New 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 that is so versatile, and report on new properties. Classically, in differential cryptanalysis, statements about the probability distribution of output differences, like mean or variance, are of interest. So far such statements where only possible for up to 4 rounds of AES. In this paper we consider the probabilistic distribution of the number of different pairs of corresponding ciphertexts that lie in certain subspaces after 5 rounds. We rigorously prove that the following two properties (independent of any key or constant additions) hold for 5 rounds of the AES permutation:

- the mean value is bigger for AES than for a random permutation;
- the variance is approximately by a factor 36 higher for AES than for a random permutation.

While the distinguisher based on the variance is (almost) independent of the details of the S-Box and of the MixColumns matrix, the mean value distinguisher does depend on the details of the S-Box and may give rise to a new design criterion for S-Boxes.
Of independent interest is the technique that we developed for this rigorous analysis. To the best of our knowledge this seems to be the first time that such a precise differential analysis was performed. Practical implementations and verification confirm our analysis.

Keywords: AES, Truncated-Differential Cryptanalysis, Distinguisher/Attack

Table of Contents

1 Introduction

AES (Advanced Encryption Standard) [15] 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 [15], so pure versions of these techniques have limited applications in attacks.

Since its conception by Biham and Shamir [7] 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 [22] - where only part of the difference between pairs of texts is considered, impossible differentials [5] - where differences with zero-probability are exploited, and higher-order differentials [22]. Differential cryptanalysis can be used to set up secret-key distinguisher and key-recovery attack. In a secret-key distinguisher, there are two oracles: one that simulates the cipher for which the key has been chosen at random and one that simulates a truly random permutation. The adversary can query both oracles and her task is to decide which oracle is the cipher. Secret-key distinguishers which are independent of the secret-key are usually starting points for key-recovery attacks.

As is state of the art, truncated differential distinguishers which are *independent* of the secret key - that exploits difference 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.

1.1 New (Truncated) Differential Distinguishers for 5-round AES

Recently at Eurocrypt 2017, a new property which is *independent* of the secret key has been found for 5-round AES [20]. 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] $\mathcal{M}$ is always a multiple of 8. Such distinguisher has then been exploited in e.g. [19] to set up new key-recovery attacks on 6-round AES. Later, at Asiacrypt 2017, Rønjom, Bardeh and Helleseth [26] presented new secret-key distinguishers for 3- to 6-round AES, which are based on the "yoyo-game", i.e. they require adaptively chosen ciphertexts in addition to chosen plaintexts.

Several open questions arise from the result provided in [20]: 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,

[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. *(Theoretical) 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 lie in a particular subspace $\mathcal{M}_I$ for $I \subseteq \{0,1,2,3\}$ fixed with $|I| = 3$ - as defined in Def. 4. *Accurate theoretical expected values mean and variance* of this distribution is given in this table for 5-round AES and for a random permutation. For this result, we need to assume that the S-Box is an APN permutation (practical results on AES - with the real non-APN S-Box - are close and are discussed in Sect. 6).

	Random Permutation	**5-round AES (APN S-Box)**
Mean (Sect. 3-4)	$2\,147\,483\,647.5 \approx 2^{31}$	$2\,147\,484\,685.6 \approx 2^{31} + 2^{10}$
Variance (Sect. 5)	$2\,147\,483\,647 \approx 2^{31}$	$76\,435\,327\,505.945 \approx 2^{36.155}$
Multiple-of-8 [20]		✓

...) affected by the multiple-of-8 property?

Diagonal Set of Plaintexts. In this paper, given a diagonal set of plaintexts - i.e. a set of plaintexts with one active diagonal, we consider the probabilistic distribution of the corresponding number of pairs of ciphertexts after 5-round AES that belong to the same coset of a particular subspace $\mathcal{M}$.

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 [13], or that each pair of ciphertexts can not be equal in any of the four anti-diagonal (as showed by Biham and Keller in [6]). For the first time, here we perform and propose a *precise theoretical differential analysis* of such distribution after 5-round AES, supported by practical implementations and verification. A numerical summary is given in Table 1.

5-round AES - Truncated Differential Distinguisher based on the Mean. As a first result, we present in Sect. 3 an analysis of the mean and formulate it as a new truncated differential distinguisher based on the mean which is independent of the secret-key. In more detail, by appropriate choice of sets of texts, we prove for the first time that *the number of times that the difference of the resulting output pairs lie in a particular subspace is (a little) bigger for 5-round AES than for a random permutation, independently of the secret key. 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.

5-round AES - (First) Truncated Differential Distinguisher based on the Variance. As a second contribution, *we theoretically compute the vari-*

ance of the probabilistic 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. Such property - whose proof is based on the result proposed at Eurocrypt 2017 - is independent of the secret key. As a result, we propose *for the first time* in Sect. 5 *a differential distinguisher that exploits the variance parameter* - and not the mean value (usually used in the literature) - in order *to distinguish AES-like cipher from a random permutation.*

Main Difference with Other Truncated Diff. Distinguishers in the Literature. Before going on, it is important to remark a crucial difference between our distinguishers and the other currently present in the literature 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 belong or not to a given subspace. In order to set up our result 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 - i.e. a full coset of a particular subspace - to appreciate a difference in the probability of the previous event. More details are given in the following.

1.2 Potential Impact of Our Results

Truncated Differential Distinguisher. Until now, truncated-differential-style attacks (and many others) rely only on the mean value to distinguish a cipher from a random permutation. To the best of our knowledge, our variance (differential) distinguisher is the first case in which the variance value is used to distinguish a random permutation from a cipher, giving in substantially better results compared to using the mean value. *As a future work, it would be interesting to see if this more direct use of the variance can find applications and give improvements elsewhere in cryptanalysis.*

Moreover, our theoretical analysis proposed in this paper is general enough to be applied to any AES-like cipher, allowing to improve/extend (almost) all the truncated differential distinguishers based on the mean for AES-like ciphers - which are independent of the key - currently present in the literature.

Round-Reduced AES as part of New Designs. Since reduced versions of AES have nice and well-studied properties, many constructions employ round-reduced AES as part of their design.

Only to cite some examples, several candidates in the on-going "Competition for Authenticated Encryption: Security, Applicability, and Robustness" (CAESAR) [1] - which is currently at its third round - are designed based on an AES-like SPN structure. Focusing only on the third-round candidates, among many others, AEGIS [29] uses five AES round-functions in the state update functions, while ELmD v1.0 [16] recommends to use round-reduced AES including 5-round AES to partially encrypt the data[2]. Focusing on this last cipher, in [4]

[2] We mention that 5-round AES has been replaced by 6-round AES in ELmD v2.0.

authors exploit known properties of round-reduced AES in a new way to set up a new attack on ELmD.

In a very different context, Mennink and Neves [24] propose a method for transforming a dedicated block-cipher design into a dedicated PRF design. The main proposal AES-PRF-128 is defined to be AES xored with the internal state after 5 rounds, that is $AES\text{-}PRF(\cdot) = AES_{10}(\cdot) \oplus AES_5(\cdot)$.

Although the security of these constructions does not directly (or only) depend on the underlying round-reduced AES primitives, we believe that a better understanding of the security of round-reduced AES can help to get insights to both the design and cryptanalysis of such algorithms.

New Key-Recovery Attacks for 5-round AES. As examples of applications of our distinguishers, in App. H we propose new (practically verified) attacks on 5-round AES, based on the properties analyzed in this paper when adapted to 4-round AES. For a diagonal set of chosen plaintexts, we propose a *(1st) truncated differential attack based on the mean* - the average number of pairs of ciphertexts with a particular difference is (a little) bigger for the right key than for wrongly guessed key - and *(2nd) one based on the mean* - the variance of the number of pairs of ciphertexts with a particular difference is higher for the right key than for wrongly guessed key.

2 Preliminaries - Description of AES

2.1 Description of AES

The Advanced Encryption Standard [15] 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 round 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[3]);
- *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.

[3] SR makes sure column values are spread, MC makes sure each column is mixed.

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[4] of AES, while we denote r rounds of AES by R^r. Finally, in the paper we often use the term "partial collision" (or "*collision*") when two texts belong to the same coset of a given subspace $\mathcal{X}$. We recall that given a subspace X, the cosets $X \oplus a$ and $X \oplus b$ (where $a \neq b$) are *equivalent* (that is $X \oplus a \sim X \oplus b$) if and only if $a \oplus b \in X$.

2.2 Subspace Trails Cryptanalysis

Let F denote a round function in a 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* of the subspace V for the function F. As shown in [21], this concept can be generalized to *trails of subspaces*.

Definition 1. *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 $a_i \in V_i^{\perp}$, there exist (unique) $a_{i+1} \in V_{i+1}^{\perp}$ such that $F(V_i \oplus a_i) \subseteq V_{i+1} \oplus a_{i+1}$, 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 $F^t(V_1 \oplus a_1) = V_{t+1} \oplus a_{t+1}$. We refer to [21] for more details about the concept of subspace trails. Our treatment here is however meant to be self-contained.

Subspace Trails of AES. Here we briefly recall the subspace trails of AES presented in [21] - we refer to App. A for more details. For 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 2. *The column spaces $\mathcal{C}_i$ are defined as $\mathcal{C}_i = \langle e_{0,i}, e_{1,i}, e_{2,i}, e_{3,i} \rangle$.*

Definition 3. *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 $\mathcal{ID}_i = SR(\mathcal{C}_i)$.*

Definition 4. *The i-th mixed spaces $\mathcal{M}_i$ are defined as $\mathcal{M}_i = MC(\mathcal{ID}_i)$.*

Definition 5. *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 [21]:

- for any coset $\mathcal{D}_I \oplus a$ there exists unique $b \in \mathcal{C}_I^{\perp}$ such that $R(\mathcal{D}_I \oplus a) = \mathcal{C}_I \oplus b$;
- 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 1 ([21]). *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. $R^2(\mathcal{D}_I \oplus a) = \mathcal{M}_I \oplus b$.*

[4] Sometimes we use the notation R_k instead of R to highlight the round 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 1. *For all x, y and for all $I \subseteq \{0,1,2,3\}$:*

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

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 [21]. It follows that:

Theorem 2 ([21]). *Let $I, J \subseteq \{0,1,2,3\}$ such that $|I| + |J| \leq 4$. For all x, y:*

$$Prob(R^4(x) \oplus R^4(y) \in \mathcal{M}_I \,|\, x \oplus y \in \mathcal{D}_J, \quad x \neq y) = 0. \quad (2)$$

We remark that all these results can be re-described using a more "classical" truncated differential notation, as formally pointed out in [8]. For example, if two texts t^1 and t^2 are equal except for the bytes in the i-th diagonal[5] for each $i \in I$, then they belong to the same coset of $\mathcal{D}_I$. A coset of $\mathcal{D}_I$ corresponds to a set of $2^{32 \cdot |I|}$ texts with $|I|$ active diagonals. Again, two texts t^1 and t^2 belong to the same coset of $\mathcal{ID}_I$ if the difference of the bytes that lie in the i-th anti-diagonal for each $i \notin I$ is equal to zero. Similar considerations hold for the column space $\mathcal{C}_I$ and the mixed space $\mathcal{M}_I$.

3 A New Truncated-Differential for 5-round AES

In this section, we describe a new truncated differential property for up to 5-round AES which is independent of the secret-key.

As already recalled in the introduction, differential attacks [7] 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 [22], 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 [8] 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 $\mathcal{X}$ can not 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 respectively for up to 3 and for up 4 rounds of AES. In both cases, the subspaces $\mathcal{X}$ and $\mathcal{Y}$ correspond respectively to $\mathcal{D}_I$ and $\mathcal{M}_J$, as showed in detail in [21].

The truncated differential property - and the corresponding distinguisher - that we are going to present works in a similar way. The main difference with

[5] 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 $r + c = i \bmod 4$.

other differential distinguishers in literature is the fact that our distinguisher works if and only if one considers entire cosets of a particular space $\mathcal{X}$ and not random pairs of texts. Our result can be summarized as follows.

Consider 2^{32} plaintexts p^i for $i = 0, 1, ..., 2^{32} - 1$ in a coset of a diagonal space $\mathcal{D}_k$, that is $\mathcal{D}_k \oplus a$ for $k \in \{0, 1, 2, 3\}$ and $a \in \mathcal{D}_k^{\perp}$, and the corresponding ciphertexts after 5-round, that is $c^i = R^5(p^i)$. The average number of different pairs[6] of ciphertexts (c^i, c^j) for $i \neq j$ that belong to the same coset of $\mathcal{M}_K$ for $K \subseteq \{0, 1, 2, 3\}$ fixed with $|K| = 3$ is approximately equal to $2\,147\,484\,685.6 \simeq 2^{31} + 2^{10}$, in contrast to an average number of $2\,147\,483\,647.5 \simeq 2^{31}$ for a random permutation (approximately $1\,038.1$ more collisions for the AES case). In other words, given a set of 2^{32} plaintexts as before, the probability that two ciphertexts belong to the same coset of $\mathcal{M}_K$ is approximately $2^{-32} + 2^{-52.9}$ versus a probability 2^{-32} for a random permutation.

If the final MixColumns is omitted, it is sufficient to replace the mixed space $\mathcal{M}_K$ with an inverse-diagonal space $\mathcal{ID}_K$. We remember that a coset of $\mathcal{D}_k$ corresponds to a set of 2^{32} texts with one active diagonal, while that two ciphertexts c^i and c^j belong to the same coset of an inverse-diagonal space $\mathcal{ID}_K$ (that is, $c^i \oplus c^j \in \mathcal{ID}_K$) if and only if they are equal in the k-th anti-diagonal where $k \equiv \{0, 1, 2, 3\} \setminus K$. For completeness, the same result holds also in the decryption direction - that is, using chosen ciphertexts instead of chosen plaintexts - and also in the case in which K is not fixed.

Even if the difference between the numbers of collisions for the two cases is very small, it is possible to set up a secret-key distinguisher which is independent of the secret key for 5-round AES that exploits such property. In particular, such distinguisher on 5-round AES requires $2^{47.4}$ chosen plaintexts (or ciphertexts) and it has a computational cost of 2^{51} table look-ups, as showed in detail in Sect. 7. In the following, we first theoretically prove the result just given, and we present our practical results on small-scale AES.

A Formal Statement. As we are going to show, the previous result about the number of collisions depends on the details of the S-Box. For this reason, we first recall some properties of the S-Box function.

Given a bijective S-Box function, let $\Delta_I, \Delta_O \in \mathbb{F}_{2^8}$. We denote by n_{Δ_I, Δ_O} the number of solutions x of the following equation

$$\text{S-Box}(x \oplus \Delta_I) \oplus \text{S-Box}(x) = \Delta_O. \tag{3}$$

For the following, we limit 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).

[6] The pairs (s, t) and (t, s) are considered to be equivalent. We use the partial order $\leq$ to formalize this concept.

Independently of the details of the S-Box, the *mean value*[7] of n_{Δ_I,Δ_O} is equal to

$$\mathbb{E}[n_{\Delta_I,\Delta_O}] = \frac{256}{255} \simeq 1.00392... \simeq 1 + 2^{-7.9944}, \tag{4}$$

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. (3). Since there are 256 different x and 255 different values of Δ_I and Δ_O, the average number of solutions is $\frac{256 \cdot 255}{255^2} = \frac{256}{255}$ independently of the details of the (bijective) S-Box.

For the following, we denote by $Var(n_{\Delta_I,\Delta_O})$ the *variance*[8] 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 127 values of $\Delta_O \neq 0$ for which equation (3) has no solution, 126 values of $\Delta_O \neq 0$ for which equation (3) 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 (3) has 4 solutions. The variance of the AES S-Box is so equal to $Var_{AES}(n_{\Delta_I,\Delta_O}) = 2^2 \cdot \frac{126}{255} + 4^2 \cdot \frac{1}{255} - \left(\frac{256}{255}\right)^2 = \frac{67\,064}{65\,025}$.

We also recall the definitions of *Maximum Differential Probability* and of *Uniform Differential S-Box*. The Maximum Differential Probability DP_{max} of an S-Box is defined as

$$DP_{max} = \max_{\Delta_I \neq 0, \Delta_O} n_{\Delta_I,\Delta_O}. \tag{5}$$

DP_{max} is always even and it is always bigger than or equal to 2 (i.e. $DP_{max} \geq 2$). *Those permutation with $DP_{max} = 2$ are called Almost Perfect Nonlinear (APN).*

Finally, given $\Delta_I \neq 0$ (resp. $\Delta_O \neq 0$), consider the probabilistic distribution of n_{Δ_I,Δ_O} w.r.t. $\Delta_O \neq 0$ (resp. $\Delta_I \neq 0$). The S-Box is *Uniform Differential* if such distribution is independent of $\Delta_I \neq 0$ (resp. $\Delta_O \neq 0$). As examples, the AES S-Box is uniform differential 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. I) is instead not uniform differential, 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. $n_{0\text{x}F,\Delta_O} \neq 4\ \forall \Delta_O$) while $Prob(n_{\Delta_I,\Delta_O} = 4) = \frac{2}{15}$ if $\Delta_I = 0\text{x}A$ (two values of Δ_O satisfy $n_{0\text{x}A,\Delta_O} = 4$).

Definition 6. *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 $t^1 \neq t^2$.*

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

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

Theorem 3. *Consider an AES-like cipher that works with texts in $\mathbb{F}_{2^8}^{4\times 4}$, s.t. the MixColumns matrix is an MDS matrix*[9] *and the S-Box is an APN permutation.*

Consider 2^{32} plaintexts p^i for $i = 0, 1, ..., 2^{32} - 1$ in a coset of a diagonal space $\mathcal{D}_k$, that is $\mathcal{D}_k \oplus a$ for $k \in \{0, 1, 2, 3\}$ and $a \in \mathcal{D}_k^{\perp}$, and the corresponding ciphertexts after 5-round, that is $c^i = R^5(p^i)$. 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 $\mathcal{M}_K$ for $K \subseteq \{0, 1, 2, 3\}$ fixed with $|K| = 3$ is equal to $2\,147\,484\,685.6 \simeq 2^{31} + 1\,038.1$, approximately 2^{10} more collisions than for a random permutation.

The proof of this Theorem is given in the next section.

We emphasize that if the S-Box doesn't satisfy the required property of the Theorem, then the number of collisions can be different than the one previously given. To be more concrete, in Sect. 8 we provide several practical examples of the dependency of the number of collisions for small-scale AES-like ciphers with respect to the properties of the S-Box, and we provide theoretical argumentations 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.

Generic Considerations. Before presenting the proof, we discuss the assumptions of the Theorem, focusing on the S-Box one. 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, it is known that there is no (invertible) APN permutation in dimension 4 [23], while it is known that there is at least one APN permutation in dimension 6 - called the Dillon's permutation [11]. 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 dimension equals to a power of 2 (e.g. $\mathbb{F}_{2^4}$ or $\mathbb{F}_{2^8}$), it is a common choice to use permutations which are not APN but are differentially 4-uniform in order to design the S-Box function [25]. For this reason, in the following we apply the results of the previous Theorem to S-Boxes that satisfy a "relaxed" assumption with respect to the one given in the Theorem. In particular, we apply the result of Theorem 3 to S-Boxes which are *(1st)* uniform differential, *(2nd)* have DP_{max} equal to 4 and for which *(3rd)* $Var(n_{\Delta_I,\Delta_O})$ is "low"[10]. This is equivalent to ask that the solutions n_{Δ_I,Δ_O} of (3) are uniform distributed with respect to $\Delta_I \neq 0$ and $\Delta_O \neq 0$.

Without going into the details (which are out of the scope of this paper), it turns out the *the only (known) S-Box that (approximately) matches all the assumptions of the Theorem in dimensions 4 or 8 is the one generated by the multiplicative-inverse permutation - unless affine equivalence relations*[11] (e.g. the

[9] 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 of M 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.

[10] If the S-Box is an APN permutation ($DP_{max} = 2$), then $Var = 65\,024/65\,025$.

[11] Uniform 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

AES S-Box). Our practical results on small-scale AES (for which the S-Box has the same property of the full-size AES one) confirm that the practical number of collisions for this case is very close to the one predicted by the previous Theorem.

4 Proof of Theorem 3

4.1 Reduction to the Middle Round

In order to prove Theorem 3, the idea is to prove an equivalent result on a single round. Since each coset of a diagonal space is mapped into a mixed space after 2 rounds - see Theorem 1 - and since $\text{Prob}(t^1 \oplus t^2 \in \mathcal{D}_J \mid R^2(t^1) \oplus R(t^2) \in \mathcal{M}_J) = 1$, observe that for any $I, J \subseteq \{0,1,2,3\}$:

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

Working on the middle round, the idea is to prove the following equivalent result.

Lemma 2. *Consider an AES-like cipher that works with texts in $\mathbb{F}_{2^8}^{4\times 4}$, such that the MixColumns matrix is MDS and the S-Box is an APN permutation.*

Consider 2^{32} plaintexts p^i for $i = 0, 1, ..., 2^{32}-1$ in a coset of a mixed space $\mathcal{M}_k$, that is $\mathcal{M}_k \oplus a$ for $k \in \{0,1,2,3\}$ and $a \in \mathcal{M}_k^{\perp}$, and the corresponding ciphertexts after 1-round, that is $c^i = R(p^i)$. 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 $\mathcal{D}_K$ for $K \subseteq \{0,1,2,3\}$ fixed with $|K| = 3$ is equal to $2\,147\,484\,685.6 \simeq 2^{31} + 2^{10}$, w.r.t. approximately $2\,147\,483\,647.5 \simeq 2^{31}$ collisions for a random permutation.

Idea of the Proof. For simplicity, we limit to consider plaintexts in the same coset of $\mathcal{M}_0$ and to count the collisions in the same coset of a 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. For 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) - 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 $w^1 = w^2$;

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 $\mathcal{S}'(x) = B \cdot [\mathcal{S}(A \cdot x + a)] + b \ \forall x \in \mathbb{F}_2^n$.

– all variables are different, e.g. $x^1 \neq x^2$, $y^1 \neq y^2$, $z^1 \neq z^2$, $w^1 \neq w^2$.

In the first case - if 3 variables are equal (e.g. $y^1 = y^2$, $z^1 = z^2$ and $w^1 = w^2$), then $p^1 \oplus p^2 \in \mathcal{C}_k$ and $R(p^1) \oplus R(p^2) \in \mathcal{M}_k$ for a certain $k \in \{0, 1, 2, 3\}$. By Theorem 2, it follows that $R(p^1) \oplus R(p^2) \notin \mathcal{D}_J$ for each J. Thus, for the following we limit to consider the case in which at least 2 generating variables are different.

4.2 Case: Two Equal Generating Variables

As first case, we consider the case in which 2 generating variables are different, e.g. $x^1 \neq x^2$, $y^1 \neq y^2$, $z^1 = z^2$ and $w^1 = w^2$. This is equivalent to consider 2^{16} plaintexts in the same coset of $\mathcal{C}_{0,1} \cap \mathcal{M}_0$ (the other cases are equivalent).

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 and by definition of the diagonal space $\mathcal{D}_{1,2,3}$, $R(p^1) \oplus R(p^2) \in \mathcal{D}_{1,2,3}$ if and only if the following four equations are satisfied

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

Equivalently, four equations of the form

$$\begin{aligned}
A \cdot \left[\text{S-Box}(B \cdot x^1 \oplus a) \oplus \text{S-Box}(B \cdot x^2 \oplus a)\right] \oplus \\
\oplus C \cdot \left[\text{S-Box}(D \cdot y^1 \oplus c) \oplus \text{S-Box}(D \cdot y^2 \oplus c)\right] = 0
\end{aligned} \tag{6}$$

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. Consider one of these four equations. By simple observation, equation (6) is satisfied if and only if[12] the following system of equations is satisfied

$$\begin{aligned}
\text{S-Box}(\hat{x} \oplus \Delta_I) \oplus \text{S-Box}(\hat{x}) &= \Delta_O \\
\text{S-Box}(\hat{y} \oplus \Delta'_I) \oplus \text{S-Box}(\hat{y}) &= \Delta'_O \\
\Delta'_O &= A^{-1} \cdot C \cdot \Delta_O
\end{aligned} \tag{7}$$

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 $\Delta'_I = D \cdot (y^1 \oplus y^2)$.

[12] Observe that the equality $\Delta'_O = A^{-1} \cdot C \oplus \Delta_O$ 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.*

What is the number of different (not null) solutions $[(x^1, y^1), (x^2, y^2)]$ *of eq. (6)?* Given Δ_O, each one of the first two equations of (7) admits $\frac{256}{255} \cdot 255 = 256$ different solutions $(\hat{x}, \Delta_I)$ (resp. $(\hat{y}, \Delta'_I)$) - note that there are 255 different values of $\Delta_I, \Delta'_I \neq 0$ and that the average number of solutions is 256/255. It follows that the number of different (not null) solutions $[(x^1, y^1), (x^2, y^2)]$ of eq. (6) - 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. Equivalently, a solution $[(x^1, y^1), (x^2, y^2)]$ is considered to be valid if $x^2 \neq x^1$ and $y^1 < y^2$.

Knowing the number of solutions of one eq. (6), *what is the number of common - different (not null) - solutions* $[(x^1, y^1), (x^2, y^2)]$ *of 4 equations of the form (6)?* We have just seen that each equation of the form (6) 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

$$(256 \cdot 255)^{-1} \cdot (255 \cdot 128)^{-1} = 255^{-2} \cdot 2^{-15}. \tag{8}$$

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. $\hat{x}^2$) can not 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[13], 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. $(128 \cdot 255)^{-1}$. We highlight that *probability (8) (strongly) depends on the assumption that the S-Box is APN, or equivalently on the fact that the number of solutions* n_{Δ_I, Δ_O} *are uniform distributed for each* (Δ_I, Δ_O).

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 (6) is given by

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

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

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

[13] 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.

4.3 Case: One Equal Generating Variable

As second case, we consider the case in which only 1 generating variable is equal, e.g. $x^1 \neq x^2$, $y^1 \neq y^2$, $z^1 \neq z^2$ and $w^1 = w^2$. This is equivalent to consider 2^{24} plaintexts in the same coset of $\mathcal{C}_{0,1,2} \cap \mathcal{M}_0$ (the other cases are equivalent).

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

$$\begin{aligned} & A \cdot [\text{S-Box}(B \cdot x \oplus b) \oplus \text{S-Box}(B \cdot x' \oplus b)] \\ \oplus & C \cdot [\text{S-Box}(D \cdot y \oplus d) \oplus \text{S-Box}(D \cdot y' \oplus d)] \oplus \\ \oplus & E \cdot [\text{S-Box}(F \cdot z \oplus f) \oplus \text{S-Box}(F \cdot z' \oplus f)] = 0 \end{aligned} \tag{9}$$

are satisfied, where A, B, C, D, E, F depend only on the MixColumns matrix, while b, d, f depend on the secret key and on the initial constant that defines the coset. As before, each one of these equations is equivalent to a system of equations like (7), that is

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

together with one of the two following conditions[14]:

1. $\Delta''_O = 0$ and $\Delta'_O = A^{-1} \cdot C \cdot \Delta_O$, 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. Since this first case is analogous to the case in which two generating variables are equal, we simply re-use the previous computation.

First of all note that if $\Delta''_O = 0$, then the third equation admits solutions if and only if $\Delta''_I = 0$. In other words, if $\Delta''_O = 0$, the only possible solutions of the third equation are $(z, \Delta''_I = 0)$ for each z. Using the same computation as before, the average number of (not null) common solutions for this first case is

$$\binom{3}{1} \cdot 256 \cdot \frac{2^{15}}{255^2} = \frac{2^{23}}{21\,675} \simeq 387.018.$$

Second Case. Consider now the case $\Delta_O, \Delta'_O, \Delta''_O \neq 0$ (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 $A^{-1} \cdot C \cdot \Delta_O$ (if $\Delta'_O = A^{-1} \cdot C \cdot \Delta_O$, then $\Delta''_O = 0$ which is excluded by assumption). Finally, the value of Δ''_O depends on Δ_O and Δ'_O.

Using the same argumentation as before, for each equation (9) 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 $\frac{1}{2} \cdot$

[14] A solution of the first case can not be equal to a solution of the second case. Indeed, $\Delta''_O = 0$ implies $\Delta''_I = 0$ in the first case, while in the second one $\Delta_I, \Delta'_I, \Delta''_I \neq 0$.

$255 \cdot 254 \cdot \left(255 \cdot \frac{256}{255}\right)^3 = 32\,385 \cdot 2^{24}$, while the probability that two equations of the form (9) have a common solution is given by $(256 \cdot 255)^{-2} \cdot (128 \cdot 255)^{-1} = 2^{-23} \cdot 255^{-3}$ *under the assumption of uniform distribution of the solutions* n_{Δ_I, Δ_O} (equivalently, under the assumption that the S-Box is APN). It follows that for this second case we expect on average

$$(32\,385 \cdot 2^{24})^4 \cdot (2^{-23} \cdot 255^{-3})^3 = \frac{127^4 \cdot 2^{27}}{255^5} \simeq 32\,383.506$$

different - not null - common solutions for the 4 equations of the form (9).

Total Number of Different - not null - Common Solutions. By simple calculation, given plaintexts in the same coset of $\mathcal{C}_{0,1,2} \cap \mathcal{M}_0$, the average number of different pairs of ciphertexts that belong to the same coset of $\mathcal{D}_{1,2,3}$ is

$$32\,383.506 + 387.018 \simeq 32\,770.524 \simeq 2^{15} + 2^{1.336}$$

For comparison, if the ciphertexts are generated by a random permutation, the average number of different pairs of ciphertexts that belong to the same coset of $\mathcal{M}_J$ is approximately given by

$$\binom{2^{24}}{2} \cdot 2^{-32} \simeq 32\,767.998 \simeq 2^{15} - 2^{-9}$$

4.4 Case: No Equal Generating Variables

Finally, we consider the case in which all the generating variables are different, that is $x^1 \neq x^2$, $y^1 \neq y^2$, $z^1 \neq z^2$ and $w^1 \neq w^2$.

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

$$\begin{aligned} &A \cdot [\text{S-Box}(B \cdot x \oplus b) \oplus \text{S-Box}(B \cdot x' \oplus b)] \oplus \\ \oplus &C \cdot [\text{S-Box}(D \cdot y \oplus d) \oplus \text{S-Box}(D \cdot y' \oplus d)] \oplus \\ \oplus &E \cdot [\text{S-Box}(F \cdot z \oplus f) \oplus \text{S-Box}(F \cdot z' \oplus f)] \oplus \\ \oplus &G \cdot [\text{S-Box}(H \cdot w \oplus h) \oplus \text{S-Box}(H \cdot w' \oplus h)] = 0 \end{aligned} \tag{10}$$

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 (7), that is:

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

together with one of the following conditions

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

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

First Case. In the case $\Delta_O^{'''} = \Delta_O^{''} = 0$ and $\Delta_O' = A^{-1} \cdot C \cdot \Delta_O \neq 0$, 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

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

Second Case. Similarly, in the second case $\Delta_O^{'''} = 0$, $\Delta_O, \Delta_O^{'}, \Delta_O^{''} \neq 0$ and using the same computations as before, it follows that the average number of (not null) common solutions of this case is

$$\binom{4}{1} \cdot 256 \cdot \frac{127^4 \cdot 2^{27}}{255^5} = \frac{127^4 \cdot 2^{37}}{255^4} \simeq 33\,160\,710.047.$$

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 the previous equation - and so $\Delta_O^{'''}$ - 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 $\Delta_O^{'}$ 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 $\Delta_O^{''}$ 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 (10) 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 assumption of uniform distribution of the solutions*, the average number of (non null) common solutions (with no equal generating variables) is

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

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 belong to the same coset of $\mathcal{D}_{1,2,3}$ 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 belong to the same coset of $\mathcal{M}_J$ 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} of the random case.

Remark - 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 $t^1 \oplus t^3 \in \mathcal{M}_J$ and $t^2 \oplus t^3 \in \mathcal{M}_J$, then (t^1, t^2) belong to the same coset of $\mathcal{M}_J$ 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 $\mathcal{M}_J$ is different than $2^{-32 \cdot (4 - |J|)}$. In App. C, 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 $\mathcal{M}_J$ is exactly* $2^{-32 \cdot (4 - |J|)}$ (that is 2^{-32} if $|J| = 3$ for J fixed).

4.5 Generic Case

For completeness, we briefly discuss the case in which one considers the number of different pairs of ciphertexts that belong to the same coset of a mixed space $\mathcal{M}_K$ for an arbitrary $K \subset \{0, 1, 2, 3\}$ with $|K| = 3$. Since there are 4 different values of K with $|K| = 3$, the number of collisions is (approximately) obtained by multiplying by a factor 4 the number obtained in the previous case.

Proposition 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 3 hold.*

Consider 2^{32} plaintexts p^i for $i = 0, 1, ..., 2^{32} - 1$ in a coset of a diagonal space $\mathcal{D}_k$, that is $\mathcal{D}_k \oplus a$ for $i \in \{0, 1, 2, 3\}$ and $a \in \mathcal{D}_k^{\perp}$, and the corresponding ciphertexts after 5-round, that is $c^i = R^5(p^i)$. The probability that a pair of ciphertexts (c^i, c^j) with $c^i \leq c^j$ for $i \neq j$ belong to the same coset of $\mathcal{M}_K$ for any $K \subseteq \{0, 1, 2, 3\}$ with $|K| = 3$ is equal to $2^{-30} + 2^{-50.9803} - 2^{61.415} + ...$

A proof of this proposition can be found in App. D. For comparison, the same event has probability $2^{-30} - 2^{61.415} + 2^{-94}$ in the case of a random permutation.

5 5-round Secret-Key Distinguisher based on Variance

The previous result on 5-round AES can be used to set up a secret-key distinguisher which is independent of the secret key. Given $2^{32 \cdot n}$ plaintexts distributed in n cosets of $\mathcal{D}_k$ and the corresponding ciphertexts, for each coset one simply counts the number of pairs of ciphertexts that belong to the same coset of $\mathcal{M}_K$ for a fixed K with $|K| = 3$. Due to the fact that this number is on average bigger for AES than for a random permutation, one can distinguish the two cases.

Since the difference between the average number of collisions for the random permutation and the AES one is very small, *what is the minimum number of initial cosets $\mathcal{D}_k \oplus a$ necessary to guarantee that the distinguisher works with high probability?* To solve this problem, we study the probabilistic distributions of the number of collisions in the AES and in the random cases.

In the random case, note that given n cosets $\mathcal{D}_k \oplus a$ of approximately 2^{63} pairs of plaintexts and the corresponding ciphertexts generated by random oracle, the probabilistic distribution of the number of collisions is simply described by a *binomial distribution*. 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. Thus, this model is well described by a binomial distribution. We remember that for a random variable Z that follows the binomial distribution, that is $Z \sim \mathcal{B}(n, p)$, the mean μ and the variance σ^2 are respectively given by $\mu = n \cdot p$ and $\sigma^2 = n \cdot p \cdot (1 - p)$.

To derive concrete numbers for our distinguisher, we approximate the binomial distribution with a normal one. In particular, the distribution probability of the random case is well described by a normal distribution $X \sim \mathcal{N}(\mu \equiv n \cdot p, \sigma^2 \equiv n \cdot p \cdot (1-p))$, where the mean value is $\mu = n \cdot p = 2^{31} \cdot (2^{32} - 1) \cdot 2^{-32} = 2^{31} - 2^{-1} = 2\,147\,483\,647.5$ and the variance $\sigma^2 = n \cdot p \cdot (1-p) = 2\,147\,483\,647$, or equivalently the standard deviation is $\sigma = 46\,340.95$.

For 5-round AES, we are going to prove the following result.

Theorem 4. *Consider an AES-like cipher that works with texts in $\mathbb{F}_{2^8}^{4\times 4}$ and for which the assumptions of Theorem 3 hold.*

Consider 2^{32} plaintexts p^i for $i = 0, 1, ..., 2^{32} - 1$ in a coset of a diagonal space $\mathcal{D}_i$, that is $\mathcal{D}_i \oplus a$ for $i \in \{0, 1, 2, 3\}$ and $a \in \mathcal{D}_i^{\perp}$, and the corresponding ciphertexts after 5-round, that is $c^i = R^5(p^i)$. The distribution probability of the 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 $\mathcal{M}_J$ for $J \subseteq \{0, 1, 2, 3\}$ fixed with $|J| = 3$ has mean value $\mu = 2\,147\,484\,685.6$ and standard deviation $\sigma = 276\,469.4$.

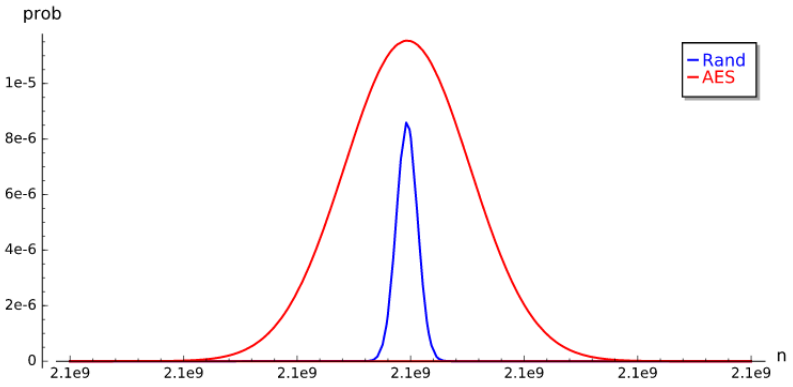

Fig. 1. Comparison between the theoretical probabilistic distribution of the number of collisions between 5-round AES and a random permutation.

The probability to have n collisions where $n \in \mathbb{N}$ is well approximated by:

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

Roughly speaking, the distribution Y of the number of collisions for the AES case is approximated by $Y = 8 \times X$ where X is a normal distribution with mean value and variance as given in Theorem 4. In particular, 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:*

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

Fig. 5 proposes a comparison between the probabilistic distribution of the number of collisions for the AES case in red - 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. The theoretical computation of the variance is largely based on the result presented in [20], which is recalled in the following.

5.1 "Multiple-of-8" Secret-Key Distinguisher for 5-round AES

Consider a set of plaintexts in the same coset of the diagonal space $\mathcal{D}_I$ and the corresponding ciphertexts after 5 rounds, that is a set of plaintexts/ciphertexts (p^i, c^i) for $i = 0, ..., 2^{32 \cdot |I|} - 1$ where all plaintexts are equal in $4 - |I|$ diagonals. The 5-round AES distinguisher proposed in [20] exploits the fact that the number of different pairs of ciphertexts (c^i, c^j) that belong to the same coset of $\mathcal{M}_J$

for a fixed J (i.e. $c^i \oplus c^j \in \mathcal{M}_J$) is always a multiple of 8 with probability 1 independently of the secret key, of the details of the S-Box and of the MixColumns matrix (assuming branch number equal to 5).

Since each coset of $\mathcal{D}_I$ is mapped into a coset of $\mathcal{M}_I$ after 2 rounds with prob. 1 (by Theorem 1) and viceversa, in order to prove the result given in [20] it is sufficient to show that given plaintexts in the same coset of $\mathcal{M}_I$, then the number of collisions after one round in the same coset of $\mathcal{D}_J$ is a multiple of 8.

Theorem 5. *Let $\mathcal{M}_I$ and $\mathcal{D}_J$ be the subspaces defined as before for certain fixed I and J with $1 \leq |I| \leq 3$. Given an arbitrary coset of $\mathcal{M}_I$ - that is $\mathcal{M}_I \oplus a$ for a fixed $a \in \mathcal{M}_I^{\perp}$, consider all the $2^{32\cdot|I|}$ plaintexts and the corresponding ciphertexts after 1 round, that is (p^i, c^i) for $i = 0, ..., 2^{32\cdot|I|} - 1$ where $p^i \in \mathcal{M}_I \oplus a$ and $c^i = R(p^i)$. The number n of different pairs of ciphertexts (c^i, c^j) for $i \neq j$ such that $c^i \oplus c^j \in \mathcal{D}_J$ (i.e. that belong to the same coset of $\mathcal{D}_J$) is a multiple of 8.*

We refer to [20] for a detailed proof, and we limit ourselves here to recall and highlight the main concepts that are useful for the following.

Without loss of generality (w.l.o.g.), we focus on the case $|I| = 1$ and we assume $I = \{0\}$. Given two texts p^1 and p^2 in $\mathcal{M}_0 \oplus a$, by definition there exist $x^1, y^1, z^1, w^1 \in \mathbb{F}_{2^8}$ and $x^2, y^2, z^2, w^2 \in \mathbb{F}_{2^8}$ s.t. $p^i \equiv (x^i, y^i, z^i, w^i)$ for $i = 1, 2$. As first thing, we recall that if $1 \leq r \leq 3$ generating variables are equal, then the two texts can not belong to the same coset of $\mathcal{D}_J$ for $|J| \leq r$ after one round - this is due to the branch number of the MixColumns matrix (which is 5).

Case: Different Generating Variables. If the two elements p^1 and p^2 defined as before have different generating variables (e.g. $x^1 \neq x^2$, $y^1 \neq y^2$, ...), then they can belong to the same of $\mathcal{D}_J$ (for a certain J with $|J| \geq 1$) after one round. It is possible to prove that $p^1 \equiv (x^1, y^1, z^1, w^1)$ and $p^2 \equiv (x^2, y^2, z^2, w^2)$ satisfy $R(p^1) \oplus R(p^2) \in \mathcal{D}_J$ for $|J| \geq 1$ if and only if $\hat{p}^1$ and $\hat{p}^2$ satisfy $R(\hat{p}^1) \oplus R(\hat{p}^2) \in \mathcal{D}_J$, where $\hat{p}^1$ and $\hat{p}^2$ are generated by

1. (x^1, y^1, z^1, w^1) and (x^2, y^2, z^2, w^2);
2. (x^2, y^1, z^1, w^1) and (x^1, y^2, z^2, w^2);
3. (x^1, y^2, z^1, w^1) and (x^2, y^1, z^2, w^2);
4. (x^1, y^1, z^2, w^1) and (x^2, y^2, z^1, w^2);
5. (x^1, y^1, z^1, w^2) and (x^2, y^2, z^2, w^1);
6. (x^2, y^2, z^1, w^1) and (x^1, y^1, z^2, w^2);
7. (x^2, y^1, z^2, w^1) and (x^1, y^2, z^1, w^2);
8. (x^2, y^1, z^1, w^2) and (x^1, y^2, z^2, w^1).

Case: Equal Generating Variables. Similar results can obtained in the case in which one or two generating variables are equal. In the case in which 2 generating variables are equal, it is possible to prove that $p^1 \equiv (x^1, y^1, z, w)$ and $p^2 \equiv (x^2, y^2, z, w)$ satisfy $R(p^1) \oplus R(p^2) \in \mathcal{D}_J$ for $|J| \geq 1$ if and only if $\hat{p}^1$ and $\hat{p}^2$ satisfy $R(\hat{p}^1) \oplus R(\hat{p}^2) \in \mathcal{D}_J$ where $\hat{p}^1$ and $\hat{p}^2$ are generated by

1. (x^1, y^1, z, w) and (x^2, y^2, z, w);
2. (x^2, y^1, z, w) and (x^1, y^2, z, w)

where z and w can take any possible value in $\mathbb{F}_{2^8}$.

Similarly, it is possible to prove that $p^1 \equiv (x^1, y^1, z^1, w)$ and $p^2 \equiv (x^2, y^2, z^2, w)$ satisfy $R(p^1) \oplus R(p^2) \in \mathcal{D}_J$ for $|J| \geq 1$ if and only if $\hat{p}^1$ and $\hat{p}^2$ have the same property - that is $R(\hat{p}^1) \oplus R(\hat{p}^2) \in \mathcal{D}_J$ - where $\hat{p}^1$ and $\hat{p}^2$ are generated by

1. (x^1, y^1, z^1, w) and (x^2, y^2, z^2, w);
2. (x^2, y^1, z^1, w) and (x^1, y^2, z^2, w);
3. (x^1, y^2, z^1, w) and (x^2, y^1, z^2, w);
4. (x^1, y^1, z^2, w) and (x^2, y^2, z^1, w)

where w can take any possible value in $\mathbb{F}_{2^8}$.

5.2 Proof of Theorem 4

Exploiting the result given in [20] just recalled, here we prove Theorem 4. Since

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

as we have just seen, we focus only on the middle round, that is we consider plaintexts in the same coset of $\mathcal{M}_i$ for $i \in \{0, 1, 2, 3\}$ and the corresponding ciphertexts after one round. We recall that given two plaintexts with three equal generating variables, then they can not belong to the same coset of $\mathcal{D}_J$ for $|J| \leq 3$ after one round.

To prove the result, the idea is to consider separately the pairs of texts with n different generating variables for $0 \leq n \leq 3$. First of all, note that given a coset of $\mathcal{M}_i$ of 2^{32} chosen plaintexts, it is possible to construct $2^{31} \cdot (2^{32} - 1) \simeq 2^{63}$ different pairs. Among them, the number of pairs of texts with $0 \leq n \leq 3$ equal generating variables is given by

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

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

Different Generating Variables. As first case, we consider the case in which all the generating variables are different, i.e. $n = 0$. The number of pairs with this property is $2^{31} \cdot (2^8 - 1)^4$.

As we have just seen, these pairs are not independent. By [20], it is possible to divide them in $2^{31} \cdot (2^8 - 1)^4/8 = 2^{28} \cdot (2^8 - 1)^4$ sets of 8 pairs such that for each set only two events can happen: (1) all the pairs belong to the same coset of $\mathcal{D}_J$ after one round or (2) no one of them satisfies this property. Thus, the idea is to consider only independent pairs, i.e. one pair for each one of these sets, for a total of $2^{28} \cdot (2^8 - 1)^4$ pairs. Since these pairs are independent, the probabilistic distribution of the number of pairs that belong to the same coset of $\mathcal{D}_J$ after one round is given by a binomial distribution X with mean value $\mu = n \cdot p_{AES} =$

$2^{28} \cdot (2^8 - 1)^4 \cdot (2^{-32} + 2^{-52.9803})$ and variance $\sigma^2 = n \cdot p_{AES} \cdot (1 - p_{AES}) = 2^{28} \cdot (2^8 - 1)^4 \cdot (2^{-32} + 2^{-52.9803}) \cdot (1 - 2^{-32} - 2^{-52.9803}) \approx 264\,265\,791.745$, that is $\sigma \approx 16\,256.254$. It follows that the probabilistic distribution Y of the number of collisions for the pairs with no equal generating variables is simple given by $Y = 8 \times X$. Since $Var(Y) = 8^2 \times Var(X)$, it follows that the standard deviation σ for this case is given by $8 \cdot 16\,256.254 \approx 130\,050.031$.

One Equal Generating Variable. As second case, we consider the case in which all except one of the generating variables are different, i.e. $n = 1$. The number of pairs with this property is $4 \cdot 2^{31} \cdot (2^8 - 1)^3$.

By [20], these pairs are not independent. Thus, it is possible to divide them in $2^{33} \cdot (2^8 - 1)^3 / 2^{10} = 2^{23} \cdot (2^8 - 1)^3$ sets of 2^{10} pairs such that for each set either (1) all the pairs belong to the same coset of $\mathcal{D}_J$ after one round or (2) no one of them satisfies this property. Considering only one pair for each one of these sets (for a total of $2^{23} \cdot (2^8 - 1)^3$ pairs) and since these pairs are independent, the probabilistic distribution of the number of pairs that belong to the same coset of $\mathcal{D}_J$ after one round is given by a binomial distribution X with mean value $\mu = n \cdot p_{AES} = 2^{23} \cdot (2^8 - 1)^3 \cdot (2^{-32} + 2^{-52.9803})$ and variance $\sigma^2 = n \cdot p_{AES} \cdot (1 - p_{AES}) = 2^{23} \cdot (2^8 - 1)^3 \cdot (2^{-32} + 2^{-52.9803}) \cdot (1 - 2^{-32} - 2^{-52.9803}) \approx 32\,385.513$, that is $\sigma \approx 179.96$. It follows that the probabilistic distribution Y of the number of collisions for the pairs with no equal generating variables is simple given by $Y = 2^{10} \times X$. Since $Var(Y) = (2^{10})^2 \times Var(X)$, it follows that the standard deviation σ for this case is given by $2^{10} \cdot 179.96 \approx 184\,278.788$.

Two Equal Generating Variables. As third case, we consider the case in which all except two of the generating variables are different, i.e. $n = 2$. The number of pairs with this property is $6 \cdot 2^{31} \cdot (2^8 - 1)^2$.

Working exactly as before, it is possible to divide them in $3 \cdot 2^{32} \cdot (2^8 - 1)^2 / 2^{17} = 3 \cdot 2^{15} \cdot (2^8 - 1)^2$ sets of 2^{17} pairs. Considering only one pair for each one of these sets and since these pairs are independent, the probabilistic distribution of the number of pairs that belong to the same coset of $\mathcal{D}_J$ after one round is given by a binomial distribution X with mean value $\mu = n \cdot p_{AES} = 3 \cdot 2^{15} \cdot (2^8 - 1)^2 \cdot (2^{-32} + 2^{-52.9803})$ and variance $\sigma^2 = n \cdot p_{AES} \cdot (1 - p_{AES}) = 3 \cdot 2^{15} \cdot (2^8 - 1)^2 \cdot (2^{-32} + 2^{-52.9803}) \cdot (1 - 2^{-32} - 2^{-52.9803}) \approx 1.488$, that is $\sigma \approx 1.21984$. It follows that the probabilistic distribution Y of the number of collisions for the pairs with no equal generating variables is simple given by $Y = 2^{17} \times X$. Since $Var(Y) = (2^{17})^2 \times Var(X)$, it follows that the standard deviation σ for this case is given by $2^{17} \cdot 1.21984 \approx 159\,886.351$.

Final Result. Note that all the previous cases are independent. In other words, the behavior of a pair of texts with n equal generating variables is independent of another pair with $\hat{n}$ equal generating variables where $\hat{n} \neq n$. Moreover, we recall that given n independent variables $X_1, ..., X_n$, the variance of $Y = X_1 + ... + X_n$ is given by $Var(Y) = Var(X_1) + ... + Var(X_n)$. It follows that the total variance of the probabilistic distribution for the AES case is given by $\sigma^2 \simeq 130\,050.031^2 +$

$184\,278.788^2 + 159\,886.351^2 \simeq 76\,435\,327\,505.945 \simeq 2^{36.155}$, or equivalently the standard deviation is equal to $\sigma \simeq 276\,469.397$.

6 Practical Results on AES

In order to have a practical verification, we have practically verified the mean and the variance for 5-round AES given above using a C/C++ implementation[15]. In particular, we have verified the mean value on a small-scale AES as proposed in [12], and the variance value both on full-size and on the small-scale AES . We limit to recall the small-scale S-Box is defined in the same way as the full-size one and it has the same properties of 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 refer to [12] for a complete description of this small-scale AES. We emphasize that our verification on the 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.

Theoretical Results. To compare the practical values with the theoretical ones, we first re-propose Theorem 4 for the case of small-scale AES.

Proposition 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 3 hold.*

Consider 2^{16} plaintexts p^i for $i = 0, 1, ..., 2^{16} - 1$ in a coset of a diagonal space $\mathcal{D}_k$, that is $\mathcal{D}_k \oplus a$ for $i \in \{0, 1, 2, 3\}$ and $a \in \mathcal{D}_k^{\perp}$, and the corresponding ciphertexts after 5-round, that is $c^i = R^5(p^i)$. The distribution probability of the 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 $\mathcal{M}_K$ for $K \subseteq \{0, 1, 2, 3\}$ fixed with $|K| = 3$ is well approximated by a Normal Distribution with mean value $\mu = 32\,847.124$ and variance $\sigma^2 = 982\,466.615$ (or equivalently, standard deviation $\sigma = 991.195$).

A complete proof of this result can be found in App. F and App. G. For comparison, in the case in which the ciphertexts are generated by a random permutation, the distribution probability of the number of collisions is well approximated by a normal distribution with mean value $\mu = 32\,767.5$ and variance $\sigma^2 = 32\,767$ (or equivalently, standard deviation $\sigma = 181.017$).

Practical Results. In order to test our results, we used 320 initial cosets for full-size AES and 100 initial cosets for the small-scale one[16]. The variance results for full-size AES[17] are given in the following

$$\sigma_T^2 = 76\,435\,327\,505.945 \simeq 2^{36.155} \qquad \sigma_P^2 = 73\,288\,132\,411.36 \simeq 2^{36.093}$$

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

[16] This number of tests was chosen in accordance to Eq. (11) - Sect. 7 - adapted to the case of small-scale AES.

[17] 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.

where the subscript $\cdot_T$ denotes the theoretical value and the subscript $\cdot_P$ the practical one. The results for small-scale AES are given in the following

$$\mu^T_{AES} = 32\,847.124 \qquad \mu^T_{rand} = 32\,767.5 \qquad \sigma^T_{AES} = 991.195 \qquad \sigma^T_{rand} = 181.02$$
$$\mu^P_{AES} = 32\,848.57 \qquad \mu^P_{rand} = 32\,768.2 \qquad \sigma^P_{AES} = 1023.06 \qquad \sigma^P_{rand} = 182.42$$

where μ denotes the mean value, σ^2 the variance, the superscript $\cdot^T$ the theoretical values and the superscript $\cdot^P$ the practical ones. In Fig. 3 (see App. B), we also propose a comparison between the theoretical and the practical distributions of the number of collisions for small-scale 5-round AES. Similarly, in Fig. 4 (see App. B) we propose a comparison between the theoretical and the practical distributions of the number of collisions for a random permutation. From both figures, it is possible to observe that the practical distributions - obtained by experiments - are (very) close to the theoretical ones.

6.1 New 5-round Secret-Key Distinguisher based on the Variance

The fact that *the variance of the AES case is different from the one of the random case independently of the secret-key* allows 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 $\mathcal{D}_i$, one counts the number of different pairs of ciphertexts that belong to the same coset of $\mathcal{M}_J$ for each J 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 AES permutation and for a random permutation is similar to the same ratio in the case of small-scale AES, that is

$$\frac{276\,469.4}{46\,340.95} \approx 6 \approx \frac{991.195}{181.02},$$

we conjecture that the results obtained for the small-scale AES are applicable as well to full-size AES. In particular, by practical tests on small-scale AES, it is possible to distinguish the two cases using $n \geq 2^8$ initial cosets, or in other words 2^8 initial cosets are largely sufficient to "accurately" compute the variance for the AES case and the random one[18]. Since it is possible to compute the number of collisions in $\mathcal{M}_J$ for each J with $|J| = 3$ (4 cases in total), 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 same number of initial cosets is sufficient to distinguish (full-size) AES from a random permutation (using this distinguisher based on the variance), for a data cost of $2^6 \cdot 2^{32} = 2^{38}$ chosen plaintexts distributed in 2^6 initial cosets of $\mathcal{D}_j$. The computational cost is well approximated by the cost to compute the number of collisions. Using Algorithm 1 - described in the details in App. E, the cost is well approximated by $4 \cdot 2^6 \cdot 3 \cdot 2^{32} \simeq 2^{41.6}$ table look-ups, that is approximately 2^{35} five-round encryptions.

[18] 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 $i = 0, ..., 2^{32} - 1$ in a coset of $\mathcal{D}_I$ with $|I| = 1$.
Result: Number of Collisions n
for *all* $j \in \{0, 1, 2, 3\}$ **do**
  Let $A[0, ..., 2^{32} - 1]$ an array initialized to zero;
  for i from *0 to* $2^{32} - 1$ **do**
    $x \leftarrow \sum_{k=0}^{3} MC^{-1}(c^i)_{k,j-k} \cdot 256^k$; // $MC^{-1}(c^i)_{k,j-k}$ `denotes the byte of` $MC^{-1}(c^i)$ `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
  $n \leftarrow 0$; // $n \equiv$ `Number of Collisions`
  for i from *0 to* $2^{32} - 1$ **do**
    $n \leftarrow n + A[i] \cdot (A[i] - 1)/2$;
  end
end
return n.

Algorithm 1: *Secret-Key Distinguisher for 5 Rounds of AES* - Count the number of collisions in the same coset of $\mathcal{M}_J$

7 Truncated Differential Distinguisher for 5-round AES

Using all 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 collisions in $\mathcal{M}_J$ for each J with $|J| = 3$ is a little bigger for AES than for a random permutation.

First of all, to derive concrete numbers for our distinguisher, we approximate the binomial distributions - that describe the AES case and the random one - with normal ones. Moreover, for our goal 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 $X - Y \sim \mathcal{N}(\mu, \sigma^2) = N(\mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2)$. Indeed, in order to distinguish the two cases, note that it is sufficient to guarantee that the average number of pairs that satisfy the required property in the random case is smaller than for 5-round AES. As a result, the mean μ and the variance σ^2 of the difference between the AES and the random distributions are given by

$$\mu = |\mu_{AES} - \mu_{rand}| = n \cdot |p_{AES} - p_{rand}|$$
$$\sigma^2 = \sigma_{rand}^2 + \sigma_{AES}^2 = n \cdot \left[p_{rand} \cdot (1 - p_{rand}) + 35.593 \cdot p_{AES} \cdot (1 - p_{AES})\right]$$

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

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

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

Property	**Rounds**	**Data** (CP/CC)	**Cost**	**Ref.**
Multiple-of-8	5	2^{32}	$2^{35.6}$ M $\approx 2^{29}$ E	[20]
Variance Diff.	**5**	$\mathbf{2^{38}}$	$\mathbf{2^{41.6}}$ **M** $\approx \mathbf{2^{35}}$ **E**	**Sect. 6.1**
Truncated Diff.	**5**	$\mathbf{2^{47.38}}$	$\mathbf{2^{51}}$ **M** $\approx \mathbf{2^{44.34}}$ **E**	**Sect. 7**
Prob. Mixture Diff.	5	2^{52}	$2^{71.5}$ M $\approx 2^{64.9}$ E	[19]

where $\mathrm{erf}(x)$ 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

$$n > \frac{2 \cdot [p_{rand} \cdot (1 - p_{rand}) + 35.593 \cdot p_{AES} \cdot (1 - p_{AES})]}{(p_{rand} - p_{AES})^2} \cdot \left[\mathrm{erfinv}\big(2 \cdot prob - 1\big)\right]^2.$$

where $\mathrm{erfinv}(x)$ is the inverse error function. For the case $p_{rand}, p_{AES} \ll 1$, a good approximation of n is given by[19]

$$n > \frac{73.186 \cdot \max(p_{rand},\, p_{AES})}{(p_{rand} - p_{AES})^2} \cdot \left[\mathrm{erfinv}\big(2 \cdot prob - 1\big)\right]^2. \tag{11}$$

It follows that in order to have a probability of success bigger than 95%, the number of pairs must satisfy $n \geq 2^{78.374}$, since $p_{rand} \approx p_{AES} \approx 2^{-30}$ and $|p_{rand} - p_{AES}| \approx 2^{-50.98}$. Since each coset of $\mathcal{D}_k$ contains 2^{32} different texts and approximately 2^{63} different pairs, this means that the distinguisher requires $2^{15.374}$ different cosets for a data cost of $2^{47.374}$ chosen plaintexts.

The Computational Cost. We have just seen that $2^{47.374}$ 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 $\mathcal{M}_J$ 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

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

encryptions of AES (using the approximation[20] 20 table look-ups $\approx$ 1 round of AES). All details can be found in App. E.1.

8 Open Problems - 5-round Truncated Distinguisher

In this paper, we have presented a new truncated property for 5-round AES-like ciphers in the case in which the S-Box is an APN permutation. Even if no S-Box satisfies such assumption in $\mathbb{F}_{2^4}$ or $\mathbb{F}_{2^8}$, such theoretically result matches the practical result obtained for the AES S-Box, which approximately satisfies the assumptions of Theorem 3 as discussed in Sect. 3. Thus, natural questions arise: *What happens when the AES S-Box is changed with an S-Box that doesn't satisfy (at all) the assumptions of Theorem 3? Is it possible to naturally extend our results to any general case?*

We have studied this problem working on small-scale[21] AES, and by practical results the answer to the second question seems to be negative. In other words, our theory doesn't extend naturally to generic S-Box, but it should be modified depending on the particular properties/details of the S-Box function.

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 [10], MIDORI [3], KLEIN [18], PRESENT [9], RECTANGLE [30], NOEKON [14] and PRIDE [2] - and with some "toy" S-Boxes. For our tests, given 2^{16} plaintexts in the same coset of $\mathcal{D}_i$, we counted the average number of collision in the same coset of $\mathcal{M}_J$ 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. I) and the difference between the number of collisions found by experiments and the theoretical number 32 847.124 under the assumptions of Theorem 3. For each AES-like cipher, we used $125\,000 \simeq 2^{17}$ different initial cosets (values given in the table are the average ones) - new keys are generated at random for each test.

A first observation about the results given in this table is that *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 3 - the number of solutions n_{Δ_I,Δ_O} of equation (3) are uniform distributed) and the practical one is minimum when the S-Box almost satisfies the assumption of Theorem 3 - e.g. the AES S-Box, as expected.

To explain these results, we must refer to the proof of Theorem 3 given in Sect. 4. The idea is to consider a system of 4 equations of the generic form (10), and

[20] 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 distinguishers and the others currently present in the literature. This approximation is largely used in literature.

[21] Remember that no (invertible) APN permutation exists in dimension 4 - see [23] for a complete classification of 4 bit S-Boxes.

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 (see App. I for details) and the "Difference" between the practical and the theoretical number (= 32 847.124) of collisions - under the assumptions of Theorem 3.

AES-like Cipher	**Number Collisions**	**Difference**	$\mathbf{DP_{max}}$	**Var**	**Unif. Diff.**
AES S-Box	32 848.57	+1.45	4	344/225	✓
KLEIN S-Box	32 849.77	+2.65	4	344/225	
MIDORI SB_1 S-Box	32 843.03	−4.10	4	344/225	
PRINCE S-Box	32 852.66	+5.54	4	344/225	
Toy-6 S-Box	32 840.08	−7.05	6	392/225	
RECTANGLE S-Box	32 861.15	+14.03	4	416/225	
NOEKON S-Box	32 878.7	+31.58	4	416/225	
PRESENT S-Box	32 886.32	+35.71	4	416/225	
MIDORI SB_0 S-Box	32 882.83	+39.20	4	416/225	
PRIDE S-Box	32 806.63	−40.50	4	416/225	
Toy-8 S-Box	32 815.68	−31.45	8	464/225	
Toy-10 S-Box	32 919.0	+71.88	10	864/225	

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 (3) are uniform 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. 4. 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 (3) are not uniform distributed (e.g. if the variance of the S-Box is not "low"), then 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 (10) can be bigger or smaller with respect to the one given in Theorem 3 (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 theoretical 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 (presented in Sect. 5)? In order to compute the value of the variance, we have exploited the result given in [20] (that is, the number of collisions is a multiple of 8), the properties of the Variance (if X is a random variable and a a scalar, then $Var(a \cdot X) = a^2 \cdot Var(X)$) and the probability p_{AES} that - given a pair of plaintexts in $\mathcal{D}_i$ - two ciphertexts belong to the same coset of $\mathcal{M}_J$ after 5-round. This probability p_{AES} proposed in Sect. 3 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 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?*

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 3, then the average number of collisions is (almost) independent of the MixColumns matrix details. Instead, if the S-Box doesn't 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 (3) 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 (10) 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 (10) 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 obtain 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 an higher difference occurs for the PRESENT S-Box case (on average, there are -13.37 collisions for this new MDS matrix). As we have just said, this is due to the fact that the probability that a system of 4 equations of the generic form (10) admits a common solution 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 3). Similar results can be obtained using different MDS MixColumns matrices.

Finally, if the AES MixColumns matrix is replaced by an "almost MDS" one[22] (which doesn't satisfy the assumptions of Theorem 3), then the number of collisions can be different with respect to the one predicted by Theorem 3 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-round is on average 31 883.27 (instead of a theoretical number of 32 847.124).

Future Open Problems. In conclusion, while we provide a theoretical explanation (besides practical implementations and verifications) of our results, an *open problem* is to adapt our theoretical argumentations to the cases in which the S-Box doesn't satisfy the assumptions of Theorem 3. As first step, we con-

[22] A $n \times n$ matrix is MDS matrix if its branch number is $n + 1$, while it is "almost MDS" if its branch number is n.

jecture an explanation of our results in this last case, but more research in that sense must be done.

References

1. "CAESAR: Competition for Authenticated Encryption: Security, Applicability, and Robustness," http://competitions.cr.yp.to/caesar.html.
2. M. R. Albrecht, B. Driessen, E. B. Kavun, G. Leander, C. Paar, and T. Yalçın, "Block Ciphers - Focus on the Linear Layer (feat. PRIDE)," in *Advances in Cryptology - CRYPTO 2014*, ser. LNCS, vol. 8616, 2014, pp. 57–76.
3. S. Banik, A. Bogdanov, T. Isobe, K. Shibutani, H. Hiwatari, T. Akishita, and F. Regazzoni, "Midori: A Block Cipher for Low Energy," in *Advances in Cryptology - ASIACRYPT 2015*, ser. LNCS, vol. 9453, 2015, pp. 411–436.
4. A. Bay, O. Ersoy, and F. Karakoç, "Universal Forgery and Key Recovery Attacks on ELmD Authenticated Encryption Algorithm," in *Advances in Cryptology - ASIACRYPT 2016*, ser. LNCS, vol. 10031, 2016, pp. 354–368.
5. E. Biham, A. Biryukov, and A. Shamir, "Cryptanalysis of Skipjack Reduced to 31 Rounds Using Impossible Differentials," in *Advances in Cryptology - EUROCRYPT 1999*, ser. LNCS, vol. 1592, 1999, pp. 12–23.
6. E. Biham and N. Keller, "Cryptanalysis of Reduced Variants of Rijndael," 2001, unpublished, http://csrc.nist.gov/archive/aes/round2/conf3/papers/35-ebiham.pdf.
7. E. Biham and A. Shamir, "Differential cryptanalysis of DES-like cryptosystems," *Journal of Cryptology*, vol. 4, no. 1, pp. 3–72, Jan 1991.
8. C. Blondeau, G. Leander, and K. Nyberg, "Differential-Linear Cryptanalysis Revisited," *Journal of Cryptology*, vol. 30, no. 3, pp. 859–888, 2017.
9. 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*, ser. LNCS, vol. 4727, 2007.
10. J. Borghoff, A. Canteaut, T. Güneysu, E. B. Kavun, M. Knezevic, L. R. Knudsen, G. Leander, V. Nikov, C. Paar, C. Rechberger, P. Rombouts, S. S. Thomsen, and T. Yalçın, "PRINCE – A Low-Latency Block Cipher for Pervasive Computing Applications," in *Advances in Cryptology - ASIACRYPT 2012*, ser. LNCS, vol. 7658, 2012, pp. 208–225.
11. K. A. Browning, J. F. Dillon, M. T. McQuistan, and A. J. Wolfe, "An APN permutation in dimension six," *IEEE Transactions on Information Theory*, vol. 518, pp. 33–42, 2010.
12. C. Cid, S. Murphy, and M. J. B. Robshaw, "Small Scale Variants of the AES," in *Fast Software Encryption - FSE 2005*, ser. LNCS, vol. 9054, 2005, pp. 145–162.
13. J. Daemen, L. R. Knudsen, and V. Rijmen, "The Block Cipher Square," in *Fast Software Encryption - FSE 1997*, ser. LNCS, vol. 1267, 1997, pp. 149–165.
14. J. Daemen, M. Peeters, G. V. Assche, and V. Rijmen, "Nessie proposal: the block cipher NOEKEON," Nessie submission, 2000, http://gro.noekeon.org/.
15. J. Daemen and V. Rijmen, *The Design of Rijndael: AES - The Advanced Encryption Standard*, ser. Information Security and Cryptography. Springer, 2002.
16. N. Datta and M. Nandi, "ELmD," https://competitions.cr.yp.to/round1/elmdv10.pdf.
17. P. Derbez, "Meet-in-the-middle attacks on AES," PhD thesis, Ecole Normale Supérieure de Paris, 2013, https://tel.archives-ouvertes.fr/tel-00918146.

18. Z. Gong, S. Nikova, and Y. W. Law, "KLEIN: A New Family of Lightweight Block Ciphers," in *RFID 2011*, ser. LNCS, vol. 7055, 2012, pp. 1–18.
19. L. Grassi, "Mixture Differential Cryptanalysis: New Approaches for Distinguishers and Attacks on round-reduced AES," Cryptology ePrint Archive, Report 2017/832, 2017.
20. L. Grassi, C. Rechberger, and S. Rønjom, "A New Structural-Differential Property of 5-Round AES," in *Advances in Cryptology - EUROCRYPT 2017*, ser. LNCS, vol. 10211, 2017, pp. 289–317.
21. ——, "Subspace Trail Cryptanalysis and its Applications to AES," *IACR Transactions on Symmetric Cryptology*, vol. 2016, no. 2, pp. 192–225, 2017. [Online]. Available: http://ojs.ub.rub.de/index.php/ToSC/article/view/571
22. L. R. Knudsen, "Truncated and higher order differentials," in *Fast Software Encryption - FSE 1994*, ser. LNCS, vol. 1008, 1995, pp. 196–211.
23. G. Leander and A. Poschmann, "On the Classification of 4 Bit S-Boxes," in *Arithmetic of Finite Fields - WAIFI 2007*, ser. LNCS, vol. 4547, 2007, pp. 159–176.
24. B. Mennink and S. Neves, "Optimal PRFs from Blockcipher Designs," Cryptology ePrint Archive, Report 2017/812, to appear at FSE 2018, 2017.
25. K. Nyberg, "Perfect nonlinear S-boxes," in *Advances in Cryptology - EUROCRYPT 1991*, ser. LNCS, vol. 547, 1991, pp. 378–386.
26. S. Rønjom, N. G. Bardeh, and T. Helleseth, "Yoyo Tricks with AES," in *Advances in Cryptology - ASIACRYPT 2017*, ser. LNCS, vol. 10624, 2017, pp. 217–243.
27. T. Tiessen, "Polytopic Cryptanalysis," in *Advances in Cryptology - EUROCRYPT 2016*, ser. LNCS, vol. 9665, 2016, pp. 214–239.
28. M. Tunstall, "Improved "Partial Sums"-based Square Attack on AES," in *International Conference on Security and Cryptography - SECRYPT 2012*, ser. LNCS, vol. 4817, 2012, pp. 25–34.
29. H. Wu and B. Preneel, "A Fast Authenticated Encryption Algorithm," http://competitions.cr.yp.to/round1/aegisv11.pdf.
30. W. Zhang, Z. Bao, D. Lin, V. Rijmen, B. Yang, and I. Verbauwhede, "RECTANGLE: a bit-slice lightweight block cipher suitable for multiple platforms," *Science China Information Sciences*, vol. 58, no. 12, pp. 1–15, Dec 2015.

A Subspace Trails Cryptanalysis for AES

In this section, we give all the details about the subspace trails of AES presented in [21], and briefly recalled in Sect. 2.2.

We recall that for 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 2. *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 3. *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 4. *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

$$\mathcal{M}_0 \equiv \begin{bmatrix} 0\text{x}02 \cdot x_1 & x_4 & x_3 & 0\text{x}03 \cdot x_2 \\ x_1 & x_4 & 0\text{x}03 \cdot x_3 & 0\text{x}02 \cdot x_2 \\ x_1 & 0\text{x}03 \cdot x_4 & 0\text{x}02 \cdot x_3 & x_2 \\ 0\text{x}03 \cdot x_1 & 0\text{x}02 \cdot x_4 & x_3 & x_2 \end{bmatrix}$$

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

Definition 5. *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 [21], 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 1 ([21]). *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 [21] for a complete proof of this theorem, and we limit 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 1. *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 [21]. It follows that:

Theorem 2 ([21]). *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 Experimental Results on Small-Scale AES

In Sect. 6, we reported our practical tests on small-scale AES. We remember that the idea is the following: given 2^{16} plaintexts in the same coset of $\mathcal{D}_i$ for $i \in \{0,1,2,3\}$, one counts the number of different pairs of ciphertexts that belong to the same coset of $\mathcal{M}_J$ for $|J| = 3$ after 5-round.

In this section, we propose a comparison between the theoretical and the practical distributions of the number of collisions for small-scale 5-round AES, and a comparison between the theoretical and the practical distributions of the number of collisions for a random permutation.

B.1 A (possible) Distinguisher based on the Skew

Interestingly, it is possible to observe an *asymmetry in the (small-scale) 5-round AES distribution. A parameter that measures the asymmetry of the probabilistic distribution of a real-valued random variable about its mean is the skewness.* The skewness value can be positive or negative, or undefined.

To understand the meaning of the skewness parameter, consider the two distributions in Figure 5[23]. Within each graph, the values on the right side of the distribution taper differently from the values on the left side. These tapering sides are called tails, and they provide a visual means to determine which of the two kinds of skewness a distribution has:

- *negative skew*: the left tail is longer; the mass of the distribution is concentrated on the right of the figure. The distribution is said to be left-skewed, left-tailed, or skewed to the left, despite the fact that the curve itself appears to be skewed or leaning to the right;

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

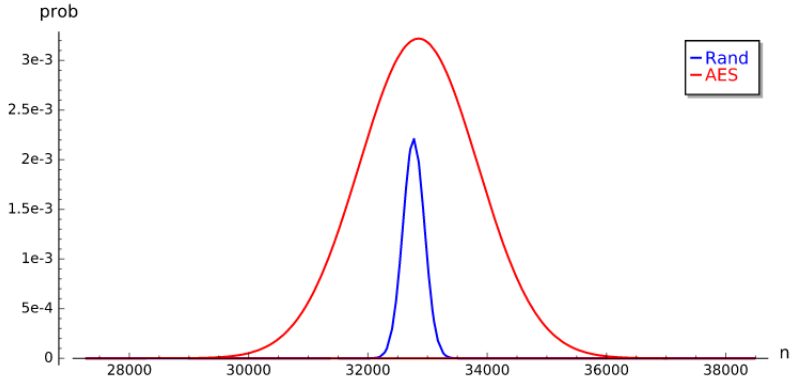

Fig. 2. Comparison between the theoretical probabilistic distribution of the number of collisions between small-scale AES and a random permutation.

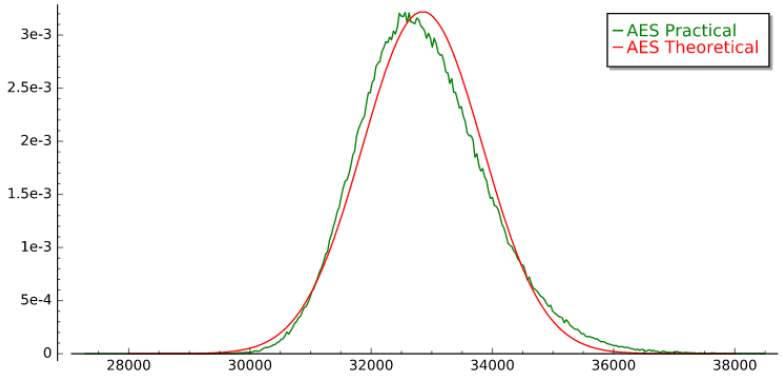

Fig. 3. Comparison between the theoretical and the practical probabilistic distributions of the number of collisions for small-scale 5-round AES.

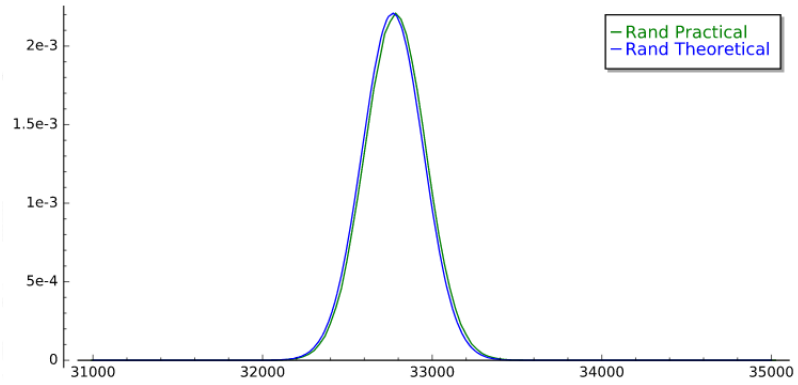

Fig. 4. Comparison between the theoretical and the practical probabilistic distributions of the number of collisions for a random permutation.

– *positive skew*: the right tail is longer; the mass of the distribution is concentrated on the left of the figure. The distribution is said to be right-skewed, right-tailed, or skewed to the right, despite the fact that the curve itself appears to be skewed or leaning to the left.

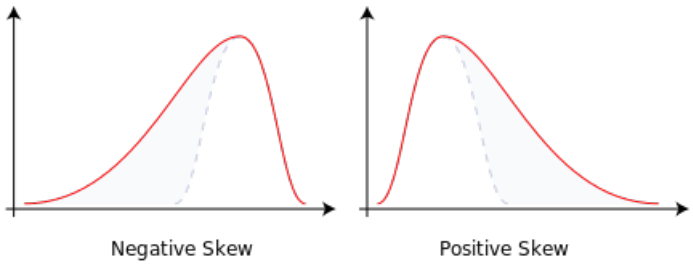

Fig. 5. Examples of negative and positive skew.

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.

For a sample of n values, an estimator z for the skewness is given by

$$z = \frac{\frac{1}{n}\sum_{i=1}^{n}[(x_i - \bar{X})^3]}{(\frac{1}{n}\sum_{i=1}^{n}[(x_i - \bar{X})^2])^{3/2}}.$$

where $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$. By Fig. 3, 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:

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

while we got that the skew of a random permutation is close to 0.

It follows that also the skew 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.

C Number of Collisions - Random Permutation

Consider 2^{32} plaintexts in the same coset of a diagonal space $\mathcal{D}_i$. In Sect. 4, 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 $|J| = 3$. 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 $(\Omega, \mathcal{F}, \mathbb{P})$

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

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

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

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.

D Proof of Proposition 3

Proposition 1. *Consider an AES-like cipher that works with texts in* $\mathbb{F}_{2^8}^{4\times 4}$ *for which the assumptions of Theorem 3 hold.*

Consider 2^{32} *plaintexts* p^i *for* $i = 0, 1, ..., 2^{32}-1$ *in a coset of a diagonal space* $\mathcal{D}_i$*, that is* $\mathcal{D}_i \oplus a$ *for* $i \in \{0,1,2,3\}$ *and* $a \in \mathcal{D}_i^{\perp}$*, and the corresponding ciphertexts after 5-round, that is* $c^i = R^5(p^i)$*. The probability that a pair of ciphertexts* (c^i, c^j) *with* $c^i \le c^j$ *for* $i \neq j$ *belong to the same coset of* $\mathcal{M}_J$ *for any* $J \subseteq \{0,1,2,3\}$ *with* $|J| = 3$ *is equal to* $2^{-30} + 2^{-50.9803} - 2^{61.415} + ...$

The proof is based on Theorem 3 and by the fact that given the events $A_1, ..., A_n$ in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$

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

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

Proof. As showed in Theorem 3, the probability that two ciphertexts belong to the same coset of a mixed space $\mathcal{M}_J$ for a fixed J with $|J| = 3$ is $p_{AES} \simeq 2^{-32} + 2^{-52.9803}$.

By definition, given the events $A_1, ..., A_n$ in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$

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

where the last sum runs over all subsets J of the indexes $1, ..., n$ which contain exactly k elements[24] and $A_I := \bigcap_{i\in I} A_i$ denotes the intersection of all those A_i with index in I.

Moreover, observe that $\mathcal{M}_I \cap \mathcal{M}_J = \mathcal{M}_{I\cap J}$ for each $I, J \subseteq \{0,1,2,3\}$ by definition, where $\mathcal{M}_I \cap \mathcal{M}_J = \emptyset$ if $I \cap J = \emptyset$. It follows that for $|J| = 3$:

$$\begin{aligned}
&Prob(\exists J \subseteq \{0,1,2,3\}\,|J| = 3 \text{ s.t. } x \in \mathcal{M}_J) = \\
&= \sum_{J\subseteq\{0,1,2,3\},\,|J|=3} Prob(x \in \mathcal{M}_J) - \sum_{I\subseteq\{0,1,2,3\},\,|J|=2} Prob(x \in \mathcal{M}_J) + \\
&+ \sum_{J\subseteq\{0,1,2,3\},\,|J|=1} Prob(x \in \mathcal{M}_J) = 4\cdot p_{AES} - 6\cdot p_{AES}^2 + 4\cdot p_{AES}^3 \simeq \\
&\simeq 2^{-30} + 2^{-50.9803} - 2^{61.415} + ...
\end{aligned}$$

□

For comparison, in the case of a random permutation, the same event has probability $2^{-30} - 2^{61.415} + 2^{-94}$.

E The Computational Cost of Algorithm 1

In this section, we explain the details of Algorithm 1 used in Sect. 6.1 and Sect. 7 to count the different number of pairs of ciphertexts that belong to the same coset of $\mathcal{M}_J$ 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 (1) construct all the possible pairs of ciphertexts and (2) count the number of collisions. First of all, note that the cost to check that a given pair of ciphertexts belong to the same coset of $\mathcal{M}_J$ is equal to the cost of a XOR operation and an inverse MixColumns operation[25].

As we are going to show, the major cost of this distinguisher regards the construction of all the possible different pairs, which corresponds to step (1). Since it is possible to construct approximately 2^{63} pairs for each coset, the simplest way to do it requires 2^{63} table look-ups for each coset. In the following, we present a way to reduce the total cost to approximately $2^{41.6}$ table look-ups, where the used tables are of size 2^{32} texts (or $2^{32}\cdot 16 = 2^{36}$ byte).

[24] For example for $n = 2$, it follows that $Prob(A_1 \cup A_2) = Prob(A_1) + Prob(A_2) - \mathbb{P}(A_1 \cap A_2)$, while for $n = 3$ it follows that $Prob(A_1 \cup A_2 \cup A_3) = Prob(A_1) + Prob(A_2) + Prob(A_3) - Prob(A_1 \cap A_2) - Prob(A_1 \cap A_3) - Prob(A_2 \cap A_3) + Prob(A_1 \cap A_2 \cap A_3)$.

[25] As example, given a pair (c^1, c^2) and for the subspace $\mathcal{M}_{\{1,2,3\}}$, this operation can be reduced to check that $MC^{-1}(c^1 \oplus c^2)_{i,i} = MC^{-1}(c^1)_{i,i} \oplus MC^{-1}(c^2)_{i,i} = 0$ for each $i = 0, ..., 3$ - note that $c^1 \oplus c^2 \in \mathcal{M}_J$ if and only if $MC^{-1}(c^1 \oplus c^2) \in \mathcal{ID}_J$.

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 $c^1 \oplus c^2 \in \mathcal{M}_J$, or equivalently

$$MC^{-1}(c^1)_{i,j-i} = MC^{-1}(c^2)_{i,j-i} \qquad \forall i = 0,1,2,3 \tag{12}$$

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 $\mathbb{N}$):

$$x = c_{0,0-j} + 256 \cdot MC^{-1}(c)_{1,1-j} + MC^{-1}(c)_{2,2-j} \cdot 256^2 + MC^{-1}(c)_{3,3-j} \cdot 256^3.$$

It's simple to observe that if two ciphertexts c^1 and c^2 satisfy (12), 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 (12). The complete pseudo-code of such an algorithm 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 $\mathcal{M}_J$, 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[26] 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.

E.1 Details - Mean Value Distinguishers

The same algorithm is used to implement the "mean value" secret-key distinguisher proposed in Sect. 7. We refer to that section for all the details, and we focus here on the details about the computational cost. As showed in Sect. 7, $2^{47.374}$ 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 $\mathcal{M}_J$ 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

[26] 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 literature.

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.

F Proof of Proposition 2 - Average Number of Collisions for small-scale AES case

In this section we provide the proof of Proposition 2 for the case of small-scale AES. Since the idea of the proof is the same of the one given in Sect. 4, we limit to adapt it to the case of small-scale AES. Since

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

the idea is to work only on the middle round. That is, for 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 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. For 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 2, it follows that $R(p^1) \oplus R(p^2) \notin \mathcal{D}_J$ for each J. For the following we limit to consider the case in which at least 2 generating variables are different.

F.1 Case: Two Equal Generating Variables. As first case, we consider the case in which 2 generating variables are equal, e.g. $x^1 \neq x^2$, $y^1 \neq y^2$, $z^1 = z^2$ and $w^1 = w^2$. This is equivalent to consider 2^8 plaintexts in the same coset of $\mathcal{C}_{0,1} \cap \mathcal{M}_0$ (the other cases are equivalent).

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

$$\begin{aligned} A \cdot (\text{S-Box}(B \cdot x^1 \oplus a) \oplus \text{S-Box}(B \cdot x^2 \oplus a)) \oplus \\ \oplus C \cdot (\text{S-Box}(D \cdot y^1 \oplus c) \oplus \text{S-Box}(D \cdot y^2 \oplus c)) = 0 \end{aligned} \tag{13}$$

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

$$\begin{aligned} \text{S-Box}(x \oplus \Delta_I) \oplus \text{S-Box}(x) &= \Delta_O \\ \text{S-Box}(y \oplus \Delta'_I) \oplus \text{S-Box}(y) &= \Delta'_O \\ \Delta_O = A^{-1} \cdot C \cdot \Delta''_O & \end{aligned} \tag{14}$$

Due to the same argumentations given in Sect. 4, the number of different (not null) solutions $[(x^1, y^1), (x^2, y^2)]$ of eq. (14) is approximately given by

$$\frac{1}{2} \cdot 15 \cdot \left(\frac{16}{15} \cdot 15\right)^2 = 15 \cdot 2^7$$

independently of the details of the S-Box. Indeed, observe that given $\Delta_O \neq 0$, each one of the two equations (14) for small-scale AES admit $\frac{16}{15} \cdot 15 = 16$ different solutions $(\hat{x}, \Delta_I)$ - resp. $(\hat{y}, \Delta'_I)$ - where $\Delta_I, \Delta'_I \neq 0$ and 16/15 is the average number of solutions. Moreover, note that there are 15 values of $\Delta_O \neq 0$ and that the condition $y^1 < y^2$ holds.

Given the number of solutions of eq. (14), *what is the number of common - different (not null) - solutions* $[(x^1, y^1), (x^2, y^2)]$ *of 4 equations of the form (13)?* Due to the same argumentation given in Sect. 4, this probability is equal to $(16 \cdot 15)^{-1} \cdot (15 \cdot 8)^{-1} = 15^{-2} \cdot 2^{-7}$.

In conclusion, the number of common - different (not null) - solutions $[(x^1, y^1), (x^2, y^2)]$ of 4 equations of the form (13) is approximately given by

$$(15 \cdot 2^7)^4 \cdot (15^{-2} \cdot 2^{-7})^3 = \frac{2^7}{15^2} \simeq 0.568888889$$

For comparison, if the ciphertexts are generated by a random permutation, the number of different pairs of ciphertexts that belong to the same coset of $\mathcal{M}_J$ is

given by

$$\binom{2^8}{2} \cdot (2^{-4})^4 = \frac{2^8 - 1}{2^9} \simeq 0.498046875$$

F.2 Case: One Equal Generating Variable. As second case, we consider the case in which 1 generating variable is equal, e.g. $x^1 \neq x^2$, $y^1 \neq y^2$, $z^1 \neq z^2$ and $w^1 = w^2$. This is equivalent to consider 2^{12} plaintexts in the same coset of $\mathcal{C}_{0,1,2} \cap \mathcal{M}_0$ (the other cases are equivalent).

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

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

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 $\Delta'_O = A^{-1} \cdot C \cdot \Delta_O$, 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

$$\binom{3}{1} \cdot 2^4 \cdot \frac{2^7}{15^2} = \frac{2048}{75} \simeq 27.306667$$

Second Case. For the second case, the idea is to work as for the cases of 1 equal generating variables. For each eq. (15) 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 $15 \cdot 14 \cdot \frac{1}{2} \cdot \left(15 \cdot \frac{16}{15}\right)^3 = 105 \cdot 2^{12}$. Moreover, using the same argumentation as before, the probability to have a common solution for two equations of the form (15) is given by $(16 \cdot 15)^{-2} \cdot (8 \cdot 15)^{-1} = 15^{-3} \cdot 2^{-11}$ under the given assumptions of the S-Box. It follows that we expect on average

$$(105 \cdot 2^{12})^4 \cdot (15^{-3} \cdot 2^{-11})^3 = \frac{7^4 \cdot 2^{15}}{15^5} \simeq 103.60621$$

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

Total Number of Different - not null - Common Solutions in the Case of One Equal Generating Variable. 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

$$103.60621 + 27.306667 \simeq 130.912877$$

For comparison, if the ciphertexts are generate by a random permutation, the number of different pairs of ciphertexts that belong to the same coset of $\mathcal{M}_J$ is approximately given by

$$\binom{2^{12}}{2} \cdot 2^{-16} \simeq 127.96875$$

F.3 Case: No Equal Generating Variables. Finally, we consider the case in which all the generating variables are different, that is $x^1 \neq x^2$, $y^1 \neq y^2$, $z^1 \neq z^2$ 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 $\Delta'_O = A^{-1} \cdot C \cdot \Delta_O \neq 0$ 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

$$\binom{4}{2} \cdot 16^2 \cdot \frac{2^7}{15^2} = \frac{65\,536}{75} \simeq 873.813$$

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

$$\binom{4}{1} \cdot 16 \cdot \frac{7^4 \cdot 2^{15}}{15^5} = \frac{7^4 \cdot 2^{21}}{15^5} \simeq 6\,630.798$$

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 $15^3 - 15 \cdot 14 = 3\,165$, since 15^3 is the total number of values and $15 \cdot 14$ 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 can not be equal zero since $\Delta_O^{''} \neq 0$). As a result, the total number of solutions for this case is

$$\frac{1}{2} \cdot 3\,165 \cdot \left(15 \cdot \frac{16}{15}\right)^4 = 3\,165 \cdot 2^{15}.$$

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 $(15 \cdot 16)^{-3} \cdot (15 \cdot 8)^{-1} = 15^{-4} \cdot 2^{-15}$, the average number of (non null) common solutions with no equal generating variables is

$$(3\,165 \cdot 2^{15})^4 \cdot (15^{-4} \cdot 2^{-15})^3 = \frac{211^4 \cdot 2^{15}}{15^8} \simeq 25\,342.513$$

Total Number of Different - not null - Common Solutions in the Case of No Equal Generating Variables. 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

$$25\,342.513 + 6\,630.798 + 873.813 \simeq 32\,847.124$$

For comparison, if the ciphertexts are generated by a random permutation, the number of different pairs of ciphertexts that belong to the same coset of $\mathcal{M}_J$ is approximately given by

$$\binom{2^{16}}{2} \cdot 2^{-16} \simeq 32\,767.5$$

In other words, on average there are

$$32\,847.124 - 32\,767.5 \simeq 79.624$$

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

Finally, since the number of possible couples of texts is $2^{15} \cdot (2^{16} - 1)$, the probability in 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

$$p_{AES} \simeq \frac{32\,874.124}{2^{15} \cdot (2^{16} - 1)} \simeq 2^{-16} + 2^{-24.68485},$$

versus 2^{-16} of the random case.

G Proof of Proposition 2 - Variance of the Number of Collisions for small-scale AES

In this section we provide the proof of Proposition 2 for the case of small-scale AES. Since the proof is similar to the one given in Sect. 5.2, we limit to adapt it to the case of small-scale AES. As before, since

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

the idea is to work only on the middle round. That is, for the following we consider 2^{16} plaintexts in the same coset of $\mathcal{M}_i$ for $i \in \{0, 1, 2, 3\}$ and we study the average number of collisions after one round in the same coset of $\mathcal{D}_J$ for $|J| = 3$ fixed.

First of all, note that given a coset of $\mathcal{M}_i$ of 2^{16} chosen plaintexts, it is possible to construct $2^{15} \cdot (2^{16} - 1)$ pairs. Among them, the number of pairs of texts with $0 \leq n \leq 3$ equal generating variables are

$$\binom{4}{n} \cdot 2^{15} \cdot (2^4 - 1)^{4-n}.$$

To prove the result, the idea is to consider separately the pairs of texts with $0 \leq n \leq 3$ different generating variables (see proof in Sect. 5.2).

Different Generating Variables. As first case, we consider the case in which all the generating variables are different, i.e. $n = 0$. The number of pairs with this property is $2^{15} \cdot (2^4 - 1)^4$.

As we have just seen, these pairs are not independent. Indeed, by [20], it is possible to divide them in $2^{15} \cdot (2^4 - 1)^4 / 8 = 2^{12} \cdot (2^4 - 1)^4$ sets of 8 pairs such that for each set only two events can happen: (1) all the pairs belong to the same coset of $\mathcal{D}_J$ after one round or (2) no one of them has this property. Thus, the idea is to consider only one pair for each one of these sets, for a total of $2^{12} \cdot (2^4 - 1)^4$ pairs. Since these pairs are independent, the probabilistic distribution of the number of

pairs that belong to the same coset of $\mathcal{D}_J$ after one round is given by a binomial distribution X with mean value $\mu = n \cdot p_{AES} = 2^{12} \cdot (2^4 - 1)^4 \cdot (2^{-16} + 2^{-24.68485})$ and variance $\sigma^2 = n \cdot p_{AES} \cdot (1 - p_{AES}) = 2^{12} \cdot (2^4 - 1)^4 \cdot (2^{-16} + 2^{-24.68485}) \cdot (1 - 2^{-16} - 2^{-24.68485}) \approx 3\,171.702$, that is $\sigma \approx 56.318$. It follows that the probabilistic distribution Y of the number of collisions for the pairs with no equal generating variables is simple given by $Y = 8 \times X$. Since $Var(Y) = 8^2 \times Var(X)$, it follows that the standard deviation σ for this case is given by $8 \cdot 56.318 \approx 450.543$.

One Equal Generating Variable. As second case, we consider the case in which all the generating variables are different, i.e. $n = 1$. The number of pairs with this property is $4 \cdot 2^{15} \cdot (2^4 - 1)^3$.

As we have just seen, these pairs are not independent. Indeed, by [20], it is possible to divide them in $2^{17} \cdot (2^4 - 1)^3 / 2^6 = 2^{11} \cdot (2^4 - 1)^3$ sets of 2^6 pairs such that for each set only two events can happen: (1) all the pairs belong to the same coset of $\mathcal{D}_J$ after one round or (2) no one of them has this property. Thus, the idea is to consider only one pair for each one of these sets, for a total of $2^{11} \cdot (2^4 - 1)^3$ pairs. Since these pairs are independent, the probabilistic distribution of the number of pairs that belong to the same coset of $\mathcal{D}_J$ after one round is given by a binomial distribution X with mean value $\mu = n \cdot p_{AES} = 2^{11} \cdot (2^4 - 1)^3 \cdot (2^{-16} + 2^{-24.68485})$ and variance $\sigma^2 = n \cdot p_{AES} \cdot (1 - p_{AES}) = 2^{11} \cdot (2^4 - 1)^3 \cdot (2^{-16} + 2^{-24.68485}) \cdot (1 - 2^{-16} - 2^{-24.68485}) \approx 105.723$, that is $\sigma \approx 10.282$. It follows that the probabilistic distribution Y of the number of collisions for the pairs with no equal generating variables is simple given by $Y = 2^6 \times X$. Since $Var(Y) = (2^6)^2 \times Var(X)$, it follows that the standard deviation σ for this case is given by $2^6 \cdot 10.282 \approx 658.06$.

Two Equal Generating Variables. As third case, we consider the case in which all the generating variables are different, i.e. $n = 2$. The number of pairs with this property is $6 \cdot 2^{15} \cdot (2^4 - 1)^2$.

As we have just seen, these pairs are not independent. Indeed, by [20], it is possible to divide them in $3 \cdot 2^{16} \cdot (2^4 - 1)^2 / 2^9 = 3 \cdot 2^7 \cdot (2^4 - 1)^2$ sets of 2^9 pairs such that for each set only two events can happen: (1) all the pairs belong to the same coset of $\mathcal{D}_J$ after one round or (2) no one of them has this property. Thus, the idea is to consider only one pair for each one of these sets, for a total of $3 \cdot 2^7 \cdot (2^4 - 1)^2$ pairs. Since these pairs are independent, the probabilistic distribution of the number of pairs that belong to the same coset of $\mathcal{D}_J$ after one round is given by a binomial distribution X with mean value $\mu = n \cdot p_{AES} = 3 \cdot 2^7 \cdot (2^4 - 1)^2 \cdot (2^{-16} + 2^{-24.68485})$ and variance $\sigma^2 = n \cdot p_{AES} \cdot (1 - p_{AES}) = 3 \cdot 2^7 \cdot (2^4 - 1)^2 \cdot (2^{-16} + 2^{-24.68485}) \cdot (1 - 2^{-16} - 2^{-24.68485}) \approx 1.322$, that is $\sigma \approx 1.15$. It follows that the probabilistic distribution Y of the number of collisions for the pairs with no equal generating variables is simple given by $Y = 2^9 \times X$. Since $Var(Y) = (2^9)^2 \times Var(X)$, it follows that the standard deviation σ for this case is given by $2^9 \cdot 1.15 \approx 588.587$.

Final Result. Finally, remember that given two plaintexts with three equal generating variables, then they can not belong to the same coset of $\mathcal{D}_J$ after

one round. Moreover, note that all the previous cases are independent. In other words, consider one pairs of texts with n equal generating variables and another one with $\hat{n}$ equal generating variables where $\hat{n} \neq n$. The fact that the first pair belong (or not) to the same coset of $\mathcal{D}_J$ after one round is independent by the fact that the second pair belong (or not) to the same coset of $\mathcal{D}_J$ after one round.

Remember that given n independent variables $X_1, ..., X_n$, the variance of $Y = X_1 + ... + X_n$ is given by $Var(Y) = Var(X_1) + ... + Var(X_n)$. It follows that the total variance of the probabilistic distribution for the AES case is given by $\sigma^2 \simeq 588.587^2 + 658.06^2 + 450.543^2 \simeq 982\,466.615$, or equivalently the standard deviation is given by $\sigma \simeq 991.195$.

H 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.

To give an overview, consider the following aspect. To construct the proposed distinguishers, one consider a full coset of a subspace $\mathcal{D}_I$ - 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 doesn't 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. Instead to have 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 guess key

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

$$... \xrightarrow{R(\cdot)} 2^{24} \text{ texts in } \mathcal{D}_{0,2,3} \cap \mathcal{C}_0 \oplus a \xrightarrow{R^4(\cdot)} \text{ distinguisher property.}$$

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

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}	[17, Sec. 7.5.1]
Imp. Polytopic	5	15	2^{70}	2^{41}	[27]
Partial Sum	5	2^{8}	2^{38}	small	[28]
Integral (EE)	5	2^{11}	$2^{45.7}$	small	[15]
Imp. Differential	5	$2^{31.5}$	2^{33} $(+2^{38})$	2^{38}	[6]
Integral (EB)	5	2^{33}	$2^{37.7}$	2^{32}	[15]
Variance	**5**	$\mathbf{2^{33}}$	$\mathbf{2^{64.6}}$	$\mathbf{2^{32}}$	**App. H.4**
Mixture Diff.	5	$2^{33.6}$	$2^{33.3}$	2^{34}	[19]
Multiple-of-n	**5**	$\mathbf{2^{33.6}}$	$\mathbf{2^{48}}$	$\mathbf{2^{32}}$	**App. H.2**
Trunc. Diff.	**5**	$\mathbf{2^{36}}$	$\mathbf{2^{67.6}}$	$\mathbf{2^{32}}$	**App. H.3**

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

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

- 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.

H.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 less or equal 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.

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

For this reason, we limit 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 $\{R_k \circ R_{\hat{k}}^{-1}(t^i)\}_i$ don't belong to the same coset of $\mathcal{D}_{0,2,3} \cap \mathcal{C}_0$, and the distinguisher property doesn't work. In this case, the behavior is the same of a random permutation, and the attacker can filter wrong keys.

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 6. *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. 6.

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.

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. 6; `// ⪯ 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.*

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).

Practical Tests on small-scale AES. All the attacks that we are going to present have been practically tested on small-scale AES[27]. The practical results are in accordance with the theoretical ones.

H.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 [20], the number of different pairs of ciphertexts that belong to the same coset of $\mathcal{M}_K$ for

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

$|K| = 3$ 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 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? 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, using $n \geq 32$ it is possible to filter all the wrong keys. Since a coset of $\mathcal{D}_j$ contains 2^{16} different cosets of $\mathcal{D}_I \cap \mathcal{C}_j$ after one round, it follows that 2^{32} chosen plaintexts in the same coset of $\mathcal{D}_j$ are sufficient to find one diagonal. Using this strategy to find three diagonals of the key (one diagonal is found by brute force), the data complexity is $2^{33.6}$ chosen plaintexts.

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.

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

In this subsection, 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. 4, 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. Assume that the goal is to find the right key with probability bigger than 95%[28], 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 given by formula (11) 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 $n \geq 2^{59.43}$. Since each coset of $\mathcal{D}_I \cap \mathcal{C}_j$ contains approximately 2^{47} different pairs after one round, one needs approximately $2^{12.43}$ different initial cosets or approximately $2^{34.43}$ 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 to find the entire key is of $2 \cdot 2^{34.43} = 2^{35.5}$ chosen plaintexts.

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 $2^{12.43} \cdot 2^{32} \cdot 2^{24} \cdot (2 + \log 2^{24}) \simeq 2^{73.1}$ table look-ups. Thus the total cost is $2 \cdot 2^{73.1} \cdot (5 \cdot 20)^{-1} + 2^{64} \simeq 2^{67.6}$ five-round encryption to find the entire key (by assuming 20 table look-ups $\approx$ 1 encryption).

H.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. 5.2. 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

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

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

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

pairs with different generating variables. The variance is given by

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

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

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

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

As for the secret-key distinguisher of Sect. 6.1, 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 $\geq 2^8$ 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 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 to find one diagonal is well approximated by $2^{32} \cdot 4 \cdot 2^6 \cdot 2^{24} \cdot (2 + \log 2^{24}) \simeq 2^{68.7}$ table look-ups, that is the total cost is well approximated by $2 \cdot 2^{68.7} \cdot (100)^{-1} + 2^{64} \simeq 2^{64.6}$ five-round encryption to find the entire key by assuming 20 table look-ups ≈ 1 encryption.

I Details of used S-Box

In this section, we recall the main information of the S-Box used in Sect. 8 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 S-Box (x)	6	B	5	4	2	E	7	A	9	D	F	C	3	1	0	8
PRINCE S-Box (x)	B	F	3	2	A	C	9	1	6	7	8	0	E	5	D	4
KLEIN S-Box (x)	7	4	A	9	1	F	B	0	C	3	2	6	8	E	D	5
Midori SB_0 S-Box(x)	C	A	D	3	E	B	F	7	8	9	1	5	0	2	4	6
Midori SB_1 S-Box(x)	1	0	5	3	E	2	F	7	D	A	9	B	C	8	4	6
PRESENT S-Box (x)	C	5	6	B	9	0	A	D	3	E	F	8	4	7	1	2
RECTANGLE S-Box (x)	6	5	C	A	1	E	7	9	B	0	3	D	8	F	4	2
NOEKON S-Box (x)	7	A	2	C	4	8	F	0	5	9	1	e	3	D	B	6
PRIDE S-Box (x)	0	4	8	F	1	5	E	9	2	7	A	C	B	D	6	3
Toy-6 S-Box(x)	1	3	6	4	2	5	9	A	0	F	7	E	C	B	D	8
Toy-8 S-Box(x)	1	3	6	4	2	5	A	C	0	F	7	8	E	B	D	9
Toy-10 S-Box(x)	6	4	C	5	0	7	2	E	1	F	3	D	8	A	9	B

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).

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