 \oplus t^2 \in \mathcal{M}_J$, $t^1 \oplus t^3 \in \mathcal{M}_J$ and $t^2 \oplus t^3 \in \mathcal{M}_J$ occurs with probability $(2^{-32})^2$;
- $t^1 \oplus t^2 \in \mathcal{M}_J$, $t^1 \oplus t^3 \notin \mathcal{M}_J$ and $t^2 \oplus t^3 \notin \mathcal{M}_J$ occurs with probability $2^{-32} \cdot (1-2^{-32})$ (similar for the other 3 cases);
- $t^1 \oplus t^2 \notin \mathcal{M}_J$, $t^1 \oplus t^3 \notin \mathcal{M}_J$ and $t^2 \oplus t^3 \notin \mathcal{M}_J$ occurs with probability $(1-2^{-32})^2 \cdot (1-2^{-32} \cdot (1-2^{-32})^{-1})$.

All the other cases have probability 0 (since $\mathcal{M}_J$ is a subspace). By simple computation, the probability of all the possible events is equal to

$$(2^{-32})^2 + 3 \cdot 2^{-32} \cdot (1-2^{-32}) + (1-2^{-32})^2 \cdot (1-2^{-32} \cdot (1-2^{-32})^{-1}) = 1,$$

as expected. In other words, if one uses the probability $(1-2^{-32})^3$ for the last case, it follows that the probability of all the possible events is equal to $1-2^{-96}$, which is obviously wrong.

Thus, *what is the probability that* $t^2 \oplus t^3 \in \mathcal{M}_J$*?* Remember that given the events $A_1, ..., A_n$ in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$

$$Prob\Big(\bigcup_{i=1}^{n} A_i\Big) = \sum_{k=1}^{n}\Big((-1)^{k-1} \sum_{J \subset \{1,...,n\},\, |J|=k} Prob\big(\bigcap_{j\in J} A_j\big)\Big),$$

where the last sum runs over all subsets J of the indexes $1, ..., n$ which contain exactly k elements. Thus:

$$\begin{aligned} Prob(t^2 \oplus t^3 \in \mathcal{M}_J) = &\underbrace{2^{-32} \cdot 2^{-32} \cdot 1}_{1st \text{ Case}} + \underbrace{2 \cdot 2^{-32} \cdot (1-2^{-32}) \cdot 0}_{2nd \text{ Case}} + \\ &+ \underbrace{(1-2^{-32})^2 \cdot 2^{-32} \cdot (1-2^{-32})^{-1}}_{3rd \text{ Case}} = 2^{-32}. \end{aligned}$$

It follows that even if the pairs are not independent, the number of collisions is well approximated by

$$\binom{2^{32}}{2} \cdot 2^{-32} \simeq 2\,147\,483\,647.5 = 2^{31} - 2^{-1}.$$

Our practical experiments made for the small-scale AES also confirm this fact.

C Truncated Differential on 5-round AES: MDS MixColumns Matrix

In Sect. 5, we briefly explained why we require the MixColumns matrix to be MDS in order to compute the average number of collisions over 5-round AES. Here we focus on the case in which the MixColumns matrix is not MDS and it has no null coefficients.

As we have already said, consider a system of four equations of the form (12) or (15). If there exists a $r \times r$ submatrix (for $2 \leq r \leq 3$) whose determinant is equal to zero, this fact can have effects on the probability that the studied system of four equations has a common solutions. In particular, it can happen that the solutions of different equations of this system are not unrelated/independent, as instead assumed in the previous proof in order to compute the probability that different equations admit the same solutions.

To better understand this fact, we show a concrete example. Consider the MixColumns matrix $MC = circ(1, 1, 1, 2) \in (\mathbb{F}_{2^n})^{4\times 4}$ for $n \geq 4$, which is not-MDS, it is invertible and all coefficients are (obviously) different from zero. Moreover, let's focus for simplicity on 2^{2n} texts in $(\mathcal{M}_0 \cap \mathcal{C}_{0,1}) \oplus a$ (similar arguments hold also for the case of 2^{4n} texts in $\mathcal{M}_0$). As shown in Sect. 5, $p^1 \equiv \langle x^1, y^1\rangle, p^2 \equiv \langle x^2, y^2\rangle \in (\mathcal{M}_0 \cap \mathcal{C}_{0,1}) \oplus a$ satisfy $R^3(p^1) \oplus R^3(p^2) \in \mathcal{ID}_J$ for a certain J with $|J| = 3$ if and only if

$$\begin{aligned}
(R(p^1) \oplus R(p^2))_{0,0} &= \text{S-Box}(x^1 \oplus a_{0,0}) \oplus \text{S-Box}(x^2 \oplus a_{0,0}) \oplus \\
&\oplus \text{S-Box}(y^1 \oplus a_{1,1}) \oplus \text{S-Box}(y^2 \oplus a_{1,1}) = 0,\\
(R(p^1) \oplus R(p^2))_{1,1} &= \text{S-Box}(2 \cdot x^1 \oplus a_{3,0}) \oplus \text{S-Box}(2 \cdot x^2 \oplus a_{3,0}) \oplus \\
&\oplus \text{S-Box}(2 \cdot y^1 \oplus a_{0,1}) \oplus \text{S-Box}(2 \cdot y^2 \oplus a_{0,1}) = 0,\\
(R(p^1) \oplus R(p^2))_{2,2} &= \text{S-Box}(x^1 \oplus a_{2,0}) \oplus \text{S-Box}(x^2 \oplus a_{2,0}) \oplus \\
&\oplus \text{S-Box}(y^1 \oplus a_{3,1}) \oplus \text{S-Box}(y^2 \oplus a_{3,1}) = 0,\\
(R(p^1) \oplus R(p^2))_{3,3} &= \text{S-Box}(x^1 \oplus a_{1,0}) \oplus \text{S-Box}(x^2 \oplus a_{1,0}) \oplus \\
&\oplus \text{S-Box}(y^1 \oplus a_{2,1}) \oplus \text{S-Box}(y^2 \oplus a_{2,1}) = 0.
\end{aligned}$$

Focusing e.g. on the last two equations, it's very simple to observe that they can be re-written as

$$\begin{aligned}
\text{S-Box}(X^2 \oplus \Delta_X) \oplus \text{S-Box}(X^2) \oplus \text{S-Box}(Y^2 \oplus \Delta_Y) \oplus \text{S-Box}(Y^2) = 0\\
\text{S-Box}(X^3 \oplus \Delta_X) \oplus \text{S-Box}(X^3) \oplus \text{S-Box}(Y^3 \oplus \Delta_Y) \oplus \text{S-Box}(Y^3) = 0
\end{aligned}$$

where

$$X^2 = x^1 \oplus a_{2,0}, X^3 = x^1 \oplus a_{1,0}, \Delta_X = x^1 \oplus x^2, Y^2 = y^1 \oplus a_{3,1}, Y^3 = y^1 \oplus a_{2,1}, \Delta_Y = y^1 \oplus y^2.$$

The crucial point here is that Δ_X (and Δ_Y) is equal for the two equations (which does not occur for a non-MDS matrix).

Now, assume that a solution of the first/second equation is given by

$$X^2 = X', X''; \quad \Delta_X = \widehat{\Delta_X}; \quad Y^2 = Y', Y''; \quad \Delta_Y = \widehat{\Delta_Y},$$

where note that for each non-null Δ_X^I, Δ_X^O there exist an even number of solutions X of $\text{S-Box}(X \oplus \Delta_X^I) \oplus \text{S-Box}(X) = \Delta_X^O$. Focusing on the x-variables (similar for y), it follows that

- solutions of $(R(p^1) \oplus R(p^2))_{2,2} = 0$:

  $$(x_1 = X' \oplus a_{2,0}, x_2 = X' \oplus \widehat{\Delta_X} \oplus a_{2,0}) \quad \text{or} \quad (x_1 = X'' \oplus a_{2,0}, x_2 = X'' \oplus \widehat{\Delta_X} \oplus a_{2,0})$$

- solutions of $(R(p^1) \oplus R(p^2))_{3,3} = 0$:

  $$(x_1 = X' \oplus a_{3,1}, x_2 = X' \oplus \widehat{\Delta_X} \oplus a_{3,1}) \quad \text{or} \quad (x_1 = X'' \oplus a_{3,1}, x_2 = X'' \oplus \widehat{\Delta_X} \oplus a_{3,1})$$

It follows that

$$X' \oplus a_{2,0} = X' \oplus a_{3,1} \quad \textit{if and only if} \quad X' \oplus \widehat{\Delta_X} \oplus a_{2,0} = X' \oplus \widehat{\Delta_X} \oplus a_{3,1}$$

and

$$X' \oplus a_{2,0} = X'' \oplus a_{3,1} \quad \textit{if and only if} \quad X' \oplus \widehat{\Delta_X} \oplus a_{2,0} = X'' \oplus \widehat{\Delta_X} \oplus a_{3,1}$$

and so on. Assuming $x_1 \neq x_2$ (we are working in $\mathcal{M}_0 \cap \mathcal{C}_{0,1}$), the consequence of this fact is that the probability that a solution (x_1, x_2) of $(R(p^1) \oplus R(p^2))_{2,2} = 0$ is also a solution of $(R(p^1) \oplus R(p^2))_{3,3} = 0$ is (a little) higher than $2^{-n} \cdot (2^n - 1)^{-1}$ as given in the proof of Sect. 5. The same happens for the y variable.

To summarize, in the case in which the MixColumns matrix is not MDS, the solutions of one equation $(R(p^1) \oplus R(p^2))_{\cdot,\cdot} = 0$ *can be related* to the solutions of the other equations. In more details, if there is a submatrix whose determinant is zero, a particular relation among the equations *can* occur, which implies that particular relations between the solutions of different equations $(R(p^1) \oplus R(p^2))_{\cdot,\cdot} = 0$ *can* occur. As a result, the probability that a solution of $(R(p^1) \oplus R(p^2))_{\cdot,\cdot} = 0$ is also a solution of another equation of the studied system of equations can be smaller/bigger than the probability given and exploited in the proof of Sect. 5. This is not the case of an MDS MixColumns matrix, where one can assume that the solutions of different equations of the studied system of equations are unrelated/independent.

Practical Example. To provide a practical example of the previous argument, we tested the average number of collisions for small-scale AES, as described in [CMR05]. In particular, we tested it using both AES MixColumns matrix and using the MixColumns matrix $MC = circ(1, 1, 1, 2)$, as given in the previous case.

Working with texts in a coset of $\mathcal{M}_0 \cap \mathcal{C}_{0,1}$, we found that

- for AES MixColumns matrix, the *practical* average number of collisions after 1 round is 0.577 (the theoretical number of collisions for this case is 0.569 – see App. E)
- for the MixColumns matrix $MC = circ(1, 1, 1, 2)$, the *practical* average number of collisions after 1 round is 0.684

(the average has been computed over 2^{17} initial cosets). This gap is justified by the arguments given in this section.

D Computational Cost of Algorithm 1

Here we explain the details of Algorithm 1 used in Sect. 9.1 and Sect. 9.3 to count the different number of pairs of ciphertexts that belong to the same coset of $\mathcal{ID}_J$ for $|J| = 3$.

Assume the final MixColumns operation is not omitted. For each initial coset of $\mathcal{D}_I$ the two steps of the distinguisher are *(1st)* construct all the possible pairs of ciphertexts and *(2nd)* count the number of collisions. First of all, note that the major cost of this distinguisher regards the construction of all the possible different pairs, which corresponds to the first step.

The basic idea is to implement the distinguisher using a *data structure*. Assume $J \subseteq \{0, 1, 2, 3\}$ is fixed. The goal is to count the number of pairs of ciphertexts (c^1, c^2) such that $c^1 \oplus c^2 \in \mathcal{ID}_J$, or equivalently

$$c^1_{i,j-i} = c^2_{i,j-i} \qquad \forall i = 0, 1, 2, 3 \tag{26}$$

where $j = \{0, 1, 2, 3\} \setminus J$, and the index is computed modulo 4. To do this, consider an array A of 2^{32} elements completely initialized to zero. The element of A in position x for $0 \leq x \leq 2^{32} - 1$ - denoted by $A[x]$ - represents the number of ciphertexts c that satisfy the following equivalence (in the integer field $\mathbb{N}$):

$$x = c_{0,0-j} + 256 \cdot c_{1,1-j} + c_{2,2-j} \cdot 256^2 + c_{3,3-j} \cdot 256^3.$$

It's simple to observe that if two ciphertexts c^1 and c^2 satisfy (26), then they increment the same element x of the array A. It follows that given $r \geq 0$ texts that increment the same element x of the array A, then it is possible to construct $\binom{r}{2} = \frac{r \cdot (r-1)}{2}$ different pairs of texts that satisfy (26). The complete pseudo-code is given in Algorithm 1.

What is the total computational cost of this procedure? Given a set of 2^{32} (plaintexts, ciphertexts) pairs, one has first to fill the array A using the strategy just described, and then to compute the number of total of pairs of ciphertexts that satisfy the property, for a cost of $3 \cdot 2^{32} = 2^{33.6}$ table look-ups - each one of these three operations require 2^{32} table look-ups. Since one has to repeat this algorithm 4 times - i.e. one time for each $\mathcal{ID}_J$, or equivalently one time for each one of the four anti-diagonal, the total cost is of $4 \cdot 2^{33.6} = 2^{35.6}$ table look-ups, or equivalently 2^{29} five-round encryptions of AES (using the approximation[23] 20 table look-ups $\approx$ 1 round of AES).

Finally, if one has to repeat this procedure for 2^n different cosets, the total cost is given by $2^n \cdot 2^{35.6} \simeq 2^{35.6+n}$ table look-ups.

D.1 Details of the Secret-Key Distinguisher of Sect. 9.3

The same algorithm is used to implement the "mean value" secret-key distinguisher proposed in Sect. 9.3. We refer to that section for all the details, and we focus here on the details about the computational cost. As shown in Sect. 9.3, $2^{47.374}$ chosen plaintexts (i.e. $2^{15.374}$ cosets of $\mathcal{D}_I$ with $|I| = 1$) are sufficient to distinguish a random permutation from 5 rounds of AES, simply counting the number of pairs of ciphertexts that belong to the same coset of $\mathcal{ID}_J$ for $|J| = 3$ and using the fact that this number is bigger for AES. Here we give an estimation of the computational cost of the distinguisher, which is approximately given by the cost to count the number of collisions. Using Algorithm 1, the total computational cost can be well approximated by 2^{51} table look-ups, or equivalently $2^{44.34}$ five-round encryptions of AES.

In more detail, given a set of 2^{32} (plaintexts, ciphertexts) pairs, one has first to fill the array A using the strategy just described, and then to compute the number of total of pairs of ciphertexts that satisfy the property, for a cost of $3 \cdot 2^{32} = 2^{33.6}$ table look-ups - each one of these three operations require 2^{32} table look-ups. Since one has to repeat this algorithm 4 times - i.e. one time for each one of the four anti-diagonal, the total cost is of $4 \cdot 2^{33.6} = 2^{35.6}$ table look-ups, or equivalently 2^{29} five-round encryptions of AES (using the approximation 20 table look-ups $\approx$ 1 round of AES). Finally, since one has to repeat this procedure for $2^{15.374}$ different cosets, the total cost is given by $2^{15.374} \cdot 2^{35.6} \simeq 2^{51}$ table look-ups, or equivalently $2^{44.34}$ five-round encryptions of AES.

E Probability distribution for 5-round AES over $(\mathbb{F}_{2^n})^{4\times 4}$

Preliminary Considerations. First of all, given a coset of $\mathcal{C}_i$ of 2^{4n} chosen texts with one active column, we compute the number of different pairs of texts with v equal generating variables for $0 \le v \le 3$. Note that given a coset of $\mathcal{C}_i$ of 2^{4n} chosen plaintexts, one can construct $2^{4n-1} \cdot (2^{4n} - 1) \simeq 2^{8n-1}$ different pairs. Among them, the number of pairs of texts with $0 \le v \le 3$ equal generating variables (and $4 - v$ different generating variables) *after one round* is given by

$$\binom{4}{v} \cdot 2^{4n-1} \cdot (2^n - 1)^{4-v}.$$

Indeed, if v variables are equal for the two texts of the couple, then they can take $(2^n)^v$ different values. For each one of the remaining $4 - v$ variables, the variables must be different for the two texts of each couple. Thus, these $4 - v$ variables can take exactly

[23] We highlight that even if this approximation is not formally correct - the size of the table of an S-Box look-up is smaller than the size of the table used for our proposed distinguisher, it allows to give a comparison between our proposed distinguisher and the others currently present in the literature. At the same time, we note that the same approximation is largely used in the literature.

$\big[2^n \cdot (2^n-1)\big]^{4-v}/2$ different values. The result follows immediately since there are $\binom{4}{v}$ different combinations of v variables.

For the follow-up:

$$n_3 = \binom{4}{0} \cdot 2^{4n-1} \cdot (2^n-1)^4, \quad n_{n+2} = \binom{4}{1} \cdot 2^{4n-1} \cdot (2^n-1)^3, \quad n_{2n+1} = \binom{4}{2} \cdot 2^{4n-1} \cdot (2^n-1)^2.$$

E.1 Average Number of Collisions

In this section we compute *the average number of collisions for 5-round AES defined over* $(\mathbb{F}_{2^n})^{4\times 4}$. Since the idea of the proof is the same of the one given in Sect. 5, we limit ourselves to adapt it to the case of AES defined over $(\mathbb{F}_{2^n})^{4\times 4}$. Since

$$\mathcal{D}_I \oplus a' \xrightarrow[\text{prob. } 1]{R^2(\cdot)} \mathcal{M}_I \oplus b' \xrightarrow{R(\cdot)} \mathcal{D}_J \oplus a'' \xrightarrow[\text{prob. } 1]{R_f^2(\cdot)} \mathcal{ID}_J \oplus b''.$$

the idea is to work only on the middle round. That is, in the following we consider 2^{32} plaintexts in the same coset of $\mathcal{M}_i$ for $i \in \{0,1,2,3\}$ and we compute the average number of collisions after one round in the same coset of $\mathcal{D}_J$ for $|J| = 3$ fixed.

For simplicity, we limit ourselves to consider plaintexts in the same coset of $\mathcal{M}_0$ and the diagonal space $\mathcal{D}_{1,2,3}$ (the other cases are analogous). By definition of $\mathcal{M}_0$ if $p^1, p^2 \in \mathcal{M}_0 \oplus b'$, there exist $x^i, y^i, z^i, w^i \in \mathbb{F}_{2^8}$ for $i = 1,2$ such that:

$$p^i = b' \oplus \begin{bmatrix} 2\cdot x^i & y^i & z^i & 3\cdot w^i \\ x^i & y^i & 3\cdot z^i & 2\cdot w^i \\ x^i & 3\cdot y^i & 2\cdot z^i & w^i \\ 3\cdot x^i & 2\cdot y^i & z^i & w^i \end{bmatrix}$$

where $2 \equiv$ 0x02 and $3 \equiv$ 0x03. In the following we say that p^1 is "generated" by the variables (x^1, y^1, z^1, w^1) and that p^2 is "generated" by the variables (x^2, y^2, z^2, w^2) - we denote it by $p^i \equiv (x^i, y^i, z^i, w^i)$ for $i = 1,2$.

The idea is to consider separately the following cases

- 3 variables are equal, e.g. $x^1 \neq x^2$ and $y^1 = y^2$, $z^1 = z^2$, $w^1 = w^2$;
- 2 variables are equal, e.g. $x^1 \neq x^2$,$y^1 \neq y^2$ and $z^1 = z^2$, $w^1 = w^2$;
- 1 variable is equal, e.g. $x^1 \neq x^2$, $y^1 \neq y^2$, $z^1 \neq z^2$ and $w^1 = w^2$;
- all variables are different, e.g. $x^1 \neq x^2$, $y^1 \neq y^2$, $z^1 \neq z^2$, $w^1 \neq w^2$.

As we have already seen, if $y^1 = y^2$, $z^1 = z^2$ and $w^1 = w^2$, then $p^1 \oplus p^2 \in \mathcal{C}_0$, that is $R(p^1) \oplus R(p^2) \in \mathcal{M}_0$. By Theorem 5, it follows that $R(p^1) \oplus R(p^2) \notin \mathcal{D}_J$ for each J. In the following we limit ourselves to consider the case in which at least 2 generating variables are different.

Case: $2^{2\cdot n}$ Texts with Two Equal Generating Variables

As first case, we consider the case of 2^{2n} plaintexts in the same coset of $\mathcal{C}_{0,1} \cap \mathcal{M}_0$ (the other cases are equivalent). This is equivalent to consider texts with (at least) 2 equal generating variables, e.g. $x^1 \neq x^2$, $y^1 \neq y^2$, $z^1 = z^2$ and $w^1 = w^2$.

Thus, consider two plaintexts p^1 generated by $(x^1, y^1, 0, 0)$ and p^2 generated by $(x^2, y^2, 0, 0)$ in $(\mathcal{C}_{0,1} \cap \mathcal{M}_0) \oplus b'$. By simple computation, $R(p^1) \oplus R(p^2) \in \mathcal{D}_{1,2,3}$ if four equations of the form

$$\begin{aligned} A \cdot (\text{S-Box}(B \cdot x^1 \oplus a) \oplus \text{S-Box}(B \cdot x^2 \oplus a)) \oplus \\ \oplus C \cdot (\text{S-Box}(D \cdot y^1 \oplus c) \oplus \text{S-Box}(D \cdot y^2 \oplus c)) = 0 \end{aligned} \tag{27}$$

are satisfied, where A, B, C, D depend only on the MixColumns matrix definition, while a, c depend on the secret key and on the initial constant that defines the coset. Equivalently, four systems of two equations as follows must be satisfied

$$\begin{aligned} \text{S-Box}(x \oplus \Delta_I) \oplus \text{S-Box}(x) = \Delta_O \\ \text{S-Box}(y \oplus \Delta'_I) \oplus \text{S-Box}(y) = \Delta'_O \\ \Delta_O = C^{-1} \cdot A{\cdot}\Delta''_O \end{aligned} \tag{28}$$

Due to the same argumentation given in Sect. 5, the number of different (not null) solutions $[(x^1, y^1), (x^2, y^2)]$ of eq. (28) is approximately given by

$$\frac{1}{2} \cdot (2^n - 1) \cdot \left(\frac{2^n}{2^n - 1} \cdot (2^n - 1) \right)^2 = (2^n - 1) \cdot 2^{2n-1}$$

independently of the details of the S-Box. Indeed, observe that given $\Delta_O \neq 0$, each one of the two equations (28) for small-scale AES admit $\frac{2^n}{2^n-1} \cdot (2^n - 1) = 2^n$ different solutions $(\hat{x}, \Delta_I)$ - resp. $(\hat{y}, \Delta'_I)$ - where $\Delta_I, \Delta'_I \neq 0$ and $2^n/(2^n - 1)$ is the average number of solutions. Moreover, note that there are $(2^n - 1)$ values of $\Delta_O \neq 0$ and that the condition $y^1 < y^2$ holds.

Given the number of solutions of eq. (28), *what is the number of common - different (not null) - solutions* $[(x^1, y^1), (x^2, y^2)]$ *of 4 equations of the form (27)?* Due to the same argumentation given in Sect. 5, this probability is equal to $[2^n \cdot (2^n - 1)]^{-1} \cdot [(2^n - 1) \cdot 2^{n-1}]^{-1} = (2^n - 1)^{-2} \cdot 2^{-2n+1}$.

In conclusion, the number of common - different (not null) - solutions $[(x^1, y^1), (x^2, y^2)]$ of 4 equations of the form (27) is approximately given by

$$[(2^n - 1) \cdot 2^{2n-1}]^4 \cdot [(2^n - 1)^{-2} \cdot 2^{-2n+1}]^3 = \frac{2^{2n-1}}{(2^n - 1)^2}$$

Case: $2^{3 \cdot n}$ Texts with One Equal Generating Variable

As second case, we consider the case of 2^{3n} plaintexts in the same coset of $\mathcal{C}_{0,1,2} \cap \mathcal{M}_0$ (the other cases are equivalent). This is equivalent to consider texts with (at least) one equal generating variable, e.g. $x^1 \neq x^2$, $y^1 \neq y^2$, $z^1 \neq z^2$ and $w^1 = w^2$.

As before, given two plaintexts $p^1, p^2 \in (\mathcal{C}_{0,1,2} \cap \mathcal{M}_0) \oplus b'$, they belong to the same coset of the diagonal space $\mathcal{D}_{1,2,3}$ if 4 equations of the form

$$\begin{aligned} A \cdot (\text{S-Box}(B \cdot x \oplus b) \oplus \text{S-Box}(B \cdot x' \oplus b)) \oplus \\ \oplus C \cdot (\text{S-Box}(D \cdot y \oplus d) \oplus \text{S-Box}(D \cdot y' \oplus d)) \oplus \\ \oplus E \cdot (\text{S-Box}(F \cdot z \oplus f) \oplus \text{S-Box}(F \cdot z' \oplus f)) = 0 \end{aligned} \tag{29}$$

are satisfied, where A, B, C, D, E, F depend only on the MixColumns matrix definition, while b, d, f depend on the secret key and on the initial constant that defines the coset. Each one of these equations is equivalent to

$$\begin{aligned} \text{S-Box}(x \oplus \Delta_I) \oplus \text{S-Box}(x) = \Delta_O \\ \text{S-Box}(y \oplus \Delta'_I) \oplus \text{S-Box}(y) = \Delta'_O \\ \text{S-Box}(z \oplus \Delta''_I) \oplus \text{S-Box}(z) = \Delta''_O \end{aligned}$$

together with one of the two following conditions:

1. $\Delta''_O = 0$ and $\Delta'_O = C^{-1} \cdot A \cdot \Delta_O$, or analogous (3 possibilities);

2. $\Delta_O, \Delta'_O, \Delta''_O \neq 0$, and $\Delta''_O = E^{-1} \cdot (A \cdot \Delta_O \oplus C \cdot \Delta'_O)$.

First Case. The first case is analogous to the case in which two generating variables are equal. For this reason, we can re-use the same calculation as before. It follows that the average number of not null - common solutions of this case is

$$\binom{3}{1} \cdot 2^n \cdot \frac{2^{2n-1}}{(2^n-1)^2} = 3 \cdot \frac{2^{3n-1}}{(2^n-1)^2}$$

Second Case. For the second case, the idea is to work as for the cases of 1 equal generating variables. For each eq. (29) the number of different solutions $[(x^1, y^1, z^1), (x^2, y^2, z^2)]$ - where $z^1 < z^2$ - is given by $(2^n-1)\cdot(2^n-2)\cdot\frac{1}{2}\cdot\left((2^n-1)\cdot\frac{2^n}{2^n-1}\right)^3 = 2^{3n}\cdot(2^n-1)\cdot(2^{n-1}-1)$. Moreover, using the same argumentation as before, the probability to have a common solution for two equations of the form (29) is given by $[2^n \cdot (2^n-1)]^{-2} \cdot [2^{n-1} \cdot (2^n-1)]^{-1} = (2^n-1)^{-3} \cdot 2^{-3n+1}$ under the given assumptions of the S-Box. It follows that we expect on average

$$\left[2^{3n} \cdot (2^n-1) \cdot (2^{n-1}-1)\right]^4 \cdot \left[(2^n-1)^{-3} \cdot 2^{-3n+1}\right]^3 = \frac{2^{3n+3} \cdot (2^{n-1}-1)^4}{(2^n-1)^5}$$

different - not null - common solutions for the 4 equations of the form (29).

Total Number of Different (not null) Common Solutions – 2^{3n} Texts in a Coset of $\mathcal{C}_{0,1,2} \cap \mathcal{M}_0$ By simple computation, given plaintexts in the same coset of $\mathcal{C}_{0,1,2} \cap \mathcal{M}_0$, the number of different pairs of ciphertexts that belong to the same coset of $\mathcal{D}_{1,2,3}$ is approximately

$$\frac{2^{3n+3} \cdot (2^{n-1}-1)^4}{(2^n-1)^5} + 3 \cdot \frac{2^{3n-1}}{(2^n-1)^2} = \frac{2^{3n+3} \cdot (2^{n-1}-1)^4 + 3 \cdot 2^{3n-1} \cdot (2^n-1)^3}{(2^n-1)^5}$$

Generic Case: $2^{4 \cdot n}$ Texts

Finally, we consider the case of 2^{4n} texts in a coset of $\mathcal{M}_0$. This is equivalent to consider texts with (at most) all different generating variables, e.g. $x^1 \neq x^2$, $y^1 \neq y^2$, $z^1 \neq z^2$ and $w^1 \neq^2$.

As before, given two plaintexts $p^1, p^2 \in \mathcal{M}_0 \oplus b'$, they belong to the same coset of $\mathcal{D}_{1,2,3}$ if four equations of the form

$$\begin{aligned} A \cdot (\text{S-Box}(B \cdot x \oplus b) \oplus \text{S-Box}(B \cdot x' \oplus b)) \oplus \\ \oplus C \cdot (\text{S-Box}(D \cdot y \oplus d) \oplus \text{S-Box}(D \cdot y' \oplus d)) \oplus \\ \oplus E \cdot (\text{S-Box}(F \cdot z \oplus f) \oplus \text{S-Box}(F \cdot z' \oplus f)) \oplus \\ \oplus G \cdot (\text{S-Box}(H \cdot w \oplus h) \oplus \text{S-Box}(H \cdot w' \oplus h)) = 0 \end{aligned}$$

are satisfied, where A, B, C, D, E, F, G, H depend only on the MixColumns matrix definition, while b, d, f, h depend on the secret key and on the constant that defined the initial coset. Each one of these equations is equivalent to:

$$\begin{aligned} \text{S-Box}(x \oplus \Delta_I) \oplus \text{S-Box}(x) &= \Delta_O \\ \text{S-Box}(y \oplus \Delta'_I) \oplus \text{S-Box}(y) &= \Delta'_O \\ \text{S-Box}(z \oplus \Delta''_I) \oplus \text{S-Box}(z) &= \Delta''_O \\ \text{S-Box}(w \oplus \Delta'''_I) \oplus \text{S-Box}(w) &= \Delta'''_O \end{aligned}$$

together with one of the following conditions

1. $\Delta_O^{'''} = \Delta_O^{''} = 0$ and $\Delta'_O = C^{-1} \cdot A \cdot \Delta_O \neq 0$ or analogous (6 possibilities);
2. $\Delta_O^{'''} = 0$, $\Delta_O, \Delta_O^{'}, \Delta_O^{''} \neq 0$ and $\Delta_O^{''} = E^{-1} \cdot (A \cdot \Delta_O \oplus C \cdot \Delta'_O)$ or analogous (4 possibilities);
3. $\Delta_O, \Delta_O^{'}, \Delta_O^{''}, \Delta_O^{'''} \neq 0$ and $\Delta_O^{'''} = G^{-1} \cdot (A \cdot \Delta_O \oplus C \cdot \Delta'_O \oplus E \cdot \Delta_O^{''})$.

Since the first two cases are analogous to the previous two cases already studied, we can re-use the same calculation.

First Case. In the first case, $\Delta_O^{''} = 0$ implies $\Delta_I^{''} = 0$ and z can take each possible value (similar for w).

Using the same computations as before, it follows that the average number of not null - common solutions of this case is

$$\binom{4}{2} \cdot (2^n)^2 \cdot \frac{2^{2n-1}}{(2^n-1)^2} = 3 \cdot \frac{2^{4n}}{(2^n-1)^2}$$

Probability p_{2n+1}. Before going on, we compute the probability p_{2n+1}:

$$p_{2n+1} = \frac{1}{n_{2n+1}} \cdot \frac{3 \cdot 2^{4n}}{(2^n-1)^2} = \frac{1}{(2^n-1)^4}.$$

Second Case. In the second case, using the same computations as before, it follows that the average number of not null - common solutions of this case is

$$\binom{4}{1} \cdot 2^n \cdot \frac{(2^{n-1}-1)^4 \cdot 2^{3n+3}}{(2^n-1)^5} = \frac{(2^{n-1}-1)^4 \cdot 2^{4n+5}}{(2^n-1)^5}$$

Probability p_{n+2}. Before going on, we compute the probability p_{n+2}:

$$p_{n+2} = \frac{1}{n_{n+2}} \cdot \frac{(2^{n-1}-1)^4 \cdot 2^{4n+5}}{(2^n-1)^5} = \frac{(2^{n-1}-1)^4 \cdot 2^4}{(2^n-1)^8}.$$

Third Case. We finally consider the case $\Delta_O, \Delta_O^{'}, \Delta_O^{''}, \Delta_O^{'''} \neq 0$. As explained in the main text, the idea is to consider the total number of values of $(\Delta_O, \Delta'_O, \Delta_O^{''})$ that satisfy the equation $C \cdot \Delta'_O \oplus A \cdot \Delta_O \oplus E \cdot \Delta_O^{''} \neq 0$ and such that $\Delta_O \neq 0, \Delta'_O \neq 0, \Delta_O^{''} \neq 0$. By simple computation, this number is equal to $(2^n-1)^3 - (2^n-1) \cdot (2^n-2) = (2^n-1) \cdot (2^{2n} - 3 \cdot 2^n + 3)$, since $(2^n-1)^3$ is the total number of values and $(2^n-1) \cdot (2^n-2)$ is the number of values for which the previous equation is equal to 0 (note that if $C \cdot \Delta'_O \oplus A \cdot \Delta_O = 0$, then the previous equation cannot be equal zero since $\Delta_O^{''} \neq 0$). As a result, the total number of solutions for this case is

$$\frac{1}{2} \cdot (2^n-1) \cdot (2^{2n} - 3 \cdot 2^n + 3) \cdot \left[(2^n-1) \cdot \frac{2^n}{2^n-1}\right]^4 = 2^{4n-1} \cdot (2^n-1) \cdot (2^{2n} - 3 \cdot 2^n + 3)$$

Since the probability that $[(x^1, y^1, z^1, w^1), (x^2, y^2, z^2, w^2)]$ and $[(\hat{x}^1, \hat{y}^1, \hat{z}^1, \hat{w}^1), (\hat{x}^2, \hat{y}^2, \hat{z}^2, \hat{w}^2)]$ is equal to $[(2^n-1) \cdot 2^n]^{-3} \cdot [(2^n-1) \cdot 2^{n-1})^{-1} = (2^n-1)^{-4} \cdot 2^{-4n+1}$, the average number of (non null) common solutions with no equal generating variables is

$$[2^{4n-1} \cdot (2^n-1) \cdot (2^{2n} - 3 \cdot 2^n + 3)]^4 \cdot [(2^n-1)^{-4} \cdot 2^{-4n+1}]^3 = \frac{2^{4n-1} \cdot (2^{2n} - 3 \cdot 2^n + 3)^4}{(2^n-1)^8}$$

Probability p_3. Before going on, we compute the probability p_3:

$$p_3 = \frac{1}{n_3} \cdot \frac{2^{4n-1} \cdot (2^{2n} - 3 \cdot 2^n + 3)^4}{(2^n-1)^8} = \frac{(2^{2n} - 3 \cdot 2^n + 3)^4}{(2^n-1)^{12}}.$$

Total Number of Different (not null) Common Solutions – 2^{4n} Texts in a Coset of $\mathcal{M}_0$

By simple computation, given plaintexts in the same coset of $\mathcal{M}_0$, the number of different pairs of ciphertexts that belong to the same coset of $\mathcal{D}_{1,2,3}$ is approximately

$$\frac{2^{4n-1} \cdot (2^{2n} - 3 \cdot 2^n + 3)^4}{(2^n - 1)^8} + \frac{(2^{n-1} - 1)^4 \cdot 2^{4n+5}}{(2^n - 1)^5} + 3 \cdot \frac{2^{4n}}{(2^n - 1)^2}$$

E.2 Variance

As already proved, the probabilistic of 5-round AES in $(\mathbb{F}_{2^n})^{4\times 4}$ is well described by

$$\text{5-AES} = 2^3 \cdot X_3 + 2^{n+2} \cdot X_{n+2} + 2^{2n+1} \cdot X_{2n+1}$$

where X_i are binomial distributions. The pairs of texts with no equal generating variables are represented by $2^3 \cdot X_3$, the pairs of texts with 1 equal generating variable are represented by $2^{n+2} \cdot X_{n+2}$ and finally the pairs of texts with 2 equal generating variables are represented by $2^{2n+1} \cdot X_{2n+1}$. We recall that given two plaintexts with three equal generating variables, then they cannot belong to the same coset of $\mathcal{D}_J$ for $|J| \leq 3$ after one round.

Note that all the previous cases (namely, X_3, X_{n+2} and X_{2n+1}) are independent. In other words, the behavior of a pair of texts with v equal generating variables is independent of another pair with $\hat{v}$ equal generating variables where $\hat{v} \neq v$. It follows that the total variance of the probability distribution for 5-round AES case is given by

$$\begin{aligned} Var(\text{5-AES}) &= Var(2^3 \cdot X_3) + Var(2^{n+2} \cdot X_{n+2}) + Var(2^{2n+1} \cdot X_{2n+1}) = \\ &= 2^6 \cdot Var(X_3) + 2^{2n+4} \cdot Var(X_{n+2}) + 2^{4n+2} \cdot Var(X_{2n+1}), \end{aligned}$$

where $Var(\alpha \cdot X) = \alpha^2 \cdot Var(X)$.

Since X_i are binomial distribution and working in the same way proposed in Sect. 6, it follows that

- $Var(X_3)$ is equal to

$$\begin{aligned} Var(X_3) = \frac{n_3}{2^3} \cdot p_3 \cdot (1 - p_3) = 2^{4n-4} \cdot (2^n - 1)^4 \cdot \frac{(2^{2n} - 3 \cdot 2^n + 3)^4}{(2^n - 1)^{12}} \cdot \\ \cdot \left(1 - \frac{(2^{2n} - 3 \cdot 2^n + 3)^4}{(2^n - 1)^{12}}\right) \simeq \frac{2^{4n-4} \cdot (2^{2n} - 3 \cdot 2^n + 3)^4}{(2^n - 1)^8} \end{aligned}$$

- $Var(X_{n+2})$ is equal to

$$\begin{aligned} Var(X_{n+2}) = \frac{n_{n+2}}{2^{n+2}} \cdot p_{n+2} \cdot (1 - p_{n+2}) = 2^{3n-1} \cdot (2^n - 1)^3 \cdot \frac{(2^{n-1} - 1)^4 \cdot 2^4}{(2^n - 1)^8} \cdot \\ \cdot \left(1 - \frac{(2^{n-1} - 1)^4 \cdot 2^4}{(2^n - 1)^8}\right) \simeq \frac{(2^{n-1} - 1)^4 \cdot 2^{3n+3}}{(2^n - 1)^5} \end{aligned}$$

- $Var(X_{2n+1})$ is equal to

$$\begin{aligned} Var(X_{2n+1}) = \frac{n_{2n+1}}{2^{2n+1}} \cdot p_{2n+1} \cdot (1 - p_{2n+1}) = 3 \cdot 2^{2n-1} \cdot (2^n - 1)^2 \cdot \frac{1}{(2^n - 1)^4} \cdot \\ \cdot \left(1 - \frac{1}{(2^n - 1)^4}\right) \simeq \frac{3 \cdot 2^{2n-1}}{(2^n - 1)^2} \end{aligned}$$

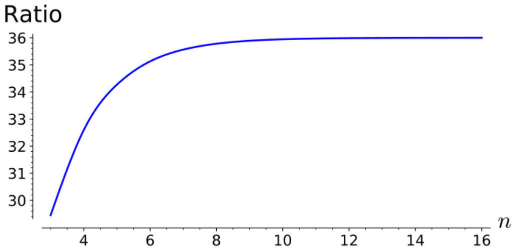

Figure 5: Ratio between the variance for 5-round AES and for the case of a random permutation given texts in $\mathbb{F}_{2^n}^{4\times 4}$ for $n \geq 3$.

By combining all previous results, it follows that $Var(5\text{-AES})$ is (approximately) equal to

$$2^6 \cdot \underbrace{\frac{2^{4n-4} \cdot (2^{2n} - 3 \cdot 2^n + 3)^4}{(2^n - 1)^8}}_{\simeq Var(X_3)} + 2^{2n+4} \cdot \underbrace{\frac{(2^{n-1} - 1)^4 \cdot 2^{3n+3}}{(2^n - 1)^5}}_{\simeq Var(X_{n+2})} + 2^{4n+2} \cdot \underbrace{\frac{3 \cdot 2^{2n-1}}{(2^n - 1)^2}}_{\simeq Var(X_{2n+1})} =$$

$$= \frac{2^{4n+2} \cdot (2^{2n} - 3 \cdot 2^n + 3)^4}{(2^n - 1)^8} + \frac{(2^{n-1} - 1)^4 \cdot 2^{5n+7}}{(2^n - 1)^5} + \frac{3 \cdot 2^{6n+1}}{(2^n - 1)^2}.$$

E.3 Observation on the Variance – Comparison with a Random Permutation

Finally, we emphasize that the ratio between the variance of 5-round AES and the one of a random permutation is (almost) constant *for each size* n *s.t.* $n \geq 8$, independently of the secret-key, of the details of the S-Box and of the MixColumns matrix – see Fig. 5.

Proposition 2. *Consider* 2^{4n} *plaintexts in* $\mathbb{F}_{2^n}^{4\times 4}$ *with one active diagonal (equivalently, a coset of a diagonal space* $\mathcal{D}_i$ *for* $i \in \{0,1,2,3\}$*), and the corresponding (cipher)texts generated by a random permutation* Π*, that is* $c^i = \Pi(p^i)$*. The probability distribution of the number of different pairs of ciphertexts* (c^i, c^j) *with* $c^i < c^j$ *that are equal in one anti-diagonal (equivalently, belong to the same coset of* $\mathcal{ID}_J$ *for* $J \subseteq \{0,1,2,3\}$ *fixed with* $|J| = 3$*) is well approximated by a binomial distribution with mean value* μ *and variance* σ^2 *given respectively by*

$$\mu = \binom{2^{4n}}{2} \times 2^{-4n} = \frac{2^{4n} - 1}{2} \quad \text{and} \quad \sigma^2 = \binom{2^{4n}}{2} \times 2^{-4n} \times (1 - 2^{-4n}) = 2^{4n-1} - 1 + 2^{-4n-1}$$

Lemma 4. *Consider* 2^{4n} *plaintexts in* $\mathbb{F}_{2^n}^{4\times 4}$ *for* $n \geq 8$ *with one active diagonal (equivalently, a coset of a diagonal space* $\mathcal{D}_i$ *for* $i \in \{0,1,2,3\}$*). The ratio between (1st) the variance* $\sigma^2_{5\text{-}AES}$ *of the probability distribution of the number of different pairs of ciphertexts* (c^i, c^j) *with* $c^i < c^j$ *that are equal in one anti-diagonal generated by 5-round AES and (2nd) the variance* σ^2_Π *of the probability distribution of the number of different pairs of ciphertexts* (c^i, c^j) *with* $c^i < c^j$ *that are equal in one anti-diagonal generated by a pseudo-random permutation* $\Pi(\cdot)$ *is (almost) constant and equal to*

$$\sigma^2_{5\text{-}AES} \approx 36 \times \sigma^2_\Pi,$$

independently of the secret-key, of the details of the S-Box and of the MixColumns matrix.

Proof. Since $\sigma^2_\Pi = 2^{4n-1} - 1 + 2^{-4n-1}$ and since $\sigma^2_{\text{5-AES}}$ is given by (23), it follows that

$$\frac{\sigma^2_{\text{5-AES}}}{\sigma^2_\Pi} = 8 \times \underbrace{\frac{1}{2^{4n} - 2 + 2^{-4n}} \times \frac{2^{4n} \cdot (2^{2n} - 3 \cdot 2^n + 3)^4}{(2^n - 1)^8}}_{\approx 1 - 4 \times 2^{-n} + \mathcal{O}(2^{-2n})} +$$
$$+16 \times \underbrace{\frac{1}{2^{4n} - 2 + 2^{-4n}} \times \frac{(2^n - 2)^4 \cdot 2^{5n}}{(2^n - 1)^5}}_{\approx 1 - 3 \times 2^{-n} + \mathcal{O}(2^{-2n})} + 12 \times \underbrace{\frac{1}{2^{4n} - 2 + 2^{-4n}} \times \frac{2^{6n}}{(2^n - 1)^2}}_{\approx 1 + 2 \times 2^{-n} + \mathcal{O}(2^{-2n})} \approx$$
$$\approx 36 - 56 \times 2^{-n} + \mathcal{O}(2^{-2n})$$

where $\mathcal{O}(\cdot)$ denotes the big O notation. As a result, the approximation $\sigma^2_{\text{5-AES}} \approx 36 \times \sigma^2_\Pi$ holds for each $n \geq 8$. □

F Key-Recovery Attacks on 5-round AES

In this section, we propose several (new) attacks on 5-round AES that exploit the secret-key distinguishers proposed in this paper revisited on 4-round AES.

Why Not an Attack on 6-round AES? To give an overview, consider the following aspect. To construct the proposed distinguishers, one consider a full coset of a subspace $\mathcal{D}_i$ for $i \in \{0, 1, 2, 3\}$ - that is, a set of 2^{32} plaintexts with one active diagonal, and exploits properties that are related to the number of ciphertexts that belong to a subspace $\mathcal{M}_J$. In order to exploit directly these distinguishers, one can guess the final key, decrypt the ciphertexts, counts the number of collisions in the same coset of $\mathcal{M}_J$ and exploits one of the proposed properties. However, since a coset of $\mathcal{M}_J$ is mapped into the full space, it seems hard to check this property one round before without guessing the entire key. Similar considerations can be done if the guessed key is the initial one. It follows that it is rather hard to set up an attack different than a brute force one that exploits directly the proposed 5-round distinguishers - this open problem is left for *future work*. For comparison, note that such a problem does not arise for the other distinguishers up to 4-round AES (e.g. the impossible differential or the integral ones), for which it is sufficient to guess only part of the secret key in order to verify if the required property is satisfied or not.

Thus, we consider round-reduced distinguishers on 4-round to propose new key-recovery attacks.

Idea of the Attack. Instead of working with 2^{32} plaintexts with one active diagonal, we consider 2^{24} texts with three active bytes in the same column, e.g. a coset of $\mathcal{D}_{0,2,3} \cap \mathcal{C}_0$. As we are going to show, the properties presented in this paper hold after 4-round in the same way. To set up the attacks, the idea is to extend the distinguishers at the beginning and to partially guess the initial key. In more details, consider 2^{32} plaintexts in $\mathcal{D}_0 \oplus a$. After one round, they are mapped into a coset of $\mathcal{C}_0$ with prob. 1. However, the way in which they are divided in cosets of $\mathcal{D}_{0,2,3} \cap \mathcal{C}_0$ depends on the guessed key

$$2^{32} \text{ plaintexts in } \mathcal{D}_0 \oplus b \xrightarrow[\text{(partially) key guess}]{R(\cdot)} 2^{24} \text{ texts in } \mathcal{D}_{0,2,3} \cap \mathcal{C}_0 \oplus a \xrightarrow{R^4(\cdot)} \ldots$$
$$\ldots \xrightarrow{R(\cdot)} 2^{24} \text{ texts in } \mathcal{D}_{0,2,3} \cap \mathcal{C}_0 \oplus a \xrightarrow{R^4(\cdot)} \text{distinguisher property.}$$

We exploit this fact to set up new key-recovery attacks on 5-round AES.

In more details, the attacks that we are going to present are based on the following properties:

Table 4: *Comparison of attacks on 5-round AES-128.* Data complexity is measured in number of required chosen plaintexts/ciphertexts (CP/CC). Time complexity is measured in round-reduced AES encryption equivalents (E) - the number in the brackets denotes the precomputation cost (if not negligible). Memory complexity is measured in texts (16 bytes). Attacks presented in this paper are in bold.

Attack	Rounds	Data	Computation	Memory	Ref.
MitM	5	8	2^{64}	2^{56}	[Der13, Sec. 7.5.1]
Imp. Polytopic	5	15	2^{70}	2^{41}	[Tie16]
Partial Sum	5	2^{8}	2^{38}	small	[Tun12]
Integral (EE)	5	2^{11}	$2^{45.7}$	small	[DR02]
Mixture Diff.	5	$2^{22.25}$	$2^{22.5}$	2^{20}	[BDK+18]
Imp. Differential	5	$2^{31.5}$	2^{33} $(+2^{38})$	2^{38}	[BK01]
Integral (EB)	5	2^{33}	$2^{37.7}$	2^{32}	[DR02]
Variance	**5**	$\mathbf{2^{33}}$	$\mathbf{2^{64.2}}$	$\mathbf{2^{32}}$	**App. F.4**
Mixture Diff.	5	$2^{33.6}$	$2^{33.3}$	2^{34}	[Gra18]
Multiple-of-n	**5**	$\mathbf{2^{33.6}}$	$\mathbf{2^{48}}$	$\mathbf{2^{32}}$	**App. F.2**
Trunc. Diff.	**5**	$\mathbf{2^{35}}$	$\mathbf{2^{69.2}}$	$\mathbf{2^{32}}$	**App. F.3**

MitM: Meet-in-the-Middle, EE: Extension at End, EB: Extension at Beginning

- the number of collisions is a multiple of 2/4/8;
- the average number of collisions is (a little) bigger for AES than for a random permutation;
- the variance of the number of collisions is higher for AES than for a random permutation.

In the following, we first present the generic strategy to set up these attacks (which is common for all the previous cases), and then we give all the details. The results are summarized in Table 4.

F.1 Generic Strategy

In order to exploit one of the previous properties, the idea is the following. Consider 2^{24} texts in the same coset of $\mathcal{D}_I \cap \mathcal{C}_j$ for $|j| = 1$ and $|I| = 2$ or $|I| = 3$, e.g.

$$\mathcal{D}_{0,2,3} \cap \mathcal{C}_0 \oplus a \equiv \begin{bmatrix} A & C & C & C \\ A & C & C & C \\ A & C & C & C \\ C & C & C & C \end{bmatrix},$$

and the corresponding ciphertexts after 4-round. The idea of the attack is to guess 4 bytes of the key (i.e. the j-th diagonal), to partially compute 1-round decryption of $\mathcal{D}_I \cap \mathcal{C}_j \oplus a$ and to ask for the corresponding ciphertexts after 5-round. Exploiting one of the previous properties that hold on the ciphertexts only if the guessed key is the right one, it is possible to filter wrong keys and to find the right one. In particular, this is due to the fact that if the guessed key is not the right one, the behavior is the same of a random permutation - *Wrong-Key Randomization Hypothesis*.

In more details, consider $2^{24 \cdot n}$ texts in n cosets of $\mathcal{D}_I \cap \mathcal{C}_j$. The idea is to compute 1-round decryption with respect to a guessed key and ask for the corresponding ciphertexts. The following properties holds

- *the number of collisions is always a multiple of 2 if* $|I| = 2$ *and of 4 if* $|I| = 3$ *for the right key*, while it can assume any value for a wrong guessed key;
- *the average number of collisions in the same coset of* $\mathcal{M}_J$ *for* J *fixed with* $|J| = 3$ *is approximately equal to* $32\,770.524$ *for the right key*, while it is approximately $32\,767.998$ for a wrong guessed key;
- *the variance of the number of collisions is approximately equal to* $2^{17.8}$ *for the right key*, while it is approximately 2^{15} for a wrong guessed key.

Note that if $n \leq 2^8$ initial cosets are sufficient to set up the attack, then the data cost of this step is at most of 2^{32} chosen plaintexts in the same coset of $\mathcal{D}_i$, since $\mathcal{D}_I \cap \mathcal{C}_j \oplus b \subseteq \mathcal{C}_j \oplus b = R(\mathcal{D}_i \oplus a)$. When one diagonal of the key is found, the other ones can be found using the same strategy or by brute force.

F.1.1 Wrong-Key Randomization Hypothesis

One assumption of the attack is the wrong-key randomization hypothesis. This hypothesis states that decrypting one or several rounds with a wrong key guess creates a function that behaves like a random one. This assumption is very common and used for classical/truncated/impossible differentials key-recovery attacks.

For this reason, we limit ourselves to show that it holds also in our case. Consider 2^{24} texts t^i in a coset of $\mathcal{D}_{0,2,3} \cap \mathcal{C}_0$ for $i = 0, ..., 2^{24} - 1$, and let k the secret subkey and $\hat{k}$ the guessed key. The decryption under the guessed key $\hat{k}$ is simply given by:

$$R_{\hat{k}}^{-1}(t^i) = \hat{k} \oplus \text{ S-Box}^{-1} \circ SR^{-1} \circ MC^{-1}(t^i).$$

To implement the attack, one asks the corresponding ciphertexts after 5-round (with respect to the right key k). By simple computation, after one round

$$R_k \circ R_{\hat{k}}^{-1}(t^i) = MC \circ SR \circ \text{ S-Box} \left[\hat{k} \oplus k \oplus \text{ S-Box}^{-1} \circ SR^{-1} \circ MC^{-1}\left(t^i\right)\right].$$

Thus, if $\hat{k} = k$, then $R_k \circ R_{\hat{k}}^{-1}(t^i) = t^i$ for each i, and the distinguisher property holds. On the other hand, if $\hat{k} \neq k$, then $R_k \circ R_{\hat{k}}^{-1}(t^i) \neq t^i$ for each i since the S-Box is a non-linear operation. It follows that $\{R_k \circ R_{\hat{k}}^{-1}(t^i)\}_i$ do not belong to the same coset of $\mathcal{D}_{0,2,3} \cap \mathcal{C}_0$, and the distinguisher property does not work. In this case, the behavior is the same of a random permutation, and the attacker can filter wrong keys.

F.1.2 Implementation Strategy

In the following we give the details of the attack. We highlight that in all cases the attacker has to count the number of collisions in the same coset of $\mathcal{M}_J$ in order to filter wrong keys. Even if it is possible to use the strategy proposed in Algorithm 1, another strategy is more competitive in this case.

The basic idea is to re-order the texts with respect to a partial order $\preceq$ and to work only on consecutive ordered texts. In particular, since our goal is to check if two texts belong to the same coset of $\mathcal{M}_J$ for $|J| = 3$, the idea is to re-order the texts using a particular numerical order which depends by J. Then, given a set of ordered texts, the idea is to work only on two consecutive elements in order to count the total number of collisions. In other words, given ordered ciphertexts, one can work only on approximately 2^{32} different pairs (composed of consecutive elements with respect to the used order) instead of 2^{63} for each coset of $\mathcal{D}_I$. For this reason, we define the following *partial order* $\preceq$:

Definition 7. Let $I \subset \{0,1,2,3\}$ with $|I| = 3$ and let $l \in \{0,1,2,3\} \setminus I$. Let $t^1, t^2 \in \mathbb{F}_{2^8}^{4\times 4}$ with $t^1 \neq t^2$. The text t^1 is less or equal than the text t^2 with respect to the partial order $\preceq$ (i.e. $t^1 \preceq t^2$) if and only if one of the two following conditions is satisfied (the indexes are taken modulo 4):

- there exists $j \in \{0,1,2,3\}$ such that for all $i < j$:

$$MC^{-1}(t^1)_{i,l-i} = MC^{-1}(t^2)_{i,l-i} \qquad \text{and} \qquad MC^{-1}(t^1)_{j,l-j} < MC^{-1}(t^2)_{j,l-j};$$

- for all $i = 0, ..., 3$:

$$MC^{-1}(t^1)_{i,l-i} = MC^{-1}(t^2)_{i,l-i} \qquad \text{and} \qquad MC^{-1}(t^1) \leq MC^{-1}(t^2),$$

  where $\leq$ defined as in Def. 1.

Thus, as first step, one must re-order the 2^{32} ciphertexts of each coset with respect to the partial order relationship $\preceq$ defined before.

After the re-ordering process, in order to count the number of pairs of texts that belong to the same coset of $\mathcal{M}_J$, one can work only on consecutive ordered elements. Indeed, consider r consecutive elements $c^l, c^{l+1}, ..., c^{l+r-1}$, with $r \geq 2$. Suppose that for each k with $l \leq k \leq l+r-2$: $c^k \oplus c^{k+1} \in \mathcal{M}_J$. Since $\mathcal{M}_J$ is a subspace, it follows immediately that for each s, t with $l \leq s, t \leq l + r - 2$ $c^s \oplus c^t \in \mathcal{M}_J$. Thus, given $r \geq 2$ consecutive elements that belong to the same coset of $\mathcal{M}_J$, it follows that $\binom{r}{2} = \frac{r\cdot(r-1)}{2}$ different pairs belong to the same coset of $\mathcal{M}_J$. In the same way, consider r consecutive elements $c^l, c^{l+1}, ..., c^{l+r-1}$ with $r \geq 2$, such that $c^k \oplus c^{k+1} \notin \mathcal{M}_J$ for each$k$ with $l \leq k \leq l + r - 2$. Since $\mathcal{M}_J$ is a subspace, it follows immediately that $c^s \oplus c^t \notin \mathcal{M}_J$ for each s, t with $l \leq s, t \leq l + r - 2$.

In other words, thanks to the ordering algorithm, it is possible to work only on $2^{32} - 1$ pairs (i.e. the pairs composed of two consecutive elements), but at the same time to have information on all the $2^{31} \cdot (2^{32} - 1) \simeq 2^{63}$ different pairs. The pseudo-code of such algorithm is given in Algorithm 2.

What is the total computational cost of this procedure? Given a set of n ordered elements, the computational cost to count the number of pairs that belong to the same coset of $\mathcal{M}_J$ is well approximated by n look-ups table, since one works only on consecutive elements. Using the *merge sort* algorithm to order this set (which has a computational cost of $O(n \log n)$ memory access), the total computational cost for the verifier is approximately of $n \cdot (1 + \log n)$ table look-ups. In our case, since the verifier has to consider a single coset of $\mathcal{D}_I$ of 2^{32} elements and to repeat this procedure four times (i.e. one for each $\mathcal{M}_J$ with $|J| = 3$), the cost is well approximated by $4 \cdot 2^{32} \cdot (1 + \log 2^{32}) = 2^{39}$ table look-ups, or equivalently $2^{32.4}$ five-round encryptions of AES (using the approximation 20 table look-ups $\approx$ 1 round of AES).

F.1.3 Practical Tests on small-scale AES

All the attacks that we are going to present have been practically tested on small-scale AES. The practical results are in accordance with the theoretical ones.

F.2 Multiple-of-n Key-Recovery Attack

Consider 2^{16} plaintexts in the same coset of $\mathcal{D}_I \cap \mathcal{C}_j$ for $|j| = 1$ and $|I| = 2$ - e.g. $\mathcal{D}_{0,1} \cap \mathcal{C}_0$, and the corresponding ciphertexts after 4-round. As proved in [GRR17a], the number of different pairs of ciphertexts that belong to the same coset of $\mathcal{M}_K$ for $|K| = 3$ is always a multiple of 2 (or 4 if $|I| = 3$), while it can take any possible value for a random permutation.

The idea of the attack is to guess 4 bytes of the key (i.e. the j-th column), to partially decrypt $\mathcal{D}_I \cap \mathcal{C}_j$ and to ask for the corresponding ciphertexts. Since for a wrong key, the

```
Data: $2^{32}$ (plaintext, ciphertext) pairs $(p^i, c^i)$ for $i = 0, ..., 2^{32} - 1$ in a single coset
      of $\mathcal{D}_I$ with $|I| = 1$.
Result: Number of collisions $n$
for all $J$ with $|J| = 3$ do
    Re-order the $2^{32}$ (plaintexts, ciphertexts) pairs using the partial order
      relationship $\preceq$ defined in Def. 7;                    // $\preceq$ depends on $J$
    Let $(\tilde{p}^i, \tilde{c}^i)$ for $i = 0, ..., 2^{32} - 1$ the order (plaintext, ciphertext) pairs;
    $n \leftarrow 0$;                 // $n$ denotes the number of collisions in $\mathcal{M}_J$
    $i \leftarrow 0$;
    while $i < 2^{32}$ do
        $r \leftarrow 1$;
        $j \leftarrow i$;
        while $\tilde{c}^j \oplus \tilde{c}^{j+1} \in \mathcal{M}_J$ do
            $r \leftarrow r + 1$;
            $j \leftarrow j + 1$;
        end
        $i \leftarrow j + 1$;
        $n \leftarrow n + r \cdot (r - 1)/2$;
    end
end
return $n$.
```

Algorithm 2: *Goal of the Algorithm is to count the number of collisions.*

behavior is similar to the one of a random permutation - the number of collisions is not a multiple of 2 with prob. 1, it is possible to filter wrong keys and to find the right one.

What is the data complexity?

Data Cost. Given a single coset of $\mathcal{D}_I \cap \mathcal{C}_j$, the probability that the number of collisions is a multiple of 2 is 1/2 for a wrong key. Thus, the probability that a wrong key survives n tests is 2^{-n}. Since there are 2^{32} different keys to test, $n \geq 32$ tests are sufficient to filter all the wrong keys with high probability. Since a coset of $\mathcal{C}_j$ contains 2^{16} different cosets of $\mathcal{D}_I \cap \mathcal{C}_j$, it follows that 2^{32} chosen plaintexts in the same coset of $\mathcal{D}_j$ are sufficient to find one diagonal (remember that a diagonal subspace $\mathcal{D}_j$ is mapped into a column one $\mathcal{C}_j$ after one round). Using this strategy to find three diagonals of the key (one diagonal is found by brute force), the data complexity is of $2^{33.6}$ chosen plaintexts.

Computational Cost. Using Algorithm 2, the computational cost of the attack is well approximated by the cost of the re-ordering step for each possible key. In particular, in order to find one diagonal of the key, the cost can be approximated by $2^{32} \cdot 2^{16} \cdot (2 + \log 2^{16}) \cdot (1 + 1/2 + 1/4 + 1/8 + ...) \simeq 2^{53.1}$ table look-ups. Thus, the total cost is $3 \cdot 2^{53.1} \cdot (5 \cdot 20)^{-1} + 2^{32} \simeq 2^{48}$ five-round encryption to find the entire key (by assuming 20 table look-ups $\approx$ 1 encryption). The term $1 + 1/2 + 1/4 + 1/8 + ...$ is due to the fact that after the 1st test only 1/2 of the possible keys survived, after the 2nd test only 1/4 of the possible keys survived and so on. Indeed, note that the number of collisions is a multiple of 2 only with probability 1/2. In other words, after the 1st test one repeats the process for $2^{32}/2 \simeq 2^{31}$ keys, after the 2nd test one repeats the process for $2^{32}/4 \simeq 2^{30}$ keys and so on. This result has been checked also by practical tests.

F.3 Truncated Diff. Attack based on the *Mean*

Here we exploit the fact the average number of collisions is (a little) bigger for the right key than for a wrong guessed key, i.e. *we propose the first truncated differential attack on*

5-round AES (that exploits a differential trail with probability different from zero).

Consider 2^{24} plaintexts in the same coset of $\mathcal{D}_I \cap \mathcal{C}_j$ for $|j| = 1$ and $|I| = 3$ - e.g. $\mathcal{D}_{0,1,2} \cap \mathcal{C}_0$, and the corresponding ciphertexts after 4-round. As we have just seen in Sect. 5, the average number of different pairs of ciphertexts that belong to the same coset of $\mathcal{M}_K$ for $|K| = 3$ is approximately 32 770.524 versus 32 767.998 in the random case. In other words, the probability that a pair of ciphertexts belongs to the same coset of $\mathcal{M}_K$ for $|K| = 3$ is $2^{-32} + 2^{-45.6625}$ for AES versus 2^{-32} for the random case/wrong guessed key.

The idea of the attack is to guess 4 bytes of the key (i.e. the j-th diagonal), to partially decrypt $\mathcal{D}_I \cap \mathcal{C}_j$ and to ask for the corresponding ciphertexts. Exploiting the previous property that holds on the ciphertexts, it is possible to filter wrong keys and to find the right one. We expect that the number of collisions is bigger for the right key of AES than for a wrong one. Indeed, if the key is wrong, then the texts are distributed in several cosets of $\mathcal{D}_I \cap \mathcal{C}_j$ after one round (not in only one), and one gets the same behavior that occurs for a random permutation. In particular, we emphasize that our truncated differential distinguisher proposed in this paper works if and only if one consider an entire initial coset of $\mathcal{D}_I \cap \mathcal{C}_j$.

What is the data complexity? To compute the data cost of the attack, we use the same strategy proposed for the 5-round secret-key distinguisher.

Data Cost. Assume that the goal is to find the right key with probability bigger than 95%[24], and assume that the behavior for a wrong guessed key is the same of a random permutation. Since one works on 4 bytes of the key, one has to use the secret-key distinguisher $4 \cdot 2^{32} = 2^{34}$ different times. In other words, the data cost is approximately given by formula (25) where $prob = 0.95^{1/2^{34}}$. It follows that for $p_{rand} \simeq 2^{-30} - 3 \cdot 2^{-63}$ and $p_{AES} \simeq 2^{-30} + 2^{-43.6625}$, the number of different pairs that one needs to use in order to set up the attack is $n \geq 3 \cdot 2^{59.43}$ (where the factor 3 is due to the observations given in Sect. 9.3). Since there are 4 different subspace $\mathcal{D}_I \cap \mathcal{C}_j$ and since each coset of $\mathcal{D}_I \cap \mathcal{C}_j$ contains approximately $\binom{2^{24}}{2} \simeq 2^{47}$ different pairs after one round, one needs approximately $2^{12.02}$ different initial cosets or approximately $2^{34.02}$ chosen plaintexts in the same coset of $\mathcal{D}_j$ in order to find one diagonal of the key. If two diagonals are found by brute force, the cost of finding the entire key is of $2 \cdot 2^{34.02} = 2^{35}$ chosen plaintexts.

Computational Cost. Using Algorithm 2, the computational cost of the attack is well approximated by the cost of the re-ordering step for each possible key. In particular, in order to find one diagonal, the cost can be approximated by $4 \cdot 2^{12.02} \cdot 2^{32} \cdot 2^{24} \cdot (2 + \log 2^{24}) \simeq 2^{74.7}$ table look-ups. Thus the total cost is $2 \cdot 2^{74.7} \cdot (5 \cdot 20)^{-1} + 2^{64} \simeq 2^{69.2}$ five-round encryption to find the entire key (by assuming 20 table look-ups ≈ 1 encryption).

F.4 Truncated Diff. Attack based on the *Variance*

In this subsection, we exploit the fact the variance is higher for the right key than for a wrong guessed key.

Consider 2^{24} plaintexts in the same coset of $\mathcal{D}_I \cap \mathcal{C}_j$ for $|j| = 1$ and $|I| = 3$ - e.g. $\mathcal{D}_{0,1,2} \cap \mathcal{C}_0$, and the corresponding ciphertexts after 4-round. What is the variance of the number of collisions in the same coset of $\mathcal{M}_K$ for $|K| = 3$ after 4 rounds? To compute a good approximation of the variance, we re-use the same calculation proposed in Sect. 6. For this reason, we refer to that section for all the details and we give here only the final result.

Assume K fixed. For a wrong guessed key, the variance is well approximated by

$$Var_{wrongKey} = 2^{23} \cdot (2^{24} - 1) \cdot 2^{-32} \cdot (1 - 2^{-32}) \simeq 2^{15},$$

[24] In other words, we assume that the maximum number of collisions occurs for the right key with probability 95%.

that is the standard deviation is equal to $\delta_{wrongKey} = 2^{7.5}$. What about right key guessed? Given 2^{24} plaintexts, there are $3 \cdot 2^{23} \cdot (2^8 - 1)^2 = 2^{40.58}$ different pairs with one equal generating variable and $2^{23} \cdot (2^8 - 1)^3 = 2^{46.99}$ different pairs with different generating variables. The variance is given by

$$\begin{aligned} Var_{rightKey} = {} & 4^2 \cdot 2^{44.99} \cdot (2^{-32} - 2^{-45.6625}) \cdot (1 - 2^{-32} + 2^{-45.6625}) + \\ & + (2^9)^2 \cdot 2^{30.58} \cdot (2^{-32} - 2^{-45.6625}) \cdot (1 - 2^{-32} + 2^{-45.6625}) \simeq 2^{17.8}, \end{aligned}$$

that is the standard deviation is equal to $\delta_{rightKey} = 2^{8.9}$. This difference can be exploited to find the right key. In order to derive concrete number for data and computational complexity, as for the secrete-key distinguisher, we consider the results on small-scale AES.

For small-scale AES - denoted in the following by symbol *, consider as before 2^{12} plaintexts in the same coset of $\mathcal{D}_I \cap \mathcal{C}_j$ for $|j| = 1$ and $|I| = 3$ and assume K fixed. For a wrong guessed key, the variance is well approximated by

$$Var^\star_{wrongKey} = 2^{11} \cdot (2^{12} - 1) \cdot 2^{-16} \cdot (1 - 2^{-16}) \simeq 2^7,$$

that is the standard deviation is equal to $\delta^\star_{wrongKey} = 2^{3.5}$. What about right key guessed? Given 2^{12} plaintexts, there are $3 \cdot 2^{11} \cdot (2^4 - 1)^2 = 2^{20.4}$ different pairs with one equal generating variable and $2^{11} \cdot (2^4 - 1)^3 = 2^{22.7}$ different pairs with different generating variables. The variance is given by

$$\begin{aligned} Var^\star_{rightKey} = {} & 4^2 \cdot 2^{20.7} \cdot (2^{-16} - 2^{-24.67}) \cdot (1 - 2^{-16} + 2^{-24.67}) + \\ & + (2^5)^2 \cdot 2^{15.4} \cdot (2^{-16} - 2^{-24.67}) \cdot (1 - 2^{-16} + 2^{-24.67}) \simeq 2^{10.1}, \end{aligned}$$

that is the standard deviation is equal to $\delta^\star_{rightKey} = 2^{5.05}$.

Data and Computational Costs. As for the secret-key distinguisher of Sect. 9.1, the ratio between the standard deviation is similar for the small scale AES and full-size AES

$$\frac{2^{8.9}}{2^{7.5}} \approx 2.75 \approx \frac{2^{5.05}}{2^{3.5}}.$$

Thus, we use our results on small-scale AES to derive concrete numbers for the full-size AES case. By practical tests, we have found that $\geq 2^6$ initial cosets are sufficient to have a good estimation of the variance/standard deviation. Since for each initial coset it is possible to compute the number of collisions in $\mathcal{M}_J$ for each J with $|J| = 3$, at least 2^6 initial cosets are largely sufficient to set up the distinguisher. Due to the relation between small-scale AES and full-size AES previously discussed, we claim that the data cost to distinguish to find one diagonal of the key is of 2^{32} chosen plaintexts in the same coset of $\mathcal{D}_j$ (observe that after one round, it contains $4 \cdot 2^8$ different cosets of $\mathcal{D}_I \cap \mathcal{C}_j$). If two diagonals are found by brute force, the total data cost is well approximated by 2^{33} chosen plaintexts.

The computational cost is well approximated by the cost to compute the number of collisions for each possible key. Using Algorithm 2, the cost of finding one diagonal is well approximated by $2^{32} \cdot 2^6 \cdot 2^{24} \cdot (2 + \log 2^{24}) \simeq 2^{66.7}$ table look-ups, that is the total cost is well approximated by $2 \cdot 2^{66.7} \cdot (100)^{-1} + 2^{64} \simeq 2^{64.2}$ five-round encryption to find the entire key by assuming 20 table look-ups $\approx$ 1 encryption.

G Details of used S-Box

In this section, we recall the main information of the S-Box used in Sect. 10 to test our theory.

Table 5: *S-Box definitions.* All the values in the table are exadecimal.

x	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
AES	6	B	5	4	2	E	7	A	9	D	F	C	3	1	0	8
PRINCE	B	F	3	2	A	C	9	1	6	7	8	0	E	5	D	4
KLEIN	7	4	A	9	1	F	B	0	C	3	2	6	8	E	D	5
Midori	C	A	D	3	E	B	F	7	8	9	1	5	0	2	4	6
Midori	1	0	5	3	E	2	F	7	D	A	9	B	C	8	4	6
PRESENT	C	5	6	B	9	0	A	D	3	E	F	8	4	7	1	2
RECTANGLE	6	5	C	A	1	E	7	9	B	0	3	D	8	F	4	2
NOEKEON	7	A	2	C	4	8	F	0	5	9	1	e	3	D	B	6
PRIDE	0	4	8	F	1	5	E	9	2	7	A	C	B	D	6	3
Toy-6 S-Box	1	3	6	4	2	5	9	A	0	F	7	E	C	B	D	8
Toy-8 S-Box	1	3	6	4	2	5	A	C	0	F	7	8	E	B	D	9
Toy-10 S-Box	6	4	C	5	0	7	2	E	1	F	3	D	8	A	9	B
Toy-12 S-Box	1	3	6	4	A	F	D	8	2	9	5	E	0	7	B	C

In the following we recall the *differential spectrum* of the S-Box, that is probability that given an arbitrary $\Delta_I \neq 0$ and $\Delta_O \neq 0$, the equation

$$\text{S-Box}(x \oplus \Delta_{IN}) \oplus \text{S-Box}(x) = \Delta_{OUT}$$

has n different solutions x (remember n is even).

S-Box	**0 sol.**	**2 sol.**	**4 sol.**	**6 sol.**	**8 sol.**	**10 sol.**	**12 sol.**
AES	8/15	6/15	1/15	0	0	0	0
PRINCE	8/15	6/15	1/15	0	0	0	0
KLEIN	8/15	6/15	1/15	0	0	0	0
MIDORI SB_1	8/15	6/15	1/15	0	0	0	0
MIDORI SB_0	43/75	24/75	8/75	0	0	0	0
PRESENT	43/75	24/75	8/75	0	0	0	0
RECTANGLE	43/75	24/75	8/75	0	0	0	0
NOEKEON	43/75	24/75	8/75	0	0	0	0
PRIDE	43/75	24/75	8/75	0	0	0	0
Toy-6 S-Box	125/225	81/225	18/225	1/225	0	0	0
Toy-8 S-Box	130/225	74/225	18/225	2/225	1/225	0	0
Toy-10 S-Box	140/225	60/225	17/225	7/225	0	1/225	0
Toy-12 S-Box	162/225	24/225	27/225	8/225	3/225	0	1/225