A New Structural-Differential Property of 5-Round AES

Lorenzo Grassi[1], Christian Rechberger[1,3] and Sondre Rønjom[2,4]

[1] IAIK, Graz University of Technology, Austria
[2] Nasjonal sikkerhetsmyndighet, Norway
[3] DTU Compute, DTU, Denmark
[4] Department of Informatics, University of Bergen, Norway
firstname.lastname@iaik.tugraz.at, Sondre.Ronjom@ii.uib.no

Abstract. AES is probably the most widely studied and used block cipher. Also versions with a reduced number of rounds are used as a building block in many cryptographic schemes, e.g. several candidates of the CAESAR competition are based on it.
So far, non-random properties which are independent of the secret key are known for up to 4 rounds of AES. These include differential, impossible differential, and integral properties.
In this paper we describe a *new structural property for up to 5 rounds of AES*, differential in nature and which is independent of the secret key, of the details of the MixColumns matrix (with the exception that the branch number must be maximal) and of the SubBytes operation. It is very simple: By appropriate choices of difference for 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 is *always* a multiple of 8.
We not only observe this property experimentally (using a small-scale version of AES), we also give a detailed proof as to why it has to exist. As a first application of this property, we describe a way to distinguish the 5-round AES permutation (or its inverse) from a random permutation with only 2^{32} chosen texts that has a computational cost of $2^{35.6}$ look-ups into memory of size 2^{36} bytes which has a success probability greater than 99%.

Keywords: Block cipher, Permutation, AES, Secret-Key Distinguisher

1 Introduction

Block ciphers play an important role in symmetric cryptography providing the basic tool for encryption. They are the oldest and most scrutinized cryptographic tools. Consequently, they are the most trusted cryptographic algorithms that are often used as the underlying tool to construct other cryptographic algorithms,

This is the extended version of the article which appears in the proceedings of EUROCRYPT 2017. It includes a more formal description of the main result based on the subspace trail notation [15] recently introduced at FSE 2017.

whose proofs of security are performed under the assumption that the underlying block cipher is ideal.

While the security of public-key encryption schemes are related to the hardness of well-defined mathematical problems, informally a block cipher is considered secure if an (efficient) adversary, with access to the encryptions of messages of its choice, cannot tell apart those encryptions from the values of a truly random permutation. In other words, this means that an (efficient) adversary, with access to the encryptions of messages of its choice, cannot tell the difference between the block cipher (equipped with a random key) and a truly random permutation. This notion of block cipher security was introduced and formally modeled by Luby and Rackoff [20] in 1988, and it was motivated by the design of DES. To be a bit more precise (but without going into the details), a secret key distinguisher is one of the weakest cryptographic attacks that can be launched against a secret-key cipher. In this attack, there are two oracles: one that simulates the cipher for which the cryptographic key has been chosen at random and the other simulates a truly random permutation. The adversary can query both oracles and his task is to decide which oracle is the cipher and which is the random permutation. The attack is considered to be successful if the number of queries required to make a correct decision is below a well defined level.

The Rijndael block cipher [9] has been designed by Daemen and Rijmen in 1997 and was chosen as the AES (Advanced Encryption Standard) by NIST in 2000. Nowadays, it is probably the most used and studied block cipher. The possibility to set up a *secret key distinguisher for 5-round of AES* that exploits a property which is *independent of the secret key* was already considered in [22] and improved in [15]. However, only partial solutions have been proposed and the problem is still open. As we will argue below, the solutions so far are partial because the distinguishers are derived from a key-recovery attack and they actually exploit as property the existence of a sub-key for which a property on 4 rounds holds.

In this paper, we present (and practical verify) the *first secret-key distinguisher for 5-round AES* which exploits a new structural/differential property which is independent of the secret key, that is a property that can be verified without needing to know or to get to know any information of the secret key. As we are going to show, it requires 2^{32} chosen plaintexts/ciphertexts and has a computational cost of $2^{35.6}$ table look-ups.

1.1 Secret-Key Distinguishers for AES-128

In the usual security model, the adversary is given a *black box* (oracle) access to an instance of the encryption function associated with a random secret key and its inverse. The goal is to find the key or more generally to efficiently distinguish the encryption function from a random permutation.

More formally, a block cipher is a family of functions $E : \mathcal{K} \times \mathcal{S} \to \mathcal{S}$, with $\mathcal{K}$ a finite set called the key space and $\mathcal{S}$ a finite set called the domain or message space. For every $k \in \mathcal{K}$, the function $E_k(\cdot) = E(k, \cdot)$ is a permutation. The inverse of the block cipher E is defined as a function $E^{-1} : \mathcal{K} \times \mathcal{S} \to \mathcal{S}$ that

satisfies $E_k^{-1}(E_k(s)) = s$ for each $k \in \mathcal{K}$ and for each $s \in \mathcal{S}$. A block cipher $E_k(\cdot)$ with key space $\mathcal{K}$ is a (q, t, ε)-pseudorandom permutation (PRP) if any adversary making at most q oracle queries and running in time at most t can distinguish E_k (for a random key k) from a uniformly random permutation with advantage at most ε.

Definition 1. *Let E be block cipher defined as before, and $Perm(\mathcal{S})$ be the set of all permutations of $\mathcal{S}$. Let D be a distinguisher with oracle access to a permutation and its inverse, and returning a single bit. The (Strong PseudoRandom Permutation) SPRP-advantage of D against E is defined as*

$$\mathbf{Adv}_E^{sprp}(D) = |Prob(\pi \leftarrow Perm(\mathcal{S}) : D^{\pi(\cdot),\pi^{-1}(\cdot)} = 1) - Prob(k \leftarrow \mathcal{K} : D^{E_k(\cdot),E_k^{-1}(\cdot)} = 1)|.$$

For integers q and t, the SPRP-advantage of E is defined as

$$\mathbf{Adv}_E^{sprp}(q, t) = \max_D \mathbf{Adv}_E^{sprp}(D),$$

where the maximum is taken over all distinguishers making at most q oracle queries and running in time at most t. E is a (q, t, ε)-SPRP if $\mathbf{Adv}_E^{sprp}(q, t) \leq \varepsilon$.

Note that if $Adv_E(D) \simeq 0$, then the $E_k(\cdot)$ behaves (exactly) like a random permutation from the distinguisher point of view.

Before we focus on the 5-round distinguisher, we briefly summarize in Sect. 3 the properties exploited by the secret key distinguisher on AES-like permutations up to 4 rounds. We stress that, even if a key-recovery attack can also be used as a secret key distinguisher in this paper we focus only on secret-key distinguisher that are independent of the secret key.

The most competitive secret-key distinguishers up to 3-round are based on the differential [5] and on the truncated differential cryptanalysis [18]. These distinguishers exploit the fact that some r-round differential characteristics exist with higher probability for an AES permutation than for a random one. In [8], Daemen *et al.* proposed an attack vector that uses a 3-round distinguisher to attack up to 6 rounds of the cipher and later became known as integral attacks. In an integral distinguisher, given inputs with particular properties, one exploits the fact that the sum of the corresponding ciphertexts is zero with probability 1 for an AES permutation, while this happens with a (much) lower probability for a random permutation. Finally, another possible distinguisher exploits the impossible-differential cryptanalysis, which was independently proposed by Knudsen [19] and by Biham *et al.* [3]. In impossible-differential cryptanalysis, the idea is to exploit the fact that some differential trails hold with probability 0 for an AES permutation (i.e. impossible differential trails), while they have probability greater than 0 for a random permutation.

5-Round "Distinguisher" for AES-128: State of Art. A distinguisher for five rounds of AES-128 has been recently proposed by Sun, Liu, Guo, Qu,

and Rijmen at Crypto 2016 [22]. This distinguisher - which requires the *whole* input-output space to work - has been improved in [15], where authors set up a secret key distinguisher in the same setting of the one proposed in [22], but which requires only $2^{98.2}$ chosen plaintexts.

Both these two distinguishers are derived by a key-recovery attack on AES-128 with a secret S-Box. In particular, they are able to distinguish a random permutation from an AES one exploiting the existence of a (secret) key for which a property on 4-round is verified. In more details, the property on 4-round used in [22] is the balance property, while the one used in [15] is the impossible differential one. With respect to a classical key-recovery attack, these distinguishers require the knowledge only of a single byte of the secret subkey to distinguish an AES permutation with a secret S-Box from a random one.

For a complete comparison with the distinguisher presented in this paper, we briefly recall how they are set up, and we refer to [22] and [15] for a complete discussion. In both cases, authors first assume to know the difference of two bytes (i.e. 1 byte) of one secret subkey. Using this knowledge, they are able to extend a four rounds distinguisher to five rounds. In order to turn these distinguishers into secret-key ones, the idea is simply to iterate these distinguishers on all the 2^8 possible values of the difference of these two bytes of the secret subkey. The idea is that for an AES permutation there exists one difference of these two bytes for which a property (which is independent of the secret key) on 4-round is satisfied, while for a random permutation this property on 4-round is never satisfied (with high probability) for any of the 2^8 possible values.

We stress that both these distinguishers require to find part of the secret key in order to verify a property on 4 rounds, i.e. they work as key-recovery attacks. Note that the research of a secret-key distinguisher which is independent of the secret key is of particular interest and importance since it (theoretically) allows to set up key recovery attacks, as it already happened for the secret-key distinguishers up to 4 rounds just described. Moreover, we highlight that both these distinguishers are independent of the details of the S-Box, but they depend on the details of the MixColumns matrix (in particular, they exploit the fact that for at least one column of the MixColumns matrix or its inverse two elements are identical).

1.2 Our Result: the First 5-Round Secret-Key Distinguisher for AES-128 Independent of the Secret Key

The results presented in the previous two papers don't solve the problem to set up *a 5-round secret key distinguisher of AES which exploits a property which is independent of the secret key.* In Sect. 4 of this paper, we provide a solution to this problem, that is we propose the *first* secret-key distinguisher on 5-round AES which exploits a new property which is independent of the secret key and of the details of the S-Box. To present this new distinguisher in an easy and

natural way, we use the subspace trail notation[1] introduced at FSE 2017 in [15], which is briefly recalled in Sect. 3.

The high-level idea is very easily described. By appropriate choices of difference for 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 is *always* a multiple of 8.

More concretely, suppose to use a coset of a particular subspace $\mathcal{D}$ of the plaintexts space, and the corresponding ciphertexts after 5 rounds. Let $\mathcal{M}$ be a particular subspace of the ciphertexts space. The idea is to count the total number of different ciphertext pairs that belong to the same coset of this subspace $\mathcal{M}$. As we show in detail in the paper, for an AES permutation this number can only be a multiple of 8 (independently of the dimensions of $\mathcal{D}$ and of $\mathcal{M}$), while it does not have any particular property for the case of a random permutation. As we will see in the comparison, the resulting distinguisher proposed in this paper is much more efficient than those proposed earlier, it *works both in the encryption and in the decryption mode of AES and it does not depend on the details of the MixColumns matrix (with the exception that the branch number must be five) or/and of the SubBytes operation.* A formal statement of this property used by our distinguisher is given in Theorem 3 in Sect. 4.1, and its detailed proof is given in Sect. 6.

Comparison with 4-Round Secret-Key Distinguishers. These last properties also highlight a difference between our new distinguisher and the others currently used in literature. In most cases, especially in the cryptanalysis of AES, one does not have the necessity to investigate the details of the S-Boxes. Consider for example the 4-round secret-key distinguishers, based on the integral [14] and on the impossible-differential [4] properties. In the first one, given a set of chosen plaintexts of which part is held constant and another part varies through all possibilities, it is possible to prove that their XOR-sum after 4-round is always equal to 0. In the second one, given a set of chosen plaintexts with analogous properties, it is possible to prove that the difference of each possible pair of ciphertexts after 4-round can not take some values (some differences have prob. 0, i.e. they are impossible). In both cases, the corresponding results are independent of the key and of the non-linear components. That is, if some other S-Boxes with similar differential/linear properties are chosen in a cipher, the corresponding cryptanalytic results remain the same.

Although there are already 4-round impossible differentials and zero-correlation linear hulls for AES, the effort to find new impossible differentials and zero-correlation linear hulls that could cover more rounds has never been stopped. In Eurocrypt 2016, Sun *et al.* [23] proved that, unless the details of the S-Boxes are

[1] *Our choice to use the subspace trail notation* to present our new distinguisher on 5-round AES is *motivated* by the fact that such notation allows to describe it in an easier and more formal way than using the "classical" one. An example of this fact is given in [15], where all the secret-key distinguisher up to 4-round AES are re-described using the subspace trail notation.

Table 1. *5-round Secret-Key Distinguishers for AES with a Single Secret S-Box.* In this table, we limit to consider the distinguishers that exploit a property which is independent of the key, or which are derived by a key-recovery attack but are independent of the S-Box and require the knowledge of only part of the key. The complexity is measured in minimum number of chosen plaintexts CP or/and chosen ciphertexts CC which are needed to distinguish the AES permutation from a random one with probability higher than 99%. Time complexity is measured in memory accesses (M) or XOR operations (XOR). The case in which the final MixColumns operation is omitted is denoted by "r.5 rounds", that is r full rounds and the final round. "Key-Independence" denotes a distinguisher which is able to distinguish 5-round AES from a random permutation without discovering any information of the secret key or of part of it.

Property	**Rounds**	**Data**	**CP**	**CC**	**Cost**	**Key-Independence**	**Ref.**
Subspace Trail	$\mathbf{4.5-5}$	$\mathbf{2^{32}}$	✓	✓	$\mathbf{2^{35.6}}$ **M**	✓	**Sect. 4**
Impossible Diff.	$4.5-5$	$2^{98.2}$	✓		2^{107} M		[15]
Integral	5	2^{128}		✓	2^{128} XOR		[22]

exploited, one cannot find any impossible differential or zero-correlation linear hull of the AES that covers 5 or more rounds. Moreover, due to the link among impossible differential, integral and zero correlation linear cryptanalysis [24], an analogous result holds also for the integral case. On the other hand, our new property presented in this paper holds up to 5-round of AES independently of the key and of the details of the S-Box (and of the MixColumns operation), and allows to answer an almost 20-year old problem: given a set of chosen plaintexts similar to the one used by the integral and impossible differential distinguishers just recalled, is there any property which is independent of the secret key after 5-round AES?

Comparison of 5-Round Secret-Key Distinguishers. For a better comparison between this new secret-key distinguisher proposed in this paper and earlier ones, we propose to classify the secret-key distinguishers in the following way (from strongest to weakest):

1. a distinguisher which is completely independent of the secret key (e.g., it exploits properties that are not related to the existence of a key) and independent of the details of the S-Box;
2. a distinguisher which depends on the existence of a key and is derived by a key-recovery attack.

A comparison between our new distinguisher and the ones proposed in [22] and [15] is given in Table 1, where "Key-Independence" denotes a secret-key distinguisher which is derived by a key-recovery attack, i.e. that does not exploit a property which is independent of the secret key. Moreover, with respect to the previous classification, a complete comparison of all the secret-key distinguishers and key recovery attacks (used as distinguishers) for 5-round AES is provided in Table 2 - App. C.

2 Preliminary - Description of AES

The Advanced Encryption Standard [9] 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 fields $\mathbb{F}_{256}$, 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 (it 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 M_{MC} (MC and SR provide diffusion in the cipher[2]);
- *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 could be omitted.

Finally, as we don't use the details of the AES key schedule in this paper, we refer to [9] for a complete description.

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 k^r the key of the r-th round, where k^0 is the secret key. If only the key of the final round is used, then we denote it by k to simplify the notation. Finally, we denote by R one round of AES, while we denote r rounds of AES by R^r. We sometimes use the notation R_K instead of R to highlight the round key K. As last thing, in the paper we often use the term "partial collision" (or "*collision*") when two texts belong to the same coset of a given subspace X.

3 Subspace Trails Cryptanalysis

Subspace trails [15] - recently introduced at FSE 2017 - are a generalization of invariant subspaces and allow to express techniques such as truncated differentials, impossible differentials, or integral properties in the same framework.

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. This concept can be generalized to *trails of subspaces*.

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

Definition 2. *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. If all the previous relations hold with equality, the trail is called a constant-dimensional subspace trail.

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 [15] for more details about the concept of subspace trails. Our treatment here is however meant to be self-contained.

3.1 Subspace Trails of AES

In this section, we recall the subspace trails of AES presented in [15]. 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). We also 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 $a \oplus b \in X$.

Definition 3. *The column spaces $\mathcal{C}_i$ are defined as*

$$\mathcal{C}_i = \langle e_{0,i}, e_{1,i}, e_{2,i}, e_{3,i} \rangle.$$

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

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

Definition 4. *The diagonal spaces $\mathcal{D}_i$ are defined as*

$$\mathcal{D}_i = SR^{-1}(\mathcal{C}_i) = \langle e_{0,i \pmod 4}, e_{1,(i+1) \pmod 4}, e_{2,(i+2) \pmod 4}, e_{3,(i+3) \pmod 4} \rangle.$$

Similarly, the inverse-diagonal spaces $\mathcal{ID}_i$ are defined as

$$\mathcal{ID}_i = SR(\mathcal{C}_i) = \langle e_{0,i \pmod 4}, e_{1,(i-1) \pmod 4}, e_{2,(i-2) \pmod 4}, e_{3,(i-3) \pmod 4} \rangle.$$

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

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

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

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

$$\mathcal{M}_i = MC(\mathcal{ID}_i).$$

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

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

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

Definition 6. *Let $I \subseteq \{0,1,2,3\}$. The subspaces $\mathcal{C}_I$, $\mathcal{D}_I$, $\mathcal{ID}_I$ and $\mathcal{M}_I$ are defined as follow:*

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

As shown in detail in [15]:

- 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 $R(\mathcal{C}_I \oplus a) = \mathcal{M}_I \oplus b$.

This simply states that a coset of a sum of diagonal spaces $\mathcal{D}_I$ encrypts to a coset of a corresponding sum of column spaces. Similarly, a coset of a sum of column spaces $\mathcal{C}_I$ encrypts to a coset of the corresponding sum of mixed spaces. It follows that:

Theorem 1. *For each I and for each $a \in \mathcal{D}_I^{\perp}$, there exists one and only one $b \in \mathcal{M}_I^{\perp}$ such that*

$$R^2(\mathcal{D}_I \oplus a) = \mathcal{M}_I \oplus b. \tag{1}$$

We refer to [15] for a complete proof of this theorem. Observe that b depends on a and on the secret key k, and that this theorem does not depend on the particular choice of the S-Box (i.e. it is independent of the details of the S-Box).

Observe that if X is a generic 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 $I \subseteq \{0,1,2,3\}$:*

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

We finally recall that for each $I, J \subseteq \{0,1,2,3\}$:

$$\mathcal{M}_I \cap \mathcal{D}_J = \{0\} \qquad \text{if and only if} \qquad |I| + |J| \leq 4, \tag{3}$$

as demonstrated in [15]. It follows that:

Theorem 2. *Let $I, J \subseteq \{0,1,2,3\}$ such that $|I| + |J| \leq 4$. For all x, y with $x \neq y$:*

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

For the following, note that two texts t^1 and t^2 belong in the same coset of $\mathcal{D}$ if the bytes of their difference $t^1 \oplus t^2$ that lie on n diagonals[3] for $n \leq 3$ (depending on the dimension of $\mathcal{D}$) are equal to zero. As example, $t^1 \oplus t^2 \in \mathcal{D}_i$ if and only if t^1 and t^2 have equal bytes on the i-th diagonal, or in other words if $t^1_{j,i+j} = t^2_{j,i+j}$ for each $j = 0, 1, 2, 3$ where the index $i + j$ is computed modulo 4. In a similar way, two texts t^1 and t^2 belong in the same coset of $\mathcal{M}$ if the bytes of their difference $MixColumns^{-1}(t^1 \oplus t^2)$ that lie on n anti-diagonals $n \leq 3$ (depending on the dimension of $\mathcal{M}$) are equal to zero. As example, $t^1 \oplus t^2 \in \mathcal{M}_i$ if and only if $MC^{-1}(t^1)$ and $MC^{-1}(t^2)$ have equal bytes on the i-th anti-diagonal, or in other words if $MC^{-1}(t^1 \oplus t^2)_{j,i-j} = 0$ for each $j = 0, 1, 2, 3$ where the index $i - j$ is computed modulo 4.

4 New 5-round Secret Key Distinguisher for AES

4.1 Statement of the Property

Consider a set of plaintexts in the same coset of the diagonal space $\mathcal{D}_I$, that is $\mathcal{D}_I \oplus a$ for a certain $a \in \mathcal{D}_I^{\perp}$, and the corresponding ciphertexts after 5 rounds. In order to set up the distinguisher on 5 rounds of AES, the idea is to count the number of different pairs of ciphertexts that belong to the same coset of $\mathcal{M}_J$ for a fixed J, and to exploit the property that only for an AES permutation this number is a multiple of 8 with probability 1.

In more detail, given a set of plaintexts/ciphertexts (p^i, c^i) for $i = 0, ..., 2^{32 \cdot |I|} - 1$ - where all the plaintexts belong to the same coset of $\mathcal{D}_I$, the idea is to construct all the possible pairs of ciphertexts (c^i, c^j) for $i \neq j$ and to count the number of different pairs[4] of ciphertexts (c^i, c^j) such that $c^i \oplus c^j \in \mathcal{M}_J$ for a certain fixed $J \subset \{0, 1, 2, 3\}$. It is possible to prove that for 5-round AES this number has the special property to be a multiple of 8 independently of the dimension of $\mathcal{M}_J$ (i.e. $|J|$) or of $\mathcal{D}_I$ (i.e. $|I|$). Instead, for a random permutation the same number does not have any special property (e.g. it has the same probability to be even or odd). This allows to distinguish 5-round AES from a random permutation.

Before we go on, we formalize the concept of different pairs of ciphertexts, defining the *partial order*[5] $\leq$:

Definition 7. *Given two different texts t^1 and t^2, we say that $t^1 \leq t^2$ if $t^1 = t^2$ or if there exists $i, j \in \{0, 1, 2, 3\}$ such that (1) $t^1_{k,l} = t^2_{k,l}$ for all $k, l \in \{0, 1, 2, 3\}$ with $k + 4 \cdot l < i + 4 \cdot j$ and (2) $t^1_{i,j} < t^2_{i,j}$. Moreover, we say that $t^1 < t^2$ if $t^1 \leq t^2$ (with respect to the definition just given) and $t^1 \neq t^2$.*

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

[4] Two pairs (c^i, c^j) and (c^j, c^i) are considered equivalent. We formalize this concept in the following using a partial order $\leq$.

[5] If P is an order set with respect to the relation $\leq$, then the following relationships hold: (1) *reflexivity* $\forall a \in P$ then $a \leq a$; (2) *antisymmetry* $\forall a, b \in P$ s.t. $a \leq b$ and $b \leq a$, then $a = b$; (3) *transitivity* $\forall a, b \in P$ s.t. $a \leq b$ and $b \leq c$, then $a \leq c$.

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

$$n := |\{(p^i, c^i), (p^j, c^j) \,|\, \forall p^i, p^j \in \mathcal{D}_I \oplus a,\, p^i < p^j \text{ and } c^i \oplus c^j \in \mathcal{M}_J\}| \qquad (5)$$

is a multiple of 8, that is $\exists\, n' \in \mathbb{N}$ such that $n = 8 \cdot n'$.

Only for completeness, if the final MixColumns operation is omitted, then the above theorem holds in the same way with $\mathcal{ID}_J$ instead of $\mathcal{M}_J$.

Idea of the Proof - Lemma 2. As we have seen in the previous section, a coset of $\mathcal{D}_I$ is always mapped into a coset of $\mathcal{M}_I$ after two rounds, that is for each $a \in \mathcal{D}_I^{\perp}$ there exists unique $b \in \mathcal{M}_I$ such that $R^2(\mathcal{D}_I \oplus a) = \mathcal{M}_I \oplus b$. This statement holds also in the same way in the reverse direction, that is for each $b' \in \mathcal{M}_I^{\perp}$ there exists unique $a' \in \mathcal{D}_I$ such that $R^{-2}(\mathcal{M}_I \oplus b') = \mathcal{D}_I \oplus a'$. 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 focus only on the central round $\mathcal{M}_I \oplus b \to \mathcal{D}_J \oplus a'$ in order to prove the statement of Theorem 3. In particular, this theorem on 5 rounds of AES (and its proof) is related to the following lemma on 1-round AES.

Lemma 2. *Let $\mathcal{M}_I$ and $\mathcal{D}_J$ the subspaces defined as before for certain fixed I and J, and assume $|I| = 1$. Given an arbitrary coset of $\mathcal{M}_I$, consider all the 2^{32} plaintexts and the corresponding ciphertexts after 1 round, that is $(\hat{p}^i, \hat{c}^i)$ for $i = 0, ..., 2^{32} - 1$ where $\hat{c}^i = R(\hat{p}^i)$. The number n of different pairs of ciphertexts $(\hat{c}^i, \hat{c}^j)$ for $i \neq j$ such that $\hat{c}^i \oplus \hat{c}^j \in \mathcal{D}_J$ (i.e. $\hat{c}^i$ and $\hat{c}^j$ belong to the same coset of $\mathcal{D}_J$) is a multiple of 8, that is $\exists\, n' \in \mathbb{N}$ s.t. $n = 8 \cdot n'$.*

The complete proof is provided in the next section - Sect. 6. We emphasize that the proof of Theorem 3 follows immediately by the proof of Lemma 2. Indeed, note that considering 2^{32} plaintexts in the same coset of $\mathcal{D}_I$ is equivalent to consider 2^{32} texts in the same coset of $\mathcal{M}_I$ after two rounds. Moreover, note that the number of collisions (i.e. a pair of texts that belong to the same coset of a given subspace) in the same coset of $\mathcal{M}_J$ is the same of the number of collisions in the same coset of $\mathcal{D}_J$ two rounds before.

To prove the lemma, the idea is show that if one pair of ciphertexts satisfies the requirement to belong to the same coset of $\mathcal{D}_J$, then also other pairs of ciphertexts have the same property with probability 1. The complete proof is given in Sect. 6. We highlight that the statement given in Theorem 3 (or Lemma 2) does not depend on the details of the MixColumns matrix (with the exception that the branch number must be five) or/and of the SubBytes operation. In other words, the only property that the proof - given in the next section - exploits is the branch number of the MixColumns matrix.

4.2 Setting Up the Distinguisher

Our 5-round distinguisher exploits the property just described that the above defined number of collisions n is a multiple of 8 for 5-round AES, while it can take any possible value in the case of a random permutation. Thus, assume $J \subseteq \{0, 1, 2, 3\}$ fixed with $|J| = 3$. First of all, since the probability that two ciphertexts belong to the same coset of $\mathcal{M}_J$ is $2^{-128+32\cdot|J|} = 2^{-32}$ for $|J| = 3$, we expect that *on average*

$$\binom{2^{32}}{2} \cdot 2^{-32} = 2^{31} \cdot (2^{32} - 1) \cdot 2^{-32} \simeq 2^{31}$$

different pairs of ciphertexts belong to the same coset of $\mathcal{M}_J$ both for an AES permutation and for a random one. However, while for an AES permutation this number is a multiple of 8 with probability 1, for a random permutation this happens only with probability $0.125 \equiv 2^{-3}$. In particular, consider s initial arbitrary cosets of $\mathcal{D}_I$ and for each of them count the number of different pairs of ciphertexts that belong to the same coset of $\mathcal{M}_J$ for $|J| = 3$ fixed. For an AES permutation, each of these numbers is a multiple of 8, while the probability that this happens for a random permutation is only $2^{-3\cdot s}$. In order to distinguish the AES permutation from the random one with probability at least pr, it is sufficient that for a random permutation at least one of these numbers is not a multiple of 8, which happens with probability pr:

$$pr = 1 - 2^{-3\cdot s}.$$

Thus, the probability of success of this distinguisher is greater than 99% (i.e. $pr \geq 0.99$) for $s \geq 3$. Note that for each initial coset $\mathcal{D}_I$ with $|I| = 1$, it is possible to count the number of collisions for at most 4 different subspaces $\mathcal{M}_J$ for $|J| = 3$ (note that there are $\binom{4}{3} = 4$ different J with $|J| = 3$). It follows that using a single initial coset $\mathcal{D}_I$ with $|I| = 1$ (for a total of 4 different subspaces $\mathcal{M}_J$ in the ciphertexts space), it is possible to distinguish 5-round AES from a random permutation with a probability of success of approximately 99.975%.

In conclusion, a single initial arbitrary coset of $\mathcal{D}_I$ with $|I| = 1$ in the plaintexts space are sufficient to distinguish a random permutation from an AES one, for a total data complexity of 2^{32} chosen plaintexts. An approximation of the computational cost is given in the following. For completeness, it is also possible to set up a distinguisher for the cases $|J| = 2$ or $|J| = 1$. However, it should be noticed that the average number of collisions in these cases are respectively $2^{31} \cdot (2^{32} - 1) \cdot 2^{-64} \simeq 2^{-1}$ and $2^{31} \cdot (2^{32} - 1) \cdot 2^{-96} \simeq 2^{-33}$. As a consequence, the data and computational cost of these cases is not lower than for the case $|J| = 3$.

4.3 The Computational Cost

We have just seen that 2^{32} chosen plaintexts (i.e. one coset 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 $\mathcal{M}_J$ and checking if it is a multiple of 8 or not. Here we give an estimation of the computational cost of the distinguisher, which is approximately given by the sum of the cost to construct all the pairs and of the cost to count the number of collisions. As a result, the total computational cost can be well approximated by $2^{35.6}$ table look-ups.

Assume the final MixColumns operation is not omitted. As we have just said, 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[6].

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. In the following, we present a way to reduce the total cost to approximately $2^{35.6}$ table look-ups, where the used tables are of size 2^{32} texts (or $2^{32} \cdot 16 = 2^{36}$ byte).

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

$$MC^{-1}(c^1)_{i,j-i} = MC^{-1}(c^2)_{i,j-i} \qquad \forall i = 0,1,2,3 \tag{6}$$

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 (6), 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 (6). 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

[6] 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 $MC^{-1}(c^1 \oplus c^2) \in \mathcal{ID}_J$.

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: 1 for an AES permutation, 0 otherwise (prob. $\geq 99\%$)
Let (p^i, c^i) for $i = 0, ..., 2^{32} - 1$ the (plaintext, ciphertext) pairs;
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**
    $x \leftarrow 0$;
    for k from *0 to* 3 **do**
      $x \leftarrow x + MC^{-1}(c^i)_{k,j-k} \cdot 256^k$; // $MC^{-1}(c^i)_{k,j-k}$ denotes the byte of $MC^{-1}(c^i)$ in row k and column $j - k \mod 4$
    end
    $A[x] \leftarrow A[x] + 1$; // $A[x]$ denotes the value stored in the x-th address of the array A
  end
  $n \leftarrow 0$;
  for i from *0 to* $2^{32} - 1$ **do**
    $n \leftarrow n + A[i] \cdot (A[i] - 1)/2$;
  end
  if $(n \mod 8) \neq 0$ **then**
    return 0;
  end
end
return 1.

Algorithm 1: *Secret-Key Distinguisher for 5 Rounds of AES* which exploits a property which is independent of the secret key - probability of success: $\geq 99\%$.

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[7] 1 table look-up $\approx$ 1 round of AES).

Another possible way to implement our distinguisher exploits a re-ordering algorithm. In order to count the number of pairs of ciphertexts that belong to the same coset of $\mathcal{D}_J$, the idea is to re-order the texts using a particular numerical order $\preceq$ which depends on 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 pairs of ciphertexts with the required property. 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 initial diagonal set. All the details of this method are given in App. D. This second implementation could be in some cases more efficient than the one proposed in

[7] We highlight that even if this approximation is not formally correct - the size of the table of an S-Box look-up is lower 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 literature. At the same time, we note that the same approximation is largely used in literature.

details in this section when e.g. it is required to do further operations on the pairs of ciphertexts (c^1, c^2) such that $c^1 \oplus c^2 \in \mathcal{M}_J$.

4.4 Practical Verification

Using a C/C++ implementation[8], we have practically verified the distinguisher on a small scale variant of AES, as presented in [6]. While in "real" AES, each word is composed of 8 bits, in this variant each word is composed of 4 bits. We refer to [6] for a complete description of this small-scale AES, and we limit ourselves to describe the results of our 5-round distinguisher in this case.

First of all, note that Theorem 3 holds exactly in the same way also for this small-scale variant of AES (the proof is independent by the fact that each word of AES is of 4 or 8 bits). Thus, our verification on the small-scale variant of AES is strong evidence for it to hold for the real AES.

We have verified the theorem for each possible $|J|$ (i.e. for $|J| = 1, 2, 3$) and for $|I| = 1$. For the verification of the secret-key distinguisher, we have chosen $|I| = 1$ and $|J| = 3$ fixed. As result, we have verified that for 5-round AES the number of collisions is a multiple of 8, while this number does not have any particular property for a random permutation. Moreover, we have found that a single initial coset is largely sufficient to distinguish a random permutation from an AES permutation also from a practical point of view, as predicted.

The differences between this small-scale AES and the real AES regard the total number of pairs of ciphertexts that satisfy the required property (equal bytes in 1 fixed diagonal), which in this case is well approximated by $2^{15} \cdot (2^{16} - 1) \cdot 2^{-16} \approx 2^{15}$ for each diagonal set, and the lower computational cost, which can be approximated by $2^{17.6} \cdot 4 \approx 2^{19.6}$ memory look-ups for each initial diagonal set, besides the memory costs. The *average* practical results of our experiments are in accordance with these numbers.

4.5 Generalizations of the Central Theorem

Until now we have considered only a particular case in order to set up our distinguisher. However, here we show that it is possible to generalize Theorem 3 as follows.

Firstly, note that the same distinguisher works also in the reverse direction (i.e. in the decryption mode) with the same complexity. In this case, the strategy is to choose a coset of $\mathcal{M}_I$, and (as before) to count the number of different pairs of plaintexts that belong to the same coset of $\mathcal{D}_J$. This number has the same properties given in Theorem 3, while for a random permutation it can take any possible value. A formal statement for this case (i.e. in the decryption direction) is provided in App. A.

Secondly, Theorem 3 can be generalized for the cases $|I| = 2$ and $|I| = 3$. In particular, it is possible to prove that the result given in Theorem 3 is completely

[8] The source code is available at `https://github.com/Krypto-iaik/AES_5round_SKdistinguisher`

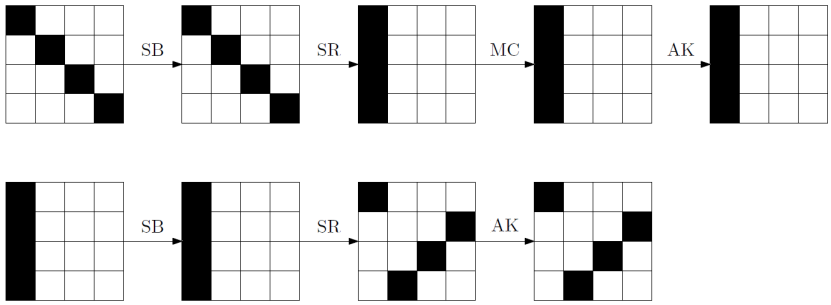

Fig. 1. Differential Trail over 2-round AES.

independent of $|I|$, i.e. given a coset of $\mathcal{D}_I$ for an arbitrary $I \subseteq \{0,1,2,3\}$ with $1 \leq |I| \leq 3$, then the number of collisions after 5 rounds in the same coset of $\mathcal{M}_J$ is a multiple of 8. A formal statement is the following:

Theorem 4. *Let $\mathcal{D}_I$ and $\mathcal{M}_J$ the subspaces defined as before, where $1 \leq |I| \leq 3$ and J are fixed. Given an arbitrary coset of $\mathcal{D}_I$ - that is $\mathcal{D}_I \oplus a$ for a fixed $a \in \mathcal{D}_I^{\perp}$, consider all the $2^{32\cdot|I|}$ plaintexts and the corresponding ciphertexts after 5 rounds, that is (p^i, c^i) for $i = 0, ..., 2^{32\cdot|I|} - 1$ where $p^i \in \mathcal{D}_I \oplus a$ and $c^i = R^5(p^i)$. The number n of different pairs of ciphertexts (c^i, c^j) for $i \neq j$ such that $c^i \oplus c^j \in \mathcal{M}_J$ (i.e. c^i and c^j belong to the same coset of $\mathcal{M}_J$)*

$$n := |\{(p^i, c^i), (p^j, c^j) \,|\, \forall p^i, p^j \in \mathcal{D}_I \oplus a,\, p^i < p^j \text{ and } c^i \oplus c^j \in \mathcal{M}_J\}| \qquad (7)$$

is a multiple of 8, that is $\exists\, n' \in \mathbb{N}$ such that $n = 8 \cdot n'$.

The proof of this theorem is given in App. B - it is simply a generalization of the proof of Theorem 3 given in the next section.

5 Description of the 5-round Secret-Key Distinguisher using a Classical Notation

For sake of completeness, we re-describe the 5-round secret-key distinguisher just presented using a classical notation. Before to do this, we recall the 2-round truncated differential trail of AES illustrated in Fig. 1 (see [10] or [11] for details) using a classical notation.

5.1 Differential Trail over 2-round AES and the Subspace Trail Notation

Let $R^2(\cdot)$ denote two AES rounds with fixed random round keys. Consider two plaintexts which are equal in all bytes except for the ones in the i-th diagonal for a certain $i = 0, 1, 2, 3$, i.e. for the bytes in row j and column $i + j$ for each

$j = 0, 1, 2, 3$ (the index $i + j$ is taken modulo 4). After one round, the two texts are equal in all bytes except for the ones in the i-th column, i.e. for the bytes in row j and column i for each j. After the second and last round - assuming the final MixColumns is omitted, the two texts are equal in all bytes except for the ones in the i-th anti-diagonal, i.e. for the bytes in row j and column $i - j$ for each j (the index $i - j$ is taken modulo 4).

For the following, we work with *diagonal sets* of 2^{32} plaintexts, defined as sets of texts which are equal in 3 diagonals, i.e. texts with active bytes in the i-th diagonal for a certain $i = 0, 1, 2, 3$ and with constant bytes in the other three:

$$\begin{bmatrix} A & C & C & C \\ C & A & C & C \\ C & C & A & C \\ C & C & C & A \end{bmatrix} \xrightarrow{R(\cdot)} \begin{bmatrix} A & C & C & C \\ A & C & C & C \\ A & C & C & C \\ A & C & C & C \end{bmatrix} \xrightarrow{R_f(\cdot)} \begin{bmatrix} A & C & C & C \\ C & C & C & A \\ C & C & A & C \\ C & A & C & C \end{bmatrix},$$

where A denotes an active byte (i.e. a byte in which every value in $\mathbb{F}_{2^8}$ appears the same number of times) and C denotes a constant byte (i.e. a byte in which the value is fixed to a constant for all texts). For completeness, we label the last set by *inverse-diagonal set*, i.e. a set of texts where the bytes in one (or more) anti-diagonal(s) are active while the others are constant.

If the final MixColumns is not omitted, certain linear relations - which are given by the definition of the MixColumns matrix - hold between the bytes of the texts that lie in the same column:

$$\begin{bmatrix} A & C & C & C \\ C & A & C & C \\ C & C & A & C \\ C & C & C & A \end{bmatrix} \xrightarrow{R(\cdot)} \begin{bmatrix} A & C & C & C \\ A & C & C & C \\ A & C & C & C \\ A & C & C & C \end{bmatrix} \xrightarrow{R(\cdot)} MC \times \begin{bmatrix} A & C & C & C \\ C & C & C & A \\ C & C & A & C \\ C & A & C & C \end{bmatrix},$$

In this case, we label the last set by *mixed set*. As an example, consider two plaintexts p^1 and p^2 which are equal in all bytes except for the ones in the 0-th diagonal, i.e. except for the bytes in positions (j, j) for each $j = 0, 1, 2, 3$. After 2 (complete) rounds, there exist $x, y, z, w \in F_{2^8}$ such that their difference $R^2(p^1) \oplus R^2(p^2)$ can be re-written as:

$$R^2(p^1) \oplus R^2(p^2) = \begin{bmatrix} 0\text{x}02 \cdot x & y & z & 0\text{x}03 \cdot w \\ x & y & 0\text{x}03 \cdot z & 0\text{x}02 \cdot w \\ x & 0\text{x}03 \cdot y & 0\text{x}02 \cdot z & w \\ 0\text{x}03 \cdot x & 0\text{x}02 \cdot y & z & w \end{bmatrix}. \tag{8}$$

Finally, the same truncated differential analysis of 2-round can be generalized to the cases of an initial diagonal set with more than a single active diagonal, i.e. a set of plaintexts which are equal in all bytes except for the ones that lie in two or three diagonals (instead of only one).

5.2 Description of the 5-round Secret-Key Distinguisher using a Classical Notation

Consider a diagonal set of plaintexts - i.e. a set of 2^{32} plaintexts which are equal in all bytes except for the ones in i-diagonal for a certain $i = 0, 1, 2, 3$,

and the corresponding ciphertexts after 5 rounds. Assume the final MixColumns operation is omitted. In order to set up the distinguisher on 5 rounds of AES, the idea is to count the number of different pairs of ciphertexts which are equal in d anti-diagonals for a certain $1 \leq d \leq 3$ - that is the number of pairs of ciphertexts with zero-difference in the bytes in positions $(i, j-i)$ for all $i = 0, 1, 2, 3$ and $j \in J$ for a certain $J \subseteq \{0, 1, 2, 3\}$ with $|J| = d$ - and to exploit the property that for an AES-like permutation this number is a multiple of 8 with probability 1.

In more detail, given a set of plaintexts/ciphertexts (p^i, c^i) for $i = 0, ..., 2^{32}-1$ - where all the plaintexts are in the same diagonal set, the idea is to construct all the possible pairs of ciphertexts (c^i, c^j) for $i \neq j$ and to count the number of different pairs[9] of ciphertexts (c^i, c^j) for which the bytes of the difference $c^i \oplus c^j$ that lie in d anti-diagonals are equal to zero (where $1 \leq d \leq 3$ and the anti-diagonals are fixed in advance). It is possible to prove that for 5-round AES this number has the special property to be a multiple of 8 independently of d - that is on the number of considered anti-diagonals. Instead, for a random permutation the same number does not have any special property (e.g. it has the same probability to be even or odd). This allows to distinguish 5-round AES from a random permutation.

Proposition 1. *Given 2^{32} plaintexts in the same diagonal set defined as before, consider the corresponding ciphertexts after 5 rounds, that is (p^i, c^i) for $i = 0, ..., 2^{32} - 1$ where $c^i = R^5(p^i)$ The number n of different pairs of ciphertexts (c^i, c^j) for $i \neq j$ for which the bytes of the difference $c^i \oplus c^j$ that lie in d anti-diagonals are equal to zero (where $1 \leq d \leq 3$ and the anti-diagonals are fixed in advance) is a multiple of 8, that is $\exists\, n' \in \mathbb{N}$ such that $n = 8 \cdot n'$.*

Idea of the Proof - Lemma 3. As we have seen in the previous section, a diagonal set is always mapped after two rounds into a mixed set. In other words, if two plaintexts have equal bytes expect for the ones in one diagonal, then after two rounds some particular linear relationships (given in (8)) hold among the bytes of the difference of these two texts that lie in the same column with probability 1. In the same way, if two ciphertexts have equal bytes in d anti-diagonals, then these two texts have equal bytes in d diagonals two rounds before (due to the 2-round differential trail described in Sect. 5.1). In other words, a inverse-diagonal set is mapped into a diagonal set two rounds before (assuming the final MixColumns operation is omitted).

Assume for simplicity that the 2^{32} plaintexts are chosen in a diagonal set with the active bytes in the first diagonal (analogous for the other cases). Due to these two previous considerations, Proposition 3 on 5 rounds of AES (and its proof) is strongly related to the following lemma on 1-round AES.

[9] The two pairs (c^i, c^j) and (c^j, c^i) are considered equivalent. To formalize this concept, one can consider the number of ciphertexts (c^i, c^j) *with* $i < j$ for which the bytes of the difference $c^i \oplus c^j$ that lie in d anti-diagonals are equal to zero.

Lemma 3. *Given 2^{32} plaintexts in a mixed set of the form*

$$MC \cdot \begin{bmatrix} A & C & C & C \\ C & C & C & A \\ C & C & A & C \\ C & A & C & C \end{bmatrix}, \tag{9}$$

consider the corresponding ciphertexts after 1 round, that is $(\hat{p}^i, \hat{c}^i)$ for $i = 0, ..., 2^{32} - 1$ where $\hat{c}^i = R(\hat{p}^i)$. The number n of different pairs of ciphertexts $(\hat{c}^i, \hat{c}^j)$ for $i \neq j$ for which the bytes of the difference $c^i \oplus c^j$ that lie in d diagonals are equal to zero (where $1 \leq d \leq 3$ and the diagonals are fixed in advance) is a multiple of 8, that is $\exists\, n' \in \mathbb{N}$ s.t. $n = 8 \cdot n'$.

We emphasize that the proof of Proposition 1 follows immediately by the proof of Lemma 3, due to the 2-round truncated differential trail described in Sect. 5.1. In particular, note that considering 2^{32} plaintexts in the same diagonal set (that is 2^{32} plaintexts which are equal in three diagonals and with active bytes in the other one) is equivalent to consider 2^{32} texts in the same mixed set as defined in (9) after two rounds. In other words, all 2^{32} plaintexts of Lemma 3 are definitely reachable in 2 rounds from the initial plaintext (diagonal) structure defined in Proposition 1. We highlight that the statement given in Proposition 1 (or Lemma 3) does not depend on the details of the MixColumns matrix (with the exception that the branch number must be five) or/and of the SubBytes operation. In other words, the only property that the proof - given in the next section - exploits is the branch number of the MixColumns matrix.

5.3 Generalizations of the Central Theorem

Until now we have considered only a particular case in order to set up our distinguisher. However, here we show that it is possible to generalize Proposition 1 as follows.

Firstly, note that the same distinguisher works also in the reverse direction (i.e. in the decryption mode) with the same complexity. Assume that the final MixColumns operation is omitted. In this case the strategy is to choose 2^{32} ciphertexts in a single initial inverse-diagonal set, i.e. a set of 2^{32} ciphertexts which are equal in all the bytes expect for the ones in the i-th anti-diagonal for a certain $i = 0, 1, 2, 3$ (similar definition of the diagonal set). As before, the idea is to count the number of different pairs of plaintexts for which the bytes that lie in d diagonals are equal, for d fixed diagonals with $1 \leq d \leq 3$. This number has the same properties given in Proposition 1, while for a random permutation it can take any possible value. A formal statement for this case (i.e. in the decryption direction) is provided in App. A.

Secondly, Proposition 1 can be generalized for the cases of diagonal sets in which more than a single diagonal is active. As an example, diagonal sets with

2 or 3 active diagonals can be

$$\begin{bmatrix} A & A & C & C \\ C & A & A & C \\ C & C & A & A \\ A & C & C & A \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} A & A & A & C \\ C & A & A & A \\ A & C & A & A \\ A & A & C & A \end{bmatrix}.$$

It is possible to prove that the result given in Proposition 1 is completely independent of the number of active diagonals. In other words, independently of the number of active diagonals of the initial diagonal set of the plaintexts, then the number of pairs of ciphertexts for which the bytes that lie in d anti-diagonals are equal (for d fixed anti-diagonals with $1 \leq d \leq 3$) is a multiple of 8. A formal statement is the following:

Proposition 2. *Given $2^{32 \cdot D}$ plaintexts in the same diagonal set with $1 \leq D \leq 3$ active diagonals defined as before, consider the corresponding ciphertexts after 5 rounds, that is (p^i, c^i) for $i = 0, ..., 2^{32} - 1$ where $c^i = R^5(p^i)$ The number n of different pairs of ciphertexts (c^i, c^j) for $i \neq j$ for which the bytes of the difference $c^i \oplus c^j$ that lie in d anti-diagonals are equal to zero (where $1 \leq d \leq 3$ and the anti-diagonals are fixed in advance) is a multiple of 8, that is $\exists\, n' \in \mathbb{N}$ such that $n = 8 \cdot n'$.*

6 A Detailed Proof of Theorem 3 - Lemma 2

In this section we give a detailed and formal proof of Theorem 3. As we have already said, since it is sufficient to prove Lemma 2 in order to prove the Theorem, we focus on this Lemma, which is recalled in the following.

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

$$n := |\{(p^i, c^i), (p^j, c^j) \,|\, \forall p^i, p^j \in \mathcal{M}_I \oplus a,\, p^i < p^j, \text{ and } c^i \oplus c^j \in \mathcal{D}_J\}| \quad (10)$$

is a multiple of 8, that is $\exists\, n' \in \mathbb{N}$ such that $n = 8 \cdot n'$.

Proof. Consider two elements p^1 and p^2 in the same coset of $\mathcal{M}_i \oplus a$ for $a \in \mathcal{M}_i^{\perp}$. Without loss of generality (W.l.o.g.), assume $i = 0$ (it is analogous for the other cases). By definition of $\mathcal{M}_i$, there exist $x, y, z, w \in \mathbb{F}_{2^8}$ and $x', y', z', w' \in \mathbb{F}_{2^8}$ such that:

$$p^1 = a \oplus \begin{bmatrix} 2 \cdot x & y & z & 3 \cdot w \\ x & y & 3 \cdot z & 2 \cdot w \\ x & 3 \cdot y & 2 \cdot z & w \\ 3 \cdot x & 2 \cdot y & z & w \end{bmatrix}, \qquad p^2 = a \oplus \begin{bmatrix} 2 \cdot x' & y' & z' & 3 \cdot w' \\ x' & y' & 3 \cdot z' & 2 \cdot w' \\ x' & 3 \cdot y' & 2 \cdot z' & w' \\ 3 \cdot x' & 2 \cdot y' & z' & w' \end{bmatrix}$$

where $2 \equiv$ 0x02 and $3 \equiv$ 0x03. For the following, we say that p^1 is "generated" by the variables $\langle x, y, z, w \rangle$ and that p^2 is "generated" by the variables $\langle x', y', z', w' \rangle$.

First case. First, we consider the case in which three variables are equal. W.l.o.g. we assume for example that $y = y'$, $z = z'$, $w = w'$ and $x \neq x'$ (the other cases are analogous). In other words, we suppose that the two texts p^1 and p^2 belong to the same coset of $\mathcal{M}_0 \cap \mathcal{C}_0 \oplus a$, where $a \in (\mathcal{M}_0 \cap \mathcal{C}_0)^\perp$.

Since $\mathcal{M}_0 \cap \mathcal{C}_0 \subseteq \mathcal{C}_0$, it follows that if $p^1 \oplus p^2 \in \mathcal{C}_0$, then $R(p^1) \oplus R(p^2) \in \mathcal{M}_0$. Since $\mathcal{M}_I \cap \mathcal{D}_J = \{0\}$ for each I and J with $|I| + |J| \leq 4$ (see (3)), it follows that $R(p^1) \oplus R(p^2) \notin \mathcal{D}_J$ for each $J \subseteq \{0, 1, 2, 3\}$ with $|J| \leq 3$. In other words, with the given hypothesis for this case, it is not possible that the two texts belong to the same coset of a diagonal space $\mathcal{D}_J$ for each $|J| \leq 3$ after one round.

For completeness, it is also possible to show the same result in a different way. By definition, $R(p^1) \oplus R(p^2) \in \mathcal{D}_J$ for a certain J with $|J| = 3$ if and only if $(R(p^1) \oplus R(p^2))_{i,j+i} = 0$ for each $i = 0, ..., 3$ (i.e. the four bytes of the j-th diagonal of $R(p^1) \oplus R(p^2)$ are equal to zero), where the indexes are taken modulo 4 and $j = \{0, 1, 2, 3\} \setminus J$. As we are going to show, due to the given hypothesis of this case and since the branch number of the MixColumns operation is equal to five, it follows that $R(p^1) \oplus R(p^2) \notin \mathcal{D}_J$ for all J with $|J| = 3$. In other words, $R(p^1) \oplus R(p^2) \in \mathcal{D}_J$ for $|J| = 3$ if and only if $x = x'$, that is $p^1 = p^2$.

In more details, by simple computation the first column (analogues for the other ones) of $SR \circ$ S-Box$(p^1) \oplus SR \circ$ S-Box(p^2) - denoted by $(SR \circ$ S-Box$(p^1) \oplus SR \circ$ S-Box$(p^2))_{\cdot,0}$ - is equal to:

$$(SR \circ \text{ S-Box}(p^1) \oplus SR \circ \text{ S-Box}(p^2))_{\cdot,0} = \begin{bmatrix} \text{S-Box}(2 \cdot x \oplus a_{0,0}) \oplus \text{S-Box}(2 \cdot x' \oplus a_{0,0}) \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

After the MixColumns operation (note $R(p^1) \oplus R(p^2) = MC(SR \circ \text{ S-Box}(p^1) \oplus SR \circ \text{ S-Box}(p^2)) = MC \circ SR \circ \text{ S-Box}(p^1) \oplus MC \circ SR \circ \text{ S-Box}(p^2))$, since only one input byte[10] is different from zero, it follows that at least four output bytes must be different from zero, that is all the output bytes are different from zero. This simply implies that it is not possible that $R(p^1) \oplus R(p^2) \in \mathcal{D}_J$ for $|J| \leq 3$.

Second case. Secondly, we consider the case in which two variables are equal, that is w.l.o.g. we assume for example that $z = z'$ and $w = w'$, while $x \neq x'$ and $y \neq y'$ (the other cases are analogous). That is, we suppose that the two texts p^1 and p^2 belong to the same coset of $\mathcal{M}_0 \cap \mathcal{C}_{0,1} \oplus a$, where $a \in (\mathcal{M}_0 \cap \mathcal{C}_{0,1})^\perp$.

Assume that - for certain $z = z'$ and $w = w'$ - there exist two elements p^1 (generated by $\langle x, y \rangle$) and p^2 (generated by $\langle x', y' \rangle$) defined as before in the same coset of $\mathcal{M}_0$ that belong to the same coset of $\mathcal{D}_J$ for a certain J with $|J| = 3$ after one round. In other words, let $j = \{0, 1, 2, 3\} \setminus J$ and assume that there exist x, y and x', y' and j such that the generated elements p^1 and p^2 satisfy $(R(p^1) \oplus R(p^2))_{i,i+j} = 0$ for each $i = 0, 1, 2, 3$, where the indexes are taken modulo 4.

[10] Note that S-Box$(2 \cdot x \oplus a_{0,0}) \oplus$ S-Box$(2 \cdot x' \oplus a_{0,0}) = 0$ if and only if $x = x'$, which can never happen for hypothesis.

This implies that the two elements $\hat{p}^1$ (generated by $\langle x, y' \rangle$) and $\hat{p}^2$ (generated by $\langle x, y' \rangle$)

$$\hat{p}^1 = a \oplus \begin{bmatrix} 2 \cdot x' & y & 0\ 0 \\ x' & y & 0\ 0 \\ x' & 3 \cdot y & 0\ 0 \\ 3 \cdot x' & 2 \cdot y & 0\ 0 \end{bmatrix}, \qquad \hat{p}^2 = a \oplus \begin{bmatrix} 2 \cdot x & y' & 0\ 0 \\ x & y' & 0\ 0 \\ x & 3 \cdot y' & 0\ 0 \\ 3 \cdot x & 2 \cdot y' & 0\ 0 \end{bmatrix}$$

belong to the same coset of $\mathcal{D}_J$ after one round. To prove this fact, it is sufficient to compute $R(p^1) \oplus R(p^2)$ and $R(\hat{p}^1) \oplus R(\hat{p}^2)$, and to prove that they are equal, i.e.

$$R(p^1) \oplus R(p^2) = R(\hat{p}^1) \oplus R(\hat{p}^2).$$

Since $R(p^1) \oplus R(p^2) \in \mathcal{D}_J$, it also follows that $R(\hat{p}^1) \oplus R(\hat{p}^2) \in \mathcal{D}_J$. In particular, by simple computation the first column of $R(p^1) \oplus R(p^2)$ is given by:

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

Due to the definition of $\hat{p}^1$ and $\hat{p}^2$, it follows immediately that $(R(p^1) \oplus R(p^2))_{\cdot,0} = (R(\hat{p}^1) \oplus R(\hat{p}^2))_{\cdot,0}$. The same holds for the other columns. Note that the existence of the two elements $\hat{p}^1$ and $\hat{p}^2$ is guaranteed by the fact that we are working with the entire coset of $\mathcal{M}_0$. This implies that the number of collisions must be even, that is a multiple of 2.

Question: given p^1 and p^2 as before, is it possible that x, y, x', y' exist such that $R(p^1) \oplus R(p^2) \in \mathcal{D}_J$ for $|J| = 3$? Yes, again because the branch number of the MixColumns operation is five. Indeed, compute $SR \circ \text{S-Box}(p^1) \oplus SR \circ \text{S-Box}(p^2)$ and analyze the first column (the others are analogous):

$$(SR \circ \text{S-Box}(p^1) \oplus SR \circ \text{S-Box}(p^2))_{\cdot,0} = \begin{bmatrix} \text{S-Box}(2 \cdot x \oplus a_{0,0}) \oplus \text{S-Box}(2 \cdot x' \oplus a_{0,0}) \\ \text{S-Box}(y \oplus a_{1,1}) \oplus \text{S-Box}(y' \oplus a_{1,1}) \\ 0 \\ 0 \end{bmatrix}.$$

After the MixColumns operation (note $R(p^1) \oplus R(p^2) = MC(SR \circ \text{S-Box}(p^1) \oplus SR \circ \text{S-Box}(p^2))$), since two input bytes[11] are different from zero, it follows that

[11] Note that $\text{S-Box}(2 \cdot x \oplus a_{0,0}) \oplus \text{S-Box}(2 \cdot x' \oplus a_{0,0}) = 0$ if and only if $x = x'$, which can never happen for hypothesis. In the same way, $\text{S-Box}(y \oplus a_{1,1}) \oplus \text{S-Box}(y' \oplus a_{1,1}) = 0$ if and only if $y = y'$, which can never happen for hypothesis.

at least three output bytes must be different from zero, or at most one output byte could be equal to zero (similar for the other columns). In other words, it is possible that p^1 and p^2 exist such that $R(p^1) \oplus R(p^2) \in \mathcal{D}_J$ for $|J| = 3$. Moreover, this also implies that it is not possible that two or more output bytes in the same column are equal to zero, or in other words that $R(p^1) \oplus R(p^2) \in \mathcal{D}_J$ for $|J| \leq 2$, with the previous conditions.

Moreover, observe that $R(p^1) \oplus R(p^2) \in \mathcal{D}_J$ for $|J| = 3$ if and only if four bytes (one per column) of $R(p^1) \oplus R(p^2)$ are equal to zero. Since there are four "free" variables (i.e. x, y, x', y') and a system of four equations, such a system can have a non-negligible solution.

Finally, since the previous result is independent of the values of $z = z'$ and $w = w'$, it follows that the number of collisions for this case must be a multiple of 2^{17}. Indeed, assume that for certain $\hat{z}$ and $\hat{w}$ there exist x, y, x', y' such that the two elements p^1 and p^2 in $\mathcal{M}_0 \cap \mathcal{C}_{0,1} \oplus a$ generated respectively by $\langle x, y \rangle$ and by $\langle x', y' \rangle$ satisfy the condition $R(p^1) \oplus R(p^2) \in \mathcal{D}_J$ for a certain J. By simple computation, the difference $R(p^1) \oplus R(p^2)$ doesn't depend on $z = z'$ and on $w = w'$, that is for each byte of $(R(p^1) \oplus R(p^2))_{k,l}$ for $k, l = 0, 1, 2, 3$ there exist constant A_i, B_i, C_i for $i = 0, 1, 2, 3$ - that depend only on the coefficients of the MixColumns matrix or/and of the secret-key - such that

$$\begin{aligned}
(R(p^1) \oplus R(p^2))_{i,j} =& A_0 \cdot (\text{S-Box}(B_0 \cdot x \oplus C_0) \oplus \text{S-Box}(B_0 \cdot x' \oplus C_0)) \oplus \\
& \oplus A_1 \cdot (\text{S-Box}(B_1 \cdot y \oplus C_1) \oplus \text{S-Box}(B_1 \cdot y' \oplus C_1) \oplus \\
& \oplus A_2 \cdot (\text{S-Box}(B_2 \cdot z \oplus C_2) \oplus \text{S-Box}(B_2 \cdot z' \oplus C_2) \oplus \\
& \oplus A_3 \cdot (\text{S-Box}(B_3 \cdot w \oplus C_3) \oplus \text{S-Box}(B_3 \cdot w' \oplus C_3) = \\
=& A_0 \cdot (\text{S-Box}(B_0 \cdot x \oplus C_0) \oplus \text{S-Box}(B_0 \cdot x' \oplus C_0)) \oplus \\
& \oplus A_1 \cdot (\text{S-Box}(B_1 \cdot y \oplus C_1) \oplus \text{S-Box}(B_1 \cdot y' \oplus C_1).
\end{aligned}$$

It follows that - under the previous hypothesis - each pair of elements p^1 and p^2 respectively generated by (1) $\langle x, y, z, w \rangle$ and by $\langle x', y', z, w \rangle$ or (2) $\langle x, y', z, w \rangle$ and by $\langle x', y, z, w \rangle$ *for each possible value of* z *and* w satisfy the condition $R(p^1) \oplus R(p^2) \in \mathcal{D}_J$. Thus, the number of collisions for this case must be a multiple of $2 \cdot (2^8)^2 = 2^{17}$. As before, the existence of all these elements is guaranteed by the fact that we are working with the entire coset of $\mathcal{M}_0$.

Third case. Thirdly, we consider the case in which only one variable is equal, that is w.l.o.g. we assume for example $w = w'$, while $x \neq x'$, $y \neq y'$ and $z \neq z'$ (the other cases are analogous). That is, we suppose that the two texts p^1 and p^2 belong to the same coset of $\mathcal{M}_0 \cap \mathcal{C}_{0,1,2} \oplus a$, where $a \in (\mathcal{M}_0 \cap \mathcal{C}_{0,1,2})^{\perp}$.

Assume there exist two elements p^1 (generated by $\langle x, y, z \rangle$) and p^2 (generated by $\langle x', y', z' \rangle$) defined as before in the same coset of $\mathcal{M}_0$ that belong to the same coset of $\mathcal{D}_J$ for a certain J with $|J| \geq 2$ after one round. In other words, assume there exist x, y, z and x', y', z' such that the generated elements p^1 and p^2 satisfy $R(p^1) \oplus R(p^2) \in \mathcal{D}_J$ for a certain J with $|J| \geq 2$ Similar to before, it follows that also the following three pairs of elements in the same coset of $\mathcal{M}_0$ generated by:

– $\langle x', y, z\rangle$ and $\langle x, y', z'\rangle$
– $\langle x, y', z\rangle$ and $\langle x', y, z'\rangle$
– $\langle x, y, z'\rangle$ and $\langle x', y', z\rangle$

belong after one round in the same coset of $\mathcal{D}_J$ for the same J of p^1 and p^2, for a total of four different pairs. As before, in order to prove this fact it is sufficient to show that $R(p^1) \oplus R(p^2) = R(\hat{p}^1) \oplus R(\hat{p}^2)$, where $\hat{p}^1$ and $\hat{p}^2$ are generated by the previous combinations of variables. Note that the existence of these elements is guaranteed by the fact that we are working with the entire coset of $\mathcal{M}_0$. This implies that the number of collisions must be a multiple of 4.

Finally, we have only to prove that such x, y, z and x', y', z' can exist. As before, we compute $SR\circ$ S-Box$(p^1) \oplus SR\circ$ S-Box(p^2) and analyze the first column (the others are analogous):

$$(SR\circ \text{ S-Box}(p^1) \oplus SR\circ \text{ S-Box}(p^2))_{\cdot,0} = \begin{bmatrix} \text{S-Box}(2 \cdot x \oplus a_{0,0}) \oplus \text{S-Box}(2 \cdot x' \oplus a_{0,0}) \\ \text{S-Box}(y \oplus a_{1,1}) \oplus \text{S-Box}(y' \oplus a_{1,1}) \\ \text{S-Box}(2 \cdot z \oplus a_{2,2}) \oplus \text{S-Box}(2 \cdot z' \oplus a_{2,2}) \\ 0 \end{bmatrix}.$$

After the MixColumns operation, since three input bytes[12] are different from zero, it follows that at least two output bytes must be different from zero, or at most two output bytes could be equal to zero. This implies that the event $R(p^1) \oplus R(p^2) \in \mathcal{D}_J$ for $|J| \geq 2$ is possible. Moreover, this also implies that it is not possible that three output bytes (of the same column) are equal to zero, or in other words that $R(p^1) \oplus R(p^2) \in \mathcal{D}_J$ for $|J| = 1$, with the previous hypothesis. Also in this case, variables x, y, z and x', y', z' can exist since the number of equations is less or equal than the number of variables.

Finally, since the previous result is independent of the values of $w = w'$, it follows that the number of collisions for this case must be a multiple of $4 \cdot 2^8 = 2^{10}$. As before, assume that for a certain $\hat{w}$ there exist x, y, z, x', y', z' such that the two elements p^1 and p^2 in $\mathcal{M}_0 \cap \mathcal{C}_{0,1,2} \oplus a$ generated respectively by $\langle x, y, z\rangle$ and by $\langle x', y', z'\rangle$ satisfy the condition $R(p^1) \oplus R(p^2) \in \mathcal{D}_J$ for a certain J. Also in this case, the idea is to show that the difference $R(p^1) \oplus R(p^2)$ doesn't depend on $w = w'$, that is for each byte of $(R(p^1) \oplus R(p^2))_{i,j}$ there exist constant A_i, B_i, C_i for $i = 0, 1, 2$ - that depend only on the coefficients of the MixColumns matrix or/and of the secret-key - such that

$$\begin{aligned}(R(p^1) \oplus R(p^2))_{i,j} = {} & A_0 \cdot (\text{S-Box}(B_0 \cdot x \oplus C_0) \oplus \text{S-Box}(B_0 \cdot x' \oplus C_0)) \oplus \\ & \oplus A_1 \cdot (\text{S-Box}(B_1 \cdot y \oplus C_1) \oplus \text{S-Box}(B_1 \cdot y' \oplus C_1) \oplus \\ & \oplus A_2 \cdot (\text{S-Box}(B_2 \cdot z \oplus C_2) \oplus \text{S-Box}(B_2 \cdot z' \oplus C_2).\end{aligned}$$

It follows that - under the previous hypothesis - each pair of elements p^1 and p^2 respectively generated by one of the four different combinations of the variables

[12] Note that S-Box$(2 \cdot x \oplus a_{0,0}) \oplus$S-Box$(2 \cdot x' \oplus a_{0,0}) =$ S-Box$(y \oplus a_{1,1}) \oplus$S-Box$(y' \oplus a_{1,1}) =$ S-Box$(2 \cdot z \oplus a_{2,2}) \oplus$ S-Box$(2 \cdot z' \oplus a_{2,2}) = 0$ if and only if $x = x'$, $y = y'$ and $z = z'$, which can never happen for hypothesis.

$\langle x, y, z, w\rangle$ and $\langle x', y', z', w\rangle$ *for each possible value of* w satisfy the condition $R(p^1) \oplus R(p^2) \in \mathcal{D}_J$. As before, the existence of all these elements is guaranteed by the fact that we are working with the entire coset of $\mathcal{M}_0$.

Fourth case. Fourthly, we consider the case in which all the variables are different, that is w.l.o.g. we assume that $x \neq x'$, $y \neq y'$, $z \neq z'$ and $w \neq w'$. That is, we suppose that the two texts p^1 and p^2 belong to the same coset of $\mathcal{M}_0 \oplus a$, where $a \in \mathcal{M}_0^{\perp}$ and where $p^1 \oplus p^2 \notin \mathcal{C}_J$ for each $|J| \leq 3$.

Assume there exist two elements p^1 (generated by $\langle x, y, z, w\rangle$) and p^2 (generated by $\langle x', y', z', w'\rangle$) defined as before in the same coset of $\mathcal{M}_0$ that belong to the same coset of $\mathcal{D}_J$ for a certain J with $|J| \geq 1$ after one round. In other words, assume there exist x, y, z, w and x', y', z', w' such that the generated elements p^1 and p^2 satisfy $R(p^1) \oplus R(p^2) \in \mathcal{D}_J$ for a certain J with $|J| \geq 1$. Similar to before, it follows that also the following seven pairs of elements in the same coset of $\mathcal{M}_0$ generated by:

- $\langle x', y, z, w\rangle$ and $\langle x, y', z', w'\rangle$
- $\langle x, y', z, w\rangle$ and $\langle x', y, z', w'\rangle$
- $\langle x, y, z', w\rangle$ and $\langle x', y', z, w'\rangle$
- $\langle x, y, z, w'\rangle$ and $\langle x', y', z', w\rangle$
- $\langle x', y', z, w\rangle$ and $\langle x, y, z', w'\rangle$
- $\langle x', y, z', w\rangle$ and $\langle x, y', z, w'\rangle$
- $\langle x', y, z, w'\rangle$ and $\langle x, y', z', w\rangle$

belong after one round in the same coset of $\mathcal{D}_J$ for the same J of p^1 and p^2, for a total of eight different pairs. As before, in order to prove this fact it is sufficient to show that $R(p^1) \oplus R(p^2) = R(\hat{p}^1) \oplus R(\hat{p}^2)$. Moreover, as before note that existence of these elements is guaranteed by the fact that we are working with all the coset of $\mathcal{M}_0$. This implies that the number of collisions must be a multiple of 8.

Finally, we have only to prove that such x, y, z, w and x', y', z', w' can exist. As before, we compute $SR\circ$ S-Box$(p^1) \oplus SR\circ$ S-Box(p^2) and analyze the first column (the others are analogous):

$$(SR\circ \text{ S-Box}(p^1) \oplus SR\circ \text{ S-Box}(p^2))_{\cdot,0} = \begin{bmatrix} \text{S-Box}(2 \cdot x \oplus a_{0,0}) \oplus \text{S-Box}(2 \cdot x' \oplus a_{0,0}) \\ \text{S-Box}(y \oplus a_{1,1}) \oplus \text{S-Box}(y' \oplus a_{1,1}) \\ \text{S-Box}(2 \cdot z \oplus a_{2,2}) \oplus \text{S-Box}(2 \cdot z' \oplus a_{2,2}) \\ \text{S-Box}(w \oplus a_{3,3}) \oplus \text{S-Box}(w' \oplus a_{3,3}) \end{bmatrix}.$$

After the MixColumns operation, since four input bytes[13] are different from zero, it follows that at least one output byte must be different from zero, or at most three output bytes could be equal to zero. This implies that the event $R(p^1) \oplus R(p^2) \in \mathcal{D}_J$ for $|J| \geq 1$ is possible. Also in this case, variables x, y, z, w and x', y', z', w' can exist since the number of equations is less or equal than the number of variables.

[13] Note that S-Box$(2 \cdot x \oplus a_{0,0}) \oplus$S-Box$(2 \cdot x' \oplus a_{0,0}) =$ S-Box$(y \oplus a_{1,1}) \oplus$S-Box$(y' \oplus a_{1,1}) =$ S-Box$(2 \cdot z \oplus a_{2,2}) \oplus$S-Box$(2 \cdot z' \oplus a_{2,2}) =$ S-Box$(w \oplus a_{3,3}) \oplus$S-Box$(w' \oplus a_{3,3}) = 0$ if and only if $x = x'$, $y = y', z = z'$ and $w = w'$, which can never happen for hypothesis.

Conclusion. We summarize the previous results and we prove the lemma. Given a coset of $\mathcal{M}_i$, we analyze the number of collisions in the same coset of $\mathcal{D}_J$ after one round.

If $|J| = 1$, it is possible to have a collision only in the case in which all the variables that generate the two texts are different, that is $x \neq x'$, $y \neq y'$, and so on. In this case, the number of collisions n must be a multiple of 8, that is there exists $n' \in \mathbb{N}$ such that $n = 8 \cdot n'$.

If $|J| = 2$, it is possible to have a collision only if at least three variables that generate the two texts are different (i.e. at most one variable can be equal). If all the variables are different, the number of collisions is a multiple of 8, while if one is equal then the number of collisions is a multiple of $1024 \equiv 2^{10}$. In other words, there exist $n', n'_2 \in \mathbb{N}$ such that the total number of collisions n is equal to $n = 8 \cdot n' + 1024 \cdot n'_2 = 8 \cdot (n' + 128 \cdot n'_2)$, i.e. it is a multiple of 8.

If $|J| = 3$, it is possible to have a collision only if at least two variables that generate the two texts are different (i.e. at most two variables can be equal). If all the variables are different, the number of collisions is a multiple of 8, if one is equal then the number of collisions is a multiple of $1024 \equiv 2^{10}$, while if two are equal then the number of collisions is a multiple of $131072 \equiv 2^{17}$. In other words, there exist $n', n'_2, n'_3 \in \mathbb{N}$ such that the total number of collisions n is equal to $n = 8 \cdot n' + 2^{10} \cdot n'_2 + 2^{17} \cdot n'_3 = 8 \cdot (n' + 2^7 \cdot n'_2 + 2^{14} \cdot n'_3)$, i.e. it is a multiple of 8.

This proves the lemma. □

For completeness, we briefly recall why the proof of Lemma 2 implies Theorem 3. As we have already said, consider the following description of 5-round of AES:

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

By Lemma 2 and focusing in the middle round, we know that the number of collision n must a multiple of 8. Then, the backward extension is simply given by the fact that a coset of $\mathcal{M}_I$ is mapped into a coset of $\mathcal{D}_I$ two rounds before. About the forward extension, for the same reason note that if two texts belong to the same coset of $\mathcal{D}_J$, then they belong to the same coset of $\mathcal{M}_J$ after two rounds. Since these two events hold with probability 1, this finally proves the theorem.

7 Conclusion, applications and open problems

In this paper, we have presented a new non-random property for 5 rounds of AES. Additionally, we showed how to set up an efficient 5-round secret-key distinguisher for AES which exploits this property, which is independent of the secret key, improving the very recent results [22] and providing answers to the questions posed in [22]. This distinguisher is structural in the sense that it is independent of the details of the MixColumns matrix (with the exception that the branch number must be five) and also independent of the SubBytes operation.

As such it will be straightforward to apply to many other AES-like constructions. Starting from our results, a range of new questions arise for future investigations:

Application to schemes that directly use round-reduced AES. Round-reduced AES is a popular construction to build different schemes. For example, in the on-going "Competition for Authenticated Encryption: Security, Applicability, and Robustness" (CAESAR) [1], which is currently at its third round, several candidates are designed based on an AES-like SPN structure. Focusing only on the third-round candidates[14], among many others, AEGIS [16] uses four AES round-functions in the state update functions while ELmD [21] recommends to use round-reduced AES including 5-round AES to partially encrypt the data. Although the security of these candidates does not completely depend on the underlying primitives, we believe that a better understanding of the security of round-reduced AES can help get insights to both the design and cryptanalysis of authenticated encryption algorithms.

Further Extensions. Is it possible to set up a secret-key distinguisher for 6-round of AES which exploits a property which is independent of the secret key? Is it possible to set up efficient key recovery attacks for 6- or more rounds of AES that exploits this new 5-round secret-key distinguisher proposed in this paper or a modified version of it?

Permutation and Known-Key Distinguishers. The new 5-round property (or its approach to derive it) might find applications to permutation distinguishers or known-key distinguishers. Permutation distinguisher are usually set up by combining two secret-key distinguishers in an inside-out fashion. It is not immediately clear how the 5-round secret-key distinguisher presented in this paper used in an inside-out approach would be able to maintain the property in both directions simultaneously, but it seems interesting to investigate this direction also.

References

1. "CAESAR: Competition for Authenticated Encryption: Security, Applicability, and Robustness," http://competitions.cr.yp.to/caesar.html.
2. A. Biryukov, D. Khovratovich, "PAEQ v1," http://competitions.cr.yp.to/round1/paeqv1.pdf.
3. E. Biham, A. Biryukov, and A. Shamir, "Cryptanalysis of Skipjack Reduced to 31 Rounds Using Impossible Differentials," in *Advances in Cryptology - EUROCRYPT 1999: International Conference on the Theory and Application of Cryptographic Techniques, Czech Republic. Proceedings*, ser. LNCS, vol. 1592, 1999, pp. 12–23.

[14] Among previous-round candidates, it is also possible to include PRIMATEs [13] which design is based on an AES-like SPN structure, while 4-round AES is adopted by Marble [17] and used to build the AESQ permutation in PAEQ [2].

4. E. Biham and N. Keller, "Cryptanalysis of Reduced Variants of Rijndael," unpublished, 2001, http://csrc.nist.gov/archive/aes/round2/conf3/papers/35-ebiham.pdf.
5. E. Biham and A. Shamir, *Differential Cryptanalysis of the Data Encryption Standard*. Springer-Verlag, 1993.
6. C. Cid, S. Murphy, and M. J. B. Robshaw, "Small Scale Variants of the AES," in *Fast Software Encryption - FSE 2005: 12th International Workshop, France. Revised Selected Papers*, ser. LNCS, vol. 9054, 2005, pp. 145–162.
7. T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms, Third Edition*, 3rd ed. The MIT Press, 2009.
8. J. Daemen, L. R. Knudsen, and V. Rijmen, "The Block Cipher Square," in *Fast Software Encryption - FSE 1997: 4th International Workshop, Israel. Proceedings*, ser. LNCS, vol. 1267, 1997, pp. 149–165.
9. J. Daemen and V. Rijmen, *The Design of Rijndael: AES - The Advanced Encryption Standard*, ser. Information Security and Cryptography. Springer, 2002.
10. ——, "Two-round aes differentials," Cryptology ePrint Archive, Report 2006/039, 2006, http://eprint.iacr.org/2006/039.
11. ——, "Understanding Two-Round Differentials in AES," in *Security and Cryptography for Networks - SCN 2006: 5th International Conference, Italy, 2006, Proceedings*, ser. LNCS, vol. 4116, 2006, pp. 78–94.
12. P. Derbez, "Meet-in-the-middle attacks on AES," Ph.D. thesis, Ecole Normale Supérieure de Paris - ENS Paris, (Dec 2013), https://tel.archives-ouvertes.fr/tel-00918146.
13. E. Andreeva, B. Bilgin, A. Bogdanov, A. Luykx, F. Mendel, B. Mennink, N. Mouha, Q. Wang, K. Yasuda, "PRIMATEs v1.02 Submission to the CAESAR Competition," http://competitions.cr.yp.to/round2/primatesv102.pdf.
14. N. Ferguson, J. Kelsey, S. Lucks, B. Schneier, M. Stay, D. Wagner, and D. Whiting, "Improved Cryptanalysis of Rijndael," in *Fast Software Encryption - FSE 2000: 7th International Workshop, USA, 2000, Proceedings*, ser. LNCS, vol. 1978, 2001, pp. 213–230.
15. L. Grassi, C. Rechberger, and S. Rønjom, "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
16. H. Wu, B. Preneel, "A Fast Authenticated Encryption Algorithm," http://competitions.cr.yp.to/round1/aegisv1.pdf.
17. J. Guo, "Marble Version 1.1," https://competitions.cr.yp.to/round1/marblev11.pdf.
18. L. R. Knudsen, "Truncated and higher order differentials," in *Fast Software Encryption - FSE 1994: Second International Workshop, Belgium. Proceedings*, ser. LNCS, vol. 1008, 1995, pp. 196–211.
19. ——, "DEAL - a 128-bit block cipher," Technical Report 151, Department of Informatics, University of Bergen, Norway, Feb. 1998.
20. M. Luby and C. Rackoff, "How to Construct Pseudorandom Permutations from Pseudorandom Functions," *SIAM J. Comput.*, vol. 17, no. 2, pp. 373–386, 1988.
21. N. Datta, M. Nandi, "ELmD v2.0," http://competitions.cr.yp.to/round2/elmdv20.pdf.
22. B. Sun, M. Liu, J. Guo, L. Qu, and V. Rijmen, "New Insights on AES-Like SPN Ciphers," in *Advances in Cryptology – CRYPTO 2016: 36th Annual International Cryptology Conference, Santa Barbara, CA, USA. Proceedings, Part I*, ser. LNCS, vol. 9814, 2016, pp. 605–624.

23. B. Sun, M. Liu, J. Guo, V. Rijmen, and R. Li, "Provable Security Evaluation of Structures Against Impossible Differential and Zero Correlation Linear Cryptanalysis," in *Advances in Cryptology - EUROCRYPT 2016: 35th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Austria. Proceedings, Part I*, ser. LNCS, vol. 9665, 2016, pp. 196–213.
24. B. Sun, Z. Liu, V. Rijmen, R. Li, L. Cheng, Q. Wang, H. AlKhzaimi, and C. Li, "Links Among Impossible Differential, Integral and Zero Correlation Linear Cryptanalysis," in *Advances in Cryptology - CRYPTO 2015: 35th Annual Cryptology Conference, Santa Barbara, CA, USA, 2015, Proceedings, Part I*, ser. LNCS, vol. 9215, 2015, pp. 95–115.
25. T. Tiessen, "Polytopic Cryptanalysis," in *Advances in Cryptology - EUROCRYPT 2016: 35th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Austria. Proceedings, Part I*, ser. LNCS, vol. 9665, 2016, pp. 214–239.
26. T. Tiessen, L. R. Knudsen, S. Kölbl, and M. M. Lauridsen, "Security of the AES with a Secret S-Box," in *Fast Software Encryption - FSE 2015: 22nd International Workshop, Turkey. Revised Selected Papers*, ser. LNCS, vol. 9054, 2015, pp. 175–189.

A Secret-Key Distinguisher on 5 Rounds AES - Decryption Direction

The secret-key distinguisher on 5-round AES presented in Sect. 4 works also in the decryption direction. Here we give a formal theorem for this case.

Theorem 5. *Let $\mathcal{M}_I$ and $\mathcal{D}_J$ the subspaces defined as before for certain fixed I and J, and assume $|I| = 1$. 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} ciphertexts and the corresponding plaintexts 5 rounds before, that is (p^i, c^i) for $i = 0, ..., 2^{32} - 1$ where $p^i \in \mathcal{M}_I \oplus a$ and $c^i = R^{-5}(p^i)$. 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. c^i and c^j belong to the same coset of $\mathcal{M}_J$)*

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

is a multiple of 8, that is $\exists\, n' \in \mathbb{N}$ s.t. $n = 8 \cdot n'$.

The proof of this theorem is completely analogous to the one proposed for Theorem 3. Thus, we limit ourselves to give the sketch of the proof, and we refer to the previous case for all the details.

As before, the idea is to focus on the middle round, that is given 2^{32} texts in the same coset of $\mathcal{D}_I$ with $|I| = 1$, the idea is to prove that the number of collisions n in the same coset of $\mathcal{M}_J$ one round before is a multiple of 8.

In particular, consider two element p^1 and p^2 in the same coset of $\mathcal{D}_i \oplus a$ for $a \in \mathcal{D}_i^{\perp}$. W.l.o.g., assume $i = 0$ (analogous for the other cases). By definition of $\mathcal{D}_i$, there exist $x, y, z, w \in \mathbb{F}_{2^8}$ and $x', y', z', w' \in \mathbb{F}_{2^8}$ such that:

$$p^1 = a \oplus \begin{bmatrix} x & 0 & 0 & 0 \\ 0 & y & 0 & 0 \\ 0 & 0 & z & 0 \\ 0 & 0 & 0 & w \end{bmatrix}, \qquad p^2 = a \oplus \begin{bmatrix} x' & 0 & 0 & 0 \\ 0 & y' & 0 & 0 \\ 0 & 0 & z' & 0 \\ 0 & 0 & 0 & w' \end{bmatrix},$$

where $a \in \mathcal{D}_i^{\perp}$ fixed. We say that p^1 is "generated" by $\langle x, y, z, w \rangle$ and p^2 is "generated" by $\langle x', y', z', w' \rangle$.

As before, the idea is to analyze in details the following four cases: (1) only one variable is different (e.g. $x \neq x'$ and $y = y'$, $z = z'$, $w = w'$), (2) two variables are different, (3) three variables are different and (4) all the variables are different.

For completeness, we analyze only the case (2) - the other cases are analogous. W.l.o.g. we assume for example that $x \neq x'$, $y \neq y'$ and $z = z'$, $w = w'$. Assume that there exist $x \neq x'$, $y \neq y'$ and $z = z'$, $w = w'$ such that p^1 generated by $\langle x, y, z, w \rangle$ and p^2 generated by $\langle x', y', z', w' \rangle$ belong to the same coset of $\mathcal{M}_J$ one round before. As before, it is possible to prove that also the elements $\hat{p}^1$ and $\hat{p}^2$ in $\mathcal{D}_i \oplus a$ generated by $\langle x', y \rangle$ and $\langle x, y' \rangle$ belong one round before in the same coset of $\mathcal{M}_J$ for the same J of p^1 and p^2. To prove this, it is sufficient to show that $R^{-1}(p^1) \oplus R^{-1}(p^2) = R^{-1}(\hat{p}^1) \oplus R^{-1}(\hat{p}^2)$. Moreover, showing that all the bytes of $R^{-1}(p^1) \oplus R^{-1}(p^2) = R^{-1}(\hat{p}^1) \oplus R^{-1}(\hat{p}^2)$ are independently of $z = z'$ and $w = w'$, it follows that the number of collisions must be a multiple of $2 \cdot (2^8)^2 = 2^{17}$. Note that the existence of all these elements $\hat{p}^1$ and $\hat{p}^2$ is guaranteed by the fact that we are working with the entire coset of $\mathcal{D}_0$.

We finally prove that the variables x, y, x' and y' can exist. By simple computation - where for the following $b \equiv MC^{-1}(a \oplus k)$ and k is the secret key of the round, the first column of $R^{-1}(p^1) \oplus R^{-1}(p^2)$ is given by (analogous for the others):

$$[R^{-1}(p^1) \oplus R^{-1}(p^2)]_{\cdot,0} = \begin{bmatrix} \text{S-Box}^{-1}(E \cdot x \oplus b_{0,0}) \oplus \text{S-Box}^{-1}(E \cdot x' \oplus b_{0,0}) \\ 0 \\ 0 \\ \text{S-Box}^{-1}(D \cdot y \oplus b_{3,1}) \oplus \text{S-Box}^{-1}(D \cdot y' \oplus b_{3,1}) \end{bmatrix},$$

where $E \equiv 0\text{x}0E$, $B \equiv 0\text{x}0B$, $D \equiv 0\text{x}0D$ and $9 \equiv 0\text{x}09$ (analogous for the other columns). Note that two bytes of the first column are different from zero (since $x \neq x'$ and $y \neq y'$). Since $R^{-1}(p^1) \oplus R^{-1}(p^2) \in \mathcal{M}_J$ if and only if $MC^{-1}(R^{-1}(p^1) \oplus R^{-1}(p^2)) \in \mathcal{ID}_J$ and since the InverseMixColumns matrix has branch number 5, it follows that at most one output byte of each column can be equal to zero, that is J must satisfy $|J| \geq 3$. Moreover, $R^{-1}(p^1) \oplus R^{-1}(p^2) \in \mathcal{M}_J$ implies only one condition for each column, for a total of four conditions. Since there are four variables, the variables x, x', y and y' can exist.

The complete proof of the theorem is obtained working in a similar way also for the other cases, as for Theorem 3 - see Sect. 6 for details.

B Generalization of Theorem 3

In Theorem 3 given in Sect. 4, we only considered the case $|I| = 1$. A natural question arises: is it possible to generalize the theorem also for $|I| = 2$ or/and $|I| = 3$? The answer is yes, and it is given in Theorem 4 recalled in the following. In particular, we prove in this section that the result obtained in Theorem 3 is

independent of $|I| = 1$, or, in other words, the property of n to be a multiple of 8 is independent of I.

Theorem 4. *Let $\mathcal{D}_I$ and $\mathcal{M}_J$ the subspaces defined as before, where $1 \leq |I| \leq 3$ and J are fixed. Given an arbitrary coset of $\mathcal{D}_I$ - that is $\mathcal{D}_I \oplus a$ for a fixed $a \in \mathcal{D}_I^{\perp}$, consider all the $2^{32\cdot|I|}$ plaintexts and the corresponding ciphertexts after 5 rounds, that is (p^i, c^i) for $i = 0, ..., 2^{32\cdot|I|} - 1$ where $p^i \in \mathcal{D}_I \oplus a$ and $c^i = R^5(p^i)$. The number n of different pairs of ciphertexts (c^i, c^j) for $i \neq j$ such that $c^i \oplus c^j \in \mathcal{M}_J$ (i.e. c^i and c^j belong to the same coset of $\mathcal{M}_J$)*

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

is a multiple of 8, that is $\exists\, n' \in \mathbb{N}$ s.t. $n = 8 \cdot n'$.

Since the proof for the case $|I| = 1$ is given in Sect. 6, we focus on the cases $|I| = 2$ and $|I| = 3$. Also for these cases, the idea is to analyze the middle round and to study each possible case, as done in Sect. 6. Thus, given pair of texts in the same coset of $\mathcal{M}_I$, we analyze the property of the number of collisions in the same coset of $\mathcal{D}_J$ after one round.

Since the idea of the proof for $|I| = 2$ and $|I| = 3$ is analogous to that given for $|I| = 1$, we limit ourselves to do some considerations which justify the theorem. A complete proof can be easily obtained exploiting the following considerations and using the same strategy proposed in Sect. 6.

First Consideration. As first consideration, note that we are considering pairs of plaintexts/ciphertexts (p^1, c^1) and (p^2, c^2) such that $p^1 \oplus p^2 \in \mathcal{M}_I$ for $|I| = 2$ and $|I| = 3$ (note that we analyze the middle round). Since $\mathcal{M}_I$ can be seen as the union of set of $\mathcal{M}_{\hat{I}}$ for each $|\hat{I}| = 1$ and $|I| \geq 2$ such that $\hat{I} \subseteq I$

$$\mathcal{M}_I \equiv \bigcup_{x \in \mathcal{M}_{I \setminus \hat{I}}} \mathcal{M}_{\hat{I}} \oplus x,$$

then if n is a multiple of 2^m then m must satisfy $m \leq 3$. This follows immediately by Theorem 3 (which can be applied to each coset $\mathcal{M}_{\hat{I}} \oplus x$ defined previously) and the corresponding proof of Sect. 6.

Thus, we have to prove that n is a multiple of 2^m and that $m = 3$ also for the cases $|I| = 2$ and $|I| = 3$.

B.1 Case $|I| = 2$

We start studying the case $|I| = 2$. As we show in details in the following, the same analysis can be simply modified and adapted for the case $|I| = 3$.

W.l.o.g we assume $I = \{0, 1\}$ (the other cases are analogous). Consider two texts p^1 and p^2 in the same coset of $\mathcal{M}_I$, that is $\mathcal{M}_I \oplus a$ for a given $a \in$

$\mathcal{M}_I^{\perp}$. By definition, there exist $x_0, x_1, y_0, y_1, z_0, z_1, w_0, w_1 \in \mathbb{F}_{2^8}$ and $x'_0, x'_1, y'_0, y'_1, z'_0, z'_1, w'_0, w'_1 \in \mathbb{F}_{2^8}$ such that:

$$p^1 = a \oplus MC \cdot \begin{bmatrix} x_0 & y_0 & 0 & 0 \\ x_1 & 0 & 0 & w_0 \\ 0 & 0 & z_0 & w_1 \\ 0 & y_1 & z_1 & 0 \end{bmatrix}, \qquad p^2 = a \oplus MC \cdot \begin{bmatrix} x'_0 & y'_0 & 0 & 0 \\ x'_1 & 0 & 0 & w'_0 \\ 0 & 0 & z'_0 & w'_1 \\ 0 & y'_1 & z'_1 & 0 \end{bmatrix}.$$

For the following, let $2 \equiv 0\text{x}02$ and $3 \equiv 0\text{x}03$.

Following the same strategy of Sect. 6, the idea is to consider all the possible cases in which some or no-one variables of p^1 are equal to the ones of p^2. Note that the case $x_1 = x'_1$, $y_1 = y'_1$, $z_1 = z'_1$ and $w_1 = w'_1$ (i.e. two texts that belong to the same coset of $\mathcal{M}_I$ for $|I| = 1$) has already been considered. In particular, by Theorem 3 it follows that in this case the number n is a multiple of 8.

First Case. W.l.o.g. we consider the case $y_1 = y'_1$, $w_i = w'_i$ and $z_i = z'_i$ for $i = 0, 1$, while $y_0 \neq y'_0$ and $x_i \neq x'_i$ for $i = 0, 1$ (the other cases are analogous).

Assume that for certain there exist x_0, x_1, y_0 and x'_0, x'_1, y'_0 such that the generated elements p^1 and p^2 satisfy $R(p^1) \oplus R(p^2) \in \mathcal{D}_J$ for a certain J for $|J| = 3$. First of all, we show that such variables can exist if $|J| = 3$. The condition $R(p^1) \oplus R(p^2) \in \mathcal{D}_J$ for a certain J with $|J| = 3$ implies that four bytes (one per column) of $R(p^1) \oplus R(p^2)$ must be equal to 0. Since there are six independent variables, a solution can exist (note that the number of variables is higher than the number of equations, so two variables are still "free"). Moreover, this is also due to the branch number of the MixColumns operation, which is five. Indeed, by simple computation the first column of $SR(\text{S-Box}(p^1) \oplus \text{S-Box}(p^2))$ (analogous for the others) is given by:

$$\begin{aligned} SR(\text{S-Box}(p^1) \oplus \text{S-Box}(p^2))_{0,0} &= \text{S-Box}(2 \cdot x_0 \oplus 3 \cdot x_1 \oplus a_{0,0} \oplus a_{1,0}) \oplus \\ &\quad \oplus \text{S-Box}(2 \cdot x'_0 \oplus 3 \cdot x'_1 \oplus a_{0,0} \oplus a_{1,0}), \\ SR(\text{S-Box}(p^1) \oplus \text{S-Box}(p^2))_{1,0} &= \text{S-Box}(y_0 \oplus a_{1,1}) \oplus \text{S-Box}(y'_0 \oplus a_{1,1}), \\ SR(\text{S-Box}(p^1) \oplus \text{S-Box}(p^2))_{2,0} &= SR(\text{S-Box}(p^1) \oplus \text{S-Box}(p^2))_{3,0} = 0. \end{aligned}$$

Thus, if we compute $MC \circ SR(\text{S-Box}(p^1) \oplus \text{S-Box}(p^2))$ (that is, $R(p^1) \oplus R(p^2)$), since at most two input bytes are different from zero, then it follows that at least three output bytes must be different from zero, or equivalently at most one output byte can be equal to zero. As a consequence, it is possible that $R(p_1) \oplus R(p^2) \in \mathcal{D}_J$ for $|J| = 3$, but not for $|J| \leq 2$. We emphasize that with respect to the case $|I| = 1$, it is possible that one input byte of the MixColumns operation can be equal to zero. Indeed, it is possible that exist x_0 and x'_0 such that $SR(\text{S-Box}(p^1) \oplus \text{S-Box}(p^2))_{0,0}$ (analogous for the others columns).

As before, the idea is to consider the pairs of texts generated by all the possible combinations of these six variables, as for example $\langle x_0, x_1, y'_0 \rangle$ and $\langle x'_0, x'_1, y_0 \rangle$, $\langle x_0, x'_1, y_0 \rangle$ and $\langle x'_0, x_1, y'_0 \rangle$, $\langle x'_0, x_1, y_0 \rangle$ and $\langle x_0, x'_1, y'_0 \rangle$, $\langle x_1, x_0, y'_0 \rangle$ and $\langle x'_0, x'_1, y_0 \rangle$ (note that the elements generated by $\langle x_0, x_1, y'_0 \rangle$ and by $\langle x_1, x_0, y'_0 \rangle$ are different) and so on.

We analyze these cases. It is simple to observe that if p^1 generated by $\langle x_0, x_1, y_0\rangle$ and p^2 generated by $\langle x'_0, x'_1, y'_0\rangle$ belong to the same coset of $\mathcal{M}_J$ for $|J| = 3$ after one round, then also the elements generated by $\langle x_0, x_1, y'_0\rangle$ and $\langle x'_0, x'_1, y_0\rangle$ have the same property. To prove this fact, it is sufficient to show that $R(p^1) \oplus R(p^2) = R(\hat{p}^1) \oplus R(\hat{p}^2)$. As an example, by simple computation, it is simple to observe that for the first column:

$$SR(\text{S-Box}(\hat{p}^1) \oplus \text{ S-Box}(\hat{p}^2))_{i,0} = SR(\text{S-Box}(p^1) \oplus \text{ S-Box}(p^2))_{i,0} \qquad \forall i,$$

which implies the statement.

Consider now the elements $\hat{p}^1$ generated by $\langle x_0, x'_1, y_0\rangle$ and $\hat{p}^2$ generated by $\langle x'_0, x_1, y'_0\rangle$ (similar for the elements generated by $\langle x'_0, x_1, y_0\rangle$ and $\langle x_0, x'_1, y'_0\rangle$). By simple computation, the first column of $SR(\text{S-Box}(\hat{p}^1) \oplus \text{ S-Box}(\hat{p}^2))$ (analogous for the others) is given by:

$$\begin{aligned} SR(\text{S-Box}(\hat{p}^1) \oplus \text{ S-Box}(\hat{p}^2))_{0,0} = \text{S-Box}(2 \cdot x_0 \oplus 3 \cdot x'_1 \oplus a_{0,0} \oplus a_{1,0}) \oplus \\ \oplus \text{S-Box}(2 \cdot x'_0 \oplus 3 \cdot x_1 \oplus a_{0,0} \oplus a_{1,0}) \end{aligned}$$

and for $i = 1, 2, 3$

$$SR(\text{S-Box}(\hat{p}^1) \oplus \text{ S-Box}(\hat{p}^2))_{i,0} = SR(\text{S-Box}(p^1) \oplus \text{ S-Box}(p^2))_{i,0}.$$

Since the S-Box is a non-linear operation, three different cases can happen:

1. $SR(\text{S-Box}(\hat{p}^1) \oplus \text{ S-Box}(\hat{p}^2))_{0,0} = 0$;
2. $SR(\text{S-Box}(\hat{p}^1) \oplus \text{ S-Box}(\hat{p}^2))_{0,0} \neq 0$ and the elements $\hat{p}^1$ and $\hat{p}^2$ belong to the same coset of $\mathcal{D}_J$ after one round (for the same J of p^1 and p^2);
3. $SR(\text{S-Box}(\hat{p}^1) \oplus \text{ S-Box}(\hat{p}^2))_{0,0} \neq 0$ and the elements $\hat{p}^1$ and $\hat{p}^2$ don't belong to the same coset of $\mathcal{D}_J$ after one round (for the same J of p^1 and p^2).

We analyze in details these three cases, starting from the first one. As first thing, note that this case can happen since $R(p_1) \oplus R(p^2) \in \mathcal{D}_J$ imposes a condition only on four out of six variables, that is two variables are still "free". If $SR(\text{S-Box}(\hat{p}^1) \oplus \text{ S-Box}(\hat{p}^2))_{0,0} = 0$, it follows that only one byte (i.e. the second one) of the first column of $SR(\text{S-Box}(\hat{p}^1) \oplus \text{ S-Box}(\hat{p}^2))$ is different from 0 (since $y_0 \neq y'_0$). Thus, since MixColumns operation has branch number 5, all the bytes of the first column of $R(\hat{p}^1) \oplus R(\hat{p}^2)$ must be different from zero, that is $R(\hat{p}^1) \oplus R(\hat{p}^2) \notin \mathcal{D}_J$ for $|J| \leq 3$. However, note that also in this case it is possible to deduce something. Indeed, by the previous consideration, it follows that the elements generated by $\langle x_0, x'_1, y'_0\rangle$ and by $\langle x'_0, x_1, y_0\rangle$ can not belong to the same coset of $R(\hat{p}^1) \oplus R(\hat{p}^2) \notin \mathcal{D}_J$.

Consider now the other two cases. Since the S-Box is a non-linear operation, it is not possible to guarantee that

$$SR(\text{S-Box}(\hat{p}^1) \oplus \text{ S-Box}(\hat{p}^2))_{0,0} = SR(\text{S-Box}(p^1) \oplus \text{ S-Box}(p^2))_{0,0}.$$

In other words, they can be equal (which implies that the elements $\hat{p}^1$ and $\hat{p}^2$ belong to the same coset of $\mathcal{D}_J$ after one round for the same J of p^1 and p^2) or

different. In this second case, one can not say anything about the fact that the elements $\hat{p}^1$ and $\hat{p}^2$ belong or not to the same coset of $\mathcal{D}_J$ after one round for the same J of p^1 and p^2. However, suppose that $\hat{p}^1$ and $\hat{p}^2$ belong to the same coset of $\mathcal{D}_J$ after one round for the same J of p^1 and p^2 (which is independent by the previous condition). In the same way of before, note that also the elements generated by $\langle x_0, x'_1, y'_0\rangle$ and $\hat{p}^2$ generated by $\langle x'_0, x_1, y_0\rangle$ have the same property.

Thus, assume that p^1 generated by $\langle x_0, x_1, y_0\rangle$ and p^2 generated by $\langle x'_0, x'_1, y'_0\rangle$ belong or not to the same coset of $\mathcal{D}_J$ after one round. By previous considerations, it follows that also the $\hat{p}^1$ generated by $\langle x_0, x'_1, y_0\rangle$ and $\hat{p}^2$ generated by $\langle x'_0, x_1, y'_0\rangle$ have the same property. Thus, even if we can not do any claim for the other texts generated by a different combination of these six variables, it is possible to conclude that - *for fixed* $y_1 = y'_1$, $w_i = w'_i$ and $z_i = z'_i$ for $i = 0, 1$ - the number of collisions must be a multiple of 2 for this case.

Finally, since we are working with the entire coset of $\mathcal{M}_{0,1}$ - that is, $y_1 = y'_1$, $w_i = w'_i$ and $z_i = z'_i$ for $i = 0, 1$ can take any possible value - and due to the same considerations of Sect. 6, it follows that the number of collisions must be a multiple of $2 \cdot (2^8)^5 = 2^{41}$ for this case.

Second Case. Similar considerations can be done for the case $w_i = w'_i$ and $z_i = z'_i$ for $i = 0, 1$, while $x_i \neq x'_i$ and $y_i \neq y'_i$ for $i = 0, 1$ (the other cases are analogous).

Assume there exist x_0, x_1, y_0, y_1 and x'_0, x'_1, y'_0, y'_1 such that the generated elements p^1 and p^2 satisfy $R(p_1) \oplus R(p^2) \in \mathcal{D}_J$ for a certain J with $|J| = 3$. As before, note that this is possible since this implies that four bytes of $R(p_1) \oplus R(p^2)$ (one per column) must be equal to 0. Since there are eight independent variables, a solution can exist (note that the number of variables is higher than the number of equations, so four variables are still "free"). Due to the branch number of the MixColumns operation, even if four variables are still "free" it is not possible that $R(p_1) \oplus R(p^2) \in \mathcal{M}_J$ for $|J| \leq 2$. Indeed, the first column of $SR(\text{S-Box}(p^1) \oplus \text{S-Box}(p^2))$ (analogous for the others) is given by:

$$\begin{aligned}
SR(\text{S-Box}(p^1) \oplus \text{S-Box}(p^2))_{0,0} &= \text{S-Box}(2 \cdot x_0 \oplus 3 \cdot x_1 \oplus a_{0,0} \oplus a_{1,0}) \oplus \\
&\quad \oplus \text{S-Box}(2 \cdot x'_0 \oplus 3 \cdot x'_1 \oplus a_{0,0} \oplus a_{1,0}), \\
SR(\text{S-Box}(p^1) \oplus \text{S-Box}(p^2))_{1,0} &= \text{S-Box}(y_0 \oplus y_1 \oplus a_{0,1} \oplus a_{3,0}) \oplus \\
&\quad \oplus \text{S-Box}(y'_0 \oplus y'_1 \oplus a_{0,1} \oplus a_{3,0}), \\
SR(\text{S-Box}(p^1) \oplus \text{S-Box}(p^2))_{2,0} &= SR(\text{S-Box}(p^1) \oplus \text{S-Box}(p^2))_{3,0} = 0.
\end{aligned}$$

After the MixColumns operation $MC \circ SR(\text{S-Box}(p^1) \oplus \text{S-Box}(p^2))$, since at most two input bytes are different from zero, then it follows that at least three output bytes must be different from zero.

Thus, given x_0, x_1, y_0, y_1 and x'_0, x'_1, y'_0, y'_1, the idea is to consider all the possible combinations as before. Also in this case, we can do a claim only on one of them. In particular, if two elements p^1 generated by $\langle x_0, x_1, y_0, y_1\rangle$ and p^2 generated by $\langle x'_0, x'_1, y'_0, y'_1\rangle$ satisfies $R(p_1) \oplus R(p^2) \in \mathcal{D}_J$, we can only claim that also the elements $\hat{p}^1$ generated by $\langle x'_0, x'_1, y_0, y_1\rangle$ and $\hat{p}^2$ generated by $\langle x_0, x_1, y'_0, y'_1\rangle$

have the same property. Considerations for the other combinations are similar to the previous case. Thus, we can claim that - *for fixed* $w_i = w'_i$ and $z_i = z'_i$ for $i = 0, 1$ - also for this case the number of collisions is a multiple of 2.

Finally, since we are working with the entire coset of $\mathcal{M}_{0,1}$ - that is, $w_i = w'_i$ and $z_i = z'_i$ for $i = 0, 1$ can take any possible value - and due to the same considerations of Sect. 6, it follows that the number of collisions must be a multiple of $2 \cdot (2^8)^4 = 2^{33}$ for this case.

Second Consideration. What can we deduce by the previous two cases? Suppose to have two texts p^1 generated by $\langle x \equiv (x_0, x_1), y \equiv (y_0, y_1)\rangle$ and p^2 generated by $\langle x' \equiv (x'_0, x'_1), y' \equiv (y'_0, y'_1)\rangle$ that satisfy $R(p_1) \oplus R(p^2) \in \mathcal{D}_J$ for $|J| = 3$ and where $x, y \in \mathbb{F}_{2^8} \times \mathbb{F}_{2^8} \equiv \mathbb{F}_{2^8}^2$. We have seen that given these two elements, one can only claim that also the texts $\hat{p}^1$ generated by $\langle x' \equiv (x'_0, x'_1), y \equiv (y_0, y_1)\rangle$ and $\hat{p}^2$ generated by $\langle x \equiv (x_0, x_1), y' \equiv (y'_0, y'_1)\rangle$ have the same property, that is $R(\hat{p}_1) \oplus R(\hat{p}^2) \in \mathcal{D}_J$ for the same J of p^1 and p^2. In the same way, if $R(p_1) \oplus R(p^2) \notin \mathcal{D}_J$ for $|J| = 3$ one can claim that $R(\hat{p}_1) \oplus R(\hat{p}^2) \notin \mathcal{D}_J$, where $p^1, p^2, \hat{p}^1$ and $\hat{p}^2$ are defined as before.

As a consequence, the idea for the case $|I| = 2$ is not to consider the variables that generate the texts and that are in the same column as independent. In other words, the idea is to work with variables in $\mathbb{F}_{2^8}^2$ and not in $\mathbb{F}_{2^8}$, i.e. to consider only all the possible combinations of $x \equiv (x_0, x_1), y \equiv (y_0, y_1)$ and $x' \equiv (x'_0, x'_1), y' \equiv (y'_0, y'_1)$, and not of x_0, x_1, y_0, y_1 and x'_0, x'_1, y'_0, y'_1. Using this strategy and working in the same way of Sect. 6, it is possible to analyze all the possible cases.

For example, consider the case in which $w_i = w'_i$ for $i = 0, 1$ and $x \equiv (x_0, x_1) \neq x' \equiv (x'_0, x'_1)$, $y \equiv (y_0, y_1) \neq y' \equiv (y'_0, y'_1)$ and $z \equiv (z_0, z_1) \neq z' \equiv (z'_0, z'_1)$. In the same way of before, it is only possible to prove that if there exist p^1 generated by $\langle x, y, z\rangle$ and p^2 generated by $\langle x', y', z'\rangle$ such that $R(p^1) \oplus R(p^2) \in \mathcal{D}_J$ for $|J| \geq 2$, then a total of four elements generated by

- $\langle x, y, z\rangle$ and $\langle x', y', z'\rangle$
- $\langle x', y, z\rangle$ and $\langle x, y', z'\rangle$
- $\langle x, y', z\rangle$ and $\langle x', y, z'\rangle$
- $\langle x, y, z'\rangle$ and $\langle x', y', z\rangle$

have the same property. No claim can be made about other combinations of variables (as before, this is due to the fact that the S-Box is non-linear). It follows that - *for fixed* $w_i = w'_i$ for $i = 0, 1$- the number of collisions must be a multiple of 4 for this case. As before, since we are working with the entire coset of $\mathcal{M}_{0,1}$ it follows that the number of collisions must be a multiple of $4 \cdot (2^8)^2 = 2^{18}$. Moreover, since the branch number of the MixColumns operation is five, note that it is not possible that $R(p^1) \oplus R(p^2) \in \mathcal{D}_J$ for $|J| = 1$ if $w_i = w'_i$ for $i = 0, 1$ (even if $R(p^1) \oplus R(p^2) \in \mathcal{D}_J$ for $|J| = 2$ imposes only 8 conditions while the number of variables is 12, so 4 variables are still "free").

Similar considerations can be done for the case in which all the variables are different. As a consequence, the theorem is proved for the case $|I| = 2$.

B.2 Case $|I| = 3$

The case $|I| = 3$ is analogous to the case $|I| = 2$ and to the proof given in Sect. 6. For this reason, we limit ourselves to show how to adapt the proof of the case $|I| = 2$ for this case.

W.l.o.g assume $I = \{0, 1, 2\}$ and consider two texts p^1 and p^2 in the same coset of $\mathcal{M}_I$, i.e. $\mathcal{M}_I \oplus a$ for $a \in \mathcal{M}_I^{\perp}$. By definition, there exist $x_0, x_1, x_2, y_0, y_1, y_2, z_0, z_1, z_2, w_0, w_1, w_2 \in \mathbb{F}_{2^8}$ and $x'_0, x'_1, x'_2, y'_0, y'_1, y'_2, z'_0, z'_1, z'_2, w'_0, w'_1, w'_2 \in \mathbb{F}_{2^8}$ such that:

$$p^1 = a \oplus MC \cdot \begin{bmatrix} x_0 & y_0 & z_0 & 0 \\ x_1 & y_1 & 0 & w_0 \\ x_2 & 0 & z_1 & w_1 \\ 0 & y_2 & z_2 & w_2 \end{bmatrix}, \qquad p^2 = a \oplus MC \cdot \begin{bmatrix} x'_0 & y'_0 & z'_0 & 0 \\ x'_1 & y'_1 & 0 & w'_0 \\ x'_2 & 0 & z'_1 & w'_1 \\ 0 & y'_2 & z'_2 & w'_2 \end{bmatrix}.$$

Similarly to the case $|I| = 2$, the idea is to work with variables in $\mathbb{F}_{2^8}^3 \equiv \mathbb{F}_{2^8} \times \mathbb{F}_{2^8} \times \mathbb{F}_{2^8}$, e.g. $x \equiv (x_0, x_1, x_2), y \equiv (y_0, y_1, y_2)$ and so on. In other words, the idea is to consider the variables in the same column as not independent, that is to consider the possible combinations only of variables in $\mathbb{F}_{2^8}^3$ and not in $\mathbb{F}_{2^8}$.

C Comparison of 5-Round Secret-Key Distinguishers

Table 2. *Properties of 5-round secret-key Distinguishers for AES.* In this table, we consider all the possible secret-key distinguishers for AES (included the key-recovery attacks), and we highlight the major properties. In particular, based on the previous categorization: "(1)" denotes a distinguisher which exploits a property which is independent of the secret key; "(2)" denotes a distinguisher which requires the knowledge of the entire secret key, while "(2a)" denotes a distinguisher which requires only the knowledge of part of the secret key; "*MC*" denotes a distinguisher which is independent of the final MixColumns; "Secret S-Box" denotes the case of AES with a secret S-Box and "(b)" denotes a distinguisher which does not find/exploit any information of the secret S-Box.

Property	**(1)**	**(2)**	**(2a)**	**MC**	**Secret S-Box**	**(b)**	**Reference**
Subspace Trail	×			×	×	×	**Sect. 4**
Impossible Differential			×		×	×	[15]
Integral			×	×	×	×	[22]
Impossible Differential			×	×			[4] - [15, App. I]
Integral		×		×	×		[26]
Integral		×		×			[8]
Polytopic		×		×			[25]
MitM		×		×			[12, Sec. 7.5.1]

We recall the categorization of secret-key distinguishers proposed in Sect. 1.1:

1. a distinguisher which is completely independent of the secret key (that is, it exploits property that are not related to the existence of a key) and independent of the details of the S-Box;
2. a distinguisher which depends on the existence of a key and is derived by a key recovery attack; in particular, we highlight two properties that the distinguisher of this category can have, which are
   (a) a distinguisher which requires the knowledge only of a part (e.g. one byte) of the key;
   (b) a distinguisher which is independent of the details of the S-Box, i.e. which does not find or/and exploit any information of the S-Box.

   We stress that these two properties are not mutually exclusive.

A complete comparison of all the secret key distinguishers and key recovery attacks (used as distinguishers) is provided in Table 2.

D Implementation of the Distinguisher using a re-Ordering Algorithm

In Sect. 4 we have presented an implementation of the distinguisher using data structures. In this appendix, we propose another way to implement the distinguisher which exploits a re-ordering algorithm. This implementation could be in some cases more efficient than the one proposed in Sect. 4 when e.g. it is required to do further operations on the pairs of ciphertexts (c^1, c^2) such that $c^1 \oplus c^2 \in \mathcal{M}_J$. For simplicity (and in order to have a direct comparison with the other implementation), we present this strategy based on the re-ordering algorithm for the case $|J| = 3$, which has a total cost of approximately 2^{39} table look-ups for each of the four subspaces $\mathcal{M}_J$ for $|J| = 3$, where the used tables are of size 2^{32} texts (or $2^{32} \cdot 16 = 2^{36}$ byte).

The basic idea to do this is to re-order the ciphertexts. 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 8. *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* $i = 0, ..., 3$*:*

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

where $\leq$ *defined as in Def. 7.*

To better explain this definition and the re-ordering algorithm, we provide a concrete example in App. D.2. Thus, as first step, one must re-order the 2^{32} ciphertexts of each coset with respect to the partial order relationship $\preceq$ defined before.

After the re-ordering process, in order to count the number of pairs of texts that belong to the same coset of $\mathcal{M}_J$, one can work only on consecutive ordered elements. Indeed, consider r consecutive elements $c^l, c^{l+1}, ..., c^{l+r-1}$, with $r \geq 2$. Suppose that for each k with $l \leq k \leq l + r - 2$:

$$c^k \oplus c^{k+1} \in \mathcal{M}_J.$$

Since $\mathcal{M}_J$ is a subspace, it follows immediately that for each s, t with $l \leq s, t \leq l + r - 2$

$$c^s \oplus c^t \in \mathcal{M}_J.$$

Thus, given $r \geq 2$ consecutive elements that belong to the same coset of $\mathcal{M}_J$, it follows that

$$\binom{r}{2} = \frac{r \cdot (r-1)}{2}$$

different pairs belong to the same coset of $\mathcal{M}_J$. In the same way, consider r consecutive elements $c^l, c^{l+1}, ..., c^{l+r-1}$ with $r \geq 2$, such that $c^k \oplus c^{k+1} \notin \mathcal{M}_J$ for each$k$ with $l \leq k \leq l + r - 2$. Since $\mathcal{M}_J$ is a subspace, it follows immediately that $c^s \oplus c^t \notin \mathcal{M}_J$ for each s, t with $l \leq s, t \leq l + r - 2$.

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

What is the total computational cost of this procedure? Given a set of n ordered elements, the computational cost to count the number of pairs that belong to the same coset of $\mathcal{M}_J$ is well approximated by n look-ups table, since one works only on consecutive elements. Using the *merge sort* algorithm to order this set (which has a computational cost of $O(n \log n)$ memory access), the total computational cost for the verifier is approximately of

$$n \cdot (1 + \log n) \quad \text{table look-ups.}$$

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[15] 1 table look-up $\approx$ 1 round of AES).

[15] We highlight that even if this approximation is not formally correct - the size of the table of an S-Box look-up is lower than the size of the table used for our proposed

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: 1 for an AES permutation, 0 otherwise (prob. of success: $\geq 99\%$)
for *all* J *with* $|J| = 3$ **do**
  Re-order the 2^{32} (plaintexts, ciphertexts) pairs using the *partial order relationship* $\preceq$ defined in Def. 8; `// remember that the order` $\preceq$ `depends on` J
  Let $(\tilde{p}^i, \tilde{c}^i)$ for $i = 0, ..., 2^{32} - 1$ the order (plaintext, ciphertext) pairs;
  $n \leftarrow 0$; `//` n `denotes the number of collisions in` $\mathcal{M}_J$
  $i \leftarrow 0$;
  while $i < 2^{32}$ **do**
    $r \leftarrow 1$;
    $j \leftarrow i$;
    while $\tilde{c}^j \oplus \tilde{c}^{j+1} \in \mathcal{M}_J$ **do**
      $r \leftarrow r + 1$;
      $j \leftarrow j + 1$;
    end
    $i \leftarrow j + 1$;
    $n \leftarrow n + r \cdot (r - 1)/2$;
  end
  if $(n \bmod 8) \neq 0$ **then**
    return 0;
  end
end
return 1.

Algorithm 2: *Secret-Key Distinguisher for 5 Rounds of AES* which exploits a property which is independent of the secret key - probability of success: $\geq 99\%$.

D.1 Practical Verification

Using a C/C++ implementation, we have practically verified the distinguisher implemented using a re-ordering algorithm as described in this section on a small scale variant of AES, as presented in [6].

We refer to Sect. 4 for a complete discussion about the implementation on small-scale AES and the results, and we limit here to focus on the computational cost. The differences between this small-scale AES and the real AES regard the total number of collisions, which in this case is well approximated by $2^{15} \cdot (2^{16} - 1) \cdot 2^{-16} \approx 2^{15}$ for each coset, and the lower computational cost, which can be approximated by $4 \cdot 2^{16} \cdot (\log 2^{16} + 1) = 2^{21}$ memory look-ups for each coset, besides the memory costs. The *average* practical computational cost found in our experiments is approximately 2^{22} memory look-ups. This difference (a factor 2) can be simply justified by the fact that the cost of the merge sort algorithm is $O(n \cdot \log n)$ and by the definition of the big O notation (recalled in App. D.2).

---

distinguisher, it allows to give a comparison between our proposed distinguisher and the others currently present in literature. At the same time, we note that the same approximation is largely used in literature.

D.2 The Merge Sort Algorithm: a Concrete Example

In Sect. 4 we use the merge sort algorithm to reduce the computational complexity of our new secret-key distinguisher on 5 rounds of AES. In this section, we recall some concepts of this sort algorithm, and we provide an example for our case.

The *merge sort algorithm* is a sort algorithm for rearranging lists (or any other data structure that can only be accessed sequentially) into a specified order. Assume a sequence of n elements A is given, which we assume is stored in an array $A[1, ..., n]$. The objective is to output a permutation of this sequence, sorted in increasing order. This is normally done by permuting the elements within the array A. Given a list of n elements, merge sort has an average and worst-case performance of[16] $O(n \cdot \log(n))$.

Merge sort algorithm is an example of "divide-and-conquer" algorithm, which major elements are:

- *Divide*: Split A down the middle into two subsequences, each of size roughly $n/2$;
- *Conquer*: Sort each subsequence (by calling MergeSort recursively on each subsequence);
- *Combine*: Merge the two sorted subsequences into a single sorted list.

The dividing process ends when we have split the subsequences down to a single item. A sequence of length one is trivially sorted. The key operation where all the work is done is in the combine stage, which merges together two sorted lists into a single sorted list.

We refer to [7] for a complete explanation of the merge sort algorithm, and we limit here to give an example for our case. Assume to have four texts $A, B, C, D \in \mathbb{F}_{2^8}^{4\times 4}$:

$$A = \begin{bmatrix} 0\text{x}27 & 0\text{x}a3 & 0\text{x}46 & 0\text{x}01 \\ 0\text{x}12 & 0\text{x}55 & 0\text{x}a6 & 0\text{x}bc \\ 0\text{x}46 & 0\text{x}30 & 0\text{x}d4 & 0\text{x}93 \\ 0\text{x}65 & 0\text{x}f2 & 0\text{x}07 & 0\text{x}21 \end{bmatrix}, \quad B = \begin{bmatrix} 0\text{x}27 & 0\text{x}03 & 0\text{x}10 & 0\text{x}aa \\ 0\text{x}66 & 0\text{x}55 & 0\text{x}32 & 0\text{x}bc \\ 0\text{x}52 & 0\text{x}a3 & 0\text{x}27 & 0\text{x}01 \\ 0\text{x}f2 & 0\text{x}97 & 0\text{x}ff & 0\text{x}23 \end{bmatrix},$$

$$C = \begin{bmatrix} 0\text{x}27 & 0\text{x}76 & 0\text{x}22 & 0\text{x}7d \\ 0\text{x}08 & 0\text{x}a3 & 0\text{x}00 & 0\text{x}bc \\ 0\text{x}26 & 0\text{x}a3 & 0\text{x}d4 & 0\text{x}35 \\ 0\text{x}17 & 0\text{x}f2 & 0\text{x}0c & 0\text{x}2b \end{bmatrix}, \quad D = \begin{bmatrix} 0\text{x}64 & 0\text{x}14 & 0\text{x}15 & 0\text{x}03 \\ 0\text{x}32 & 0\text{x}17 & 0\text{x}5c & 0\text{x}b1 \\ 0\text{x}23 & 0\text{x}88 & 0\text{x}d4 & 0\text{x}37 \\ 0\text{x}bb & 0\text{x}f3 & 0\text{x}43 & 0\text{x}96 \end{bmatrix}.$$

Our goal is to re-order them, using the merge sort algorithm and the *partial order* relationship $\preceq$ defined in Sect. 4, where we assume $l = \{0\}$ and $I = \{1, 2, 3\}$ (the

[16] Let f and g be two functions defined on some subset of the real numbers. One writes $f(x) = O(g(x))$ if and only if there exists a positive real number C and a real number x_0 such that $|f(x)| \leq C \cdot |g(x)|$ for all $x \geq x_0$. Only as an example, $f(x) = 3x^2 + x \cdot \log x + 2x^{-1} = O(x^2)$ and the constant C is different from 1. The notation can also be used to describe the behavior of f near some real number x_0, that is one writes $f(x) = O(g(x))$ if and only if there exists a positive numbers δ and C such that $|f(x)| \leq C \cdot |g(x)|$ for $|x - a| < \delta$.

example can be easily generalized for each number of texts and for each possible I and l). The final goal is to count the number of collisions among the ciphertexts in the same coset of $MC^{-1}(\mathcal{M}_{1,2,3}) = \mathcal{ID}_{1,2,3}$. For simplicity, we assume that an InverseMixColumns operation has been already applied to the four ciphertexts.

By definition 8, we are only interested in the bytes in positions - (row, column): $(0,0), (1,3), (2,2), (3,1)$. Indeed, two texts p and q belong to the same coset of $\mathcal{ID}_{1,2,3}$ (that is $p \oplus q \in \mathcal{ID}_{1,2,3}$) if and only if $p_{0,0} = q_{0,0}$, $p_{1,3} = q_{1,3}$, $p_{2,2} = q_{2,2}$ and $p_{3,1} = q_{3,1}$.

Using the merge sort algorithm, as first step one works on the pair A and B. Since $A_{0,0} = B_{0,0}$, $A_{1,3} = B_{1,3}$ and $B_{2,2} < A_{2,2}$, we can deduce that $B \preceq A$ with respect to the defined partial order $\preceq$. Thus, after the first step, the elements are re-ordered as B, A, C, D. In a similar way, one then works on the pair C and D. In this case, $C \preceq D$ since $C_{0,0} < D_{0,0}$. Thus, after the second step, the elements are re-ordered as B, A, C, D. At the third step, note that $A_{i,-i} = C_{i,-i}$ for each $i = 0, ..., 3$. However, since $C \leq A$ with respect to $\leq$ defined in Def. 7, one obtains the final sequence B, C, A, D[17].

Given an ordered array, in order to count the number of pairs whose texts belong to the same coset of $\mathcal{ID}_{1,2,3}$, one can work only on consecutive elements, that is on the pairs (B, C), (C, A) and (A, D). In this case, only one pair of texts (that is, (C, A)) belongs to the same coset of $\mathcal{ID}_{1,2,3}$.

As second example, consider the previous case in which the element D is defined as follow:

$$D = \begin{bmatrix} 0\text{x}27 & 0\text{x}14 & 0\text{x}15 & 0\text{x}03 \\ 0\text{x}32 & 0\text{x}17 & 0\text{x}5c & 0\text{x}bc \\ 0\text{x}23 & 0\text{x}88 & 0\text{x}d4 & 0\text{x}37 \\ 0\text{x}bb & 0\text{x}f2 & 0\text{x}43 & 0\text{x}96 \end{bmatrix}.$$

In this case, the re-ordered array is given by B, C, A, D[18]. In this case, working again on consecutive pairs (B, C), (C, A) and (A, D), two pairs of texts (that is, (C, A) and (A, D)) belongs to the same coset of $\mathcal{ID}_{1,2,3}$). Thus, one can conclude that there are $2 \cdot (2+1)/2 = 3$ pairs of texts (that is, also (C, D)) that belongs to the same coset of $\mathcal{ID}_{1,2,3}$.

We stress that after the re-ordering process, it is sufficient to work on consecutive texts in order to count the total number of texts that belong to the same coset of $\mathcal{ID}_{1,2,3}$ - in other words, it is not necessary to construct all the possible pairs.

[17] Only for completeness, we highlight that since $A_{i,-i} = C_{i,-i}$ for each $i = 0, ..., 3$, the final sequence B, A, C, D is equivalent to B, C, A, D for our goal.

[18] As before, we highlight that for our goal the elements A, C, D can be ordered in any possible way.