Structural Truncated Differential Attacks on round-reduced AES

Lorenzo Grassi

IAIK, Graz University of Technology, Austria
lorenzo.grassi@iaik.tugraz.at

Abstract. At Eurocrypt 2017 the first secret-key distinguisher for 5-round AES has been presented. Although it allows to distinguish a random permutation from an AES like one, it seems (rather) hard to exploit such a distinguisher in order to implement a key-recovery attack different than brute-force like.
In this paper, we propose new secret-key distinguishers for 4 and 5 rounds of AES that exploit properties which are independent of the secret key and of the details of the S-Box. While the 4-round distinguisher exploits in a different way the same property presented at Eurocrypt 2017, the new proposed 5-round one is obtained by combining our new 4-round distinguisher with a modified version of a truncated differential distinguisher. As a result, while a "classical" truncated differential distinguisher exploits the probability that a couple of texts satisfies or not a given differential trail independently of the others couples, our distinguisher works with sets of 2^{17} (related) couples of texts. In particular, our new 5-round AES distinguisher exploits the fact that the probability that at least one couple of texts of such a set satisfies a given differential trail is lower for 5-round AES than for a random permutation in order to distinguish the two cases. These probabilities exploited by the distinguishers have been practically verified on a small-scale AES.
Even if such a 5-round distinguisher has higher complexity than the one present in the literature, it allows to set up the *first* key-recovery attack on 6-round AES that exploits *directly* a 5-round secret-key distinguisher. The goal of this paper is indeed to present and explore new approaches, showing that even a distinguisher like the one presented at Eurocrypt - believed to be hard to exploit - can be used to set up a key-recovery attack. Finally we show how to exploit the proposed 4-round distinguisher to set up new (practically verified) key-recovery attacks on 5-round AES with a single secret S-Box.

Keywords: AES · Secret-Key Distinguisher · Key-Recovery Attack · Truncated Differential · Secret S-Box · Subspace Trail Cryptanalysis

Contents

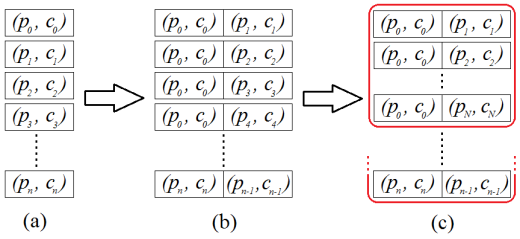

Figure 1: *New Differential Secret-Key Distinguishers up to 5 rounds of AES.* Consider N (plaintext, ciphertext) pairs (a). In a "classical" differential attack (b), one works independently on each couple of two (plaintext, ciphertext) pairs and exploits the probability that it satisfies a certain differential trail. In our attack (c), one divides the couples into non-random sets, and exploits particular relationships (based on differential trails) that hold among the couples that belong to the same set in order to set up a distinguisher.

1 Introduction

One of the weakest attacks that can be launched against a secret-key cipher is a secret-key distinguisher. In this attack, there are two oracles: one that simulates the cipher for which the cryptographic key has been chosen at random and one that simulates a truly random permutation. The adversary can query both oracles and her task is to decide which oracle is the cipher 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.

At Eurocrypt 2017, Grassi, Rechberger and Rønjom [GRR17a] presented the first 5-round secret-key distinguisher for AES which exploits a property which is independent of the secret key (it isn't a key-recovery attack) and of the details of the S-Box. This distinguisher is based on a new structural property for up to 5 rounds of AES: by appropriate choices of a number of input pairs it is possible to make sure that the number of times that the difference of the resulting output pairs lie in a particular subspace is *always* a multiple of 8. This distinguisher allows to distinguish an AES permutation from a random one with a success probability greater than 99% using 2^{32} chosen texts and a computational cost of $2^{35.6}$ look-ups. On the other hand, no key-recovery attack that exploits this distinguisher has been presented yet.

1.1 New Class of Secret-Key Distinguisher up to 5-round AES

In this paper, we present new secret-key distinguishers for 4- and 5-round AES which exploit in a different way the property presented in [GRR17a]. Such distinguishers - presented in detail in Sect. 5 and 7 - can be seen as a generalization of "classical" truncated differential attacks, as introduced by Knudsen in [Knu95].

Differential attacks exploit the fact that couples of plaintexts with certain differences yield other differences in the corresponding ciphertexts with a non-uniformity probability distribution. Such a property can be used both to distinguish an AES permutation from a random one, and to recover the secret key. A variant of this attack/distinguisher is the truncated differential attack [Knu95], in which the attacker considers only part of the difference between pairs of texts, i.e. a differential attack where only part of the difference in the ciphertexts can be predicted. We emphasize that in these cases the attacker focuses on the probability that single pairs of plaintexts with certain differences yield other differences in the corresponding ciphertexts *independently* of the other pairs.

Our new distinguishers proposed in this paper are also differential in nature. Instead of

Table 1: *Secret-Key Distinguishers for AES.* 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 95%. Time complexity is measured in equivalent encryptions (E), memory accesses (M) or XOR operations (XOR) - using the common approximation 20 M $\approx$ 1 Round of Encryption. The distinguishers of this paper are in bold.

Property	**Rounds**	**Data** (CP/CC)	**Cost**	**Ref.**
Impossible Differential	4	$2^{16.25}$	$2^{22.3}$ M $\approx 2^{16}$ E	[BK01]
Diff. Structural	**4**	$\mathbf{2^{17}}$	$\mathbf{2^{23.1}}$ **M** $\approx \mathbf{2^{16.75}}$ **E**	**Sect. 5**
Integral	4	2^{32}	2^{32} XOR	[DKR97]
Diff. Structural	4	2^{33}	2^{40} M $\approx 2^{33.7}$ E	[GRR17a]
Diff. Structural	5	2^{32}	$2^{35.6}$ M $\approx 2^{29}$ E	[GRR17a]
Prob. Diff. Struc.	**5**	$\mathbf{2^{51.2}}$	$\mathbf{2^{77.3}}$ **M** $\approx \mathbf{2^{70.7}}$ **E**	**Sect. 7 - App. D**

Prob. Diff. Struc.: Probabilistic Differential Structure

working on each couple[1] of two (plaintext, ciphertext) pairs independently of the others as in the previous case, in our case *one works on the relations that hold among the couples. In other words, given a couple of two (plaintext, ciphertext) pairs with a certain input/output differences, one focuses on how it influences other couples of two (plaintext, ciphertext) pairs to satisfy particular input/output differences.*

Referring to Fig. 1, given n chosen (plaintext, ciphertext) pairs, in a "classical" attack one works on each couple independently of the others - case (b). In our distinguishers/attacks, one first divides the couples in (non-random) sets of $N \geq 2$ couples - case (c). These sets are defined such that particular relationships (that involve differential trails and linear relationships) hold among the plaintexts of the couples that belong to the same set. Thus, consider a pair of plaintexts that belong to the same coset[2] of a particular subspace $\mathcal{C}$, such that the corresponding pair of ciphertexts belong to the same coset of another particular subspace $\mathcal{M}$. Our 4-round secret-key distinguisher proposed in Sect. 5 exploits the fact that for an AES permutation other couples of (plaintext, ciphertext) pairs have the same property with probability 1. All these couples make the sets just described and depicted in Fig. 1. Another possibility is to consider the probability that a given set contains at least one couple that satisfies a particular differential trail. Our proposed 5-round secret-key distinguisher exploits the fact that this probability is (a little) lower for 5-round AES than for a random permutation - independently of the key. All details are given in Sect. 7.

1.2 New Key-Recovery Attacks on 5- and 6-round AES-128

Even if our 5-round secret-key distinguisher is worse than the one presented in [GRR17a], it allows to set up *the first 6 rounds key-recovery attack on AES that exploits directly a 5-round secret-key distinguisher* (which exploits a property which is independent of the secret key). In particular, we propose in Sect. 5.3 an attack on 5-round AES that exploits the distinguisher on 4 rounds proposed in Sect. 5 (with the *lowest computational cost* among the attacks currently present in the literature), while in Sect. 7.4 we propose the first attack on 6 rounds of AES that exploits the distinguisher on 5 rounds presented in Sect. 7. The idea of both these attacks is to choose plaintexts in the same coset of

[1] We use the term "pair" to denote a plaintext and its corresponding ciphertext. A "couple" denotes a set of two such pairs.

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

Table 2: *Comparison of attacks on round-reduced AES-128.* Data complexity is measured in number of required chosen plaintexts/ciphertexts (CP/CC). Time complexity is measured in round-reduced AES encryption equivalents (E) - the number in the brackets denotes the precomputation cost (if not negligible). Memory complexity is measured in texts (16 bytes). R_{Dist} denotes the number of rounds of the secret-key distinguisher exploited to set up the attack. Attacks presented in this paper are in bold.

Attack	Rounds	Data	Computation	Memory	R_{Dist}	Ref.
MitM	5	8	2^{64}	2^{56}	-	[Der13, Sec. 7.5.1]
Imp. Polytopic	5	15	2^{70}	2^{41}	3	[Tie16]
Partial Sum	5	2^{8}	2^{38}	small	4	[Tun12]
Integral (EE)	5	2^{11}	$2^{45.7}$	small	4	[DR02]
Imp. Differential	5	$2^{31.5}$	2^{33} $(+2^{38})$	2^{38}	4	[BK01]
Integral (EB)	5	2^{33}	$2^{37.7}$	2^{32}	4	[DR02]
Diff. Struc.	**5**	$\mathbf{2^{33.6}}$	$\mathbf{2^{33.3}}$	$\mathbf{2^{34}}$	**4**	**Sect. 5.3 - App. C**
MitM	6	2^{8}	$2^{106.2}$	$2^{106.2}$	-	[DF13]
Partial Sum	6	2^{32}	2^{42}	2^{40}	4	[Tun12]
Integral	6	2^{35}	$2^{69.7}$	2^{32}	4	[DR02]
Prob. Diff. Struc.	**6**	$\mathbf{2^{72.8}}$	$\mathbf{2^{106}}$	$\mathbf{2^{35.5}}$	**5**	**Sect. 7.4**
Imp. Differential	6	$2^{91.5}$	2^{122}	2^{89}	4	[CKK+02]

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

a particular subspace $\mathcal{D}$ which is mapped after one round into a coset of $\mathcal{C}$. Using the distinguishers just introduced and the fact that the behavior for a wrongly guessed key is (approximately) the same of a random permutation, it is possible to deduce the right key.

Generic Considerations. Before we go on, we would like to do some preliminary considerations about our work, in particular about the fact that our distinguishers and key-recovery attacks presented in this paper have higher complexities than the ones currently present in the literature. Even if all the attacks on AES-like ciphers currently present in the literature are constantly improved, they seem not be able to break full-AES - with the only exception of the Biclique attack [BKR11], which can be considered as brute force[3]. Thus, besides improving the known attacks present in the literature, we believe that it is important and crucial to propose new idea and techniques. Even if they are not initially competitive, *they can provide new directions of research and can lead to new competitive attacks*. Only to provide an example, consider the impossible differential attack on AES. When it was proposed in 2001 by Biham and Keller [BK01], it was an attack on ("only") 5 rounds of AES and it was not competitive with respect to others attacks, as the integral one. It took approximately 6 years before that such attack was extended and set up against 7-round AES-128 [ZWF07], becoming one of the few attacks (together with Meet-in-the-Middle [DFJ13]) on such number of rounds. We believe that similar considerations can be done for the attacks/distinguisher proposed in this paper. In particular, the main contribution and merit of our paper is to show *for the first time* that even a distinguisher of the type [GRR17a] - *believed to be hard to exploit* - can be used to set up key-recovery attacks.

[3]The biclique attack on 10-round AES-128 requires 2^{88} chosen texts and it has a computational cost of approximately $2^{126.2}$ encryptions.

Table 3: *Comparison of attacks on round-reduced AES-128 with secret S-Box.* Data complexity is measured in number of required chosen plaintexts/ciphertexts (CP/CC). Time complexity is measured in round-reduced AES encryption equivalents (E), memory accesses (M) or XOR operations. Memory complexity is measured in texts (16 bytes). The case in which the final MixColumns operation is omitted is denoted by "r.5 rounds" - r full rounds + the final one. The symbol * denotes an attack that can *not* work independently on the S-Box and on the key. New attacks are in bold.

Attack	Rounds	Data	Computation	Memory	Reference
I*	4.5 - 5	2^{40} CC	$2^{38.7}$ E	2^{40}	[TKKL15]
I*	4.5 - 5	2^{40} CP	$2^{54.7}$ E	2^{40}	[TKKL15, Sect. 3.5]
Diff. Struc.	**4.5 – 5**	$\mathbf{2^{53.25}}$ **CP**	$\mathbf{2^{59.25}}$ **M** $\approx \mathbf{2^{52.6}}$ **E**	$\mathbf{2^{16}}$	**Sect. 6.3**
Diff. Struc.	**4.5 – 5**	$\mathbf{2^{53.6}}$ **CP**	$\mathbf{2^{55.6}}$ **M** $\approx \mathbf{2^{48.96}}$ **E**	$\mathbf{2^{40}}$	**Sect. 6.2**
ImD	**4.5 – 5**	$\mathbf{2^{76.37}}$ **CP**	$\mathbf{2^{81.54}}$ **M** $\approx \mathbf{2^{74.9}}$ **E**	$\mathbf{2^{8}}$	**App. G.1**
ImD	4.5 - 5	2^{102} CP	2^{107} M $\approx 2^{100.4}$ E	2^{8}	[GRR17b]
I	5	2^{128} CC	$2^{129.6}$ XOR	small	[SLG+16]

I: Integral, ImD: Impossible Differential

1.3 Key-Recovery Attacks on AES-128 with a Single Secret S-Box

Recently, new key-recovery attacks on AES-128 with a single secret S-Box have been presented in [TKKL15] and in [GRR17b]. In this setting, the AES S-Box is replaced by a secret 8-bit one chosen uniformly at random from all the 8-bit permutation[4], with the goal to increase the security from 128-256 bits (i.e. the key size in AES) to 1812-1940.

In [TKKL15], the authors presented attacks up to 6-round AES with identical and secret S-Box using techniques from integral cryptanalysis. For such attacks, the attacker first determines the secret S-Box up to additive constants (that is, S-Box$(x \oplus a) \oplus b$ for unknown a and b), and then she uses this knowledge to derive the whitening key up to 2^8 variants. The strategy presented in [GRR17b] (and in [SLG+16]) is instead quite different. Instead of finding the secret S-Box up to additive constants, authors exploits a particular property of the MixColumns matrix (i.e. two equal elements for each row of the matrix) in order to find directly the secret key up to 2^{32} variants. Such a strategy is so generic that can be applied to integral, truncated differential and impossible differential attack.

In this paper we exploit this second strategy, and in Sect. 6.2 we adapt the attack on 5-round AES proposed in Sect. 5.3 to the case of secret S-Box. The idea of the attack is to choose a set of plaintexts that depends on some guessed bytes of the key. If the guessed bytes are the right ones, then it is possible to guarantee that the number of ciphertexts that belong to the same coset of a particular subspace $\mathcal{M}$ is a multiple of 2 or 4 with probability 1, while this happens only with probability strictly less than 1 for wrong guessed keys.

Moreover, in Sect. 6.1 we generalize the strategy proposed in [GRR17b]. While attacks proposed in [GRR17b] exploit the fact that two coefficients of each row of the MixColumns matrix are equal, we show that the same attacks can also be mounted in the case in which a XOR-sum of more than two coefficients of each row of the MixColumns matrix is equal to zero. As main result, the strategy proposed in [GRR17b] works for a bigger class of MixColumns matrices. We apply such strategy for our new 5-round attack presented in this paper in Sect. 6.3, while in App. G.1 we improve the impossible differential attack on 5-round AES proposed in [GRR17b].

[4]For completeness, we mention that a randomly chosen S-Box is very likely to be highly resistant against differential and linear, as shown in [TKKL15].

2 Preliminary - Description of AES

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

- *SubBytes* (S-Box) - applying the same 8-bit to 8-bit invertible S-Box 16 times in parallel on each byte of the state (provides non-linearity in the cipher);
- *ShiftRows* (SR) - cyclic shift of each row (i-th row is shifted by i bytes to the left);
- *MixColumns* (MC) - multiplication of each column by a constant 4×4 invertible matrix over the field $GF(2^8)$ (together with the ShiftRows operation, it provides diffusion in the cipher);
- *AddRoundKey* (ARK) - XORing the state with a 128-bit subkey.

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

The Notation Used in the Paper

Let x denote a plaintext, a ciphertext, an intermediate state or a key. Then $x_{i,j}$ with $i, j \in \{0, ..., 3\}$ denotes the byte in the row i and in the column j. We denote by 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[5] of AES, while we denote r rounds of AES by R^r. 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 $\mathcal{X}$.

3 Subspace Trails

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* [GRR17b], which has been recently introduced at FSE 2017 as generalization of the invariant subspace cryptanalysis.

Definition 1. Let $(V_1, V_2, ..., V_{r+1})$ denote a set of $r + 1$ subspaces with $\dim(V_i) \leq \dim(V_{i+1})$. If for each $i = 1, ..., r$ and for each $a_i \in V_i^{\perp}$, there exist (unique) $a_{i+1} \in V_{i+1}^{\perp}$ such that $F(V_i \oplus a_i) \subseteq V_{i+1} \oplus a_{i+1}$, then $(V_1, V_2, ..., V_{r+1})$ is *subspace trail* of length r for the function F. 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 [GRR17b] for more details about the concept of subspace trails. Our treatment here is however meant to be self-contained.

[5] Sometimes we use the notation R_k instead of R to highlight the round key k.

3.1 Subspace Trails of AES

In this section, we recall the subspace trails of AES presented in [GRR17b], working with vectors and vector spaces over $\mathbb{F}_{2^8}^{4\times 4}$. For the following, 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 recall that given a subspace $\mathcal{X}$, the cosets $\mathcal{X} \oplus a$ and $\mathcal{X} \oplus b$ (where $a \neq b$) are *equivalent* (that is $\mathcal{X} \oplus a \sim \mathcal{X} \oplus b$) if and only if $a \oplus b \in \mathcal{X}$.

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

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

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

Definition 3. The *diagonal spaces* $\mathcal{D}_i$ and the *inverse-diagonal spaces* $\mathcal{ID}_i$ are defined as $\mathcal{D}_i = SR^{-1}(\mathcal{C}_i)$ and $\mathcal{ID}_i = SR(\mathcal{C}_i)$.

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

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

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

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

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

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

Definition 5. For $I \subseteq \{0, 1, 2, 3\}$, let $\mathcal{C}_I$, $\mathcal{D}_I$, $\mathcal{ID}_I$ and $\mathcal{M}_I$ defined as

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

As shown in detail in [GRR17b]:

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

Theorem 1. *For each I and for each $a \in \mathcal{D}_I^{\perp}$, there exists one and only one $b \in \mathcal{M}_I^{\perp}$ (which depends on a and on the secret key k) such that*

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

We refer to [GRR17b] for a complete proof of the Theorem. Observe that if $\mathcal{X}$ is a generic subspace, $\mathcal{X} \oplus a$ is a coset of $\mathcal{X}$ and x and y are two elements of the (same) coset $X \oplus a$, then $x \oplus y \in \mathcal{X}$. 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 [GRR17b]. It follows that:

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

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

We remark that all these results can be re-described using a more "classical" - but equivalent - truncated differential notation. To be more concrete, if two texts t^1 and t^2 are equal expect for the bytes in the i-th diagonal[6] for each $i \in I$, then they belong in the same coset of $\mathcal{D}_I$. A coset of $\mathcal{D}_I$ corresponds to a set of $2^{32 \cdot |I|}$ texts with $|I|$ active diagonals. Again, two texts t^1 and t^2 belong in the same coset of $\mathcal{M}_I$ if the bytes of their difference $MC^{-1}(t^1 \oplus t^2)$ in the i-th anti-diagonal for each $i \notin I$ are equal to zero. Similar considerations hold for the column space $\mathcal{C}_I$ and the inverse-diagonal space $\mathcal{ID}_I$. Our choice to use the subspace trail notation in order to present our new distinguishers and key-recovery attacks is motivated by the fact that it allows to describe them in a more formal way than using the "classical" notation.

We finally introduce some notations that we largely use in the following.

Definition 6. Given two different texts $t^1, t^2 \in \mathbb{F}_{2^8}^{4\times 4}$, we say that $t^1 \leq t^2$ if $t^1 = t^2$ or if there exists $i, j \in \{0,1,2,3\}$ 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 $t^1 \neq t^2$.

Definition 7. Let $\mathcal{X}$ be one of the previous subspaces, that is $\mathcal{C}_I$, $\mathcal{D}_I$, $\mathcal{ID}_I$ or $\mathcal{M}_I$. Let $x_0, ..., x_n \in \mathbb{F}_{2^8}^{4\times 4}$ be a basis of $\mathcal{X}$ - i.e. $\mathcal{X} \equiv \langle x_0, x_1, ..., x_n \rangle$ where $n = 4 \cdot |I|$ - s.t. $x_i < x_{i+1}$ for each $i = 0, ..., n-1$. Let t be an element of an arbitrary coset of $\mathcal{X}$, that is $t \in \mathcal{X} \oplus a$ for arbitrary $a \in \mathcal{X}^{\perp}$. We say that t is "generated" by the *generating variables* $(t^0, ..., t^n)$ - for the following, $t \equiv (t^0, ..., t^n)$ - if and only if

$$t \equiv (t^0, ..., t^n) \quad \text{iff} \quad t = a \oplus \bigoplus_{i=0}^{n} t^i \cdot x_i.$$

As an example, let $\mathcal{X} = \mathcal{M}_0 \equiv \langle MC(e_{0,0}), MC(e_{3,1}), MC(e_{2,2}), MC(e_{1,3}) \rangle$, and let $p \in \mathcal{M}_0 \oplus a$. Then $p \equiv (p^0, p^1, p^2, p^3)$ if and only if

$$p \equiv p^0 \cdot MC(e_{0,0}) \oplus p^1 \cdot MC(e_{1,3}) \oplus p^2 \cdot MC(e_{2,2}) \oplus p^3 \cdot MC(e_{3,1}) \oplus a. \tag{5}$$

Similarly, let $\mathcal{X} = \mathcal{C}_0 \equiv \langle e_{0,0}, e_{1,0}, e_{2,0}, e_{3,0} \rangle$, and let $p \in \mathcal{C}_0 \oplus a$. Then $p \equiv (p^0, p^1, p^2, p^3)$ if and only if $p \equiv a \oplus p^0 \cdot e_{0,0} \oplus p^1 \cdot e_{1,0} \oplus p^2 \cdot e_{2,0} \oplus p^3 \cdot e_{3,0}$.

3.2 Intersections of Subspaces and Useful Probabilities

Here we list some useful probabilities largely used in the following[7]. For our goal, we focus on the mixed space $\mathcal{M}$, but the same results can be easily generalized for the other subspaces $\mathcal{D}, \mathcal{C}$ and $\mathcal{ID}$.

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

[7] We mention that the following probabilities are "sufficiently good" approximations useful for the target of the paper, that is the error of this approximations can be considered negligible for the target of this paper. For a complete discussion, we refer to App. A.

Let $I, J \subseteq \{0,1,2,3\}$. We first recall that a random element x belongs to the subspace $\mathcal{M}_I$ with probability $Prob(x \in \mathcal{M}_I) \simeq 2^{-32\cdot(4-|I|)}$. Moreover, as shown in details in [GRR17b], given two random elements $x \neq y$ in the same coset of $\mathcal{M}_I$, they belong after one round to the same coset of $\mathcal{M}_J$ with probability:

$$Prob(R(x) \oplus R(y) \in \mathcal{M}_J \,|\, x \oplus y \in \mathcal{M}_I) \simeq 2^{-4\cdot|I|+|I|\cdot|J|}.$$

By definition, it's simple to observe that $\mathcal{M}_I \cap \mathcal{M}_J = \mathcal{M}_{I\cap J}$ (where $\mathcal{M}_I \cap \mathcal{M}_J = \emptyset$ if $I \cap J = \emptyset$). Thus, the probability $p_{|I|}$ that a random text x belongs to the subspace $\mathcal{M}_I$ for a certain $I \subseteq \{0,1,2,3\}$ with $|I| = l$ fixed is well approximated by

$$p_{|I|} \equiv Prob(\exists I \ |I| = l \text{ s.t. } x \in \mathcal{M}_I) = (-1)^{|I|} \cdot \sum_{i=4-|I|}^{3} (-1)^i \cdot \binom{4}{i} \cdot 2^{-32\cdot i}. \tag{6}$$

Let x, y be two random elements with $x \neq y$. Assume there exists $I \subseteq \{0,1,2,3\}$ such that $x \oplus y \in \mathcal{M}_I$ and $x \oplus y \notin \mathcal{M}_L$. The probability $p_{|J|,|I|}$ that there exists $J \subseteq \{0,1,2,3\}$ - with $|J| = l$ fixed - such that $R(x) \oplus R(y) \in \mathcal{M}_J$ is well approximated by

$$\begin{aligned} p_{|J|,|I|} &\equiv Prob(\exists J \,|J| = l \text{ s.t. } R(x) \oplus R(y) \in \mathcal{M}_J \,|\, x \oplus y \in \mathcal{M}_I) = \\ &= (-1)^{|J|} \cdot \sum_{i=4-|J|}^{3} (-1)^i \cdot \binom{4}{i} \cdot 2^{8\cdot i\cdot|I|\cdot(|J|-4)}. \end{aligned} \tag{7}$$

Assume that for each $I \subseteq \{0,1,2,3\}$ $x \oplus y \notin \mathcal{M}_I$. Then, the probability $\hat{p}_{|J|,3}$ that $\exists J \subseteq \{0,1,2,3\}$ with $|J| = l$ fixed such that $R(x) \oplus R(y) \in \mathcal{M}_J$ is well approximated by

$$\hat{p}_{|J|,3} \equiv Prob(\exists J \text{ s.t. } R(x) \oplus R(y) \in \mathcal{M}_J \,|\, x \oplus y \notin \mathcal{M}_I \,\forall I) = \frac{p_{|J|} - p_{|J|,3} \cdot p_3}{1 - p_3}. \tag{8}$$

Finally, assume that for each $I \subseteq \{0,1,2,3\}$ $x \oplus y \notin \mathcal{M}_I$. Then, the probability that $\exists J \subseteq \{0,1,2,3\}$ with $|J| = l$ fixed and with $|I| + |J| \leq 4$ such that $R^2(x) \oplus R^2(y) \in \mathcal{M}_J$ is well approximated by

$$\tilde{p}_{|J|,3} \equiv Prob(\exists J \text{ s.t. } R^2(x) \oplus R^2(y) \in \mathcal{M}_J \,|\, x \oplus y \notin \mathcal{M}_I) = \frac{p_{|J|}}{1 - p_3}. \tag{9}$$

Note that the inequality[8] $\hat{p}_{|J|,3} < p_{|J|} < \tilde{p}_{|J|,3}$ holds for each J.

A complete proof of the previous probabilities is provided in App. A. To give an example, if $|I| = |J| = 3$ the previous probabilities are well approximated by

$$p_3 = 2^{-30} - 3 \cdot 2^{-63} + 2^{-94}, \qquad p_{3,3} = 2^{-22} - 3 \cdot 2^{-47} + 2^{-70}$$
$$\hat{p}_{3,3} = 2^{-30} - 2043 \cdot 2^{-63} + 390661 \cdot 2^{-94} + ...$$

where p_3 and $\hat{p}_{3,3}$ are usually approximated by 2^{-30} and $p_{3,3}$ by 2^{-22}.

4 5-round Secret-Key Distinguisher proposed in [GRR17a]

The starting point of our secret-key distinguisher is the property proposed and exploited in [GRR17a] to set up the first 5-round secret-key distinguisher of AES (independent of the secret key). For this reason, in this section we recall the main idea of that paper, and we refer to [GRR17a] for a complete discussion.

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. The 5-round AES

[8] Since $p_{|J|,3} > p_{|J|}$, it follows that $\hat{p}_{|J|,3} \equiv \frac{p_{|J|} - p_{|J|,3}\cdot p_3}{1-p_3} < \frac{p_{|J|} - p_{|J|}\cdot p_3}{1-p_3} = p_{|J|}$.

distinguisher proposed in [GRR17a] exploits the fact that the number of different pairs of ciphertexts that belong to the same coset of $\mathcal{M}_J$ for a fixed J is always a multiple of 8 with probability 1 independently of the secret key, of the details of the S-Box and of the MixColumns matrix. In more details, 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 number of different pairs[9] of ciphertexts (c^i, c^j) that satisfy $c^i \oplus c^j \in \mathcal{M}_J$ for a certain fixed $J \subset \{0, 1, 2, 3\}$ has the special property to be a multiple of 8 with prob. 1. Since for a random permutation the same number doesn't 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.

Since each coset of $\mathcal{D}_I$ is mapped into a coset of $\mathcal{M}_I$ after 2 rounds with prob. 1 - see Theorem 1 - and viceversa, in order to prove the result given in [GRR17a] it is sufficient to show that given plaintexts in the same coset of $\mathcal{M}_I$, then the number of collisions after one round in the same coset of $\mathcal{D}_J$ is a multiple of 8 (see [GRR17a] for details).

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

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

satisfies the property to be a multiple of 8 with prob. 1, i.e. $\exists\, n' \in \mathbb{N}$ s.t. $n = 8 \cdot n'$.

We refer to [GRR17a] for a detailed proof, and we limit here to recall and to highlight the main concepts that are useful for the following.

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

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

Definition 8. Let $\mathcal{X}$ be a fixed coset of $\mathcal{C}_I$ or $\mathcal{M}_I$ for $I \in \{0, 1, 2, 3\}$ with $|I| = 1$. Let p and q be two different elements in $\mathcal{X} \oplus a$ - a coset of $\mathcal{X}$ - with $p \equiv (p^0, p^1, p^2, p^3)$ and $q \equiv (q^0, q^1, q^2, q^3)$, such that $p^i \neq q^i$ for each $i = 0, ..., 3$. Moreover, let $R^r(p)$ and $R^r(q)$ be the corresponding ciphertexts after r rounds.

We define the set $\mathcal{S}_{p,q}^{\mathcal{X}\oplus a}$ as the set of eight couples $(\hat{p}^i, R^r(\hat{p}^i))$ and $(\hat{q}^i, R^r(\hat{q}^i))$ where $\hat{p}^i, \hat{q}^i \in \mathcal{X} \oplus a$ for $i = 1, ..., 8$ are respectively generated by the following combinations of

[9] Two pairs (c^i, c^j) and (c^j, c^i) are considered equivalent.

variables

1. (p^0, p^1, p^2, p^3) and (q^0, q^1, q^2, q^3);
2. (q^0, p^1, p^2, p^3) and (p^0, q^1, q^2, q^3);
3. (p^0, q^1, p^2, p^3) and (q^0, p^1, q^2, q^3);
4. (p^0, p^1, q^2, p^3) and (q^0, q^1, p^2, q^3);
5. (p^0, p^1, p^2, q^3) and (q^0, q^1, q^2, p^3);
6. (q^0, q^1, p^2, p^3) and (p^0, p^1, q^2, q^3);
7. (q^0, p^1, q^2, p^3) and (p^0, q^1, p^2, q^3);
8. (q^0, p^1, p^2, q^3) and (p^0, q^1, q^2, p^3).

Lemma 2. *Let $\mathcal{S}_{p,q}^{\mathcal{M}_I \oplus a}$ be an arbitrary set defined as in Def. 8*

$$\mathcal{S}_{p,q}^{\mathcal{M}_I \oplus a} \equiv \{[(p_i^1, c_i^1 \equiv R(p_i^1)), (p_i^2, c_i^2 \equiv R(p_i^2))]_i \quad \forall i = 1, ..., 8\}.$$

For each fixed $J \subseteq \{0, 1, 2, 3\}$, only on of the two following events can happen:

- $c_i^1 \oplus c_i^2 \notin \mathcal{D}_J$ *for all* $i = 1, ..., 8$*;*
- $c_i^1 \oplus c_i^2 \in \mathcal{D}_J$ *for all* $i = 1, ..., 8$.

In other words, given a set $\mathcal{S}_{p^1,p^2}^{\mathcal{M}_I \oplus a}$, consider the eight couples of two (plaintext, ciphertext) pairs (p_i^1, c_i^1) and (p_i^2, c_i^2) for $i = 1, ..., 8$ in such set. Two ciphertexts c^1 and c^2 belong (or not) to the same coset of $\mathcal{D}_J$ for a certain J if and only if the ciphertexts of all the other couples in the set $\mathcal{S}_{p^1,p^2}^{\mathcal{M}_I \oplus a}$ have the same property.

Case: Equal Generating Variables. Similar definitions of the set $\mathcal{S}_{p,q}^{\mathcal{M}_I \oplus a}$ can be given if one or two variables are equal. For the following, we focus on the case in which two variables are equal (e.g. $x^1 = x^2$ and $y^1 = y^2$).

Definition 9. Let $\mathcal{X}$ be a fixed coset of $\mathcal{C}_I$ or $\mathcal{M}_I$ for $I \in \{0, 1, 2, 3\}$ with $|I| = 1$. Let p and q be two different elements in a coset of $\mathcal{X}$, that is $\mathcal{X} \oplus a$, with $p \equiv (p^0, p^1, p^2, p^3)$ and $q \equiv (q^0, q^1, q^2, q^3)$, s.t. $p^i = q^i$ for $i = 0, 1$ and $p^i \neq q^i$ for $i = 2, 3$ (the set $\mathcal{Z}_{p,q}^{\mathcal{X} \oplus a}$ is defined in a similar way for the other cases). Moreover, let $R^r(p)$ and $R^r(q)$ be the corresponding ciphertexts after r rounds.

We define the set $\mathcal{Z}_{p,q}^{\mathcal{X} \oplus a}$ as the set of 2^{17} couples $(\hat{p}^i, R^r(\hat{p}^i))$ and $(\hat{q}^i, R^r(\hat{q}^i))$ where $\hat{p}^i, \hat{q}^i \in \mathcal{X} \oplus a$ for $i = 1, ..., 2^{17}$ are respectively generated by the following combinations of variables

1. (z^0, z^1, p^2, p^3) and (z^0, z^1, q^2, q^3);
2. (z^0, z^1, q^2, p^3) and (z^0, z^1, p^2, q^3);

where z^0 and z^1 can take any possible value in $\mathbb{F}_{2^8}$.

As before, it is possible to prove the following Lemma (see [GRR17a] for details).

Lemma 3. *Let $\mathcal{Z}_{p,q}^{\mathcal{M}_I \oplus a}$ be an arbitrary set defined as in Def. 9*

$$\mathcal{Z}_{p,q}^{\mathcal{M}_I \oplus a} \equiv \{[(p_i^1, c_i^1 \equiv R(p_i^1)), (p_i^2, c_i^2 \equiv R(p_i^2))]_i \quad \forall i = 1, ..., 2^{17}\}.$$

For each fixed $J \subseteq \{0, 1, 2, 3\}$, only on of the two following events can happen:

- $c_i^1 \oplus c_i^2 \notin \mathcal{D}_J$ *for all* $i = 1, ..., 2^{17}$*;*
- $c_i^1 \oplus c_i^2 \in \mathcal{D}_J$ *for all* $i = 1, ..., 2^{17}$.

In other words, given a set $\mathcal{Z}_{p^1,p^2}^{\mathcal{M}_I \oplus a}$, consider the 2^{17} couples of two (plaintext, ciphertext) pairs (p_i^1, c_i^1) and (p_i^2, c_i^2) for $i = 1, ..., 2^{17}$. Two ciphertexts c^1 and c^2 belong (or not) to the same coset of $\mathcal{D}_J$ for a certain J if and only if the ciphertexts of all the other couples in the set $\mathcal{Z}_{p^1,p^2}^{\mathcal{M}_I \oplus a}$ have the same property. It follows that for the case $x^1 = x^2$, $y^1 = y^2$, $z^1 \neq z^2$ and $w^1 \neq w^2$ or analogous (i.e. two variables that generate p^1 and p^2 are equal), the number of collisions must be a multiple of 2^{17} (the cardinality of each set $\mathcal{Z}_{p^1,p^2}^{\mathcal{M}_I \oplus a}$ is 2^{17}).

For completeness, in the case in which two plaintexts p^1 and p^2 have exactly one equal generating variable, the set $\mathcal{T}$ - analogous of the sets $\mathcal{S}$ and $\mathcal{Z}$ - can be defined.

Definition 10. Let $\mathcal{X}$ be a fixed coset of $\mathcal{C}_I$ or $\mathcal{M}_I$ for $I \in \{0,1,2,3\}$ with $|I| = 1$. Let p and q be two different elements in a coset of $\mathcal{X}$, that is $\mathcal{X} \oplus a$, with $p \equiv (p^0, p^1, p^2, p^3)$ and $q \equiv (q^0, q^1, q^2, q^3)$, such that $p^0 = q^0$ and $p^j \neq q^j$ for each $j = 1, 2, 3$ (the set $\mathcal{T}_{p,q}^{\mathcal{X} \oplus a}$ is defined in a similar way for the other cases). Moreover, let $R^r(p)$ and $R^r(q)$ be the corresponding ciphertexts after r rounds.

We define the set $\mathcal{T}_{p,q}^{\mathcal{X} \oplus a}$ as the set of 2^{10} couples $(\hat{p}^i, R^r(\hat{p}^i))$ and $(\hat{q}^i, R^r(\hat{q}^i))$ where $\hat{p}^i, \hat{q}^i \in \mathcal{X} \oplus a$ for $i = 1, ..., 1024$ are respectively generated by the following combinations of variables

1. (z^0, p^1, p^2, p^3) and (z^0, q^1, q^2, q^3);
2. (z^0, q^1, p^2, p^3) and (z^0, p^1, q^2, q^3);
3. (z^0, p^1, q^2, p^3) and (z^0, q^1, p^2, q^3);
4. (z^0, p^1, p^2, q^3) and (z^0, q^1, q^2, p^3).

where z^0 can take any possible value in $\mathbb{F}_{2^8}$.

We refer to App. D.2 for all the details about this case.

Finally, given texts in the same cosets of $\mathcal{C}_I$ or $\mathcal{M}_I$ for $I \subseteq \{0,1,2,3\}$, the number of couples of texts with n equal generating variable(s) for $0 \leq n \leq 3$ is given by

$$\binom{4}{n} \cdot 2^{32 \cdot |I| - 1} \cdot (2^{8 \cdot |I|} - 1)^{4-n} \tag{11}$$

as proved in App. A.

Case $|\mathbf{I}| = 2$ and $|\mathbf{I}| = 3$. For the following, we mention that similar considerations can be done for the cases $|I| \geq 2$. W.l.o.g consider $|I| = 2$ and assume $I = \{0,1\}$ (the other cases are analogous). Given 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 $a \in \mathcal{M}_I^{\perp}$, 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:

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

As for the case $|I| = 1$, the idea is to consider all the possible combinations of the variables $x \equiv (x_0, x_1), y \equiv (y_0, y_1), z \equiv (z_0, z_1), w \equiv (w_0, w_1)$ and $x' \equiv (x'_0, x'_1), y' \equiv (y'_0, y'_1), z' \equiv (z'_0, z'_1), w' \equiv (w'_0, w'_1)$. In other words, the idea is to consider variables in $\mathbb{F}_{2^8}^2 \equiv \mathbb{F}_{2^8} \times \mathbb{F}_{2^8}$ and not in $\mathbb{F}_{2^8}$. For $|I| = 3$, the idea is similar, working with variables in $\mathbb{F}_{2^8}^3$. Note that the definitions of $\mathcal{S}_{p,q}^{\mathcal{X}}$, $\mathcal{Z}_{p,q}^{\mathcal{X}}$ and $\mathcal{T}_{p,q}^{\mathcal{X}}$ given before can be easily adapted to all these cases.

Why is it (rather) *hard* to set up key-recovery attacks that exploit such distinguisher?

Given this 5-round distinguisher, a natural question regards the possibility to exploit it in order to set up a key-recovery attack on 6-round AES-128 which is better than a brute force one. A possible way is the following. Consider 2^{32} chosen plaintexts in the same coset of a diagonal space $\mathcal{D}_i$, and the corresponding ciphertexts after 6 rounds. A possibility is to guess the final key, decrypt the ciphertexts and check if the number of collisions in the same coset of $\mathcal{M}_J$ is a multiple of 8. If not, the guessed key is wrong. However, since a coset of $\mathcal{M}_J$ is mapped into the full space, it seems hard to check this property one round before without guessing the entire key. It follows that it is rather hard to set up an attack different than a brute force one that exploits directly the 5-round distinguisher proposed [GRR17a]. For comparison, note that such a problem doesn't arise for the other distinguishers up to 4-round AES (e.g. the impossible differential or the integral ones), for which it is sufficient to guess only part of the secret key in order to verify if the required property is satisfied or not.

5 A New 4-round Secret-Key Distinguisher for AES

As first thing, we re-exploit the property proposed in [GRR17a] to set up a *new* 4-round secret-key distinguisher for AES. Before we go into the details, we present the general idea.

As we have just seen, given 2^{32} plaintexts in the same coset of $\mathcal{M}_I$ for $|I| = 1$ and the corresponding ciphertexts after 1 round, that is (p^i, c^i) for $i = 0, ..., 2^{32} - 1$ where $p^i \in \mathcal{M}_I \oplus a$ and $c^i = R(p^i)$, then the number n of different pairs of ciphertexts (c^i, c^j) for $i \neq j$ such that $c^i \oplus c^j \in \mathcal{D}_J$ defined as in (10) is always a multiple of 8. This is due to the fact that if one pair of texts belong to the same coset of $\mathcal{D}_J$ after one round, then other pairs of texts have the same property. Thus, consider a pair of plaintexts p^1 and p^2 such that the corresponding texts after one round belong (or not) to the same coset of $\mathcal{D}_J$. As we have seen, there exist other pairs of plaintexts $\hat{p}^1$ and $\hat{p}^2$ whose ciphertexts after one round have the same property. The pairs (p^1, p^2) and $(\hat{p}^1, \hat{p}^2)$ are *not independent* in the sense that the variables that generate the first pair of texts are the same that generate the other pairs, but in a different combination. The idea is to exploit this property in order to set up new distinguishers for round-reduced AES. That is, instead of limiting to count the number of collisions and check that it is a multiple of 8, the idea is to check if these relationships between the variables that generate the plaintexts (whose ciphertexts belong or not the same coset of a given subspace) hold or not.

A New 4-round Secret-Key Distinguisher for AES. Given the subspace $\mathcal{C}_0 \cap \mathcal{D}_{0,3} \equiv \langle e_{0,0}, e_{1,0} \rangle \subseteq \mathcal{C}_0$, consider two plaintexts p^1 and p^2 in the same coset of $\mathcal{C}_0 \cap \mathcal{D}_{0,3} \oplus a$ generated by $p^1 \equiv (z^1, w^1)$ and $p^2 \equiv (z^2, w^2)$. For 4-round AES and for each fixed $J \subseteq \{0, 1, 2, 3\}$, the following event holds with probability 1

$$R^4(p^1) \oplus R^4(p^2) \in \mathcal{M}_J \quad \textit{if and only if} \quad R^4(\hat{p}^1) \oplus R^4(\hat{p}^2) \in \mathcal{M}_J$$

where $\hat{p}^1, \hat{p}^2 \in \mathcal{D}_{0,3} \cap \mathcal{C}_0 \oplus a$ are generated by $\hat{p}^1 \equiv (z^1, w^2)$ and $\hat{p}^2 \equiv (z^2, w^1)$. For a random permutation, this happens with prob. $2^{-32 \cdot |J|}$ (i.e strictly less than 1). It follows that this probability can be used to set up a 4-round distinguisher.

Why this happens? Let p^1 and p^2 be two texts in the same coset of $\mathcal{C}_0 \cap \mathcal{D}_{0,3} \oplus a$ for fixed $a \in (\mathcal{C}_0 \cap \mathcal{D}_{0,3})^\perp$, generated by $p^1 \equiv (z^1, w^1)$ and $p^2 \equiv (z^2, w^2)$, that is $p^i \equiv a \oplus z^i \cdot e_{0,0} \oplus w^i \cdot e_{1,0}$. After one round, the two texts belong to the same coset of $\mathcal{M}_0 \cap \mathcal{C}_{0,3}$ and they are equal to

$$R(p^1) \equiv b \oplus x^1 \cdot MC(e_{0,0}) \oplus y^1 \cdot MC(e_{1,3}), \qquad R(p^2) \equiv b \oplus x^2 \cdot MC(e_{0,0}) \oplus y^2 \cdot MC(e_{1,3})$$

for a certain $b \in \mathcal{M}_0^\perp$, where

$$x^i = \text{S-Box}(z^i \oplus a_{0,0}) \qquad y^i = \text{S-Box}(w^i \oplus a_{1,0}), \qquad \text{for } i = 1, 2. \tag{12}$$

Due to Lemma 2, $R^2(p^1)$ and $R^2(p^2)$ belong in the same coset of $\mathcal{D}_J$ (i.e. $R^2(p^1) \oplus R^2(p^2) \in \mathcal{D}_J$) if and only if $R(\hat{q}^1)$ and $R(\hat{q}^2)$ belong in the same coset of D_J, where $\hat{q}^1, \hat{q}^2 \in \mathcal{C}_{0,3} \cap \mathcal{M}_0 \oplus b$ are generated by $\hat{q}^1 \equiv (x^1, y^2)$ and $\hat{q}^2 \equiv (x^2, y^1)$:

$$\hat{q}^1 \equiv b \oplus x^1 \cdot MC(e_{0,0}) \oplus y^2 \cdot MC(e_{1,3}) \qquad \hat{q}^2 \equiv b \oplus x^1 \cdot MC(e_{0,0}) \oplus y^2 \cdot MC(e_{1,3}).$$

Due to the relationships between x^1, y^1, x^2, y^2 and z^1, w^1, z^2, w^2 previously defined (12), it follows that there exist two texts $\hat{p}^1 = R^{-1}(\hat{q}^1), \hat{p}^2 = R^{-1}(\hat{q}^2) \in \mathcal{D}_{0,1} \cap \mathcal{C}_0 \oplus a$ generated by

$$\hat{p}^1 = R^{-1}(\hat{q}^1) \equiv (z^1, w^2) \;\text{ and }\; \hat{p}^2 = R^{-1}(\hat{q}^2) \equiv (z^2, w^1) \;\text{ s.t. }\; R^2(\hat{p}^1) \oplus R^2(\hat{p}^2) \in \mathcal{D}_J.$$

Finally, since $Prob(R^2(s) \oplus R^2(t) \in \mathcal{M}_J \,|\, s \oplus t \in \mathcal{D}_J) = 1$ - see (2), it is possible to set up the distinguisher on 4 rounds. Thus, the basic idea of the distinguisher is to exploit the fact that two texts p^1 and p^2 belong to the same coset of $\mathcal{D}_J$ (for J fixed) after two rounds if and only if other two texts $\hat{p}^1$ and $\hat{p}^2$ in $\mathcal{C}_0 \cap \mathcal{D}_{0,3} \oplus a$ have the same property. In particular, the idea is to exploit the fact that the relationships that hold between the variables that generate p^1 and p^2 and the variables that generate $\hat{p}^1$ and $\hat{p}^2$ are known.

5.1 A New 4-round Secret-Key Distinguisher for AES - Details

Given a coset of $\mathcal{C}_0 \cap \mathcal{D}_{0,3}$ - that is $\mathcal{C}_0 \cap \mathcal{D}_{0,3} \oplus a$ for a fixed a, the idea is to construct all the $2^{15} \cdot (2^{16} - 1) \simeq 2^{31}$ possible different couples of texts. For our goal, we eliminate all the couples of texts for which one of the two variables that generate the two plaintexts is equal. Then, one constructs all the possible sets[10] $\mathcal{S}^{\mathcal{C}_0 \cap \mathcal{D}_{0,1} \oplus a}$ defined in a similar way of Def. 8, that is

$$\begin{aligned} \mathcal{S}^{\mathcal{C}_0 \cap \mathcal{D}_{0,3} \oplus a} \equiv \{[(p^1, c^1), (p^2, c^2)]; [(\hat{p}^1, \hat{c}^1), (\hat{p}^2, \hat{c}^2)] \,|\, \forall p^1, p^2, \hat{p}^1, \hat{p}^2 \in \mathcal{C}_0 \cap \mathcal{D}_{0,3} \oplus a \\ \text{s.t. } p^1 \equiv (z^1, w^1), p^2 \equiv (z^2, w^2), \hat{p}^1 \equiv (z^1, w^2), \hat{p}^2 \equiv (z^2, w^1)\}, \end{aligned} \tag{13}$$

where the ciphertexts are the 4-round encryption of the plaintexts, that is $c = R^4(p)$.

Let J fixed with $|J| = 3$. By previous observations on AES permutation - Lemma 2, it follows that for each set $\mathcal{S}^{\mathcal{C}_0 \cap \mathcal{D}_{0,3} \oplus a} \equiv \{[(p^1, c^1), (p^2, c^2)]; [(\hat{p}^1, \hat{c}^1), (\hat{p}^2, \hat{c}^2)]\}$ only one of the two following events can happen: (1) $c^1 \oplus c^2 \in \mathcal{M}_J$ and $\hat{c}^1 \oplus \hat{c}^2 \in \mathcal{M}_J$ or (2) $c^1 \oplus c^2 \notin \mathcal{M}_J$ and $\hat{c}^1 \oplus \hat{c}^2 \notin \mathcal{M}_J$. On the other hand, for a random permutation the event $c^1 \oplus c^2 \in \mathcal{M}_J$ and $\hat{c}^1 \oplus \hat{c}^2 \notin \mathcal{M}_J$ (or viceversa) is also possible, and it occurs with probability $2 \cdot 2^{-32 \cdot (4 - |J|)} \cdot (1 - 2^{-32 \cdot (4 - |J|)})$, which is approximately equal to 2^{-31} for the case $|J| = 3$ fixed (it is higher for the other cases $|J| \leq 2$). The idea is to exploit this fact to distinguish a random permutation from 4-round AES one. Moreover, since this distinguisher is based on Theorem 1 which holds also in the reverse direction (see [GRR17a] for details), an equivalent distinguisher can be set up for 4-round AES in the decryption mode, using chosen ciphertexts instead of plaintexts.

Data and Computational Cost

Since a coset of $\mathcal{C}_0 \cap \mathcal{D}_{0,3}$ contains 2^{16} plaintexts, it is possible to construct $2^{15} \cdot (2^{16} - 1) \simeq 2^{31}$ different pairs, and $2^{14} \cdot (2^{16} - 1) \simeq 2^{30}$ different sets $\mathcal{S}^{\mathcal{C}_0 \cap \mathcal{D}_{0,3} \oplus a}$ as defined in (13). For our goal, we consider only the sets $\mathcal{S}^{\mathcal{C}_0 \cap \mathcal{D}_{0,3} \oplus a} \equiv \{[(p^1, c^1), (p^2, c^2)]; [(\hat{p}^1, \hat{c}^1), (\hat{p}^2, \hat{c}^2)]\}$ such that the two plaintexts have no common variables (i.e. if $p^1 \equiv (z^1, w^1)$ and $p^2 \equiv (z^2, w^2)$, then $z^1 \neq z^2$ and $w^1 \neq w^2$). Since the probability that one of the two variables is equal is $2 \cdot 2^{-8} = 2^{-7}$, the number of sets $\mathcal{S}^{\mathcal{C}_0 \cap \mathcal{D}_{0,3} \oplus a}$ with elements generated by different variables is approximately $(2^{30} - 2^{14}) \cdot (1 - 2^{-7}) = 2^{30} - 2^{23} - 2^{14} + 2^7 \simeq 2^{29.989}$.

In order to distinguish 4-round AES from a random permutation, for each one of these sets, one has to check that $c^1 \oplus c^2 \in \mathcal{M}_J$ if and only if $\hat{c}^1 \oplus \hat{c}^2 \in \mathcal{M}_J$. If this property is not satisfied for at least one set, then it is possible to conclude that the analyzed permutation is a random one.

What is the probability that $c^1 \oplus c^2 \in \mathcal{M}_J$ and $\hat{c}^1 \oplus \hat{c}^2 \notin \mathcal{M}_J$ - or viceversa - for a certain $J \subset \{0, 1, 2, 3\}$ with $|J| = 3$? By simple computation and since there are 4 different J with $|J| = 3$, this happens with an approximated probability of

$$2 \cdot p_3 \cdot (1 - 2^{-32}) \simeq 2 \cdot 4 \cdot 2^{-32} \cdot (1 - 2^{-32}) \simeq 2^{-29},$$

where p_3 is defined as in (6). As a result, in order to distinguish a random permutation from an AES one with probability higher than pr, it is sufficient that at least one set $\mathcal{S}^{\mathcal{C}_0 \cap \mathcal{D}_{0,3} \oplus a}$ exists for which the previous property is not satisfied with probability higher than pr in order to recognize the random permutation. It follows that one needs approximately n different sets $\mathcal{S}^{\mathcal{C}_0 \cap \mathcal{D}_{0,3} \oplus a}$ such that $pr \geq 1 - (1 - 2^{-29})^n$, that is

$$n \geq \frac{\log(1 - pr)}{\log(1 - 2^{-29})} \approx -2^{29} \cdot \log(1 - pr).$$

[10] Note that $\bigcup_{p,q} \mathcal{S}_{p,q}^{\mathcal{C}_0 \cap \mathcal{D}_{0,3} \oplus a} \subsetneq \mathcal{C}_0 \cap \mathcal{D}_{0,3} \oplus a$, since $z^1 = z^2$ or $w^1 = w^2$ is not allowed.

Data: 2 cosets of $\mathcal{D}_{0,3} \cap \mathcal{C}_0$ (e.g. $\mathcal{D}_{0,3} \cap \mathcal{C}_0 \oplus a_i$ for $a_0, a_1 \in (\mathcal{D}_{0,3} \cap \mathcal{C}_0)^{\perp}$) and corresponding ciphertexts after 4 rounds
Result: $0 \equiv$ Random permutation *or* $1 \equiv$ 4-round AES - Prob. 95%
for *each coset of* $\mathcal{D}_{0,3} \cap \mathcal{C}_0$ **do**
  for *each* $I \subseteq \{0,1,2,3\}$ *with* $|I| = 3$ **do**
    let (p^i, c^i) for $i = 0, ..., 2^{16} - 1$ be the 2^{16} (plaintexts, ciphertexts) of $\mathcal{D}_{0,3} \cap \mathcal{C}_0 \oplus a_i$;
    re-order this set of elements w.r.t. the partial order $\preceq$ described in Def. 11 s.t. $c^k \preceq c^{k+1}$ for each k; `// ⪯ depends on I`
    $i \leftarrow 0$;
    while $i < 2^{16} - 1$ **do**
      $j \leftarrow i$;
      while $c^j \oplus c^{j+1} \in \mathcal{M}_I$ **do**
        $j \leftarrow j + 1$;
      end
      for *each* k *from* i *to* j **do**
        for *each* l *from* $k + 1$ *to* j **do**
          construct the corresponding set $\mathcal{S}^{\mathcal{D}_{0,3} \cap \mathcal{C}_0 \oplus a_i}_{p^k, p^l}$ as defined in Def. 8 - Eq. (13);
          for *each couple* of (plaintexts, ciphertexts) $\{(\hat{p}^1, \hat{c}^1), (\hat{p}^2, \hat{c}^2)\} \in \mathcal{S}^{\mathcal{D}_{0,3} \cap \mathcal{C}_0 \oplus a_i}_{p^k, p^l}$ **do**
            if $\hat{c}^1 \oplus \hat{c}^2 \notin \mathcal{M}_I$ **then**
              return *0.* `// Random permutation`
            end
          end
        end
      end
      $i \leftarrow j + 1$;
    end
  end
end
return *1.* `// 4-round AES permutation - Prob. 95%`

Algorithm 1: *Secret-Key Distinguisher for 4-round of AES.*

For $pr = 95\%$, one needs approximately $n \geq 2^{31.996}$ different sets $\mathcal{S}^{\mathcal{C}_0 \cap \mathcal{D}_{0,3} \oplus a}$, that is approximately 2 different cosets $\mathcal{C}_0 \cap \mathcal{D}_{0,3}$ for a total data cost of $2^{16} \cdot 2 = 2^{17}$ chosen plaintexts.

Computational Cost. We limit here to report the computational costs of the distinguisher, and we refer to App. B for all the details. In order to implement the distinguisher, the idea is to re-order the ciphertexts using a particular partial order $\preceq$ as defined in Def. 11 (recalled in the following - see also [GRR17a]), and to work in the way described in Algorithm 1. Instead of constructing all the sets, the basic idea is to construct only the sets $\mathcal{S}$ of the couples for which the two ciphertexts belong in the same coset of $\mathcal{M}_J$. This method allows to minimize the computational cost, which is well approximated by $2^{23.09}$ table look-ups, or approximately $2^{16.75}$ four-round encryptions (assuming[11] 20 table look-ups $\approx$ 1 round of encryption).

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

Definition 11. 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 $j \in \{0,1,2,3\}$ s.t. $MC^{-1}(t^1)_{i,l-i} = MC^{-1}(t^2)_{i,l-i}$ for all $i < j$ and $MC^{-1}(t^1)_{j,l-j} < MC^{-1}(t^2)_{j,l-j}$;
- $MC^{-1}(t^1)_{i,l-i} = MC^{-1}(t^2)_{i,l-i}$ for all $i = 0, ..., 3$, and $MC^{-1}(t^1) < MC^{-1}(t^2)$ where $<$ is defined in Def. 6.

Practical Verification

Using a C/C++ implementation[12], we have practically verified the distinguisher just described. In particular, we have verified the distinguisher both for "real" AES and a small-scale variant of AES, as presented in [CMR05]. While for "real" AES each word is composed of 8 bits, in the small-scale variant each word is composed of 4 bits (we refer to [CMR05] for a complete description of this small-scale AES). We highlight that Theorem 2 holds exactly in the same way also for this small-scale variant of AES, since the previous argumentation is independent of the fact that each word of AES is of 4 or 8 bits.

The distinguisher just presented works in the same way for real AES and small scale AES, and it is able to distinguish AES from a random permutation using 2^{17} chosen plaintexts in the first case and 2^9 in the second one (i.e. 2 cosets of $\mathcal{C}_0 \cap \mathcal{D}_{0,3}$) as expected. For real AES, while the theoretical computational cost is of 2^{23} table look-ups, the practical one is on average 2^{22} in the case of a random permutation and 2^{24} in the case of an AES permutation. We emphasize that for a random permutation, it is sufficient to find *one* set $\mathcal{S}^{\mathcal{C}_i \oplus a}$ that doesn't satisfy the required properties in order to recognize the random permutation. In the case of the AES permutation, the difference between the theoretical and the practical cases (i.e. a factor 2) can be 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[13].

For the small-scale AES, using 2 different initial cosets of $\mathcal{C}_0 \cap \mathcal{D}_{0,3}$, the theoretical computational cost is well approximated by $2 \cdot 4 \cdot 2^8 \cdot (\log 2^8 + 1) \simeq 2^{14.2}$ table look-ups. The practical cost is approximately $2^{13.5}$ for the case of a random permutation and 2^{15} for the AES case.

5.2 Comparison with Other 4-round Secret-Key Distinguishers

Before we go on, we highlight the major differences with respect to the other 4-round AES secret-key distinguishers present in the literature. Omitting the integral one (which exploits a completely different property), we focus on the impossible and the truncated differential distinguishers, polytopic cryptanalysis and on the distinguisher recently proposed in [GRR17a] adapted - in a natural way - to the 4-round case.

The impossible differential distinguisher is based on Prop. 1, that is it exploits the property that $\mathcal{M}_I \cap \mathcal{D}_J = \{0\}$ for $|I| + |J| \leq 4$. In our case, we consider plaintexts in the same coset of $\mathcal{C}_0 \cap \mathcal{D}_I \subseteq \mathcal{D}_I$ with $I = \{0,1\}$ and looks for collisions in $\mathcal{M}_J$ with $|J| = 3$. Since $|I| + |J| = 5$, the property exploited by the impossible differential distinguisher can not be applied.

The truncated differential distinguisher has instead some aspects in common with our distinguisher. In this case, given pairs of plaintexts with certain difference on certain bytes (i.e. that belong to the same coset of a subspace $\mathcal{X}$), one considers the probability that the corresponding ciphertexts belong to the same coset of a subspace $\mathcal{Y}$. For 2-round AES it

[12] The source codes of the distinguishers/attacks are available at https://github.com/Krypto-iaik/Distinguisher_5RoundAES

[13] A similar difference among the theoretical and the practical cases was found also in [GRR17a].

is possible to exploit truncated differential trails with probability 1, while for the 3-round case there exist truncated differential trails with probability lower than 1 but higher than for the random case (in both cases, $\mathcal{X} \equiv \mathcal{D}_I$ and $\mathcal{Y} \equiv \mathcal{M}_J$). To the best of our knowledge, no truncated differential trails with probability higher than 0 (i.e. no impossible differential trails) on 4 or more rounds AES exist in literature. Our proposed distinguisher works in a similar way and exploits a similar property. However, instead of working with a single couple of texts, in our distinguisher one considers set of 2 "non-independent" couples of texts. In particular, while in a classical truncated differential distinguisher one focuses on a single couple of two (plaintexts, ciphertexts) pairs independently of the others, one considers sets $\mathcal{S}$ of 2 - or more generally $N \geq 2$ - couples of two (plaintexts, ciphertexts) pairs and one exploits the relationships that hold among the couples of texts that belong to the same set $\mathcal{S}$.

Polytopic cryptanalysis [Tie16] has been introduced by Tiessen at Eurocrypt 2016, and it can be viewed as a generalization of standard differential cryptanalysis. Consider a set of $d \geq 2$ couples of plaintexts $(p^0, p^0 \oplus \alpha^1), (p^0, p^0 \oplus \alpha^2), ...(p^0, p^0 \oplus \alpha^d)$ with one plaintext in common (namely p^0), called d-poly. The idea of polytopic cryptanalysis is to exploit the probability that the input set of differences $\boldsymbol{\alpha} \equiv (\alpha^1, \alpha^2, ..., \alpha^d)$ is mapped into an output set of differences $\boldsymbol{\beta} \equiv (\beta^1, \beta^2, ..., \beta^d)$ after r rounds. If this probability[14] - *which depends on the S-Box details* - is different than the corresponding probability in the case of a random permutation, it is possible to set up distinguishers or key-recovery attacks. Impossible polytopic cryptanalysis focuses on the case in which the previous probability is zero. In [Tie16], an impossible 8-polytopic is proposed for 2-round AES, which allows to set up key-recovery attacks on 4- and 5-round AES. Our proposed distinguisher works in a similar way, since also in our case we consider set of "non-independent" couples of texts and we focus on the input/output differences. However, instead to work with a set of couples of plaintexts with one plaintext in common, we consider set of couples of texts for which particular relationships between the generating variables of the texts hold. Moreover, instead to consider the probability that "generic" input differences $\boldsymbol{\alpha}$ are mapped into output differences $\boldsymbol{\beta}$, the way in which the texts are divided in sets guarantees that a particular relation holds on the ciphertexts of the same set with prob. 1 after 4-round (that is, the two ciphertexts of all couples satisfy/don't satisfy an output - truncated - difference), *independently of the S-Box details.*

Finally, the distinguisher proposed in [GRR17a] can be adapted to the 4 rounds case, e.g. considering plaintexts in the same coset of $\mathcal{C}_J$, counting the number of collisions of the ciphertexts in the same coset of $\mathcal{M}_I$ and checking if it is (or not) a multiple of 8. Since our distinguisher exploits more information (i.e. the relationships that hold among the couples in the same set $\mathcal{S}$ beside the fact that the previous number is a multiple of 8), its data and computational costs are lower than [GRR17a], that is 2^{17} chosen plaintexts/ciphertexts instead of 2^{33} and approximately 2^{23} table look-ups instead of 2^{40}.

5.3 New Key-Recovery Attack on 5-round AES

The previous 4-round secret-key distinguisher can be used as starting point to set up a new (practical verified) key-recovery attack on 5-round AES.

W.l.o.g. consider two plaintexts p^1 and p^2 in the same coset of $\mathcal{D}_0$, e.g. $\mathcal{D}_0 \oplus a$ for $a \in \mathcal{D}_0^{\perp}$, such that $p^i = x^i \cdot e_{0,0} \oplus y^i \cdot e_{1,1} \oplus z^i \cdot e_{2,2} \oplus w^i \cdot e_{3,3} \oplus a$ or equivalently

[14] We mention that the probability of polytopic trails is usually much lower than the probability of trails in differential cryptanalysis, that is simple polytopic cryptanalysis can not in general outperform standard differential cryptanalysis - see Sect. 2 of [Tie16] for details.

Data: 1 coset of $\mathcal{D}_0$ (e.g. $\mathcal{D}_0 \oplus a$ for $a \in \mathcal{D}_0^{\perp}$) and corresponding ciphertexts after 5 rounds - more generally a coset of $\mathcal{D}_i$ for $i \in \{0,1,2,3\}$
Result: 4 bytes of the secret key - $(k_{0,0}, k_{1,1}, k_{2,2}, k_{3,3})$
fix $I \subseteq \{0,1,2,3\}$ with $|I| = 3$ - e.g. $I = \{0,1,2\}$;
let (p^i, c^i) for $i = 0, ..., 2^{32} - 1$ be the 2^{32} (plaintexts, ciphertexts) of $\mathcal{D}_0 \oplus a$;
re-order - and stores five different times - this set of elements w.r.t. the partial order $\preceq$ described in Def. 11 s.t. $c^i \preceq c^{i+1}$ for each i; `// ⪯ depends on I`
do
  find indexes j and h s.t. $c^j \oplus c^h \in \mathcal{M}_I$; `// look for h = j + 1 due to ⪯`
  for *each one of the* 2^{32} *combinations of* $\hat{k} = (k_{0,0}, k_{1,1}, k_{2,2}, k_{3,3})$ **do**
    (partially) compute $q^1 = R_{\hat{k}}(p^j)$ and $q^2 = R_{\hat{k}}(p^h)$;
    construct the set $\mathcal{S}^{R(\mathcal{D}_0 \oplus a)}_{p^1,p^2}$ as defined in Def. 8;`// remember that the set` $\mathcal{S}^{R(\mathcal{D}_0 \oplus a)}_{p^1,p^2} \equiv \mathcal{S}^{\mathcal{C}_0 \oplus a}_{q^1,q^2}$ `depends on` $\hat{k}$
    $flag \leftarrow 0$;
    for *each couple* of (plaintexts, ciphertexts) $\{(\hat{p}^1, \hat{c}^1), (\hat{p}^2, \hat{c}^2)\} \in \mathcal{S}^{R(\mathcal{D}_0 \oplus a)}_{q^1,q^2}$ **do**
      if $\hat{c}^1 \oplus \hat{c}^2 \notin \mathcal{M}_I$ **then**
        $flag \leftarrow 1$;
        next combination of $(k_{0,0}, k_{1,1}, k_{2,2}, k_{3,3})$;
      end
    end
    if $flag = 0$ **then**
      identify $(k_{0,0}, k_{1,1}, k_{2,2}, k_{3,3})$ as candidate of the key;
    end
  end
while *more than one candidate of the key is found* - Repeat the procedure for *different indexes* j, h *(and* I*)* `// it is usually not necessary - only one candidate is found`;
return $(k_{0,0}, k_{1,1}, k_{2,2}, k_{3,3})$

Algorithm 2: *5-round AES Key-Recovery Attack.* The attack exploits the 4-round distinguisher presented in Sect. 5. For sake of simplicity, in this pseudo-code we limit to describe the attack of 4 bytes - 1 diagonal of the secret key. Exactly the same attack can be used to recover the entire key.

$p^i \equiv (x^i, y^i, z^i, w^i)$. By Theorem 1, there exists $b \in \mathcal{C}_0^{\perp}$ such that for $i = 1, 2$

$$R(p^i) = \begin{bmatrix} \hat{x}^i & 0 & 0 & 0 \\ \hat{y}^i & 0 & 0 & 0 \\ \hat{z}^i & 0 & 0 & 0 \\ \hat{w}^i & 0 & 0 & 0 \end{bmatrix} \oplus b \equiv M^{MC} \cdot \begin{bmatrix} \text{S-Box}(x^i \oplus k_{0,0}) & 0 & 0 & 0 \\ \text{S-Box}(y^i \oplus k_{1,1}) & 0 & 0 & 0 \\ \text{S-Box}(z^i \oplus k_{2,2}) & 0 & 0 & 0 \\ \text{S-Box}(w^i \oplus k_{3,3}) & 0 & 0 & 0 \end{bmatrix} \oplus b,$$

i.e. $R(p^i) \equiv (\hat{x}^i, \hat{y}^i, \hat{z}^i, \hat{w}^i) \equiv \hat{x}^i \cdot e_{0,0} \oplus \hat{y}^i \cdot e_{1,0} \oplus \hat{z}^i \cdot e_{2,0} \oplus \hat{w}^i \cdot e_{3,0} \oplus b$. As we are going to show, it is possible to filter wrong guessed key of the first round by exploiting the previous distinguisher. Given plaintexts in the same coset of $\mathcal{D}_0$, consider two (plaintexts, ciphertexts) pairs (p^1, c^1) and (p^2, c^2) such that the two ciphertexts belong to the same coset of $\mathcal{M}_J$ for J with $|J| = 3$ after five-round, i.e. $p^1 \oplus p^2 \in \mathcal{D}_0$ and $c^1 \oplus c^2 \in \mathcal{M}_J$. The idea of the attack is simply to guess 4 bytes of the first diagonal of the secret key k, that is $k_{i,i}$ for each $i \in \{0,1,2,3\}$, (partially) compute $R_k(p^1)$ and $R_k(p^2)$ and construct the set $\mathcal{S}^{\mathcal{C}_0 \oplus b}_{R(p^1),R(p^2)}$. As example, the couple $(\hat{p}^1, \hat{c}^1)$ and $(\hat{p}^2, \hat{c}^2)$ where $\hat{p}^1$ and $\hat{p}^2$ satisfy

$$R(\hat{p}^i) = \hat{x}^{(i+1) \bmod 2} \cdot e_{0,0} \oplus \hat{y}^i \cdot e_{1,0} \oplus \hat{z}^i \cdot e_{2,0} \oplus \hat{w}^i \cdot e_{3,0} \oplus b, \tag{14}$$

belongs to such set (analogous for the other cases/combinations). We emphasize that

the way in which the couples are divided in the sets $\mathcal{S}$ depends on the guessed key. Thus, due to the previous 4-round distinguisher - Lemma 2, for the right key the set[15] $\mathcal{S}^{\mathcal{C}_0\oplus b}_{R(p^1),R(p^2)} \equiv \mathcal{S}^{R(\mathcal{D}_0\oplus a)}_{R(p^1),R(p^2)} \equiv \{[(p_i^1, c_i^1), (p_i^2, c_i^2)]_{i=1,...,8}\}$ has the property that $c_i^1 \oplus c_i^2 \in \mathcal{M}_J$ if and only if $c_j^1 \oplus c_j^2 \in \mathcal{M}_J$ for each $i, j = 1, ..., 8$ with prob. 1. In other words, for the right key and for all J, the two ciphertexts of all the couples in the $\mathcal{S}^{\mathcal{C}_0\oplus b}$ belong or not to the same coset of $\mathcal{M}_J$ (it is not possible that only some of them - not all - have this property). If this property is not satisfied, then one can simply deduce that the guessed key is wrong (for a wrong guessed key, the behavior is similar to the one of a random permutation). The attack - practical verified on a small-scale AES - requires $2^{33.6}$ chosen plaintexts and has a computational cost of $2^{33.28}$ five-round encryptions. The pseudo-code of the attack is given in Algorithm 2, while all the details - included the results of our practical tests - can be found in App. C.

Details of the Attack: Why does this attack work? First of all, since the cardinality of a coset of $\mathcal{D}_I$ for $|I| = 1$ is 2^{32} and since $\text{Prob}(t \in \mathcal{M}_J) = 2^{-32}$ for $|J| = 3$, the average number of collisions for each coset of $\mathcal{D}_I$ is approximately $4 \cdot 2^{-32} \cdot 2^{31} \cdot (2^{32} - 1) \simeq 2^{33}$ (note that there are four J with $|J| = 3$), so it's very likely that two (plaintexts, ciphertexts) pairs (p^1, c^1) and (p^2, c^2) exist such that $c^1 \oplus c^2 \in \mathcal{M}_J$.

Given a pair of ciphertext c^1 and c^2 that belong to the same coset of $\mathcal{M}_J$, consider the corresponding plaintexts p^1 and p^2 and the set $\mathcal{S}^{\mathcal{C}_0\oplus b}_{R(p^1),R(p^2)}$ of 8 couples defined as in Def. 8. For a wrong key, the probability that the two ciphertexts of each one of the other 7 couples in that set belong to the same coset of $\mathcal{M}_J$ for fixed J is $(2^{-32})^7 = 2^{-224}$. In other words, the probability that a wrong key passes the test is 2^{-224}. Indeed, remember that (1) if the guessed key is wrong, then the couples are divided in sets $\mathcal{S}$ in a random way and (2) for an AES permutation, given a set $\mathcal{S}^{\mathcal{C}_0\oplus b}_{R(p^1),R(p^2)}$ for which the two ciphertexts of one couple belong to the same coset of $\mathcal{M}_J$, then the two ciphertexts of (all) the other 7 couples have the same property with prob. 1, while this is general not true for a random permutation.

Since there are $2^{32} - 1$ wrong candidates for the diagonal of the key, the probability that at least one of them passes the test is approximately $1 - (1 - 2^{-224})^{2^{32}-1} \simeq 2^{-192}$. Thus, one pair of ciphertexts that belong to the same coset of $\mathcal{M}_J$ and the corresponding set $\mathcal{S}^{R(\mathcal{D}_I\oplus a)}$ are (largely) sufficient to discard all the wrong candidates for a diagonal of the key. Moreover, in general only two couples of such set can be sufficient to discard all the wrong candidates, that is it is not necessary to work with the entire set $\mathcal{S}^{\mathcal{C}_0\oplus b}_{R(p^1),R(p^2)}$. Indeed, given two couples, the probability that at least one wrong key passes the test is approximately $1-(1-2^{-32\cdot 2})^{2^{32}-1} \simeq 2^{-32} \ll 1$, which means that all the wrong candidates are discarded with high probability.

Our practical tests confirm these results.

Finally, we emphasize the *impossibility to set up a 5-round distinguisher similar to the one just presented in this section* choosing plaintexts in the same coset of a diagonal space $\mathcal{D}_I$ instead of a column space $\mathcal{C}_I$. Indeed, given p^1 and p^2 as before in the same coset of $\mathcal{D}_I$ (instead of $\mathcal{C}_I$), since the key k is secret and the S-Box is non-linear, there is no way to find $\hat{p}^1$ and $\hat{p}^2$ that satisfy (14) and to construct the set $\mathcal{S}^{\mathcal{C}_0\oplus b}_{R(p^1),R(p^2)}$ without guessing the secret key.

[15] We abuse the notation $\mathcal{S}^{R(\mathcal{D}_0\oplus a)}_{R(p^1),R(p^2)}$ to denote the set $\mathcal{S}^{\mathcal{C}_0\oplus b}_{R(p^1),R(p^2)}$.

6 Key-Recovery Attack on round-reduced AES-128 with a single Secret S-Box

Recently, new key-recovery attacks on AES with a single secret S-Box have been presented in [TKKL15] and in [GRR17b]. In the first paper, authors are able to set up attacks up to 6-round AES with identical and secret S-Box using techniques from integral cryptanalysis. The attack procedure consists of two steps: in the first one, an attacker determines the secret S-Box up to additive constants (that is, S-Box$(x \oplus a) \oplus b$ for unknown a and b), while in the second step, the attacker uses this knowledge to derive the whitening key up to 2^8 variants. The strategy presented in [GRR17b] is instead quite different. Instead of finding the secret S-Box up to additive constants, authors exploits a particular property of the MixColumns matrix in order to find directly (i.e. without discovering any information of the secret S-Box) the secret key up to 2^{32} variants. Such a strategy is so generic that can be applied to integral, truncated differential and impossible differential attacks. At Crypto 2016, a similar strategy has been also exploited by Sun, Liu, Guo, Qu and Rijmen [SLG+16] to present the first 5-round key-dependent distinguisher for AES. In this paper, we focus on this second strategy, and we show that it can be adapted to the 5-round AES attack proposed in Sect. 5.3. As a result, our proposed attacks have lower data and computational costs of the ones presented in [GRR17b]. Besides that, we are able to generalize such a strategy and to apply it to a bigger class of MixColumns matrices.

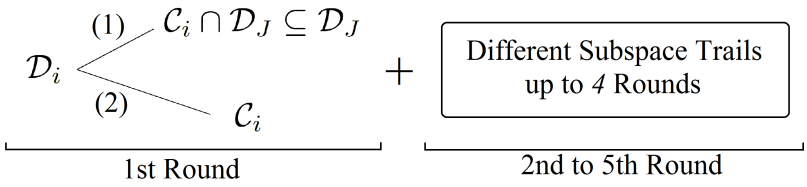

Figure 2: *Strategy of the attacks on AES with a secret S-Box* proposed in [GRR17b]. Starting with a subset of a coset of $\mathcal{D}_i$ which depends on the guessed values of the secret key, it is mapped after one round into a subset of a coset of $\mathcal{D}_J$ if the guessed values is correct - case (1), or into a subset of a coset of $\mathcal{C}_i$ if the guessed values is wrong - case (2). As a consequence, the subspace trails up to the 5-th round are different for the two cases, and this allows to set up various key-recovery attacks.

6.1 A More Generic Strategy for Key-Recovery Attacks on AES-like Ciphers with a Single Secret S-Box

The strategy proposed in [GRR17b] exploits the fact that two coefficients of each row of the MixColumns matrix are equal. The basic idea is to choose a set of plaintexts which depends by a guessed key. The attacker exploits the fact that when the guessed key is the right one a certain property holds after r rounds (in other words, a differential trail over r rounds is satisfied) with a different probability than in the case in which the guessed key is the wrong one. We limit here to recall an example and we refer to [GRR17b] for more details. Let M^{MC} be the AES MixColumns matrix, where $M^{MC}_{0,2} = M^{MC}_{0,3}$ (similar for the other rows). Let p^1 and p^2 two texts such that $p^1_{i,j} = p^2_{i,j}$ for each $(i,j) \neq \{(2,2),(3,3)\}$ and assume $p^1_{2,2} \oplus p^1_{3,3} = p^2_{2,2} \oplus p^2_{3,3}$ (note that such pair of plaintexts belong to the same coset of $\mathcal{D}_0$). Denote the secret key by k. If $p^1_{2,2} \oplus p^1_{3,3} = p^2_{2,2} \oplus p^2_{3,3} = k_{2,2} \oplus k_{3,3}$, then after one round the two texts belong to the same coset of $\mathcal{C}_0 \cap \mathcal{D}_{1,2,3} \subseteq \mathcal{D}_{1,2,3}$ with probability 1 - case (1) of Fig. 2, otherwise they belong to the same coset of $\mathcal{D}_{1,2,3}$ only with probability 2^{-8} - case (2) of Fig. 2 (note that in both cases, the two texts belong to the same coset of $\mathcal{C}_0$ after one round). Exploiting these different probabilities, it is possible to set up several

differential trails on 2-, 3-, 4- and 5-round AES that have a different probabilities between cases (1) and (2), as illustrated in Fig. 2. This allows to recover the key. As example, if the guessed key is correct one then after 3 rounds the previous texts belong to the same coset of $\mathcal{M}_{1,2,3}$ with probability 1, while this happens only with probability 2^{-8} for a wrong guessed key. We emphasize that no information on the S-Box is recovered or used.

Beside to adapt such a strategy for the attack on 5-round AES proposed in Sect. 5.3, in the following we present a way to generalize such a strategy for a large class of MixColumns matrices. Instead of exploiting the fact that two elements of each row of the MixColumns matrix M^{MC} are equal, we show that it is possible to mount similar attacks also in the case in which the XOR-sum of 2 or more elements of each row of M^{MC} is equal to zero. That is, it is possible to set up an attack also in the case in which for each row r (or for some of them) of M^{MC} there exists a set $J_r \subseteq \{0,1,2,3\}$ such that

$$\bigoplus_{j \in J_r} M^{MC}_{r,j} = 0 \tag{15}$$

As an example, each row of the AES MixColumns matrix M^{MC} satisfies this condition, e.g.

$$M^{MC}_{0,0} \oplus M^{MC}_{0,1} \oplus M^{MC}_{0,2} = 0, \qquad M^{MC}_{0,i} \neq M^{MC}_{0,j} \quad \forall i,j \in \{0,1,2\}.$$

As a special case, if two elements $M^{MC}_{r,j}$ and $M^{MC}_{r,k}$ of a row r are equal (that is $M^{MC}_{r,j} = M^{MC}_{r,k}$ for $j \neq k$), then the previous condition is obviously satisfied (vice-versa doesn't hold). It follows that the following strategy includes the one proposed in [GRR17b] as a particular case.

To explain how to exploit property (15), we show how to adapt the attacks described in [GRR17b] (just recalled) to this case. As we have already said, the idea of those attacks is to choose a set of plaintexts $\mathcal{A}_\delta$ which depends by a guessed key δ. When δ assumes the "right" value (which depends on the secret key), then the set $\mathcal{A}_\delta$ is mapped after one round into a coset of $\mathcal{D}_I$ for some I (where $|I| \leq 3$) with probability 1, while for other values of δ this happens only with probability strictly less than 1. Since the idea is to exploit the same strategy, we limit here to define the set $\mathcal{A}_\delta$ in the case in which a sum of elements of each row of M^{MC} is equal to zero.

Proposition 2. *Let M^{MC} be the AES MixColumns matrix such that*

$$M^{MC}_{i,0} \oplus M^{MC}_{i,1} \oplus M^{MC}_{i,2} = 0 \qquad i = \{0,1\}.$$

Let p^1 and p^2 be two texts, s.t. $p^1_{i,j} = p^2_{i,j}$ for all $(i,j) \neq \{(0,0),(1,1),(2,2)\}$ and

$$p^1_{i,j} \oplus p^1_{k,l} = p^2_{i,j} \oplus p^2_{k,l} \qquad \forall (i,j),(k,l) \in \{(0,0),(1,1),(2,2)\} \text{ and } (i,j) \neq (k,l).$$

If $p^1_{0,0} \oplus p^1_{1,1} = p^2_{0,0} \oplus p^2_{1,1} = k_{0,0} \oplus k_{1,1}$ and $p^1_{0,0} \oplus p^1_{2,2} = p^2_{0,0} \oplus p^2_{2,2} = k_{0,0} \oplus k_{2,2}$, then $R(p^1) \oplus R(p^2) \in \mathcal{C}_0 \cap \mathcal{D}_{2,3}$ with probability 1 (i.e. after one round, p^1 and p^2 belong to the same coset of $\mathcal{C}_0 \cap \mathcal{D}_{2,3}$). This happens with probability 2^{-16} in the other cases.

Proof. Note that the two plaintexts p^1 and p^2 belong to the same coset of $\mathcal{D}_0$. Since a coset of diagonal space $\mathcal{D}_I$ is always mapped after one round into a coset of a column space $\mathcal{C}_I$, after one round they belong to the same coset of $\mathcal{C}_0$ with probability 1. To prove the statement, it is sufficient to prove that $[R(p^1) \oplus R(p^2)]_{0,0} = [R(p^1) \oplus R(p^2)]_{1,0} = 0$.

By simple calculation

$$\begin{aligned} R(p^1)_{0,0} = &\text{0x02} \cdot \text{S-Box}(p^1_{0,0} \oplus k_{0,0}) \oplus \text{0x03} \cdot \text{S-Box}(p^1_{1,1} \oplus k_{1,1}) \oplus \\ &\oplus \text{S-Box}(p^1_{2,2} \oplus k_{2,2}) \oplus \text{S-Box}(p^1_{3,3} \oplus k_{3,3}). \end{aligned}$$

Since $p^1_{0,0} \oplus p^1_{1,1} = k_{0,0} \oplus k_{1,1}$, it follows that $\text{S-Box}(p^1_{0,0} \oplus k_{0,0}) = \text{S-Box}(p^1_{1,1} \oplus k_{1,1})$ and in a similar way $\text{S-Box}(p^1_{0,0} \oplus k_{0,0}) = \text{S-Box}(p^1_{2,2} \oplus k_{2,2})$. Since the sum of the

first three elements is equal to zero, then $R(p^1)_{0,0} = \text{S-Box}(p^1_{3,3} \oplus k_{3,3})$, and similarly $R(p^2)_{0,0} = \text{S-Box}(p^2_{3,3} \oplus k_{3,3})$. Since $p^1_{3,3} = p^2_{3,3}$, it follows that $R(p^1)_{0,0} = R(p^2)_{0,0}$. The same argumentation holds also for $R(p^1)_{1,0} = R(p^2)_{1,0}$. □

This proposition can be easily generalized for a more generic MixColumns matrix M^{MC} for which the sum of three coefficients are equal to zero. Moreover, if the sum $\bigoplus_{j\in J} M^{MC}_{r,j}$ is equal to zero for more than a single row for the same J, the following Lemma follows immediately.

Lemma 4. *Assume there exist $J \subseteq \{0,1,2,3\}$ and $r, w \in \{0,1,2,3\}$ with $r \neq w$ such that*

$$\bigoplus_{j\in J} M^{MC}_{r,j} = \bigoplus_{j\in J} M^{MC}_{w,j} = 0.$$

Let p^1 and p^2 defined as before. It follows that if $p^1_{j,j} \oplus p^1_{l,l} = p^2_{j,j} \oplus p^2_{l,l} = k_{j,j} \oplus k_{l,l}$ for each $j, l \in J$, then $p^1 \oplus p^2 \in \mathcal{C}_k \cap \mathcal{D}_{\{0,1,2,3\}\setminus\{r,w\}}$ with probability 1, otherwise this happens in general with probability 2^{-16}.

To prove this lemma, it is sufficient to exploit the previous proposition and to observe that if two plaintexts belong to the same coset of $\mathcal{C}_k \cap \mathcal{D}_{\{0,1,2,3\}\setminus\{r\}}$ and of $\mathcal{C}_k \cap \mathcal{D}_{\{0,1,2,3\}\setminus\{w\}}$, then they belong to their intersections $\mathcal{C}_k \cap \mathcal{D}_{\{0,1,2,3\}\setminus\{r,w\}}$.

What is the number of matrices that satisfy condition (15) with respect to the number of matrices with two equal coefficients in each row? As we show in details in App. H.4, if we limit to consider $n \times n$ *circulant matrix* with coefficients in $\mathbb{F}_{2^m}$, this ratio is approximately equal to

$$\frac{2^{n+1}}{n^2} \qquad \textit{if the condition} \quad 2^{m+1} \gg n^2 + 5 \cdot n \quad \textit{is fulfilled.}$$

To give an example, for the AES case (that is $m = 8$ and $n = 4$), the number of circulant matrices that satisfy property (15) is approximately *double* with respect to the number of matrices with (at least) two equal coefficients (i.e. this ratio is well approximated by 2).

In the following, we show how to adapt the attack presented in the previous section in the case of secret S-Box, by exploiting the fact that two coefficients of the MixColumns matrix are equal or that the sum of three of them is equal to zero. Moreover, in App. G.1, we show how to set up an impossible differential attack up to 5 rounds of AES that exploits (15), which improves the impossible differential attack presented in [GRR17b].

6.2 Attack on 5-round AES with a single Secret S-Box - MixColumns Matrix with Equal Coefficients

First of all, we show how to adapt the attack on 5-round AES described in the previous section in the case of a single secret S-Box. The idea is choose a particular set of plaintexts $\mathcal{A}_\delta$ (which depends on a variable δ), such that only for a particular value of δ which depends on the secret key the number of collisions among the ciphertexts in the same coset of $\mathcal{M}_I$ with $|I| = 3$ after 5 rounds is a multiple of 2 (i.e. it is an even number) with probability 1. Since for all the other values of δ this event happens only with probability $1/2$, it is possible to discover the right key. Thus, for a fixed $a \in \mathcal{D}^\perp_1$ (i.e. $a_{0,1} = a_{1,2} = 0$), let $\mathcal{A}_\delta$ be the set of plaintexts of the form:

$$\mathcal{A}_\delta \equiv \left\{ a \oplus \begin{bmatrix} y_0 & x & 0 & 0 \\ 0 & y_1 & x \oplus \delta & 0 \\ 0 & 0 & y_2 & 0 \\ 0 & 0 & 0 & y_3 \end{bmatrix} \;\middle|\; \forall x, y_0, ..., y_3 \in \mathbb{F}_{2^8} \right\}. \tag{16}$$

Given a set $\mathcal{A}_\delta$, we claim that if $\delta = k_{0,1} \oplus k_{1,2}$ then the number of collisions after 5 rounds in the same coset of $\mathcal{M}_I$ for a fixed $I \subseteq \{0,1,2,3\}$ with $|I| = 3$ is a multiple of 2 with probability 1.

Proposition 3. *Consider a set of plaintexts $\mathcal{A}_\delta$ defined as in (16), and the corresponding ciphertexts after 5 rounds. If $\delta = k_{0,1} \oplus k_{1,2}$, then the number of different pairs of ciphertexts that belong to the same coset of $\mathcal{M}_I$ for a fixed $I \subseteq \{0,1,2,3\}$ with $|I| = 3$ is a multiple of 2.*

Proof. Let $\delta = k_{0,1} \oplus k_{1,2}$. After one round, there exists b such that the set $\mathcal{A}_\delta$ is mapped into

$$R(\mathcal{A}_\delta) \equiv \left\{ b \oplus \begin{bmatrix} z_0 & w & 0 & 0 \\ z_1 & 0\text{x}03 \cdot w & 0 & 0 \\ z_2 & 0 & 0 & 0 \\ z_3 & 0\text{x}02 \cdot w & 0 & 0 \end{bmatrix} \;\middle|\; \forall w, z_0, ..., z_3 \in \mathbb{F}_{2^8} \right\}.$$

Consider two elements $z, z' \in R(\mathcal{A}_\delta)$ generated respectively by $z \equiv (z_0, z_1, z_2, z_3, w)$ and $z' \equiv (z'_0, z'_1, z'_2, z'_3, w)$, and consider separately the two cases $z_1 \neq z'_1$ and $z_1 = z'_1$. The idea is to show that in the first case (i.e. the set of all the different pairs of elements for which the condition $z_{1,1} \neq z'_{1,1}$ holds) the number of collisions is a multiple of 2, while in the second case (i.e. the set of all the different pairs of elements for which the condition $z_1 = z'_{1,1}$ holds) the number of collisions is a multiple of 256. In particular, consider two elements $z, z' \in R(\mathcal{A}_\delta)$ generated respectively by $z \equiv (z_0, z_1, z_2, z_3, w)$ and $z' \equiv (z'_0, z'_1, z'_2, z'_3, w)$ with $z_1 \neq z'_1$. For a fixed $I \in \{0,1,2,3\}$ with $|I| = 3$, the idea is to show that $R^4(z) \oplus R^4(z') \in \mathcal{M}_I$ *if and only if* $R^4(v) \oplus R^4(v') \in \mathcal{M}_I$ where the texts $v, v' \in R(\mathcal{A}_\delta)$ are generated respectively by $v \equiv (z_0, z'_1, z_2, z_3, w)$ and $v' \equiv (z'_0, z_1, z'_2, z'_3, w)$. Similarly, consider the case $z_1 = z'_1$. For this case, the idea is to prove that $z, z' \in R(\mathcal{A}_\delta)$ satisfy the condition $R^4(z) \oplus R^4(z') \in \mathcal{M}_I$ if and only if each pair of elements $v, v' \in R(\mathcal{A}_\delta)$ generated respectively by $v \equiv (z_0, v_1, z_2, z_3, w)$ and $v' \equiv (z'_0, v_1, z'_2, z'_3, w)$ for each $v_1 \in \mathbb{F}_{2^8}$ have the same property, that is $R^4(v) \oplus R^4(v') \in \mathcal{M}_I$. Since there are $2^8 = 256$ different values for v_1, then the number of collisions must be a multiple of 256. It follows that there exist $n', n'' \in \mathbb{N}$ such that the total number of collisions n can be written as $n = 2 \cdot n' + 256 \cdot n'' = 2 \cdot (n' + 128 \cdot n'')$. In other words, the total number of collisions is a multiple of 2. The details of the proof can be found in App. H. □

Consider now the case $\delta \neq k_{0,1} \oplus k_{1,2}$. In this case, the previous proposition doesn't hold and the number of collisions is a multiple of 2 only with probability 1/2. Indeed, let $\delta \neq k_{0,1} \oplus k_{1,2}$. By simple computation, there exists constants b such that the set $\mathcal{A}_\delta$ is mapped after one round into

$$R(\mathcal{A}_\delta) \equiv b \oplus \begin{bmatrix} z_{0,0} & 0\text{x}02 \cdot \text{S-Box}(x \oplus k_{0,1}) \oplus 0\text{x}03 \cdot \text{S-Box}(x \oplus \delta \oplus k_{1,1}) & 0 & 0 \\ z_{1,1} & \text{S-Box}(x \oplus k_{0,1}) \oplus 0\text{x}02 \cdot \text{S-Box}(x \oplus \delta \oplus k_{1,1}) & 0 & 0 \\ z_{2,2} & \text{S-Box}(x \oplus k_{0,1}) \oplus \text{S-Box}(x \oplus \delta \oplus k_{1,1}) & 0 & 0 \\ z_{3,3} & 0\text{x}03 \cdot \text{S-Box}(x \oplus k_{0,1}) \oplus \text{S-Box}(x \oplus \delta \oplus k_{1,1}) & 0 & 0 \end{bmatrix}$$

for each x and for each $z_{0,0}, ..., z_{3,3}$. Note that this is a subset (*not* a subspace) of a coset of $\mathcal{C}_{0,1}$. Thus, assume that two elements $z, z' \in R(\mathcal{A}_\delta)$ belong to the same coset of $\mathcal{M}_I$ after 4 rounds. Since the second column of $R(\mathcal{A}_\delta)$ can take only a limited number of values, working in the same way as before it is not possible to guarantee that other pairs of elements - defined by a different combinations of the variables - have the same property with prob. 1. It follows that in this case the number of collisions is a multiple of 2 only with probability 1/2 (this result has been practically verified).

Note that each set contains 2^{40} different texts, that is approximately $2^{39} \cdot (2^{40} - 1) \simeq 2^{79}$ different pairs of ciphertexts. Since the probability that two ciphertexts belong to the same coset of $\mathcal{M}_I$ for $|I| = 3$ is 2^{-32}, the number of collisions is approximately $2^{79} \cdot 2^{-32} = 2^{47}$. We emphasize that for the right key this number is exactly a multiple of 2 with probability 1, while for wrong guessed keys this happens only with probability 1/2. Using these considerations, it is possible to find the right key up to 2^{32} variants.

Data: 2^{10} different sets $\mathcal{A}_\delta$ defined as in (16) - 4 different sets for each δ - and corresponding ciphertexts after 5 rounds
Result: $k_{0,0} \oplus k_{1,1}$
for *each* δ *from 0 to* $2^8 - 1$ **do**
  $flag \leftarrow 0$;
  for *each set* $\mathcal{A}_\delta$ **do**
    let (p^i, c^i) for $i = 0, ..., 2^{40} - 1$ be the 2^{40} (plaintexts, ciphertexts) of $\mathcal{A}_\delta$;
    for *all* $j \in \{0, 1, 2, 3\}$ **do**
      Let $W[0, ..., 2^{32} - 1]$ be an array initialized to zero;
      for i from *0 to* $2^{40} - 1$ **do**
        $x \leftarrow \sum_{k=0}^{3} MC^{-1}(c^i)_{k,j-k} \cdot 256^k$; // $MC^{-1}(c^i)_{k,j-k}$ denotes the byte of $MC^{-1}(c^i)$ in row k and column $j-k$ mod 4
        $W[x] \leftarrow W[x] + 1$; // $W[x]$ denotes the value stored in the x-th address of the array W
      end
      $n \leftarrow 0$;
      for i from *0 to* $2^{32} - 1$ **do**
        $n \leftarrow n + W[i] \cdot (W[i] - 1)/2$;
      end
      if $(n \bmod 2) \neq 0$ **then**
        $flag \leftarrow 1$;
        next δ;
      end
    end
  end
  if $flag = 0$ **then**
    identify δ as candidate for $k_{0,0} \oplus k_{1,1}$;
  end
end
return *Candidates for* $k_{0,0} \oplus k_{1,1}$. // Only one candidate with Prob. 95%

Algorithm 3: *Key-Recovery Attack on 5 rounds of AES with a single secret S-Box.* For simplicity, the goal of the attack is to find one byte of the key - $k_{0,0} \oplus k_{1,1}$. The same attack can be used to recover the entire key up to 2^{32} variants.

Data and Computational Costs

To compute the data cost, we first analyze the case in which the goal is to discover only one byte (in particular, the difference of two bytes) of the right key with probability greater than 95%. A candidate value of δ can be claimed to be wrong if there exists at least a set $\mathcal{A}_\delta$ for which the number of collisions after five rounds is a odd number. Since there are only $2^8 - 1$ different possible values for δ, one needs that such a set $\mathcal{A}_\delta$ exists with probability higher than $(0.95)^{1/255} = 99.98\%$ (remember that since the tests for different δ are independent, the total probability of success is higher than $0.9998^{256} = 0.95$).

Since the probability that the number of collisions for a given set $\mathcal{A}_\delta$ is odd is 50%, 4 different sets $\mathcal{A}_\delta$ (note that one can count the number of collisions in $\mathcal{M}_I$ for all the 4 different I with $|I| = 3$, for a total of 16 possible tests) are sufficient to deduce the right δ with probability higher than 95%, since $2^{-16} \leq 1 - 0.9998 = 2^{-12.3}$. It follows that the cost to find 1 byte of the key is of 4 (cosets) $\cdot 2^{40}$ (number of texts in $\mathcal{A}_\delta$) $\cdot 2^8$ (values of δ) $= 2^{50}$ chosen plaintexts.

In order to find the entire key up to 2^{32} possible variants, the idea is to repeat the attack 12 times, i.e. 3 times for each column. By analogous calculation[16], it follows that

[16] In this case, one needs that for each one of the $2^8 - 1$ wrong possible values for δ, at least one set $\mathcal{A}_\delta$

16 tests (that is 4 different sets $\mathcal{A}_\delta$ - note that there are four different I with $|I| = 3$) are sufficient to deduce the right δ with total probability higher than 95%. Thus, the data cost of the attack is of $12 \cdot 2^{50} = 2^{53.6}$ chosen plaintexts.

Computational Cost. We limit here to report the computational costs of the distinguisher, and we refer to App. G.2 for all the details. In order to count the number of collisions, one can use the same procedure of the attack described in Sect. 5, i.e. one can re-oder the texts with respect to a particular partial order $\preceq$ as defined in Def. 11. However, in this case we propose an alternative strategy, which exploits *data structure* - the complete pseudo-code of such an algorithm is given in Algorithm 3. This method allows to minimize the computational cost, which is well approximated by $2^{55.6}$ table look-ups or approximately $2^{48.96}$ five-rounds encryptions (20 table look-ups $\approx$ 1 round of encryption).

Practical Verification

Using a C/C++ implementation[17], we have practically verified the attack just described on a small-scale variant of AES, as presented in [CMR05] - not on real AES due to the large computational cost of the attack. We emphasize that Prop. 3 is independent of the fact that each word is composed of 8 or 4 bits. Thus, our verification on the small-scale variant of AES is strong evidence for it to hold for the real AES. The main differences between this small-scale AES and the real AES regard the total computational cost.

For simplicity, we limit here to report the result for an attack on a single byte of the key, e.g. $k_{0,0} \oplus k_{1,1}$. For small-scale AES, since there are only $2^4 - 1$ possible candidates, it is sufficient that for each wrong candidate of $k_{0,0} \oplus k_{1,1}$ a set $\mathcal{A}_\delta$ for which the number of collisions is odd exists with probability $(0.95)^{2^{-4}} = 99.659\%$. It follows that 9 tests (that is 3 different sets $\mathcal{A}_\delta$) for each candidate of $k_{0,0} \oplus k_{1,1}$ are sufficient to find the right value. Using the same procedure just presented based on data-structure, the theoretical computational cost is well approximated by $4 \cdot 3 \cdot 2^4 \cdot (2^{20} + 2 \cdot 2^{16}) \simeq 2^{27.75}$ table look-ups. For comparison, using the re-ordering algorithm, the theoretical computational cost is well approximated by $4 \cdot 3 \cdot 2^4 \cdot 2^{20} \cdot (\log 2^{20} + 1) \simeq 2^{31.91}$ table look-ups.

Our tests confirm that 3 different sets $\mathcal{A}_\delta$ are largely sufficient to find the key. The average practical computational cost is of $2^{26.3}$ table look-ups using a data-structure, and $2^{30.5}$ table look-ups using a re-ordering algorithm. To explain the (small) difference with the theoretical value, note that the theoretical value is computed in the worst case. As an example, when a candidate of the key is found to be wrong, it is not necessary to complete the verification for all the other sets $\mathcal{A}_\delta$ or indexes I, but it is sufficient to discard it and to test the next candidate.

6.3 Attack on 5 rounds of AES with a single Secret S-Box - MixColumns Matrix with Zero-Sum of Coefficients

In this section, we show how to adapt the previous attack in order to exploits the property that the sum of three coefficients of each row of the MixColumns matrix M^{MC} is equal to zero.

For a fixed a, consider a set of plaintexts $\mathcal{A}''_\delta$ which depends on the guessed value of

for which the number of collision is odd exists with probability higher than $(0.9998)^{1/12} = 99.99835\%$.

[17] The source codes of the attacks on AES with a secret S-Box are available at https://github.com/Krypto-iaik/Attacks_AES_SecretSBox2

the key δ of the form:

$$\mathcal{A}_\delta^{''} \equiv \left\{ a \oplus \begin{bmatrix} 0 & y & 0 & 0 \\ 0 & x & y \oplus \delta_{1,2} & 0 \\ 0 & 0 & x \oplus \delta_{2,2} & y \oplus \delta_{2,3} \\ 0 & 0 & 0 & x \oplus \delta_{3,3} \end{bmatrix} \;\middle|\; \forall x, y \in \mathbb{F}_{2^8} \right\} \tag{17}$$

where $\delta \equiv (\delta_{1,2}, \delta_{2,2}, \delta_{2,3}, \delta_{3,3})$. Given a set $\mathcal{A}_\delta^{''}$, we claim that the number of collisions among the ciphertexts in the same coset of $\mathcal{M}_I$ for a fixed $I \subseteq \{0,1,2,3\}$ with $|I| = 3$ after 5 rounds is a multiple of 2. More formally:

Proposition 4. *Consider a set of plaintexts $\mathcal{A}_\delta^{''}$ defined as in (17), and the corresponding ciphertexts after 5 rounds. If $\delta_{i,i} = k_{1,1} \oplus k_{i,i}$ and $\delta_{j,j+1} = k_{0,1} \oplus k_{j,j+1}$ for $i = 2,3$ and $j = 1,2$ (the indexes are taken modulo 4), then the number of different pairs of ciphertexts that belong to the same coset of $\mathcal{M}_I$ for a fixed $I \subseteq \{0,1,2,3\}$ with $|I| = 3$ is a multiple of 2.*

Proof. Let $\delta_{i,i} = k_{i,i} \oplus k_{1,1}$ for $i = 2,3$ and $\delta_{j,j+1} = k_{j,j+1} \oplus k_{0,1}$ for $j = 1,2$. By simple computation, there exists a constant b such that a set $\mathcal{A}_\delta^{''}$ is mapped after one round into

$$R(\mathcal{A}_\delta^{''}) \equiv \left\{ b \oplus \begin{bmatrix} 0\text{x}03 \cdot z & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0\text{x}02 \cdot w & 0 & 0 \\ 0\text{x}02 \cdot z & 0\text{x}03 \cdot w & 0 & 0 \end{bmatrix} \;\middle|\; \forall z, w \in \mathbb{F}_{2^8} \right\}.$$

Consider a pair of texts $t^1, t^2 \in R(\mathcal{A}_\delta^{''})$ generated respectively by $t^1 = (z, w)$ and $t^2 = (z', w')$. The idea is to consider the following two cases separately: (1) $z = z'$ and $w \neq w'$ (or viceversa) and (2) $z \neq z'$ and $w \neq w'$, and to show that in the first case (1) the number of collisions is a multiple of 256, while in the second case (2) the number of collisions is a multiple of 2. In particular, consider a pair of texts $t^1, t^2 \in R(\mathcal{A}_\delta^{''})$ generated respectively by $t^1 = (z, w)$ and $t^2 = (z', w')$ with $z \neq z'$ and $w \neq w'$. The idea is to show that $R^4(t^1) \oplus R^4(t^2) \in \mathcal{M}_I$ *if and only if* $R^4(s^1) \oplus R^4(s^2) \in \mathcal{M}_I$ for $|I| = 3$, where the texts $s^1, s^2 \in R(\mathcal{A}_\delta^{''})$ are generated respectively by $s^1 = (z, w')$ and $s^2 = (z', w)$. Similarly, consider the case $z \neq z'$ and $w = w'$ (or viceversa). As before, the idea is to prove that $t^1, t^2 \in R(\mathcal{A}_\delta^{''})$ satisfy the condition $R^4(t^1) \oplus R^4(t^2) \in \mathcal{M}_I$ for $|I| = 3$ if and only if all the pairs of texts $s^1, s^2 \in R(\mathcal{A}_\delta^{''})$ generated respectively by $t^1 = (z, s)$ and $t^2 = (z', s)$ for all $s \in \mathbb{F}_{2^8}$ have the same property. Thus, there exist $n', n^{''} \in \mathbb{N}$ such that the total number of collisions n can be written as $n = 2 \cdot n' + 256 \cdot n^{''} = 2 \cdot (n' + 128 \cdot n^{''})$, that is n is a multiple of 2. The details of the proof can be found in App. H. □

While for $\delta_{i,i} = k_{i,i} \oplus k_{1,1}$ for $i = 2,3$ and $\delta_{j,j+1} = k_{j,j+1} \oplus k_{0,1}$ for $j = 1,2$ it is possible to guarantee that the total number of collisions is a multiple of 2 with probability 1, no analogous result holds for the other cases. That is, if $\delta_{i,i} \neq k_{i,i} \oplus k_{1,1}$ for $i = 2,3$ or/and $\delta_{j,j+1} \neq k_{j,j+1} \oplus k_{0,1}$ for $j = 1,2$, then the total number of collisions is a multiple of 2 with probability 50%.

Data and Computational Costs. Since the procedure of the attack is completely equivalent to the one described in Sect. 6.2, we refer to that section for all the details and we limit here to focus on the data and on the computational costs of this attack.

Note that each set $\mathcal{A}_\delta^{''}$ is composed of 2^{16} or equivalently $2^{15} \cdot (2^{16} - 1) = 2^{31}$ pairs. Since the probability that each pairs belong to the same coset of $\mathcal{M}_J$ for $|J| = 3$ is 2^{-32}, the average number of collision among the ciphertexts for each set is 2^{-1}, that is on average there is at least one collision in $\mathcal{M}_J$ for $|J| = 3$ for only one half of the sets $\mathcal{A}_\delta^{''}$.

With respect to the previous attack, note that in this case an attacker has to guess 4 bytes of the key instead of only 1. Thus, using the same calculation as before, in order to

Data: $19 \cdot 2^{32}$ different sets $\mathcal{A}_\delta^{'''}$ defined as in (17) - 19 different sets for each $\delta \equiv (\delta_{2,2}, \delta_{3,3}, \delta_{1,2}, \delta_{2,3})$ - and corresponding ciphertexts after 5 rounds
Result: $k_{2,2} \oplus k_{1,1}$, $k_{3,3} \oplus k_{1,1}$, $k_{0,1} \oplus k_{1,2}$ and $k_{0,1} \oplus k_{2,3}$
for *each* δ **do**
  $flag \leftarrow 0$;
  for *each set* $\mathcal{A}_\delta^{'''}$ **do**
    for *each* $I \subseteq \{0,1,2,3\}$ *with* $|I| = 3$ **do**
      let (p^i, c^i) for $i = 0, ..., 2^{16} - 1$ be the (plaintexts, ciphertexts) of $\mathcal{A}_\delta^{'''}$;
      re-order this set of elements w.r.t. the partial order $\preceq$ described in Def. 11 s.t. $c^i \preceq c^{i+1}$ for each i; // $\preceq$ depends on I
      $n \leftarrow 0$; // n denotes the number of collisions in $\mathcal{M}_I$
      $i \leftarrow 0$;
      while $i < 2^{16} - 1$ **do**
        $r \leftarrow 1$;
        $j \leftarrow i$;
        while $c^j \oplus c^{j+1} \in \mathcal{M}_I$ **do**
          $r \leftarrow r + 1$;
          $j \leftarrow j + 1$;
        end
        $i \leftarrow j + 1$;
        $n \leftarrow n + r \cdot (r-1)/2$;
      end
      if $(n \bmod 2) \neq 0$ **then**
        $flag \leftarrow 1$;
        next δ;
      end
    end
  end
  if $flag = 0$ **then**
    identify $\delta \equiv (\delta_{2,2}, \delta_{3,3}, \delta_{1,2}, \delta_{2,3})$ as candidate for the four byte of the secret key;
  end
end
return *Candidates for* $(k_{2,2} \oplus k_{1,1}, k_{3,3} \oplus k_{1,1}, k_{0,1} \oplus k_{1,2}, k_{0,1} \oplus k_{2,3})$. // Only one candidate with Prob. 95%

Algorithm 4: *Key-Recovery Attack on 5 rounds of AES with a single secret S-Box.* For simplicity, the goal of the attack is to find four bytes of the key. Exactly the same attack can be used to recover the entire key up to 2^{32} variants.

discard all the wrong candidates of 4-bytes of the key with probability higher than 95%, one needs that for each wrong candidate δ there exists at least one set $\mathcal{A}_\delta^{''}$ for which the number of collision is odd exists with probability higher than $(0.95)^{2^{-32}}$. It follows that one has to do approximately 37 different tests for each candidate δ. However, since on average there is (at least) one collision among the ciphertexts only for half of these sets, the number of tests must be double. As a result, one needs to do approximately $2 \cdot 37 = 74$ tests, that is one has to use approximately 19 different sets $\mathcal{A}_\delta^{''}$ for each wrong candidate δ (remember that there are four different subspaces $\mathcal{M}_J$ with $|J| = 3$). It follows that the data cost to find 4 bytes of the key is well approximated by $19 \cdot 2^{32} \cdot 2^{16} = 2^{52.248}$ chosen plaintexts.

Using a similar procedure, one can find the entire key. In particular, one first repeats the attack just presented on the third and on the fourth column. To find other four bytes of the key, a set $\mathcal{A}_\delta^{''}$ with the previous property must exist with probability higher than

$(0.95)^{2^{-34}}$, that is approximately $n \geq 2 \cdot 38 = 76$ different tests (i.e. 19 different sets $\mathcal{A}'_\delta$) for each δ are sufficient in order to find the right key. As before, in order to find the final four bytes of the key (one per column), the idea is to repeat the attack exploiting the knowledge of one byte of the key for each column. Since in this case the attacker has to guess only two bytes of difference of the key instead of four and using the same computation as before[18], approximately $n \geq 2 \cdot 23 = 56$ different tests (i.e. 12 different sets $\mathcal{A}''_\delta$) for each δ are sufficient to find the right key.

In conclusion, the total data cost is approximately of $2 \cdot 2^{52.248} + 12 \cdot 2^{16} \cdot 2^{16} = 2^{53.25}$ chosen plaintexts, while the computational cost using a re-ordering algorithm is well approximated by $2 \cdot 4 \cdot 19 \cdot 2^{32} \cdot 2^{16} \cdot (\log 2^{16} + 1) \simeq 2^{59.25}$ table look-ups, or approximately $2^{52.6}$ five-round encryptions. For comparison, the computational cost using data-structure as in Sect. 6.2 is approximately of $2 \cdot 4 \cdot 19 \cdot 2^{32} \cdot (2^{16} + 2 \cdot 2^{32}) \simeq 2^{72.25}$ table look-ups, that is (much) worse than using a re-ordering algorithm (besides an higher memory cost). Indeed, note that in this last case the size of the vector W - as defined in Sect. 6.2 - is (much) larger than the size of the sets $\mathcal{A}''_\delta$ (i.e. 2^{32} versus 2^{16}).

Practical Verification

Using a C/C++ implementation[19], we have practically verified the attack just described on a small-scale variant of AES, as presented in [CMR05] - not on real AES due to the large computational cost of the attack. As before, we emphasize that Prop. 4 is independent of the fact that each word is composed of 8 or 4 bits and that our verification on the small-scale variant of AES is strong evidence for it to hold for the real AES.

For simplicity, we limit here to report the result for the attack on a four bytes of the key, e.g. $k_{2,2} \oplus k_{1,1}$, $k_{3,3} \oplus k_{1,1}$, $k_{0,1} \oplus k_{1,2}$ and $k_{0,1} \oplus k_{2,3}$. For small-scale AES, since there are $(2^4)^4 = 2^{16}$ candidates for the four bytes of the key, it is sufficient that a set $\mathcal{A}''_\delta$ for which the number of collisions is odd exists for each wrong candidate with probability higher than $(0.95)^{2^{-16}}$. Thus, $22 \cdot 2 = 44$ tests (i.e. 11 different sets $\mathcal{A}_\delta$) for each candidate δ are sufficient to find the right value. Re-ordering the texts as described previously, the theoretical computational cost is well approximated by $11 \cdot 2^{16} \cdot 4 \cdot 2^8 \cdot (\log 2^8 + 1) \simeq 2^{32.6}$ table look-ups.

Our tests confirm that 2 different sets $\mathcal{A}_\delta$ are largely sufficient to find the key. The average practical computational cost is of $2^{29.7}$ table look-ups. The difference is explained by the fact that in general it is possible to discard wrong candidates without considering all the corresponding 11 sets $\mathcal{A}''_\delta$ - we found that 2 sets are usually sufficient.

7 A new 5-round Secret-Key Distinguisher for AES

Using the 4-round distinguisher of Sect. 5 as starting point, we propose a way to extend it 1 round at the end. As a result, we are able to set up a *new probabilistic 5-round secret-key distinguisher for AES which exploits a property which is independent of the secret key.* Even if such a distinguisher is worse than the deterministic one presented in [GRR17a], it can be used to set up a key-recovery attack on 6-round AES (better than a brute-force one) exploiting a distinguisher of the type [GRR17a] - *believed to be hard to exploit.* As a result, this is *the first key-recovery attack for 6-round AES set up by a 5-round secret-key distinguisher for AES.* For completeness, since the 4-round distinguisher works also in the decryption direction, this new 5-round distinguisher and the 6-round attack work also in the reverse direction using chosen ciphertexts instead of plaintexts.

[18] For each one of the 2^{16} possible candidates of the key, one needs that at least a set $\mathcal{A}''_\delta$ for which the number of collisions is not a multiple of 2 exists with probability higher than $(0.95)^{2^{-18}}$.

[19] The source codes of the attacks on AES with a secret S-Box are available at https://github.com/Krypto-iaik/Attacks_AES_SecretSBox2

7.1 5-round Secret-Key Distinguisher

To set up the previous 4-round secret-key distinguisher for AES, one considers plaintexts in the same coset of a column space $\mathcal{C}_I$ for $I \subseteq \{0,1,2,3\}$, construct all the couples and divide them in sets $\mathcal{S}^{\mathcal{C}_I \oplus a}$ as defined in Def. 8. As we have just seen, for each of these sets only one of the two following events can happen: (1) $c^1 \oplus c^2 \in \mathcal{M}_J$ or (2) $c^1 \oplus c^2 \notin \mathcal{M}_J$ for each couple (p^1, c^1) and (p^2, c^2) in $\mathcal{S}^{\mathcal{C}_I \oplus a}$. A similar property holds also for the set $\mathcal{Z}^{\mathcal{C}_I \oplus a}$ as defined in Def. 9. In order *to set up our distinguisher for 5-round of AES, the idea is to consider the number of sets $\mathcal{Z}^{\mathcal{C}_I \oplus a}$ that contains at least one couple for which the two ciphertexts belong in the same coset of $\mathcal{M}_J$ for $J \subseteq \{0,1,2,3\}$* with $|J| = 3$ (J is not fixed). As we are going to show, *the probability of the above event is (a little) lower for 5-round AES than for a random permutation.* As a result, given plaintexts in cosets of $\mathcal{C}_I$ and corresponding ciphertexts after 5 rounds, one can distinguish 5-round AES from a random permutation exploiting the fact that the number of sets $\mathcal{Z}^{\mathcal{C}_I \oplus a}$ for which two ciphertexts of at least one couple belong to the same coset of $\mathcal{M}_J$ for $|J| = 3$ is lower for 5-round AES.

Before we give the details of such a distinguisher, we emphasize the similarity with the 3-round distinguisher that exploits a truncated differential trail. In that case, the idea is to count the number of pairs of texts that satisfies the truncated differential trail. In particular, given pairs of plaintexts in the same coset of a diagonal space $\mathcal{D}_i$, one counts the number of pairs for which the corresponding ciphertexts belong in the same coset of a mixed space $\mathcal{M}_J$ for $|J| = 3$. Since the probability of this event is higher for an AES permutation than for a random one[20], one can distinguish the two cases simply counting the number of pairs that satisfy the previous property. The idea of our disitinguisher is similar. However, instead of working on single couples, one works with particular sets $\mathcal{Z}$ of couples and counts the number of sets for which at least one couple satisfies the differential trail. In App. D, we show that the same distinguisher can be set up using sets $\mathcal{S}$ or $\mathcal{T}$ instead of $\mathcal{Z}$.

Details of the new Distinguisher

In order to distinguish 5-round AES from a random permutation, the idea is to construct all the sets $\mathcal{Z}^{\mathcal{C}_I \oplus a}$ and to count the number of sets for which two ciphertexts of at least one couple belong to the same coset of $\mathcal{M}_J$ for a certain J with $|J| = 3$. As we are going to show, given a set $\mathcal{Z}^{\mathcal{C}_I \oplus a}$ for $|I| = 1$, the probability that at least one couple of ciphertexts with the previous property exists is a little lower for an AES permutation (approximately $2^{-13} - 524\,287 \cdot 2^{-46} - 22\,370\,411\,853 \cdot 2^{-77} + ...$) than for a random one (approximately $2^{-13} - 524\,287 \cdot 2^{-46} + 45\,812\,722\,347 \cdot 2^{-77} + ...$). Exploiting this small difference, it is possible to distinguish the two cases. In the following, we give all the details.

Our 5-round distinguisher is based on the following property of the previous 4-round distinguisher. As we have just seen, given plaintexts in the same coset of $\mathcal{C}_I$ and for a fixed $J \subseteq \{0,1,2,3\}$, each set $\mathcal{Z}^{\mathcal{C}_I \oplus a}$ as defined in Def. 9 has the following property after 4 rounds (by Lemma 3):

1. for each couple, the two ciphertexts belong to the same coset of $\mathcal{M}_J$;

2. for each couple, the two ciphertexts don't belong to the same coset of $\mathcal{M}_J$.

In other words, for a given set $\mathcal{Z}^{\mathcal{C}_I \oplus a}$, it is not possible that the two ciphertexts of only some - not all - couples belong to the same coset of $\mathcal{M}_J$, while this can happen for a random permutation.

[20] As recalled in Sect. 3.2, this probability is approximately equal to 2^{-22} for the AES case and 2^{-30} for the random case.

What is the probability of the two previous events for an AES permutation? Given a set $\mathcal{Z}^{\mathcal{C}_I \oplus a}$, the probability that the two ciphertexts of each couple belong to the same coset of $\mathcal{M}_J$ is approximately 2^{-30}. Indeed, let the event $\mathcal{E}_i^r$ defined as following.

Definition 12. Let $J \subseteq \{0,1,2,3\}$ fixed. Given a set $\mathcal{Z}^{\mathcal{C}_I \oplus a}$, we define $\mathcal{E}_i^r$ as the event that the i-th couple of $\mathcal{Z}^{\mathcal{C}_I \oplus a}$ for $i = 1, 2, ..., 2^{17}$ belong to the same coset of $\mathcal{M}_J$ after r rounds.

For the following, let $\overline{\mathcal{E}_i^r}$ be the complementary event of $\mathcal{E}_i^r$. It follows that

$$\begin{aligned} Prob(\mathcal{E}_1^4 \wedge \mathcal{E}_2^4 \wedge ... \wedge \mathcal{E}_{2^{17}}^4) &= Prob(\mathcal{E}_1^4) \cdot Prob(\mathcal{E}_2^4 \wedge ... \wedge \mathcal{E}_{2^{17}}^4 \,|\, \mathcal{E}_1^4) = \\ &= Prob(\mathcal{E}_1^4) \equiv p_3 = 2^{-30} - 3 \cdot 2^{-63} + 2^{-94}, \end{aligned}$$

where p_3 is defined as in (6). Indeed, note that $Prob(\mathcal{E}_i^4 \,|\, \mathcal{E}_1^4) = 1$ for each $i = 2, ..., 2^{17}$ since if two ciphertexts of one couple belong (or not) to the same coset of $\mathcal{M}_J$, then the ciphertexts of all the other couples have the same property.

Using these initial considerations as starting point, we analyze in details our proposed 5-round distinguisher. *Given a set $\mathcal{Z}^{\mathcal{C}_I \oplus a}$, what is the probability that two ciphertexts of at least one couple in that set belong to the same coset of $\mathcal{M}_J$ for a certain $J \subseteq \{0,1,2,3\}$ with $|J| = 3$?* To compute this probability, we consider separately the two cases in which for all the couples the two ciphertexts belong or not to the same coset of $\mathcal{M}_K$ for a certain K after 4 rounds. We finally obtain the desired probability using the *law* (or formula) *of total probability* $Prob(A) = \sum_i Prob(A \,|\, B_i) \cdot Prob(B_i)$ which holds for each event A such that $\bigcup_i B_i$ is the *sample space*, i.e. the set of all the possible outcomes.

Given a set $\mathcal{Z}^{\mathcal{C}_I \oplus a}$, assume first that the two plaintexts of each couple don't belong to the same coset of $\mathcal{M}_K$ for all[21] $K \subseteq \{0,1,2,3\}$ with $|K| = 3$ after 4 rounds. In this case, the probability that the two ciphertexts of at least one couple belong to the same coset of $\mathcal{M}_J$ for $|J| = 3$ after 5 rounds is well approximated by

$$1 - \left(1 - \hat{p}_{3,3}\right)^{2^{17}} = 1 - \left(1 - \frac{p_3 \cdot (1 - p_{3,3})}{1 - p_3}\right)^{2^{17}} = 2^{-13} - 526\,327 \cdot 2^{-46} + ...$$

where $\hat{p}_{3,3}$ is defined in (8). The other case is similar. Consider a set for which the two plaintexts of each couple belong to the same coset of $\mathcal{M}_K$ for $K \subseteq \{0,1,2,3\}$ with $|K| = 3$ after 4 rounds.In this case, the probability that the two ciphertexts of at least one couple belong to the same coset of $\mathcal{M}_J$ for $|J| = 3$ after 5 rounds is well approximated by

$$1 - \left(1 - p_{3,3}\right)^{2^{17}} = 2^{-5} - 524\,287 \cdot 2^{-30} + 45\,812\,722\,347 \cdot 2^{-53} + ...$$

where $p_{3,3}$ is defined in (7). Using the law of total probability and given a set $\mathcal{Z}^{\mathcal{C}_I \oplus a}$ *for* $|I| = 1$, it follows that the probability that two ciphertexts of at least one couple belong to the same coset of $\mathcal{M}_J$ is well approximated by

$$\begin{aligned} p_{AES} =& \left[1 - Prob(\overline{\mathcal{E}_1^5} \wedge \overline{\mathcal{E}_2^5} \wedge ... \wedge \overline{\mathcal{E}_{2^{17}}^5} \,|\, \mathcal{E}_i^4)\right] \cdot Prob(\mathcal{E}_i^4) + \\ &+ \left[1 - Prob(\overline{\mathcal{E}_1^5} \wedge \overline{\mathcal{E}_2^5} \wedge ... \wedge \overline{\mathcal{E}_{2^{17}}^5} \,|\, \overline{\mathcal{E}_i^4})\right] \cdot Prob(\overline{\mathcal{E}_i^4}) = \\ =& (1 - p_3) \cdot \left[1 - \left(1 - \frac{p_3 \cdot (1 - p_{3,3})}{1 - p_3}\right)^{2^{17}}\right] + p_3 \cdot \left[1 - \left(1 - p_{3,3}\right)^{2^{17}}\right] = \\ =& 2^{-13} - 524287 \cdot 2^{-46} - \underbrace{22\,370\,411\,853 \cdot 2^{-77}}_{\approx 2.604 \cdot 2^{-44}} + ... \end{aligned} \tag{18}$$

[21] Note that $\mathcal{M}_{\hat{K}} \subseteq \mathcal{M}_K$ for all $\hat{K} \subseteq K$. If two texts don't belong to the same coset of $\mathcal{M}_K$ for all $K \subseteq \{0,1,2,3\}$ with $|K| = 3$, then they don't belong to the same coset of $\mathcal{M}_{\hat{K}}$ for $\hat{K} \subseteq \{0,1,2,3\}$ with $|\hat{K}| < 3$. Viceversa, if they belong to the same coset of $\mathcal{M}_{\hat{K}}$ for $\hat{K}$ with $|\hat{K}| < 3$, then they belong to the same coset of $\mathcal{M}_K$ for all K with $|K| = 3$ and $\hat{K} \subseteq K$.

for a certain $i \in \{1, ..., 2^{17}\}$. Note that $Prob(\mathcal{E}_i^5 \wedge \mathcal{E}_j^5) = Prob(\mathcal{E}_i^5) \times Prob(\mathcal{E}_j^5)$ since the events $\mathcal{E}_i^5$ and $\mathcal{E}_j^5$ are independent for $i \neq j$. For a random permutation, the same event occurs with (approximately) probability

$$\begin{aligned} p_{rand} =& 1 - \left(1 - p_3\right)^{2^{17}} = 1 - \left[1 - \left(2^{-30} - 3 \cdot 2^{-63} + 2^{-94}\right)\right]^{2^{17}} = \\ =& 2^{-13} - 524\,287 \cdot 2^{-46} + \underbrace{45\,812\,722\,347 \cdot 2^{-77}}_{\approx 5.333 \cdot 2^{-44}} + ... \end{aligned} \tag{19}$$

We emphasize again that while a "classical" truncated differential distinguisher counts the number of pairs of texts that satisfy a particular differential trail, in our case we consider the number of sets of texts for which at least one pair satisfies a particular differential trail. This choice allows to have *a difference between the probabilities* that the previous event occurs for a random permutation p_{rand} and for 5-round AES p_{AES}.

7.2 Data and Computational Complexity

7.2.1 Data Complexity

Since the difference between the two probabilities is very small, what is the minimum number of sets $\mathcal{Z}^{\mathcal{C}_I \oplus a}$ (or equivalently of cosets $\mathcal{C}_I$) to guarantee that the distinguisher works with high probability?

First of all, given a single coset of a column space $\mathcal{C}_I$ for $|I| = 1$, the number of different couples with two generating variables is given by $6 \cdot 2^{16} \cdot 2^{15} \cdot (2^8 - 1)^2 \simeq 2^{49.6}$ (see Eq. (11)), while the number of sets $\mathcal{Z}^{\mathcal{C}_I \oplus a}$ that one can construct is well approximated by $3 \cdot 2^{15} \cdot (2^8 - 1)^2 \simeq 2^{32.574}$.

As we have just said, the difference between the number of sets that satisfy the required property for the AES case (i.e. n_{AES}) and for the random case (i.e. n_{rand}) is very small compared to the total number n_{AES} or n_{rand}:

$$\frac{|n_{AES} - n_{rand}|}{n_{AES}} \simeq \frac{|n_{AES} - n_{rand}|}{n_{rand}} \ll 1.$$

Thus, our goal is to derive a good approximation for the number of initial cosets of $\mathcal{C}_I$ that is sufficient to appreciate this difference with probability *prob*.

To solve this problem, note that given n sets $\mathcal{Z}^{\mathcal{C}_I \oplus a}$ of 2^{17} couples defined as in Def. 9, the distribution probability of our model is simply described by a *binomial distribution*. By definition, a binomial distribution with parameters n and p is the discrete probability distribution of the number of successes in a sequence of n independent yes/no experiments, each of which yields success with probability p. In our case, given n sets $\mathcal{Z}^{\mathcal{C}_I \oplus a}$, each of them satisfies or not the above property/requirement with a certain probability. Thus, this model can be described using a binomial distribution. We recall that for a random variable Z that follows the binomial distribution, that is $Z \sim \mathcal{B}(n, p)$, the mean μ and the variance σ^2 are respectively given by $\mu = n \cdot p$ and $\sigma^2 = n \cdot p \cdot (1 - p)$.

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

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

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

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

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

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

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

where $\mathrm{erfinv}(x)$ is the inverse error function. For the case $p_{rand}, p_{AES} \ll 1$, a good approximation of n is given by[22]

$$n > \frac{4 \cdot \max(p_{rand},\, p_{AES})}{(p_{rand} - p_{AES})^2} \cdot \left[\mathrm{erfinv}\big(2 \cdot prob - 1\big)\right]^2. \tag{20}$$

For a probability of success of approximately 95%, since $|p_{AES} - p_{rand}| \simeq 2^{-41.01}$ and $p_{AES} \simeq p_{rand} \simeq 2^{-13}$, it follows that n must satisfy $n > 2^{71.243}$. Since a single coset of $\mathcal{C}_I$ for $|I| = 1$ contains approximately $2^{32.574}$ different sets $\mathcal{Z}$, one needs approximately $2^{71.243} \cdot 2^{-32.574} \simeq 2^{38.669}$ different initial cosets of $\mathcal{C}_I$, that is approximately $2^{38.669} \cdot 2^{32} \simeq 2^{70.67}$ chosen plaintexts.

Another possibility is to use an initial coset of $\mathcal{C}_I$ with $|I| = 2$. In this case, using sets $\mathcal{T}^{\mathcal{C}_I \oplus a}$ - as defined in Def. 10 - instead of sets $\mathcal{Z}^{\mathcal{C}_I \oplus a}$, approximately $2^{51.17}$ chosen plaintexts in the same initial coset of $\mathcal{C}_I$ with $|I| = 2$ are sufficient to set up the distinguisher, as showed in details in App. D.2.

Before we go on, we emphasize that formula given in (20) is equivalent to the one proposed by Matsui in [Mat94] for the linear cryptanalysis case, and so it has been rigorously studied in the literature (e.g. in [BJV04], [Sel08]). Without going into the details, in linear cryptanalysis one has to construct "good" linear equations relating plaintext, ciphertext and key bits. In order to find the secret key, the idea is to exploit the fact that such linear approximation holds with probability $1/2$ for a wrong key, while they hold with probability $1/2 \pm \varepsilon$ for the right key. Exploiting this (usually small) difference between the two probabilities, one can discover the secret key. Our case is completely equivalent, since the probability p_{AES} of the AES case is related to the probability p_{rand} of the random case by $p_{AES} = p_{rand} \pm \varepsilon$, for a small difference ε.

7.2.2 Computational Complexity

Here we discuss the computational cost for the case of cosets of $\mathcal{C}_I$ with $|I| = 1$. The analysis is similar for the case $|I| = 2$, and the details are given in App. D.2. As for the 4-round distinguisher, a first possibility is to construct all the couples, to divide them in sets $\mathcal{Z}^{\mathcal{C}_I \oplus a}$ for $|I| = 1$ defined above, and to count the number of sets that satisfy the above property working on each set separately. Since just the cost to construct all the couples given $2^{38.67}$ cosets is approximately of $2^{38.67} \cdot 2^{31} \cdot (2^{32} - 1) \simeq 2^{101.67}$ table look-ups, we present a way to implement the distinguisher in a more efficient way, similar to the one proposed for the 4-round distinguisher of Sect. 5 (details are given in App. B).

[22] Observe: $p_{rand} \cdot (1 - p_{rand}) + p_{AES} \cdot (1 - p_{AES}) < p_{rand} + p_{AES} < 2 \cdot \max(p_{rand}, p_{AES})$.

Data: 1 coset of $\mathcal{C}_L$ for $|L| \subseteq \{0,1,2,3\}$ (e.g. $\mathcal{C}_L \oplus a$ with $a \in \mathcal{C}_L^{\perp}$) and corresponding ciphertexts after 5 rounds
Result: Number of sets $\mathcal{Z}^{\mathcal{C}_L \oplus a}$ with at least one couple for which the two ciphertexts in the same coset of $\mathcal{M}_J$ for a certain J with $|J| = 3$

$n \leftarrow 0$; `// number of sets` $\mathcal{Z}$ `with the required property`
for *each* j *from 0 to 3 let* $J = \{0,1,2,3\} \setminus j$ - $|J| = 3$ - **do**
  let (p^i, c^i) for $i = 0, ..., 2^{32 \cdot |L|} - 1$ be the (plaintexts, ciphertexts) in $\mathcal{C}_L \oplus a$;
  re-order this set of elements w.r.t. the partial order $\preceq$ defined in Def. 11 s.t. $c^i \preceq c^{i+1}$ for each i; `//` $\preceq$ `depends on` J
  $i \leftarrow 0$;
  while $i < 2^{32 \cdot |L|} - 1$ **do**
    $j \leftarrow i$;
    while $c^j \oplus c^{j+1} \in \mathcal{M}_J$ **do**
      $j \leftarrow j + 1$;
    end
    for *each* k *from* i *to* j **do**
      for *each* l *from* $k+1$ *to* j **do**
        if $p^k \equiv (x^0, x^1, x^2, x^3)$ *and* $p^l \equiv (y^0, y^1, y^2, y^3)$ *have only 2 common generating variables, i.e.* $\exists H \subseteq \{0,1,2,3\}$ *with* $|H| = 2$ *s.t.* $x^h = y^h$ *for* $h \in H$ *and* $x^h \neq y^h$ *for* $h \in \{0,1,2,3\} \setminus H$ **then**
          construct the set $\mathcal{Z}^{\mathcal{C}_L \oplus a}_{p^k, p^l}$ as defined in Def. 9 - Eq. (13);
          $flag \leftarrow 0$
          for *each* $I \subseteq \{0,1,2,3\}$ *with* $|I| = 3$ *and* $I > J$ *w.r.t. Def. 13* **do**
            for *each couple* of (plaintexts, ciphertexts) $\{(\hat{p}^1, \hat{c}^1), (\hat{p}^2, \hat{c}^2)\} \in \mathcal{Z}^{\mathcal{C}_L \oplus a}_{p^k, p^l}$ **do**
              if $\hat{c}^1 \oplus \hat{c}^2 \in \mathcal{M}_I$ **then**
                $flag \leftarrow 1$;
              end
            end
          end
          for *each couple* of (plaintexts, ciphertexts) $\{(\hat{p}^1, \hat{c}^1), (\hat{p}^2, \hat{c}^2)\} \in \mathcal{Z}^{\mathcal{C}_L \oplus a}_{p^k, p^l}$ **do**
            if $(\hat{p}^1, \hat{p}^2) > (p^k, p^l)$ *w.r.t. Def. 14 and* $\hat{c}^1 \oplus \hat{c}^2 \in \mathcal{M}_J$ **then**
              $flag \leftarrow 1$;
            end
          end
          if $flag = 0$ **then**
            $n \leftarrow n + 1$;
          end
        end
      end
    end
  end
end
return n.

Algorithm 5: Given (plaintexts, ciphertexts) pairs in the same coset of $\mathcal{C}_L$, *this algorithm counts the number of sets* $\mathcal{Z}^{\mathcal{C}_L \oplus a}$ *for which two ciphertext of at least one couple belong in the same coset of* $\mathcal{M}_J$ *for* $|J| = 3$.

Let $J \subseteq \{0,1,2,3\}$ with $|J| = 3$. As before, the idea is to re-order the ciphertexts with respect to the partial order $\preceq$ defined in Def. 11. Given ordered ciphertexts and

working only on consecutive ciphertexts, the idea is to look for collisions (i.e. c^1 and c^2 such that $c^1 \oplus c^2 \oplus \mathcal{M}_J$) and to construct the corresponding set $\mathcal{Z}$ only for the ciphertexts that collide. However, when a collision is found, a "problem" arises. For our scope, we are interested in the number of sets $\mathcal{Z}^{\mathcal{C}_I \oplus a}$ for which there exists J such that *at least* one couple of ciphertexts belong to the same coset of $\mathcal{M}_J$. Thus, consider the corresponding set $\mathcal{Z}^{\mathcal{C}_I \oplus a}$ of the two previous ciphertexts. In the case in which the ciphertexts of all other couples of $\mathcal{Z}^{\mathcal{C}_I \oplus a}$ don't belong to the same coset of $\mathcal{M}_J$, then one can simply increment the total number of sets for which the property is satisfied. On the other hand, if there is another couple in the set $\mathcal{Z}^{\mathcal{C}_I \oplus a}$ for which the two ciphertexts belong to the same coset of $\mathcal{M}_J$, one must guarantee that the counter is not incremented two or more times for the same set. How to do/implement this in an efficient way?

Definition 13. Let $I, J \subseteq \{0,1,2,3\}$ such that $|I| = |J| = 3$ and $I \neq J$. Let $i \in \{0,1,2,3\} \setminus I$ and $j \in \{0,1,2,3\} \setminus J$. Then $I < J$ if and only if $i < j$.

Definition 14. Let (t^1, t^2) and (s^1, s^2) be two pairs of texts - $t^1, t^2, s^1, s^2 \in \mathbb{F}_{2^8}^{4\times 4}$ - such that $s^1 < s^2$ and $t^1 < t^2$, with respect to the partial order $<$ defined in Def. 6. We say that $(t^1, t^2) < (s^1, s^2)$ if (1) $t^2 < s^2$ or (2) $s^2 = t^2$ and $t^1 < s^1$.

Working with "ordered" $J \subseteq \{0,1,2,3\}$ with $|J| = 3$ (see Algorithm 5 for details), when two ciphertexts c^1 and c^2 are found such that $c^1 \oplus c^2 \in \mathcal{M}_J$, one constructs the corresponding set $\mathcal{Z}^{\mathcal{C}_I \oplus a}$ of 2^{17} couples. The idea is to increment the number of sets unless one of the two following events occurs:

1. there exist $J' \subseteq \{0,1,2,3\}$ such that $J' > J$ and a couple of ciphertexts $\hat{c}^1$ and $\hat{c}^2$ in the set $\mathcal{Z}^{\mathcal{C}_I \oplus a}$ such that $\hat{c}^1 \oplus \hat{c}^2 \in \mathcal{M}_{J'}$;
2. there exists a couple of ciphertexts $\hat{c}^1$ and $\hat{c}^2$ in the set $\mathcal{Z}^{\mathcal{C}_I \oplus a}$ (with $\hat{c}^1 < \hat{c}^2$) such that $(c^1, c^2) < (\hat{c}^1, \hat{c}^2)$ and $\hat{c}^1 \oplus \hat{c}^2 \in \mathcal{M}_{J'}$.

This strategy - presented in details in Algorithm 5 - guarantees to not count the same set more than a single time.

What is the total computational cost? The idea is to store all the (plaintexts, ciphertexts) pairs twice, once with the plaintexts ordered w.r.t to the partial order $\leq$ and the other with the ciphertexts ordered w.r.t to the partial order $\preceq$ (see App. B for details). First of all, the cost to re-order the ciphertexts and to look for collisions is approximately of $4 \cdot 2^{38.67} \cdot 2^{32} \cdot (1 + \log 2^{32}) = 2^{77.67}$ table look-ups. In order to compute the total cost, we have to consider the average number of collisions, since for each collision one has to construct the corresponding set $\mathcal{Z}^{\mathcal{C}_I \oplus a}$. Since the probability that two texts belong to the same coset of $\mathcal{M}_J$ for $|J| = 3$ is 2^{-30} and since the number of possible pairs is approximately $2^{38.67} \cdot 2^{63} \simeq 2^{101.67}$, it follows that the average number of collisions is approximately $2^{101.67} \cdot 2^{-30} = 2^{71.67}$. On the other hands, since we are only interested in the collisions pairs for which the plaintexts have exactly 2 equal generating variables (which happens with prob. $\binom{4}{2} \cdot 2^{-16} = 3 \cdot 2^{-15}$), it follows that the number of collisions for which one has to really construct the set $\mathcal{Z}$ is approximately $2^{71.67} \cdot 3 \cdot 2^{-15} \simeq 2^{58.255}$. For each one of them, one needs $2 \cdot 2^{17} = 2^{18}$ table look-ups to construct the corresponding set $\mathcal{Z}^{\mathcal{C}_I \oplus a}$ for $|I| = 1$ and to check the required property on the ciphertexts (since this last step involves only XOR-sum, its cost is negligible w.r.t. the total cost). As a result, the total cost is well approximated by $2^{77.67} + 2^{18} \cdot 2^{58.255} = 2^{78.13}$ table look-ups, or approximately $2^{71.5}$ five-round encryptions.

7.3 Practical Verification on small-scale AES

In order to have a practical verification of the proposed distinguisher (and of the following key-recovery attack), we have practically verified the probabilities p_{AES} and p_{rand} given

above[23]. In particular, we verified them using a small-scale AES, as proposed in [CMR05]. We emphasize that our verification on the small-scale variant of AES is strong evidence for it to hold for the real AES, since the strategy used to theoretically compute such probabilities is independent of the fact that each word of AES is of 4 or 8 bits.

To compare the practical values with the theoretical ones, we list the theoretical probabilities p_{AES} and p_{rand} for the small-scale case. First of all, for small scale AES the probabilities p_3 and $p_{3,3}$ are respectively equal to $p_3 = 2^{-14} - 3 \cdot 2^{-31} + 2^{-46}$ and $p_{3,3} = 2^{-10} - 3 \cdot 2^{-23} + 2^{-34}$.

W.l.o.g. we used cosets of $\mathcal{C}_0$ to practically test the two probabilities. Using the previous procedure and formula, the (approximately) probabilities that a set $\mathcal{Z}^{\mathcal{C}_0 \oplus a}$ satisfy the required property for 5-round AES and the random case are respectively

$$p_{AES} = 2^{-5} - 2\,047 \cdot 2^{-22} - \underbrace{221\,773 \cdot 2^{-37}}_{\approx 3.384 \cdot 2^{-21}} + \ldots$$
$$p_{rand} \cong 2^{-5} - 2\,047 \cdot 2^{-22} + \underbrace{698\,027 \cdot 2^{-37}}_{\approx 10.651 \cdot 2^{-21}} + \ldots$$

As a result, using formula (20) for $p_{rand} \simeq p_{AES} \simeq 2^{-5}$ and $|p_{rand} - p_{AES}| \simeq 2^{-17.19}$, it follows that $n \geq 2^{31.6}$ different sets $\mathcal{Z}^{\mathcal{C}_0 \oplus a}$ are sufficient to set up the distinguisher with probability higher than 95%.

Note that for small-scale AES, a single coset of $\mathcal{C}_0$ contains 2^{16} (plaintexts, ciphertexts) pairs, or approximately $2^{15} \cdot (2^{16} - 1) \simeq 2^{31}$ different couples. Since the number of couples with two different generating variables is given by $6 \cdot 2^8 \cdot 2^7 \cdot (2^4 - 1)^2 \simeq 2^{25.4}$ (also tested by computer test), it is possible to construct $3 \cdot 2^7 \cdot (2^4 - 1)^2 = 86400 \simeq 2^{16.4}$ sets $\mathcal{Z}$ of 2^9 couples. As a result, it follows that $2^{31.6} \cdot 2^{-16.4} = 2^{15.2}$ different initial cosets of $\mathcal{C}_0$ must be used, for a cost of $2^{47.2}$ chosen plaintexts.

For our tests, we used 2^{16} different initial cosets of $\mathcal{C}_0$ (keys used to encrypt the plaintexts in the AES case are randomly chosen and different for each coset - the key is not fixed). For each coset we exploited Algorithm 5 to count the number of sets $\mathcal{Z}^{\mathcal{C}_0 \oplus a}$ that satisfy the required property (i.e. the number of sets for which two ciphertexts of at least one couple are in the same coset of $\mathcal{M}_J$ for certain J with $|J| = 3$). As a result, for each initial coset $\mathcal{C}_0$ the (average) theoretical numbers of sets $\mathcal{Z}^{\mathcal{C}_0 \oplus a}$ that satisfy the required property for the random and the AES cases - given by $n_X^T = 86\,400 \cdot p_X$ - and the (average) practical ones found in our experiments - denoted by n_X^P - are given are:

$$n_{rand}^T \simeq 2\,658.27 \qquad n_{AES}^T \simeq 2\,657.69$$
$$n_{rand}^P \simeq 2\,658.21 \qquad n_{AES}^P \simeq 2\,657.63$$

Note that the numbers of collisions found in our experiments are close to the theoretical ones, and that the average number of sets for AES case is lower than for the random one, as predicted.

For completeness, the probabilistic distributions of the number of collisions is given in Fig. 3 for the AES case and in Fig. 4 for the random case. In both cases, the practical distribution is obtained using $20\,000 \equiv 2^{14.3}$ initial cosets. It is possible to observe that e.g. the theoretical variance matches the practical one in both cases.

7.4 Key-Recovery Attack on 6 rounds of AES-128

Using the previous distinguisher on 5-round AES (based on a property which is independent of the secret key) as starting point, we propose the first key-recovery attack on 6 rounds of

[23] The source codes of the distinguishers/attacks are available at https://github.com/Krypto-iaik/Distinguisher_5RoundAES

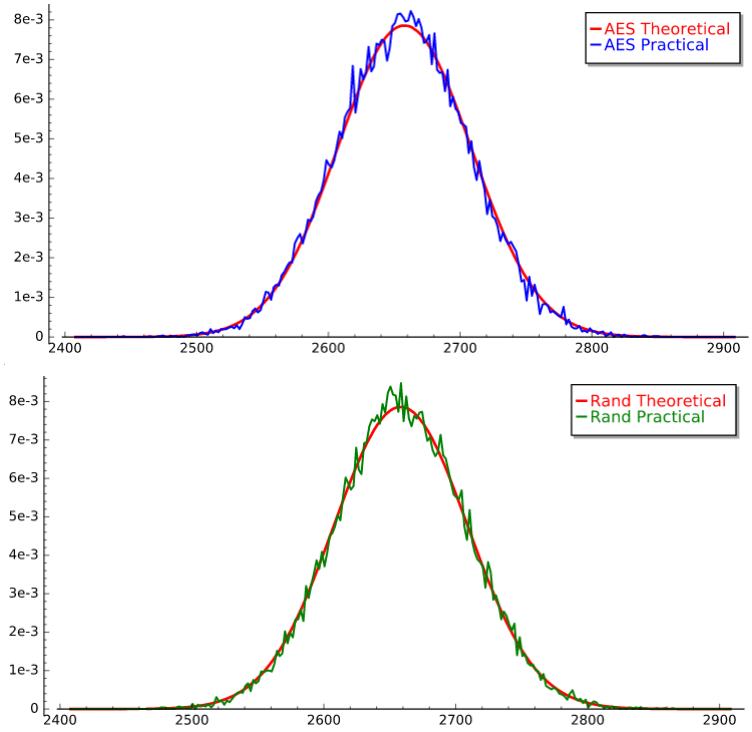

Figure 4: *Probabilistic Distribution of the Number of Sets $\mathcal{Z}$ that satisfy the required property for the Random case* - using 20 000 initial cosets.

AES that exploits a 5-round secret-key distinguisher. The strategy of the attack is similar to the one exploited by linear and differential cryptanalysis.

For the distinguisher just presented, the idea is to consider plaintexts in cosets of $\mathcal{C}_I$ for $I \subseteq \{0,1,2,3\}$, construct all the possible couples of two (plaintexts, ciphertexts) pairs (discarding the ones with common generating variables), divide them into sets $\mathcal{Z}^{\mathcal{C}_I \oplus a}$ of 2^{17} couples and count the number of sets for which at least one couple of ciphertexts belong to the same coset of $\mathcal{M}_J$ for $|J| = 3$. For the following, we limit to consider the case $|I| = 1$. To set up the key-recovery attack, the idea is simply to start with cosets of $\mathcal{D}_I$ for $I \in \{0,1,2,3\}$, and to repeat the previous procedure for each guessed combination of the I-th diagonal of the secret key. We emphasize that these *guessed 4-bytes of the key influence the way in which the couples of texts are divided into the sets* $\mathcal{Z}^{R(\mathcal{D}_I \oplus a)} := \mathcal{Z}^{\mathcal{C}_I \oplus b}$. As a consequence, if the 4 guessed bytes are different from the right ones (i.e. they are wrong), the couples are divided into set $\mathcal{Z}^{\mathcal{C}_I \oplus a}$ in a random way. As we are going to show, for wrong guessed key the probability that a set $\mathcal{Z}^{\mathcal{C}_I \oplus a}$ satisfies the required property is (approximately) equal to the probability of the random case p_{rand}, which is higher than the probability of the correct guessed key p_{AES}. As a result, the number of sets $\mathcal{Z}^{R(\mathcal{D}_I \oplus a)}$ for which two ciphertexts of at least one couple belong to the same coset of $\mathcal{M}_J$ for $|J| = 3$ is minimum for the right key. This allows to recover one diagonal of the secret key.

7.4.1 Data Complexity

As we are going to show, the behavior in the case of a wrong guessed key (for the following denoted by "AES with a wrong key") is similar to the one of a random permutation. The main difference between "AES with a wrong key" and a random permutation is given by the possibility in the first case to study the distribution of the couples after each round

- note that for a random permutation it is meaningless to consider the distribution of the texts after (e.g.) one round. In particular, a coset of a diagonal space $\mathcal{D}_I$ is always mapped into a coset of a column space $\mathcal{C}_I$ after one round independently of the key. On the other hand, we stress that *the way in which the couples are distributed in the sets* $\mathcal{Z}^{R(\mathcal{D}_I \oplus a)} := \mathcal{Z}^{\mathcal{C}_I \oplus b}$ *depends on the guessed key.*

Consider a key-recovery attack on 6-round AES

$$\mathcal{D}_I \oplus a \xrightarrow[KeyGuess]{R(\cdot)} \underbrace{\qquad\qquad\qquad\qquad}_{\text{5-round Secret-Key Distinguisher of Sect. 7}}$$
$$\bigcup_{q^1,q^2} \mathcal{Z}_{q^1,q^2}^{R(\mathcal{D}_I \oplus a)} \subseteq \mathcal{C}_I \oplus b \xrightarrow[\text{prob. 1}]{R(\cdot)} \mathcal{M}_I \oplus c \xrightarrow{R(\cdot)} \mathcal{D}_J \oplus a' \xrightarrow[\text{prob. 1}]{R^2(\cdot)} \mathcal{M}_J \oplus c' \xrightarrow{R(\cdot)} \mathcal{M}_K \oplus c''$$

and focus on the middle round $\mathcal{M}_I \oplus c \xrightarrow{R(\cdot)} \mathcal{D}_J \oplus a'$ for $|I| = 1$ and $|J| = 3$. Assume the guessed key is wrong, and consider one set $\mathcal{Z}_{R(p^1),R(p^2)}^{R(\mathcal{D}_I \oplus a)}$. For this set, the number of couples that belong to the same coset of $\mathcal{M}_J$ after four rounds can take any possible value between 0 and 2^{17} (that is, 0, 1, 2, ... or 2^{17}). Indeed, since the distribution of the couples in the sets $\mathcal{Z}^{R(\mathcal{D}_I \oplus a)}$ has the same behavior of a random one, it is not possible guarantee that the number of couples that belong to the same coset of $\mathcal{M}_J$ after 4 rounds is only 0 or 2^{17} (as for "AES with the right key"). Using same calculation of before and for a wrong guessed key, the probability $p_{AES}^{WrongKey}$ that for a set $\mathcal{Z}^{R(\mathcal{D}_I \oplus a)}$ two texts of at least one couple belong to the same coset of $\mathcal{M}_K$ for a certain $|K| = 3$ after 6 rounds is well approximated by

$$p_{AES}^{WrongKey} = \sum_{n=0}^{2^{17}} \binom{2^{17}}{n} \cdot p_3^n \cdot (1-p_3)^{2^{17}-n} \cdot \left[1 - \left(1 - p_{3,3}\right)^n \cdot \left(1 - \frac{p_3 \cdot (1-p_{3,3})}{1-p_3}\right)^{2^{17}-n}\right],$$

which is well approximated by

$$p_{AES}^{WrongKey} = 2^{-13} - 524\,287 \cdot 2^{-46} + 45\,812\,722\,347 \cdot 2^{-77} + ...$$

Note that this probability is similar - but not exactly equal - to the one of the random case, while we remember that the probability for "AES with the right key" is $p_{AES} = 2^{-13} - 524\,287 \cdot 2^{-46} - 22\,370\,411\,853 \cdot 2^{-77} + ...$ where the difference between these two probabilities is approximately $|p_{AES}^{WrongKey} - p_{AES}| \simeq 2^{-41.011}$.

What is the data cost to find one diagonal of the key? Assume we want to discover the I-th diagonal of the key with probability higher than 95%. Equivalently, this means that one has to guarantee that the number of sets $\mathcal{Z}^{R(\mathcal{D}_I \oplus a)} \equiv \mathcal{Z}^{\mathcal{C}_I \oplus b}$ that satisfy the previous required property is the lowest one for the right key with probability higher than 95%. To compute the data cost, the idea is to use the same analysis proposed for the 5-round distinguisher in Sect. 7.2. In particular, since there are 2^{32} candidates for each diagonal of the keys, one has to guarantee that the number of sets $\mathcal{Z}^{R(\mathcal{D}_I \oplus a)}$ that satisfy the previous required property is the lowest one for the right key with probability higher than $(0.95)^{2^{-32}}$ (note that the 2^{32} tests - one for each candidate - are all independent). Using formula (20), one needs approximately $2^{73.343}$ different sets $\mathcal{Z}^{R(\mathcal{D}_I \oplus a)}$ for each candidate of the i-th diagonal of the key. Since for each coset of $\mathcal{D}_I$ it is possible to construct approximately $3 \cdot 2^{15} \cdot (2^8 - 1)^2 \approx 2^{32.574}$ different sets, one needs approximately $2^{73.343} \cdot 2^{-32.573} = 2^{40.77}$ different initial cosets of $\mathcal{D}_I$ to discover the I-th diagonal of the key with probability higher than 95%, for a total cost of $2^{40.77} \cdot 2^{32} = 2^{72.77}$ chosen plaintexts. When one diagonal of the key is found and due to the computational cost of this first step, we propose to find the entire key (i.e. the other three diagonals) using a brute force attack.

7.4.2 Computational Cost

In order to implement the attack, the basic idea is to exploit Algorithm 5 for each possible guessed key, that is to count the number of sets $\mathcal{Z}$ for which the two ciphertexts of at least

Data: $2^{40.77}$ cosets of $\mathcal{D}_0$ (e.g. $\mathcal{D}_0 \oplus a_i$ for $a_i \in \mathcal{D}_0^{\perp}$) and corresponding ciphertexts after 6 rounds
Result: 4 bytes of the secret key - $(k_{0,0}, k_{1,1}, k_{2,2}, k_{3,3})$
Let $N[0, ..., 2^{32}-1]$ be an array initialized to zero; // $N[\varphi(k)]$ denotes the number of sets $\mathcal{Z}$ that satisfy the required property for the key k - $\varphi(\cdot)$ defined in (21)
/* *1st Step*: for each guessed key, count the number of sets $\mathcal{Z}$ with the required property */
for *each coset* $\mathcal{D}_0 \oplus a_i$ **do**
  re-order the coset $\mathcal{D}_0 \oplus a_i$ w.r.t. to the partial order $\preceq$ as in Def. 11 for each index J with $|J| = 3$; // the coset $\mathcal{D}_0 \oplus a_i$ is stored 5 times, one w.r.t. $<$ and four w.r.t. $\preceq$ for each J
  for *each* guessed key $\hat{k} \equiv (k_{0,0}, k_{1,1}, k_{2,2}, k_{3,3})$ **do**
    working as in Algorithm 5, count the number n of sets $\mathcal{Z}^{R_{\hat{k}}(\mathcal{D}_0 \oplus a_i)} \equiv \mathcal{Z}^{\mathcal{C}_0 \oplus b_i}$ for which the two ciphertexts of at least one couple belong to the same coset of $\mathcal{M}_J$ for a certain J with $|J| = 3$; // remember that the set $\mathcal{Z}$ is constructed only when a collision is found
    $N[\varphi(\hat{k})] \leftarrow N[\varphi(\hat{k})] + n$;
  end
end
/* *2nd Step*: look for the key with minimum number of sets $\mathcal{Z}$ */
$min \leftarrow N[0]$; // minimum number of sets
$\delta \leftarrow (0\text{x}00, 0\text{x}00, 0\text{x}00, 0\text{x}00)$;
for *each* $\hat{k}$ *from 1 to* $2^{32}-1$ **do**
  if $N[\varphi(\hat{k})] < min$ **then**
    $min \leftarrow N[\varphi(\hat{k})]$;
    $\delta \leftarrow \hat{k} \equiv (k_{0,0}, k_{1,1}, k_{2,2}, k_{3,3})$;
  end
end
return δ - *candidate of* $(k_{0,0}, k_{1,1}, k_{2,2}, k_{3,3})$

Algorithm 6: *6-round key-recovery attack on AES exploiting a 5-round secret-key distinguisher.* The goal of the attack is to find 4 bytes of the secret key. The remaining bytes (the entire key) are found by brute force.

one couple belong to the same coset of $\mathcal{M}_J$ for a certain J with $|J| = 3$ for each possible guessed key. Since the number of collision is higher for a wrong key than for the right one, it is possible to recover the right candidate of the key. An implementation of the attack is described by the pseudo-code given in Algorithm 6, where the bijective function $\varphi(\cdot) : \mathbb{F}_{2^8}^4 \equiv \mathbb{F}_{2^8} \times \mathbb{F}_{2^8} \times \mathbb{F}_{2^8} \times \mathbb{F}_{2^8} \rightarrow \mathbb{N}$ is defined as

$$\varphi(k_0, k_1, k_2, k_3) = k_0 + 256 \cdot k_1 + 256^2 \cdot k_2 + 256^3 \cdot k_3. \tag{21}$$

The data cost of the attack is of $2^{72.77}$ chosen plaintexts (distributed in $2^{40.77}$ cosets of $\mathcal{D}_I$ with $|I| = 1$), while the computational cost is approximately of $2^{112.7}$ table look-ups or approximately 2^{106} six-round encryptions, as we are going to show.

The algorithm is composed of two steps: (1) re-order the texts w.r.t. a partial order $\preceq$ and (2) construct and count the sets $\mathcal{Z}$ that satisfy the required property, when a collision is found. As for the other attacks of this paper, we remember that the way in which the texts are divided in sets $\mathcal{Z}$ depends on the guessed key (as for the attack proposed in Sect. 5.3), while the fact that two ciphertexts belong to the same coset of $\mathcal{M}_J$ is independently of the guessed key. In the following, we analyze in details the cost of the two steps, and we show that the total cost of this attack is well approximated by the cost to construct

the sets $\mathcal{Z}$ when a collision is found for each guessed key.

First of all, in order to find the ciphertexts that belong to the same coset of $\mathcal{M}_J$ (i.e. the collisions) in an efficient way, the idea is to re-order them w.r.t a partial order $\preceq$ (as defined in Def. 11) which is independent of the secret key. The cost of the re-ordering step - which is independent of the guessed value of the key - is well approximated by $4 \cdot 2^{40.77} \cdot 2^{32} \cdot \log 2^{32} \simeq 2^{79.77}$ table look-ups.

Secondly, similarly to what done for the 5-round distinguisher - Algorithm 5, the set $\mathcal{Z}$ is constructed only when a collision is found.Since each coset contains 2^{32} texts and a collision occurs with prob. 2^{-30}, we expect on average $2^{40.77} \cdot 2^{63} \cdot 2^{-30} \simeq 2^{73.77}$ collisions in total. As before, since we are interested only in the collisions pairs for which the plaintexts have exactly 2 equal generating variables (prob. $3 \cdot 2^{-15}$), the number of collisions for which one has to construct the set $\mathcal{Z}$ is approximately $2^{73.77} \cdot 3 \cdot 2^{-15} \simeq 2^{60.36}$ for each guessed key. For each one of these $2^{60.36}$ couples and for each one of the 2^{32} possible partial guessed key, the cost to construct the set $\mathcal{Z}$ is given by the following steps:

- given two ciphertexts that belong to the same coset of $\mathcal{M}_K$, one partially computes one round encryption of the corresponding plaintexts (if they have two equal generating variables), for a total cost of $2^{32} \cdot 2^{60.36} \cdot 4 \cdot 2 = 2^{95.36}$ S-Box look-ups;
- given these one round encryptions, one constructs all the 2^{17} couples given by a different combinations of the generating variables, and computes one round decryption, for a total cost of $2^{17} \cdot 2^{95.36} = 2^{112.36}$ S-Box look-ups;
- using look-ups tables (similar to before, for each coset of $\mathcal{D}_I$ one stores the (plaintexts, ciphertexts) pairs five times, one w.r.t. $\leq$ and four w.r.t. $\preceq$ for each index J with $|J| = 3$), one constructs the set $\mathcal{Z}$, for a cost of $2^{32} \cdot 2^{60.36} \cdot 2^{17} \cdot 4 = 2^{111.36}$ table look-ups.

The idea is to use the same strategy proposed in Sect. 7.2 in order to count the total number of sets $\mathcal{Z}$ with the required property for each possible guessed key. Thus, the total cost to find one diagonal of the key is well approximated by 2^{106} six-round encryptions (assuming 20 S-Box/table look-ups $\approx$ 1 round encryption), while the remaining three diagonals are found by brute force.

We emphasize that the implementation proposed in Algorithm 6 allows to minimize the memory costs. Indeed, note that each coset of $\mathcal{D}_0$ is used a single time, and that the user can work independently on each coset. Since all the (plaintexts, ciphertexts) pairs are stored in 5 different ways (i.e. one time w.r.t. to $<$ as defined in Def. 6 and for time w.r.t. $\preceq$ as defined in Def. 11 for each index J with $|J| = 3$), the memory cost is of $5 \cdot 2 \cdot 2^{32} \cdot 16 = 2^{39.32}$ bytes, or approximately $2^{35.4}$ texts.

As last thing, in App. E we explain why it is not possible to set up the key-recovery attack using cosets of $\mathcal{D}_I$ with $|I| = 2$ instead of $|I| = 1$. Without going into the details, this is due to the computational cost, since in such a case the attack requires only one coset of $\mathcal{D}_I$ with $|I| = 2$ (i.e. 2^{64} chosen plaintexts), but the total computational cost is approximately of 2^{165} table look-ups.

Acknowledgements. The author thanks Christian Rechberger for fruitful discussions and comments that helped to improve the quality of the paper. The work in this paper has been partially supported by the Austrian Science Fund (project P26494-N15).

References

[BJV04] Thomas Baignères, Pascal Junod, and Serge Vaudenay. How Far Can We Go Beyond Linear Cryptanalysis? In *Advances in Cryptology - ASIACRYPT 2004*, number 3329 in LNCS, pages 432–450, 2004.

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

[BKR11] Andrey Bogdanov, Dmitry Khovratovich, and Christian Rechberger. Biclique Cryptanalysis of the Full AES. In *Advances in Cryptology – ASIACRYPT 2011*, number 7073 in LNCS, pages 344–371, 2011.

[CKK+02] Jung Hee Cheon, MunJu Kim, Kwangjo Kim, Lee Jung-Yeun, and SungWoo Kang. Improved Impossible Differential Cryptanalysis of Rijndael and Crypton. In *Information Security and Cryptology — ICISC 2001*, volume 2288 of *LNCS*, pages 39–49, 2002.

[CMR05] Carlos Cid, Sean Murphy, and Matthew J. B. Robshaw. Small Scale Variants of the AES. In *Fast Software Encryption - FSE 2005*, volume 3557 of *LNCS*, pages 145–162, 2005.

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

[DF13] Patrick Derbez and Pierre-Alain Fouque. Exhausting demirci-selçuk meet-in-the-middle attacks against reduced-round AES. In *Fast Software Encryption - FSE 2013*, volume 8424 of *LNCS*, pages 541–560, 2013.

[DFJ13] Patrick Derbez, Pierre-Alain Fouque, and Jérémy Jean. Improved Key Recovery Attacks on Reduced-Round AES in the Single-Key Setting. In *Advances in Cryptology – EUROCRYPT 2013*, volume 7881 of *LNCS*, pages 371–387, 2013.

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

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

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

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

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

[Mat94] Mitsuru Matsui. Linear Cryptanalysis Method for DES Cipher. In *Advances in Cryptology — EUROCRYPT 1993*, number 765 in LNCS, pages 386–397, 1994.

[Sel08] Ali Aydın Selçuk. On Probability of Success in Linear andÂă DifferentialÂă Cryptanalysis. *Journal of Cryptology*, 21(1):131–147, 2008.

[SLG+16] Bing Sun, Meicheng Liu, Jian Guo, Longjiang Qu, and Vincent Rijmen. New Insights on AES-Like SPN Ciphers. In *Advances in Cryptology – CRYPTO 2016*, volume 9814 of *LNCS*, pages 605–624, 2016.

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

[TKKL15] Tyge Tiessen, Lars R. Knudsen, Stefan Kölbl, and Martin M. Lauridsen. Security of the AES with a Secret S-Box. In *Fast Software Encryption - FSE 2015*, volume 9054 of *LNCS*, pages 175–189, 2015.

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

[ZWF07] Wentao Zhang, Wenling Wu, and Dengguo Feng. New Results on Impossible Differential Cryptanalysis of Reduced AES. In *Information Security and Cryptology - ICISC 2007*, number 4817 in LNCS, pages 239–250, 2007.

A Proof - Probabilities of Sect. 3.2

In this section, we prove the probabilities given in Sect. 3.2.

Let $I, J \subseteq \{0,1,2,3\}$. We recall that

$$\mathcal{M}_I \cap \mathcal{M}_J = \mathcal{M}_{I \cap J}. \tag{22}$$

where $\mathcal{M}_I \cap \mathcal{M}_J = \emptyset$ if $I \cap J = \emptyset$. Moreover, referring to [GRR17b], we recall that the probability that a random text x belongs to $\mathcal{M}_I$ is well approximated by $Prob(x \in \mathcal{M}_I) = 2^{-32\cdot(4-|I|)}$, while given two random texts $x \neq y$

$$Prob(R(x) \oplus R(y) \in \mathcal{M}_J \,|\, x \oplus y \in \mathcal{M}_I) = 2^{-4\cdot|I|+|I|\cdot|J|}.$$

Proposition 5. *The probability $p_{|I|}$ that a random text x belongs to the subspace $\mathcal{M}_I$ for a certain $I \subseteq \{0,1,2,3\}$ with $|I| = l$ fixed is well approximated by*

$$p_{|I|} = Prob(\exists I \subseteq \{0,1,2,3\}\,|I| = l \;\; s.t. \;\; x \in \mathcal{M}_I) = (-1)^{|I|} \cdot \sum_{i=4-|I|}^{3} (-1)^i \cdot \binom{4}{i} \cdot 2^{-32\cdot i}.$$

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

$$Prob\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{k=1}^{n}\left((-1)^{k-1} \sum_{\substack{I \subset \{1,...,n\} \\ |I|=k}} Prob(A_I)\right),$$

where the last sum runs over all subsets I of the indexes $1, ..., n$ which contain exactly k elements[24] and

$$A_I := \bigcap_{i \in I} A_i$$

denotes the intersection of all those A_i with index in I.

Due to (22), it follows that for $|I| = 3$:

$$\begin{aligned} &Prob(\exists I \subseteq \{0,1,2,3\}\,|I| = 3 \text{ s.t. } x \in \mathcal{M}_I) = \\ &= \sum_{I \subseteq \{0,1,2,3\},\, |I|=3} Prob(x \in \mathcal{M}_I) - \sum_{I \subseteq \{0,1,2,3\},\, |I|=2} Prob(x \in \mathcal{M}_I) + \\ &+ \sum_{I \subseteq \{0,1,2,3\},\, |I|=1} Prob(x \in \mathcal{M}_I) = 4 \cdot 2^{-32} - 6 \cdot 2^{-64} + 4 \cdot 2^{-96}, \end{aligned}$$

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

while for $|I| = 2$

$$\begin{aligned} &Prob(\exists I \subseteq \{0,1,2,3\}\,|I| = 2 \text{ s.t. } x \oplus y \in \mathcal{M}_I) = \\ &= \sum_{I \subseteq \{0,1,2,3\},\, |I|=2} Prob(x \oplus y \in \mathcal{M}_I) - \sum_{I \subseteq \{0,1,2,3\},\, |I|=1} Prob(x \oplus y \in \mathcal{M}_I) = \\ &= 6 \cdot 2^{-64} - 4 \cdot 2^{-96}, \end{aligned}$$

and finally for $|I| = 1$

$$\begin{aligned} &Prob(\exists I \subseteq \{0,1,2,3\}\,|I| = 1 \text{ s.t. } x \oplus y \in \mathcal{M}_I) = \\ &= \sum_{I \subseteq \{0,1,2,3\},\, |I|=1} Prob(x \oplus y \in \mathcal{M}_I) = 4 \cdot 2^{-96}, \end{aligned}$$

that is the thesis. □

Proposition 6. *Let x, y be two random elements. Assume that there exists $I \subseteq \{0,1,2,3\}$ such that $x \oplus y \in \mathcal{M}_I$. The probability that $\exists J \subseteq \{0,1,2,3\}$ with $|J| = l$ fixed such that $R(x) \oplus R(y) \in \mathcal{M}_J$ is well approximated by*

$$\begin{aligned} p_{|J|,|I|} &\equiv Prob(\exists J\,|J| = l \;\; s.t. \;\; R(x) \oplus R(y) \in \mathcal{M}_J \,|\, x \oplus y \in \mathcal{M}_I) = \\ &= (-1)^{|J|} \cdot \sum_{i=4-|J|}^{3} (-1)^i \cdot \binom{4}{i} \cdot 2^{8 \cdot i \cdot |I| \cdot (|J|-4)}. \end{aligned}$$

Proof. As before, for $|J| = 3$:

$$\begin{aligned} &Prob(\exists J\,|J| = 3 \text{ s.t. } R(x) \oplus R(y) \in \mathcal{M}_J \,|\, x \oplus y \in \mathcal{M}_I) = \\ &= \sum_{J \subseteq \{0,1,2,3\},\, |J|=3} Prob(R(x) \oplus R(y) \in \mathcal{M}_J \,|\, x \oplus y \in \mathcal{M}_I) + \\ &\quad - \sum_{J \subseteq \{0,1,2,3\},\, |J|=2} Prob(R(x) \oplus R(y) \in \mathcal{M}_J \,|\, x \oplus y \in \mathcal{M}_I) + \\ &\quad + \sum_{J \subseteq \{0,1,2,3\},\, |J|=1} Prob(R(x) \oplus R(y) \in \mathcal{M}_J \,|\, x \oplus y \in \mathcal{M}_I) = \\ &= 4 \cdot 2^{8 \cdot |I| \cdot (|J|-4)} - 6 \cdot 2^{16 \cdot |I| \cdot (|J|-4)} + 4 \cdot 2^{24 \cdot |I| \cdot (|J|-4)}. \end{aligned}$$

By simple computation, it is possible to obtain similar results for $|J| = 2$ and $|J| = 1$, that is the thesis. □

Proposition 7. *Let x, y be two random elements such that $x \oplus y \notin \mathcal{M}_I$ for each $I \subseteq \{0,1,2,3\}$. Then, the probability that $\exists J \subseteq \{0,1,2,3\}$ for $|J| = l$ fixed such that $R(x) \oplus R(y) \in \mathcal{M}_J$ is well approximated by*

$$\hat{p}_{|J|,3} \equiv Prob(\exists J \;\; s.t. \;\; R(x) \oplus R(y) \in \mathcal{M}_J \,|\, x \oplus y \notin \mathcal{M}_I \,\forall I) = \frac{p_{|J|} - p_{|J|,3} \cdot p_3}{1 - p_3}.$$

Proof. Let A and B be two events, and let $A^{\perp}$ such that $A \cup A^{\perp}$ is equal to the sample space. By definition

$$Prob(B) = Prob(B \,|\, A) \cdot Prob(A) + Prob(B \,|\, A^{\perp}) \cdot Prob(A^{\perp}).$$

Thus

$$\begin{aligned} &p_{|J|} \equiv Prob(\exists J \text{ s.t. } R(x) \oplus R(y) \in \mathcal{M}_J) = \\ &= Prob(\exists J \text{ s.t. } R(x) \oplus R(y) \in \mathcal{M}_J \,|\, x \oplus y \notin \mathcal{M}_I \,\forall I) \cdot Prob(x \oplus y \notin \mathcal{M}_I \,\forall I) + \\ &+ Prob(\exists J \text{ s.t. } R(x) \oplus R(y) \in \mathcal{M}_J \,|\, \exists I \text{ s.t. } x \oplus y \in \mathcal{M}_I) \cdot Prob(\exists I \text{ s.t. } x \oplus y \in \mathcal{M}_I). \end{aligned}$$

Note that[25]

$$Prob(\exists I \text{ s.t. } x \oplus y \in \mathcal{M}_I) = Prob\Bigg(x \oplus y \in \bigcup_{\forall I \subseteq \{0,1,2,3\}} \mathcal{M}_I\Bigg) =$$

$$= Prob\Bigg(x \oplus y \in \bigcup_{I \subseteq \{0,1,2,3\},\, |I|=3} \mathcal{M}_I\Bigg) \equiv p_3.$$

It follows that

$$p_{|J|} = p_{|J|,3} \cdot p_3 + \hat{p}_{|J|,3} \cdot (1 - p_3),$$

that is the thesis. □

Proposition 8. *Let x and y such that $x \oplus y \notin \mathcal{M}_I$ for each $I \subseteq \{0,1,2,3\}$. Then, the probability that $\exists J \subseteq \{0,1,2,3\}$ with $|J| = l$ fixed and $|I|+|J| \leq 4$ such that $R^2(x) \oplus R^2(y) \in \mathcal{M}_J$ is well approximated by*

$$\tilde{p}_{|J|,3} \equiv Prob(\exists J \text{ s.t. } R^2(x) \oplus R^2(y) \in \mathcal{M}_J \,|\, x \oplus y \notin \mathcal{M}_I) = \frac{p_{|J|}}{1 - p_3}.$$

Proof. Remember that

$$Prob(\exists J \text{ s.t. } R^2(x) \oplus R^2(y) \in \mathcal{M}_J \,|\, \exists I \text{ s.t. } x \oplus y \notin \mathcal{M}_I) = 0.$$

Since

$$Prob(\exists J \text{ s.t. } R^2(x) \oplus R^2(y) \in \mathcal{M}_J) =$$
$$= Prob(\exists J \text{ s.t. } R^2(x) \oplus R^2(y) \in \mathcal{M}_J \,|\, x \oplus y \notin \mathcal{M}_I \,\forall I) \cdot Prob(x \oplus y \notin \mathcal{M}_I \,\forall I) +$$
$$+ Prob(\exists J \text{ s.t. } R^2(x) \oplus R^2(y) \in \mathcal{M}_J \,|\, \exists I \text{ s.t. } x \oplus y \in \mathcal{M}_I) \cdot Prob(\exists I \text{ s.t. } x \oplus y \in \mathcal{M}_I)$$

and using the same argumentation as before, it follows that

$$p_{|J|} = \tilde{p}_{|J|,3} \cdot (1 - p_3),$$

that is the thesis. □

As last thing, we show that given texts in the same cosets of $\mathcal{C}_I$ or $\mathcal{M}_I$ for $I \subseteq \{0,1,2,3\}$, the number of couples of texts with n equal generating variable(s) for $0 \leq n \leq 3$ is given by

$$\binom{4}{n} \cdot 2^{32 \cdot |I| - 1} \cdot (2^{8 \cdot |I|} - 1)^{4-n}.$$

W.l.o.g. consider for simplicity the case $|I| = 1$. First of all, note that there are $\binom{4}{n}$ different combinations of n variables. If $n \geq 1$, the n variables that must be equal for the two texts of the couple can take $(2^8)^n$ different values. For each one of the remaining $4 - n$ variables, the variables must be different for the two texts of each couple. Thus, these $4 - n$ variables can take exactly $\left[(2^8)^{4-n} \cdot (2^8 - 1)^{4-n}\right]/2$ different values. The result follows immediately. In particular, for $|I| = 1$ there are:

- $2^{63} \cdot (2^8 - 1)^4$ couples for which the two texts have different generating variables;
- $2^{33} \cdot (2^8 - 1)^3$ couples for which the two texts have one equal generating variable;
- $3 \cdot 2^{32} \cdot (2^8 - 1)^2$ couples for which the two texts have two equal generating variables;
- $2^{33} \cdot (2^8 - 1)$ couples for which the two texts have three equal generating variables.

The other cases are analogous. Note that the total number of all the possible couples is $2^{31} \cdot (2^{32} - 1)$.

[25] If $x \oplus y \in \mathcal{M}_I$ for $|I| < 3$, then $\exists J$ with $|J| = 3$ and $I \subseteq J$ such that $x \oplus y \in \mathcal{M}_J$.

A.1 Discussion about the Given Approximations

In Sect. 3.2, we list some useful probabilities largely used in the following. As we have already said, *all those probabilities are not the exact ones, but "good enough" approximations useful for the target of the paper.* Here we give some more details about this statement.

As first thing, consider the following simple example. Consider the probability that a pair of texts t^1 and t^2 belongs in the same coset of $\mathcal{M}_I$. This probability is usually approximated by $Prob(x \in \mathcal{M}_I) = 2^{-32\cdot(4-|I|)}$. On the other hand, in order to set up a (truncated) differential attack, one is interested to the case $t^1 \neq t^2$ (equivalently, $x \neq 0$). Thus, the "correct" probability is

$$Prob(x \in \mathcal{M}_I \,|\, x \neq 0) = \frac{2^{32\cdot|I|}-1}{2^{128}-1} = 2^{-32\cdot(4-|I|)} - 2^{-128} + 2^{-128-32\cdot(4-|I|)} + ...$$

Another interesting example regards the 4-round AES impossible differential trail. Consider plaintexts in the same coset of $\mathcal{D}_I$, and the corresponding ciphertexts after 4-round. It is well known that

$$Prob(R^4(x) \oplus R^4(y) \in \mathcal{M}_J \,|\, x \oplus y \in \mathcal{D}_I) = 0 \qquad \forall J \text{ s.t. } |I| + |J| \leq 4.$$

On the other hand, we can compute this probability using the probabilities given in Sect. 3.2. Assume for simplicity I fixed with $|I| = 1$. By Theorem 1, each coset of $\mathcal{D}_I$ is mapped into a coset of $\mathcal{M}_I$ after 2-round. Thus, the probability that

$$Prob(R(x) \oplus R(y) \in \mathcal{M}_J \,|\, x \oplus y \in \mathcal{M}_I) = (-1)^{|J|} \cdot \sum_{i=4-|J|}^{3} (-1)^i \cdot \binom{4}{i} \cdot 2^{8\cdot i\cdot(|J|-4)}.$$

Thus

$$\begin{aligned}
&Prob(R^4(x) \oplus R^4(y) \in \mathcal{M}_J \,|\, x \oplus y \in \mathcal{D}_I) = \\
=&\sum_K Prob(R^4(x) \oplus R^4(y) \in \mathcal{M}_J \,|\, R^3(x) \oplus R^3(y) \in \mathcal{M}_K \text{ and } x \oplus y \in \mathcal{D}_I) \times \\
&\times Prob(R^3(x) \oplus R^3(y) \in \mathcal{M}_K \,|\, x \oplus y \in \mathcal{D}_I) + \\
+&Prob(R^4(x) \oplus R^4(y) \in \mathcal{M}_J \,|\, R^3(x) \oplus R^3(y) \notin \mathcal{M}_K \forall K \text{ and } x \oplus y \in \mathcal{D}_I) \times \\
&\times Prob(R^3(x) \oplus R^3(y) \notin \mathcal{M}_K \forall K \,|\, x \oplus y \in \mathcal{D}_I).
\end{aligned}$$

If one approximates the probability $Prob(R^4(x) \oplus R^4(y) \in \mathcal{M}_J \,|\, R^3(x) \oplus R^3(y) \in \mathcal{M}_K$ *and* $x \oplus y \in \mathcal{D}_I)$ *with* $Prob(R^4(x) \oplus R^4(y) \in \mathcal{M}_J \,|\, R^3(x) \oplus R^3(y) \in \mathcal{M}_K)$, by simple computation *it follows that*

$$Prob(R^4(x) \oplus R^4(y) \in \mathcal{M}_J \,|\, x \oplus y \in \mathcal{D}_I) \approx 2^{-28} + 2^{-30} + ...$$

which is obviously wrong. The error arises by the fact that the probability

$$\begin{aligned}
&Prob(R^4(x) \oplus R^4(y) \in \mathcal{M}_J \,|\, R^3(x) \oplus R^3(y) \in \mathcal{M}_K \text{ and } x \oplus y \in \mathcal{D}_I) = \\
&= Prob(R^4(x) \oplus R^4(y) \in \mathcal{M}_J \,|\, R^3(x) \oplus R^3(y) \in \mathcal{M}_K \text{ and } R^2(x) \oplus R^2(y) \in \mathcal{D}_I) = 0
\end{aligned}$$

for all $|I| + |J| \leq 4$. In other words, *the assumption behind the probabilities given in Sect. 3.2 is that the elements x and y are uniform distributed, or (at least) very close to be uniform distributed - as for the events considered in this paper to set up distinguishers and key-recovery attacks on 5- and 6-round AES.*

B A New 4-round Secret-Key Distinguisher for AES - Details

In this section, we give all the details of the 4-round Secret-Key Distinguisher for AES presented in Sect. 5 about the computational cost. We refer to Sect. 5 for all the details about the distinguisher.

Computational Complexity

Given 2^{16} chosen plaintexts in the same coset of $\mathcal{C}_0 \cap \mathcal{D}_{0,1} \oplus a$ and the corresponding ciphertexts, a first possibility is to construct all the possible pairs, to divide them in sets $\mathcal{S}^{\mathcal{C}_0 \cap \mathcal{D}_{0,1} \oplus a}$ and to check for each set if the above property is satisfied (or not). First of all, given a set $\mathcal{S}^{\mathcal{C}_0 \cap \mathcal{D}_{0,1} \oplus a} = \{[(p^1, c^1), (p^2, c^2)]; [(\hat{p}^1, \hat{c}^1), (\hat{p}^2, \hat{c}^2)]\}$, the cost to check if the above property is satisfied (or not) is equal to 1 XOR and 1 MixColumns operation[26], which is negligible with respect to the total cost. For this reason, we focus on the cost to construct the sets $\mathcal{S}^{\mathcal{C}_0 \cap \mathcal{D}_{0,1} \oplus a}$. Using the previous strategy, since the number of pairs is approximately 2^{31} for each coset, the cost is of approximately $2 \cdot 2^{31} = 2^{32}$ table look-ups.

In order to reduce the computational cost, a possibility is to re-order the ciphertexts with respect to a partial order $\preceq$ as defined in Def. 11 (see also [GRR17a]). Note that $\preceq$ depends on an index J. Using a merge-sort algorithm, the cost to re-order n texts is of $O(n \cdot \log n)$ table look-ups. When the ciphertexts have been re-ordered, it is no more necessary to construct all the possible pairs. Indeed, to verifier the property, it is sufficient to work only on consecutive texts with respect to $\preceq$.

In more details, first one stores all the plaintext/ciphertext pairs twice, (1) once in which the plaintexts are ordered with respect to the partial order $\leq$ defined in Def. 6 and (2) once in which the ciphertexts are ordered with respect to the partial order $\preceq$ defined in Def. 11. Then, working on this second set, on focuses only on consecutive ciphertexts c^i and c^{i+1} for each i, and checks if $c^i \oplus c^{i+1} \in \mathcal{M}_J$ or not. Assume that $c^i \oplus c^{i+1} \in \mathcal{M}_J$ for a certain J fixed previously. The idea is to take the corresponding plaintexts $p^i \equiv (x^1, y^1)$ and $p^{i+1} \equiv (x^2, y^2)$, to construct the corresponding set $\mathcal{S}^{\mathcal{C}_0 \cap \mathcal{D}_{0,1} \oplus a}_{p^1, p^2}$ and to check if the ciphertexts $\hat{c}^1$ and $\hat{c}^2$ of the corresponding plaintexts $\hat{p}^1 \equiv (x^1, y^2)$ and $\hat{p}^2 \equiv (x^2, y^1)$ satisfy the condition $\hat{c}^1 \oplus \hat{c}^2 \in \mathcal{M}_J$ for the same J. If not, by previous observations one can simply deduce that this is a random permutation. Note that if there are r consecutive ciphertexts $c^i, c^{i+1}, ..., c^{i+r-1}$ such that $c^j \oplus c^l \in \mathcal{M}_J$ for $i \leq j, l < r$, then one has to repeat the above procedure for all these $\binom{r}{2} = r \cdot (r-1)/2$ possible pairs[27].

To optimize the computational cost, note that the plaintexts $\hat{p}^1$ and $\hat{p}^2$ are respectively in positions $x^1 + 2^8 \cdot y^2$ and $x^2 + 2^8 \cdot y^1$ in the first set of plaintext/ciphertext pairs (i.e. in the set where the plaintexts are ordered with respect to the partial order $\leq$). Thus, the cost to get these two elements is only of 2 table look-ups. Moreover, we emphasize that it is sufficient to work only on (consecutive) ciphertexts c^i and c^j such that $c^i \oplus c^j \in \mathcal{M}_J$. Indeed, consider the case in which the two ciphertexts c^i and c^j don't belong to the same coset of $\mathcal{M}_J$, i.e. $c^i \oplus c^j \notin \mathcal{M}_J$. If the corresponding ciphertexts $\hat{c}^1$ and $\hat{c}^2$ - defined as before - don't belong to the same coset of $\mathcal{M}_J$, then the property is (obviously) verified. Instead if $\hat{c}^1 \oplus \hat{c}^2 \in \mathcal{M}_J$, then this case is surely analyzed. The pseudo-code of such strategy can be found in Algorithm 1.

Using this procedure, the memory cost is well approximated by $4 \cdot 2^{17} \cdot 16 = 2^{23}$ bytes - the same plaintext/ciphertext pairs in two different ways. The cost to order the ciphertexts for each possible J with $|J| = 3$ and for each one of the two cosets is approximately of $2 \cdot 4 \cdot 2^{16} \cdot \log 2^{16} \simeq 2^{23}$ table look-ups, while the cost to construct all the possible pairs of consecutive ciphertexts is of $2 \cdot 4 \cdot 2^{16} = 2^{19}$ table look-ups. Since the probability that a

[26] Given x, y, then $x \oplus y \in \mathcal{M}_I$ if and only if $MC^{-1}(x \oplus y) \in \mathcal{ID}_I$ for each I.

[27] Since $\mathcal{M}_J$ is a subspace, given a, b, c such that $a \oplus b \in \mathcal{M}_J$ and $b \oplus c \in \mathcal{M}_J$, then $b \oplus c \in \mathcal{M}_J$.

pair of ciphertexts belong to the same coset of $\mathcal{D}_J$ for $|J| = 3$ is 2^{-30} and since each coset contains approximately 2^{31} different pairs, then one has to do on average $2 \cdot 4 \cdot 2^{-30} \cdot 2^{31} = 2^4$ table look-ups in the plaintext/ciphertext pairs ordered with respect to the plaintexts. Thus, the total cost of this distinguisher is well approximated by $2^{23} + 2^{19} + 16 \simeq 2^{23.09}$ table look-ups, or approximately $2^{16.75}$ four-round encryptions (using the approximation 20 table look-ups $\approx$ 1 round of encryption).

C Details of the Key-Recovery Attack on 5-round AES of Sect. 5.3

As we have seen in Sect. 3.1, a coset of a diagonal space is always mapped into a coset of a column space. Thus, a natural question is if it is possible to extend the 4-round distinguisher proposed in Sect. 5 to a 5-round one simply considering plaintexts in the same coset of a diagonal space $\mathcal{D}_I$ instead that in the same coset of a column space $\mathcal{C}_I$. As we are going to show, a problem arises that doesn't allow to implement the distinguisher, but a new key-recovery attack on 5-round of AES can be set up.

W.l.o.g. consider a coset of a subspace $\mathcal{C}_0$ (analogous for others $\mathcal{C}_I$ with $|I| = 1$). To set up the distinguisher on 4-round AES described in Sect. 5, one constructs all the sets $\mathcal{S}^{\mathcal{C}_0 \oplus a}$, and exploits the fact that for each given set only two events can happen in the AES case: for all the couples the two ciphertexts belong or not to the same coset of $\mathcal{M}_J$. Remember that given a couple of two pairs (p^1, c^1) and (p^2, c^2) in $\mathcal{S}^{\mathcal{C}_0 \oplus a}$ with $p^1 \equiv (x^1, y^1, z^1, w^1)$ and $p^2 \equiv (x^2, y^2, z^2, w^2)$, then the other seven couples are composed by the other possible combinations of these variables.

Consider instead two plaintexts in the same coset of $\mathcal{D}_0$ (i.e. $\mathcal{D}_0 \oplus a$ for $a \in \mathcal{D}_0^{\perp}$), that is p^1 and p^2 such that $p^i \equiv (x^i, y^i, z^i, w^i)$ for $i = 1, 2$ or equivalently:

$$p^i = x^i \cdot e_{0,0} \oplus y^i \cdot e_{1,1} \oplus z^i \cdot e_{2,2} \oplus w^i \cdot e_{3,3} \oplus a.$$

By Theorem 1, there exists $b \in \mathcal{C}_0^{\perp}$ such that for $i = 1, 2$

$$R(p^i) = \begin{bmatrix} \hat{x}^i & 0 & 0 & 0 \\ \hat{y}^i & 0 & 0 & 0 \\ \hat{z}^i & 0 & 0 & 0 \\ \hat{w}^i & 0 & 0 & 0 \end{bmatrix} \oplus b \equiv M^{MC} \cdot \begin{bmatrix} \text{S-Box}(x^i \oplus k_{0,0}) & 0 & 0 & 0 \\ \text{S-Box}(y^i \oplus k_{1,1}) & 0 & 0 & 0 \\ \text{S-Box}(z^i \oplus k_{2,2}) & 0 & 0 & 0 \\ \text{S-Box}(w^i \oplus k_{3,3}) & 0 & 0 & 0 \end{bmatrix} \oplus b,$$

i.e. $R(p^i) \equiv (\hat{x}^i, \hat{y}^i, \hat{z}^i, \hat{w}^i) \equiv \hat{x}^i \cdot e_{0,0} \oplus \hat{y}^i \cdot e_{1,0} \oplus \hat{z}^i \cdot e_{2,0} \oplus \hat{w}^i \cdot e_{3,0} \oplus b$. In order to use the previous distinguisher, one has to construct the set $\mathcal{S}^{\mathcal{C}_0 \oplus b}_{R(p^1), R(p^2)}$ defined as before[28]. As an example, the couple $(\hat{p}^1, \hat{c}^1)$ and $(\hat{p}^2, \hat{c}^2)$ such that $\hat{p}^1$ and $\hat{p}^2$ satisfy

$$R(\hat{p}^i) = \hat{x}^{i+1} \cdot e_{0,0} \oplus \hat{y}^i \cdot e_{1,0} \oplus \hat{z}^i \cdot e_{2,0} \oplus \hat{w}^i \cdot e_{3,0} \oplus b,$$

where the index $i + 1$ is taken modulo 2, belongs to such set (analogous for the other cases/combinations). However, a problem arises: since the key k is secret and the S-Box is non-linear, there is no way to find such $\hat{p}^1$ and $\hat{p}^2$ and to construct the set $\mathcal{S}^{\mathcal{C}_0 \oplus b}_{R(p^1), R(p^2)}$ if the plaintexts are in a coset of a diagonal space $\mathcal{D}_I$ instead of a column space $\mathcal{C}_I$. It follows that it is not possible to extend the 4-round distinguisher of Sect. 5 simply considering plaintexts in a coset of $\mathcal{D}_I$ instead of $\mathcal{C}_I$.

On the other hand, this allows to set up a new key-recovery attack on 5 rounds of AES. Given plaintexts in the same coset of $\mathcal{D}_I$, consider two (plaintexts, ciphertexts) pairs (p^1, c^1) and (p^2, c^2) such that the two ciphertexts belong to the same coset of $\mathcal{M}_J$ for J with $|J| = 3$ after five-round. Fixed $I \in \{0, 1, 2, 3\}$, the idea of the attack is to

[28] We abuse the notation $\mathcal{S}^{R(\mathcal{D}_0 \oplus a)}_{R(p^1), R(p^2)}$ to denote the set $\mathcal{S}^{\mathcal{C}_0 \oplus b}_{R(p^1), R(p^2)}$.

guess 4 bytes of the I-th diagonal of the secret key k, that is $k_{i,i+I}$ for each $i = 0, 1, 2, 3$, (partially) compute $R_k(p^1)$ and $R_k(p^2)$ and construct the set $\mathcal{S}^{\mathcal{C}_0 \oplus b}_{R(p^1),R(p^2)}$. Due to the previous 4-round distinguisher, such set $\mathcal{S}^{\mathcal{C}_0 \oplus b}_{R(p^1),R(p^2)}$ has the property that for all the couples $(\hat{p}^1, \hat{c}^1)$ and $(\hat{p}^2, \hat{c}^2)$, the two ciphertexts belong to the same coset of $\mathcal{M}_J$ for the previous J. If this property is not satisfied, then one simply deduces that the key is wrong. If more than one candidate of the key passes the test, one can simply repeat it with other couples of plaintexts/ciphertexts until all the wrong candidates are discarded.

Data and Computational Costs. Each coset of $\mathcal{D}_I$ with $|I| = 1$ is composed of 2^{32} texts, thus on average $2^{63} \cdot 2^{-32} = 2^{31}$ different pairs of ciphertexts belong to the same coset of $\mathcal{M}_J$ for a fixed J with $|J| = 3$. As we have just seen, it is sufficient to find one collision in order to implement the attack and to find the key. In order to find it, the best strategy is to re-ordered the ciphertexts with respect to the partial order $\preceq$ and then to work on consecutive elements. For each initial coset of $\mathcal{D}_I$ and for a fixed J, the cost to re-order the ciphertexts with respect to the partial order $\preceq$ (for $\mathcal{M}_J$ with J fixed - $|J| = 3$) and to find a collision is approximately of $2^{32} \cdot (\log 2^{32} + 1) = 2^{37}$ table look-ups. When such a collision is found, one has to guess 4 bytes of the key and to consider (at least) two different couples in the set $\mathcal{S}^{\mathcal{C}_0 \oplus b}_{R(p^1),R(p^2)}$. Since the cost to get two different couples in the set $\mathcal{S}^{\mathcal{C}_0 \oplus b}_{R(p^1),R(p^2)}$ is well approximated by 4 table look-ups (as for the 4-round distinguisher described in Sect. 5, the idea is to store the (plaintexts, ciphertexts) pairs twice, once w.r.t. the partial order $\leq$ and once w.r.t. the partial order $\preceq$), the cost of this step is of $2^{32} \cdot 2 \cdot 4 = 2^{35}$ S-Box and of $2^{32} \cdot 4 = 2^{34}$ table look-ups.

Thus, the cost to find one diagonal of the key is well approximated by 2^{35} S-Box look-ups and $2^{37.17}$ table look-ups, that is approximately $2^{30.95}$ five-round encryptions. The idea is to repeat this operation for three different diagonals, and to find the last one by brute force. As a result, the total computational cost is of $2^{32} + 3 \cdot 2^{30.95} = 2^{33.28}$ five-round encryptions, while the data cost is of $3 \cdot 2^{32} = 2^{33.6}$ chosen plaintexts.

Only for completeness, we highlight that the same attack works also in the decryption/reverse direction, using chosen ciphertexts instead of plaintexts.

C.1 Practical Verification

Using a C/C++ implementation[29], we have practically verified the attack just described on the small-scale AES presented in [CMR05]. As we have already said, while in "real" AES, each word is composed of 8 bits, in this variant each word is composed of 4 bits. We refer to [CMR05] for a complete description of this small-scale AES, and we limit ourselves to describe the results of our 5-round key-recovery in this case. Since the attack and the distinguisher are independent of the fact that each word of AES is composed of 4 or 8 bits, our verification on the small scale variant of AES is strong evidence for it to hold for the real AES.

Practical Results. We verified the key-recovery attack on small-scale AES. For the following, we limit to report the result for a single diagonal of the key. First of all, a single coset of a diagonal space $\mathcal{D}_i$ is largely sufficient to find one diagonal of the key. More in details, given two (plaintexts, ciphertexts) pairs (p^1, c^1) and (p^2, c^2), then other two different couples in the set $\mathcal{S}^{\mathcal{C}_0 \oplus b}_{R(p^1),R(p^2)}$ are sufficient to discard all the wrong candidates of the diagonal of the key, as predicted.

About the computational cost, using the same argumentation of before, the theoretical cost for the small-scale AES case is well approximated by $4 \cdot 2^{16} \cdot (\log 2^{16} + 1) + 2^{16} \cdot 4 = 2^{21}$

[29] The source codes of the distinguishers/attacks are available at https://github.com/Krypto-iaik/Distinguisher_5RoundAES

table look-ups and $2^{16} \cdot 4 \cdot 3 = 2^{19.6}$ S-Box look-ups, for a total of $2^{19.6} + 2^{21} = 2^{21.5}$ table look-ups (assuming that the cost of 1 S-Box look-up is approximately equal to the cost of 1 table look-up). The average practical computational cost is of $2^{21.5}$ table look-ups, that is approximately the same of the theoretical one.

D Details of the 5-round AES Distinguisher of Sect. 7

In this section, we list the probabilities of the 5-round secret-key distinguisher proposed in Sect. 7 for the cases of sets $\mathcal{S}$ and $\mathcal{T}$, and the details about the computational cost for the case of sets $\mathcal{Z}$. Since the way in which these probabilities are computed is the same given in Sect. 7, we refer to that section for all the details and we limit here to report the corresponding probabilities and the results of our practical implementations.

D.1 Case: $\mathcal{S}$ set

By definition Def. 8, a set $\mathcal{S}$ is composed of 8 (plaintexts, ciphertexts) couples such that the generating variables of the two plaintexts of each couple are all different. *Given a set $\mathcal{S}$, what is the probability that two ciphertexts of at least one couple belong to the same coset of $\mathcal{M}_J$?*

Using the same calculation given in Sect. 7, it follows that this probability in the AES case is well approximated by

$$\begin{aligned} p_{AES} =& \left[1 - Prob(\overline{\mathcal{E}_1^5} \wedge \overline{\mathcal{E}_2^5} \wedge ... \wedge \overline{\mathcal{E}_8^5} \mid \mathcal{E}_i^4)\right] \cdot Prob(\mathcal{E}_i^4) + \\ &+ \left[1 - Prob(\overline{\mathcal{E}_1^5} \wedge \overline{\mathcal{E}_2^5} \wedge ... \wedge \overline{\mathcal{E}_8^5} \mid \overline{\mathcal{E}_i^4})\right] \cdot Prob(\overline{\mathcal{E}_i^4}) = \\ =& (1 - p_3) \cdot \left[1 - \left(1 - \frac{p_3 \cdot (1 - p_{3,3})}{1 - p_3}\right)^8\right] + p_3 \cdot \left[1 - \left(1 - p_{3,3}\right)^8\right] = \\ =& 2^{-27} - 31 \cdot 2^{-60} - \underbrace{3\,641\,245 \cdot 2^{-91}}_{\approx 3.475 \cdot 2^{-71}} + \underbrace{20\,628\,528\,753 \cdot 2^{-124}}_{\approx 2.4 \cdot 2^{-91}} + ... \end{aligned}$$

for a certain $i \in \{1, ..., 8\}$. For a random permutation, the same event occurs with (approximately) probability

$$\begin{aligned} p_{rand} =& 1 - \left(1 - p_3^8\right) = 1 - \left[1 - \left(2^{-30} - 3 \cdot 2^{-63} + 2^{-94}\right)\right]^8 = \\ =& 2^{-27} - 31 \cdot 2^{-60} + 155 \cdot 2^{-91} + ... \end{aligned}$$

Note that $|p_{AES} - p_{rand}| \simeq 2^{-69.204}$ and $p_{AES} \simeq p_{rand} \simeq 2^{-27}$. Using (20), it follows that n must satisfy $n > 2^{113.84}$ for a prob. of success higher than 95%.

What is the data complexity? We remember that a single coset of $\mathcal{C}_I$ for $|I| = 1$ contains approximately $2^{31} \cdot (2^8 - 1)^4 \cdot 2^{-3} \simeq 2^{59.978}$ different sets $\mathcal{S}$ of eight couples, while a single coset of $\mathcal{C}_I$ for $|I| = 2$ contains approximately $2^{63} \cdot (2^{16} - 1)^4 \cdot 2^{-3} \simeq 2^{124}$ different sets $\mathcal{S}$. Thus, using a single coset of $\mathcal{C}_I$ for $|I| = 1$, one needs approximately $2^{113.84} \cdot 2^{-60} \simeq 2^{53.84}$ different initial cosets of $\mathcal{C}_I$, that is approximately $2^{85.84}$ chosen plaintexts. Using instead an initial coset of $\mathcal{C}_I$ with $|I| = 2$, it is possible to construct approximately 2^{124} different sets $\mathcal{S}$ of eight couples, which is more than one needs to set up the distinguisher. It follows that 2^{59} chosen plaintexts in the same coset of $\mathcal{C}_I$ with $|I| = 2$ are sufficient to implement the distinguisher.

What is the computational cost? The cost to re-order the set is $4 \cdot 2^{53.84} \cdot 2^{32} \cdot \log 2^{32} \simeq 2^{92.84}$ table look-ups for the case of coset $\mathcal{C}_I$ with $|I| = 1$ and $4 \cdot 2^{59} \cdot \log 2^{59} \simeq 2^{66.88}$ for the case $|I| = 2$. The number of collisions is approximately $2^{-30} \cdot 2^{53.84} \cdot 2^{63} \simeq 2^{86.84}$ for the case $|I| = 1$ and $2^{-30} \cdot 2^{117} \simeq 2^{87}$ for the case $|I| = 2$. Since the cost to construct the set $\mathcal{S}$ is of $2 \cdot 8 = 2^4$ table look-ups, the total cost is well approximated by $2^{86.84} \cdot 2^4 + 2^{92.84} = 2^{93.16}$

table look-ups, that is approximately $2^{86.52}$ five-round encryptions for the case $|I| = 1$. In a similar way, the cost for the case $|I| = 2$ is given by $2^{87} \cdot 2^4 + 2^{66.88} = 2^{91}$ table look-ups, that is approximately $2^{84.36}$ five-round encryptions

D.1.1 Key-Recovery Attack on 6-round AES

For completeness, we also give the probability $p_{AES}^{WrongKey}$ that - *when the guessed key is wrong* - for a set $\mathcal{S}^{R(\mathcal{D}_I \oplus a)}$ two texts of at least one couple belong to the same coset of $\mathcal{M}_K$ for a certain $|K| = 3$ after six rounds is approximately equal to

$$p_{AES}^{WrongKey} = \sum_{n=0}^{8} \binom{8}{n} \cdot p_3^n \cdot (1-p_3)^{8-n} \cdot \left[1 - \left(1 - p_{3,3}\right)^n \cdot \left(1 - \frac{p_3 \cdot (1-p_{3,3})}{1-p_3}\right)^{8-n}\right],$$

which is well approximated by

$$p_{AES}^{WrongKey} = 2^{-27} - 31 \cdot 2^{-60} - 3\,989 \cdot 2^{-91} + ...$$

Note that this probability is similar but not equal to the one of the random case (which is $p_{rand} = 2^{-27} - 31 \cdot 2^{-60} + 155 \cdot 2^{-91} + ...$), while we remember that the probability for "AES with the right key" is $p_{AES} = 2^{-27} - 31 \cdot 2^{-60} - 3\,641\,245 \cdot 2^{-91} + ...$, where the difference between these two probabilities is approximately $|p_{AES}^{WrongKey} - p_{AES}| \simeq 2^{-69.2053}$.

We refer to Sect. 7.4 for all the details about the attack on 6-round AES.

D.2 Case: $\mathcal{T}$ set

As first thing, we recall the definition of set $\mathcal{T}$.

Definition 10. Let $\mathcal{X}$ be a fixed coset of $\mathcal{C}_I$ or $\mathcal{M}_I$ for $I \in \{0,1,2,3\}$ with $|I| = 1$. Let p and q be two different elements in a coset of $\mathcal{X}$, that is $\mathcal{X} \oplus a$, with $p \equiv (p^0, p^1, p^2, p^3)$ and $q \equiv (q^0, q^1, q^2, q^3)$, such that $p^0 = q^0$ and $p^j \neq q^j$ for each $j = 1, 2, 3$ (the set $\mathcal{T}_{p,q}^{\mathcal{X} \oplus a}$ is defined in a similar way for the other cases). Moreover, let $R^r(p)$ and $R^r(q)$ be the corresponding ciphertexts after r rounds.

We define the set $\mathcal{T}_{p,q}^{\mathcal{X} \oplus a}$ as the set of 1024 couples $(\hat{p}^i, R^r(\hat{p}^i))$ and $(\hat{q}^i, R^r(\hat{q}^i))$ where $\hat{p}^i, \hat{q}^i \in \mathcal{X} \oplus a$ for $i = 1, ..., 1024$ respectively generated by the following combinations of variables

1. (z^0, p^1, p^2, p^3) and (z^0, q^1, q^2, q^3);
2. (z^0, q^1, p^2, p^3) and (z^0, p^1, q^2, q^3);
3. (z^0, p^1, q^2, p^3) and (z^0, q^1, p^2, q^3);
4. (z^0, p^1, p^2, q^3) and (z^0, q^1, q^2, p^3).

where z^0 can take any possible value in $\mathbb{F}_{2^8}$.

As for the cases of the sets $\mathcal{S}$ and $\mathcal{Z}$, the following Lemma holds.

Lemma 5. *Let $\mathcal{T}_{p,q}^{\mathcal{M}_I \oplus a}$ be an arbitrary set defined as in Def. 10*

$$\mathcal{T}_{p,q}^{\mathcal{M}_I \oplus a} \equiv \{[(p_i^1, c_i^1 \equiv R(p_i^1)), (p_i^2, c_i^2 \equiv R(p_i^2))]_i \quad \forall i = 1, ..., 1024\}.$$

For each fixed $J \subseteq \{0,1,2,3\}$, only on of the two following events can happen:

- $c_i^1 \oplus c_i^2 \notin \mathcal{D}_J$ *for all* $i = 1, ..., 1024$*;*
- $c_i^1 \oplus c_i^2 \in \mathcal{D}_J$ *for all* $i = 1, ..., 1024$*.*

In other words, given a set $\mathcal{T}_{p^1,p^2}^{\mathcal{M}_I \oplus a}$, consider the 1024 couples of two (plaintext, ciphertext) pairs (p_i^1, c_i^1) and (p_i^2, c_i^2) for $i = 1, ..., 1024$. If two ciphertexts c^1 and c^2 belong (or not) to the same coset of $\mathcal{D}_J$ for a certain J, then the ciphertexts of all the other couples in the set $\mathcal{S}_{p^1,p^2}^{\mathcal{M}_I \oplus a}$ have the same property.

Thus, *given a set $\mathcal{T}$, what is the probability that two ciphertexts of at least one couple belong to the same coset of $\mathcal{M}_J$?*

D.2.1 Case: $|I| = 1$

We start considering the case of cosets $\mathcal{C}_I$ with $|I| = 1$. Note that in this case one can construct $2^{23} \cdot (2^8 - 1)^3 \simeq 2^{46.983}$ sets $\mathcal{T}$, each one of $4 \cdot 2^8 = 2^{10}$ couples (note that the number of couples with one equal generating variable is $4 \cdot 2^8 \cdot 2^{23} \cdot (2^8 - 1)^3 \simeq 2^{56.983}$ - see (11)).

Using the same calculation given in Sect. 7, it follows that this probability in the AES case is well approximated by to:

$$\begin{aligned}
p_{AES} =& \left[1 - Prob(\overline{\mathcal{E}_1^5} \wedge \overline{\mathcal{E}_2^5} \wedge ... \wedge \overline{\mathcal{E}_{1024}^5} \,|\, \mathcal{E}_i^4)\right] \cdot Prob(\mathcal{E}_i^4) + \\
&+ \left[1 - Prob(\overline{\mathcal{E}_1^5} \wedge \overline{\mathcal{E}_2^5} \wedge ... \wedge \overline{\mathcal{E}_{1024}^5} \,|\, \overline{\mathcal{E}_i^4})\right] \cdot Prob(\overline{\mathcal{E}_i^4}) = \\
=& (1 - p_3) \cdot \left[1 - \left(1 - \frac{p_3 \cdot (1 - p_{3,3})}{1 - p_3}\right)^{1024}\right] + p_3 \cdot \left[1 - \left(1 - p_{3,3}\right)^{1024}\right] = \\
=& 2^{-20} - 4095 \cdot 2^{-53} - \underbrace{529\,370\,445 \cdot 2^{-84}}_{\approx 3.945 \cdot 2^{-57}} + \underbrace{374\,996\,306\,937\,593 \cdot 2^{-117}}_{\approx 2.665 \cdot 2^{-70}} + ...
\end{aligned}$$

For a random permutation, the same event occurs with (approximately) probability

$$\begin{aligned}
p_{rand} =& 1 - \left(1 - p_3\right)^{1024} = 1 - \left[1 - \left(2^{-30} - 3 \cdot 2^{-63} + 2^{-94}\right)\right]^{1024} = \\
=& 2^{-20} - 4095 \cdot 2^{-53} + \underbrace{2\,794\,155 \cdot 2^{-84}}_{\approx 2.665 \cdot 2^{-64}} + ...
\end{aligned}$$

Since $|p_{AES} - p_{rand}| \simeq 2^{-55.013}$ and $p_{AES} \simeq p_{rand} \simeq 2^{-20}$, it follows that n must satisfy $n > 2^{92.246}$ for a probability of success of approximately 95%,. Since a single coset of $\mathcal{C}_I$ for $|I| = 1$ contains approximately $2^{46.983}$ different sets $\mathcal{T}$, it follows that $2^{92.246} \cdot 2^{-46.983} \simeq 2^{45.263}$ initial cosets of $\mathcal{C}_I$ for $|I| = 1$ are sufficient, for a total data cost of $2^{32} \cdot 2^{45.263} \simeq 2^{77.263}$ chosen plaintexts.

About the computational cost, the cost to re-order them is $4 \cdot 2^{45.263} \cdot 2^{32} \cdot \log 2^{32} \simeq 2^{84.263}$ table look-ups. The number of collisions is approximately $2^{-30} \cdot 2^{63} \cdot 2^{45.263} \simeq 2^{78.263}$. Among them, the pairs for which the two plaintexts have one common variable are $2^{78.263} \cdot 2^{-6} = 2^{72.623}$. Since the cost to construct the set $\mathcal{T}$ is of $2 \cdot 2^{10} = 2^{11}$ table look-ups, the total cost is well approximated by $2^{72.623} \cdot 2^{11} + 2^{84.263} = 2^{84.98}$ table look-ups, that is approximately $2^{78.33}$ five-round encryptions.

D.2.2 Case: $|I| = 2$

Consider now the case of cosets $\mathcal{C}_I$ with $|I| = 2$. Note that in this case one can construct $2^{47} \cdot (2^{16} - 1)^3 \simeq 2^{95}$ sets $\mathcal{T}$ of $4 \cdot (2^8)^2 = 2^{18}$ couples (note that the number of couples with one equal generating variable is $4 \cdot 2^{16} \cdot 2^{47} \cdot (2^{16} - 1)^3 \simeq 2^{113}$).

Using the same calculation given in Sect. 7, it follows that this probability in the AES case is well approximated by

$$\begin{aligned}
p_{AES} =& \left[1 - Prob(\overline{\mathcal{E}_1^5} \wedge \overline{\mathcal{E}_2^5} \wedge ... \wedge \overline{\mathcal{E}_{2^{18}}^5} \,|\, \mathcal{E}_i^4)\right] \cdot Prob(\mathcal{E}_i^4) + \\
&+ \left[1 - Prob(\overline{\mathcal{E}_1^5} \wedge \overline{\mathcal{E}_2^5} \wedge ... \wedge \overline{\mathcal{E}_{2^{18}}^5} \,|\, \overline{\mathcal{E}_i^4})\right] \cdot Prob(\overline{\mathcal{E}_i^4}) = \\
=& (1 - p_3) \cdot \left[1 - \left(1 - \frac{p_3 \cdot (1 - p_{3,3})}{1 - p_3}\right)^{2^{18}}\right] + p_3 \cdot \left[1 - \left(1 - p_{3,3}\right)^{2^{18}}\right] = \\
=& 2^{-12} - 1048575 \cdot 2^{-45} + \underbrace{46\,884\,625\,075 \cdot 2^{-76}}_{\approx 2.73 \cdot 2^{-42}} + ...
\end{aligned} \tag{23}$$

For a random permutation, the same event occurs with (approximately) probability

$$\begin{aligned}
p_{rand} =& 1 - \left(1 - p_3\right)^{2^{18}} = 1 - \left[1 - \left(2^{-30} - 3 \cdot 2^{-63} + 2^{-94}\right)\right]^{2^{18}} = \\
=& 2^{-12} - 1048575 \cdot 2^{-45} + \underbrace{183\,251\,413\,675 \cdot 2^{-76}}_{\approx 10.667 \cdot 2^{-42}} + ...
\end{aligned} \tag{24}$$

Since $|p_{AES} - p_{rand}| \simeq 2^{-39.011}$ and $p_{AES} \simeq p_{rand} \simeq 2^{-12}$, it follows that n must satisfy $n > 2^{68.243}$ for a probability of success of approximately 95%,. Since a single coset of $\mathcal{C}_I$ for $|I| = 2$ contains approximately 2^{95} different sets $\mathcal{T}$, less than a single coset is sufficient to implement the distinguisher. In particular, a set of the form

$$\left\{ a \oplus \begin{bmatrix} x_0 & y_1 & 0 & 0 \\ z_0 & x_1 & 0 & 0 \\ w_0 & z_1 & 0 & 0 \\ y_0 & w_1 & 0 & 0 \end{bmatrix} \middle| \forall x_0, x_1, y_0, y_1, z_0, z_1 \in \mathbb{F}_{2^8}^2, \forall w_0, w_1 \in \{0x00, 0x01, 0x02\} \right\}$$

for a certain constant a is sufficient (note that this is a subset of the coset $\mathcal{C}_{0,1} \oplus a$). Indeed, for such a set it is possible to construct approximately $3 \cdot [(2^{16})^2 \cdot 9 \cdot (2^{16}-1)^2 \cdot (9-1)] \cdot 2^{-3} \simeq 2^{68.75}$ different sets $\mathcal{Z}$ (remember that we are working with variables in $\mathbb{F}_{2^8}^2$), for a total of $(2^8)^6 \cdot 3^2 \simeq 2^{51.17}$ chosen plaintexts.

The cost to re-order it is $4 \cdot 2^{51.17} \cdot \log 2^{51.17} \simeq 2^{58.85}$ table look-ups. The number of collisions is approximately $2^{-30} \cdot 2^{102.34} \simeq 2^{72.34}$. Among them, the number of pairs for which the two plaintexts have one common variable is approximately $2^{72.34} \cdot 2^{-14} \simeq 2^{58.34}$ (the probability that two variables in $\mathbb{F}_{2^8}^2$ are equal is $4 \cdot 2^{-16} = 2^{-14}$). Since the cost to construct the set $\mathcal{T}$ is of $2 \cdot 2^{18} = 2^{19}$ table look-ups, the total cost is well approximated by $2^{58.34} \cdot 2^{19} + 2^{58.85} = 2^{77.34}$ table look-ups, that is approximately $2^{70.7}$ five-round encryptions.

D.3 Practical Verification on small-scale AES

In order to have a practical verification of the proposed distinguisher[30] (and of the following key-recovery attack), we have practically verified the probabilities p_{AES} and p_{rand} given above. In particular, we verified them using a small-scale AES, proposed in [CMR05]. We emphasize that our verification on the small-scale variant of AES is strong evidence for it to hold for the real AES, since the strategy used to theoretically compute such probabilities is independent of the fact that each word of AES is of 4 or 8 bits.

Thus, in order to compare the practical values with the theoretical ones, we compute the theoretical probabilities p_{AES} and p_{rand} for the small-scale case. First of all, for small scale AES the probabilities p_3 and $p_{3,3}$ are respectively equal to $p_3 = 2^{-14} - 3 \cdot 2^{-31} + 2^{-46}$ and $p_{3,3} = 2^{-10} - 3 \cdot 2^{-23} + 2^{-34}$.

For the following, we *limit to consider cosets of* $\mathcal{C}_I$ *for* $|I| = 1$.

D.3.1 Case: Set $\mathcal{S}$

W.l.o.g. we used cosets of $\mathcal{C}_0$ to practically test the two probabilities. Using the previous procedure and formula, the (approximately) probabilities that a set $\mathcal{S}^{\mathcal{C}_0 \oplus a}$ satisfy the required property for 5-round AES and the random case are respectively

$$p_{AES} = 2^{-11} - 31 \cdot 2^{-28} - \underbrace{12\,445 \cdot 2^{-43}}_{\approx 3.05 \cdot 2^{-31}} + \underbrace{4\,848\,753 \cdot 2^{-60}}_{\approx 37 \cdot 2^{-43}} + \ldots$$

$$p_{rand} = 2^{-11} - 31 \cdot 2^{-28} + 155 \cdot 2^{-43} + \ldots$$

As a result, using formula (20) for $p_{rand} \simeq p_{AES} \simeq 2^{-11}$ and $|p_{rand} - p_{AES}| \simeq 2^{-29.379}$, it follows that $n \geq 2^{50.194}$ different sets $\mathcal{S}^{\mathcal{C}_0 \oplus a}$ are sufficient to set up the distinguisher with probability higher than 95%.

Since we work with small-scale AES, a single coset of $\mathcal{C}_0$ contains 2^{16} (plaintexts, ciphertexts) pairs, or approximately $2^{15} \cdot (2^{16} - 1) \simeq 2^{31}$ different couples. Since the number of couples with different generating variables is given by $2^{16} \cdot (2^4 - 1)^4$ (also tested

[30] The source codes of the distinguishers/attacks are available at https://github.com/Krypto-iaik/Distinguisher_5RoundAES

by computer test), it is possible to construct $8^{-1} \cdot 2^{16} \cdot (2^4-1)^4 = 207\,360\,000 \simeq 2^{27.628}$ sets $\mathcal{S}$ such that all the generating variables of the couples of each of these sets are different. As a result, it follows that $2^{50.194} \cdot 2^{-27.628} = 2^{22.566}$ different initial cosets of $\mathcal{C}_0$ must be used, for a cost of $2^{38.566}$ chosen plaintexts.

For our tests, we used 2^{23} different initial cosets of $\mathcal{C}_0$ (keys used to encrypt the plaintexts in the AES case are randomly chosen and different for each coset - the key is not fixed). For each coset we exploited Algorithm 5 to count the number of sets $\mathcal{S}^{\mathcal{C}_0 \oplus a}$ that satisfy the required property (i.e. the number of sets for which two ciphertexts of at least one couple are in the same coset of $\mathcal{M}_J$ for certain J with $|J| = 3$). As a result, for each initial coset $\mathcal{C}_0$ the (average) theoretical numbers of sets $\mathcal{Z}^{\mathcal{C}_0 \oplus a}$ that satisfy the required property for the random and the AES cases - given by $n_X^T = 207\,360\,000 \cdot p_X$ - and the (average) practical ones found in our experiments - denoted by n_X^P - are given are:

$$n_{rand}^T \simeq 101\,226.057 \qquad n_{AES}^T \simeq 101\,225.76$$
$$n_{rand}^P \simeq 101\,226.105 \qquad n_{AES}^P \simeq 101\,225.68$$

Note that these two numbers are close to the theoretical ones, and that the average number of sets for AES case is lower than for the random one, as predicted.

D.3.2 Case: Set $\mathcal{T}$

W.l.o.g. we used cosets of $\mathcal{C}_0$ to practically test the two probabilities. Using the previous procedure and formula, (approximately) the probabilities that a set $\mathcal{S}^{\mathcal{C}_0 \oplus a}$ satisfy the required property for 5-round AES and the random case are respectively

$$p_{AES} = 2^{-8} - 255 \cdot 2^{-25} - 102\,605 \cdot 2^{-40} + ...$$
$$p_{rand} = 2^{-8} - 255 \cdot 2^{-25} + 10\,795 \cdot 2^{-40} + ...$$

As a result, using formula (20) for $p_{rand} \simeq p_{AES} \simeq 2^{-8}$ and $|p_{rand} - p_{AES}| \simeq 2^{-23.21}$, it follows that $n \geq 2^{40.64}$ different sets $\mathcal{T}^{\mathcal{C}_0 \oplus a}$ are sufficient to set up the distinguisher with probability higher than 95%.

Since we work with small-scale AES, a single coset of $\mathcal{C}_0$ contains $4 \cdot 2^4 \cdot 2^{11} \cdot (2^4-1)^3 \simeq 2^{29.71}$ couples for which the two plaintexts have only one different generating variable (also tested by computer test). Thus, it is possible to construct $2^{11} \cdot (2^4-1)^3 = 6\,912\,000 \simeq 2^{23.721}$ sets $\mathcal{T}$ such that all the generating variables of the couples of each of these sets are different. As a result, it follows that $2^{40.64} \cdot 2^{-23.721} = 2^{16.92}$ different initial cosets of $\mathcal{C}_0$ must be used, for a cost of $2^{38.566}$ chosen plaintexts.

For our tests, we used 2^{17} different initial cosets of $\mathcal{C}_0$ (keys used to encrypt the plaintexts in the AES case are randomly chosen and different for each coset - the key is not fixed). For each coset we exploited Algorithm 5 to count the number of sets $\mathcal{T}^{\mathcal{C}_0 \oplus a}$ that satisfy the required property (i.e. the number of sets for which two ciphertexts of at least one couple are in the same coset of $\mathcal{M}_J$ for certain J with $|J| = 3$). As a result, for each initial coset $\mathcal{C}_0$ the (average) theoretical numbers of sets $\mathcal{T}^{\mathcal{C}_0 \oplus a}$ that satisfy the required property for the random and the AES cases - given by $n_X^T = 6\,912\,000 \cdot p_X$ - and the (average) practical ones found in our experiments - denoted by n_X^P - are given are:

$$n_{rand}^T \simeq 26\,497.54 \qquad n_{AES}^T \simeq 26\,496.83$$
$$n_{rand}^P \simeq 26\,497.57 \qquad n_{AES}^P \simeq 26\,496.91$$

Note that these two numbers are close to the theoretical ones, and that the average number of sets for AES case is lower than for the random one, as predicted.

E Key-Recovery Attack on 6-round AES of Sect. 7.4 - Chosen Plaintexts in Cosets of $\mathcal{D}_I$ with $|I| = 2$

Referring to the key-recovery attack on 6-round AES of Sect. 7.4, here we explain why it is not possible to use cosets of $\mathcal{D}_I$ with $|I| = 2$ for a key-recovery attack - for the following we use set $\mathcal{S}$ which allows to minimize the data complexity (however, it is completely analogous for the set $\mathcal{T}$). In this case and using the same strategy of before, since 2^{64} different combinations of 8 bytes of the key (i.e. 2 diagonals) must be tested, one has to use the 5-round distinguisher with a probability higher $(0.95)^{2^{-64}}$. This requires approximately $2^{118.9}$ sets for each guessed combination of the key, that is a single coset of $\mathcal{D}_I$ with $|I| = 2$ for a total cost of 2^{64} chosen plaintexts (each coset of $\mathcal{D}_I$ with $|I| = 2$ has approximately 2^{127} different sets). On the other hand, using the previous argumentation, the total cost of the attack is approximately of 2^{166} table look-ups, which is worse than a brute-force attack. Indeed, the cost of the re-order process is of $4 \cdot 2^{64} \cdot (\log 2^{64} + 1) = 2^{72}$ table look-ups, while the cost to construct the set $\mathcal{S}$ when a collision is found is approximately of $2^{64} \cdot 2^{97} \cdot 16 = 2^{165}$ table look-ups (note the average number of collisions is $2^{64} \cdot 2^{63} \cdot 2^{-30} \approx 2^{97}$ and that one has to repeat the procedure 2^{64} times, i.e. the number of guessed key). It follows that it is not possible to use cosets of $\mathcal{D}_I$ with $|I| = 2$ for this attack. Since we don't exclude the possibility of a different and better implementation of the attack just described (with the goal to minimize the computational and the data costs), we leave its research as an open problem for future work.

F A 6-round Secret-Key Distinguisher for AES

In Sect. 7 we have proposed a probabilistic 5-round distinguisher for AES obtained extending (at the end) the deterministic 4-round distinguisher of Sect. 5. Here we propose a probabilistic 6-round distinguisher for AES obtained extending at the end the probabilistic 5-round distinguisher for AES, or equivalently extending at the end the 4-round distinguisher by two rounds. However, as we are going to show, this 6-round secret-key distinguisher for AES (which exploits a property which is independent of the secret key) can not be used in practice, since it requires more than the full codebook to distinguish a 6-round AES from a random permutation with non-negligible property.

To explain how this 6-round distinguisher works, we briefly recall the 4-round and the 5-round ones. In order to set up the 4-round secret-key distinguisher for AES, the idea is to consider cosets of a column space $\mathcal{C}_I$ for $I \subseteq \{0, 1, 2, 3\}$, to construct all the couples and to divide them in sets $\mathcal{S}^{\mathcal{C}_I \oplus a}$ as defined in Def. 8. As we have already seen, in the case of 4-round AES only two events can happen for each set $\mathcal{S}^{\mathcal{C}_I \oplus a}$: for all the couples, the two ciphertexts belong or not to the same coset of $\mathcal{M}_J$.

The idea of the 5-round distinguisher is to consider the probability that a set $\mathcal{Z}^{\mathcal{C}_I \oplus a}$ contains at least one couple for which the two ciphertexts belong to the same coset of $\mathcal{M}_J$ with $|J| = 3$. Referring to Sect. 7, it is possible to prove that this probability is lower for 5-round AES than for a random permutation. As for the 5-round distinguisher, in order to set up our distinguisher for 6-round AES, the idea is to count the number of sets $\mathcal{Z}^{\mathcal{C}_I \oplus a}$ *with* $|I| = 3$ for which two ciphertexts of at least one couple belong to the same coset of $\mathcal{M}_J$ for a certain $J \subseteq \{0, 1, 2, 3\}$ with $|J| = 3$. As we are going to prove, also in this case the probability of the above event is *lower* for 6-round AES than for a random permutation. On the other hand, this difference is so small (much smaller than for the 5-rounds case) that this distinguisher can not be used in practice, since it requires more than the full codebook to work.

F.1 Details and Data Cost

As for 5-round distinguisher, in order to set up the 6-round distinguisher the idea is to exploit the property of the 4-round secret-key distinguisher proposed in Sect. 5. Consider plaintexts in the same coset of $\mathcal{C}_I$ with $|I| = 3$, construct all the couples of two (plaintexts, ciphertexts) pairs (skipping the ones with common generating variables) and divide them in sets $\mathcal{Z}^{\mathcal{C}_I \oplus a}$. For an AES permutation and for a fixed $J \subseteq \{0, 1, 2, 3\}$ with $|J| = 3$, only two event can occur after four rounds:

1. for all the couples, the two ciphertexts belong to the same coset of $\mathcal{M}_J$ - probability $p_3 \simeq 2^{-30}$;
2. for all the couples, the two ciphertexts don't belong to the same coset of $\mathcal{M}_J$ - probability $1 - p_3 \simeq 1 - 2^{-30}$.

For a random permutation instead, it is possible that the two ciphertexts of only some - not all - couples belong to the same coset of $\mathcal{M}_J$.

Before we go on, we remember the following facts. By the impossible-differential trail (see Prop. 1), for all $J, K \subseteq \{0, 1, 2, 3\}$ with $|J| + |K| \leq 4$ (e.g. $|J| = 3$ and $|K| = 1$) the following probability holds

$$Prob(R^2(x) \oplus R^2(y) \in \mathcal{M}_K \,|\, x \oplus y \in \mathcal{M}_J) = 0,$$

while in general two texts belong to the same coset of $\mathcal{M}_K$ for $|K| = 1$ with probability $Prob(x \oplus y \in \mathcal{M}_K) = 2^{-94}$. The idea is to use these considerations and the same argumentation of the 5-round distinguisher of Sect. 7 in order to set up a 6-round distinguisher for AES which is independent of the secret key. The idea is to show that given a set $\mathcal{Z}^{\mathcal{C}_I \oplus a}$ for $|I| = 3$, the probability that two ciphertexts of at least one couple belong to the same coset of $\mathcal{M}_K$ for $|K| = 1$ after 6 rounds is lower for an AES permutation than for a random one.

Let's start with the AES permutation, and remember that for the following we consider only cosets of $\mathcal{C}_I$ with $|I| = 3$. First of all, note that each coset of $\mathcal{C}_I \oplus a$ with $|I| = 3$ contains approximately $3 \cdot 2^{47} \cdot (2^{24} - 1)^2 \simeq 2^{96.585}$ sets $\mathcal{Z}$, and that each set $\mathcal{Z}$ contains $(2 \cdot 2^{24})^2 = 2^{49}$ couples. Given a set $\mathcal{Z}^{\mathcal{C}_I \oplus a}$, consider the case in which the two texts of each couple belong to the same coset of $\mathcal{M}_J$ for $|J| = 3$ after 4 rounds. By Prop. 1, it follows immediately that in this case the two ciphertexts of all the couples can not can belong to the same coset of $\mathcal{M}_K$ for $|K| = 1$ after 6 rounds. In other words, the probability of this case is 0. Consider now the other case, in which for each couple, the two texts don't belong to the same coset of $\mathcal{M}_J$ for $|J| = 3$ after 4 rounds. By simple calculation, the probability that two ciphertexts of at least one couple belong to the same coset of $\mathcal{M}_J$ for $|J| = 1$ after 6 rounds is given by $1 - (1 - \tilde{p}_{1,3})^{2^{49}} \simeq 2^{-45}$. Thus, given a set $\mathcal{Z}^{\mathcal{C}_I \oplus a}$ and using analogous calculation of Sect. 7 (i.e. for the 5-round distinguisher), it follows that the probability that the two ciphertexts of at least one couple belong to the same coset of $\mathcal{M}_J$ for $|J| = 1$ is well approximated by

$$\begin{aligned}
p_{AES} =& \big[1 - Prob(\overline{\mathcal{E}_1^6} \wedge \overline{\mathcal{E}_2^6} \wedge ... \wedge \overline{\mathcal{E}_{2^{49}}^6} \,|\, \mathcal{E}_i^4)\big] \cdot Prob(\mathcal{E}_i^4) + \\
&+ \big[1 - Prob(\overline{\mathcal{E}_1^6} \wedge \overline{\mathcal{E}_2^6} \wedge ... \wedge \overline{\mathcal{E}_{2^{49}}^6} \,|\, \overline{\mathcal{E}_i^4})\big] \cdot Prob(\overline{\mathcal{E}_i^4}) = \\
=& \big[1 - Prob(\overline{\mathcal{E}_1^6} \wedge \overline{\mathcal{E}_2^6} \wedge ... \wedge \overline{\mathcal{E}_{2^{49}}^6} \,|\, \overline{\mathcal{E}_i^4})\big] \cdot Prob(\overline{\mathcal{E}_i^4}) = \\
=& (1 - p_3) \cdot \big[1 - \big(1 - \tilde{p}_{1,3}\big)^{2^{49}}\big] = \\
=& 2^{-45} - 2^{-91} - 2^{-121} + 3 \cdot 2^{-139} + ...
\end{aligned} \tag{25}$$

for a certain $i = 1, ..., 2^{49}$. Instead, by simple computation, for a random permutation the same event occurs with (approximately) probability

$$p_{rand} = 1 - (1 - p_1)^{2^{49}} = 1 - (1 - 2^{-94})^{2^{49}} = 2^{-45} - 2^{-91} + 3 \cdot 2^{-139} + ... \tag{26}$$

As for the 5-round distinguisher, the idea is to exploit this small difference ($|p_{AES} - p_{rand}| \simeq 2^{-121}$) in order to distinguish the random permutation from an AES one.

Using formula (20), it follows that to distinguish the two cases with probability higher than 95%, one needs more than $2^{199.22}$ different sets. On the other hands, the maximum number of available sets using initial cosets of $\mathcal{C}_I$ with $|I| = 3$ is approximately $2^{32} \cdot 2^{96.584} \simeq 2^{128.585}$. Since similar results occur using different values of $|I|$, $|J|$ and $|K|$ with $|J| + |K| \leq 4$ and using sets $\mathcal{S}$ and $\mathcal{T}$, it follows that the distinguisher requires more than the full codebook to work. As a result, *the problem to set up a distinguisher for 6 rounds of AES which exploits a property which is independent of the secret key is still open for future research.*

Only for completeness, note that this distinguisher on 6 rounds has something in common with the 4-round distinguisher based on impossible differential trails (we refer e.g. to [BK01] for details), in the same way in which the 5-round distinguisher just presented in Sect. 7 has something in common with the 3-round distinguisher based on the truncated differential cryptanalysis. For an impossible differential trail, the idea is to exploit the given two plaintexts in the same coset of $\mathcal{D}_I$, then they belong to different cosets of $\mathcal{M}_J$ after four rounds for each $I, J \in \{0, 1, 2, 3\}$ with $|I| + |J| \leq 4$ - see Prop. 1. Here we use the same technique but working on sets and not on single pairs of texts to set up our 6-round distinguisher.

G Key-Recovery Attack on AES with a single secret S-Box

G.1 Impossible Differential Attack on 5-round AES with a single Secret S-Box

In this section, we show how to set up an impossible differential attack on 5-round AES that exploits the fact that a sum of coefficients of the MixColumns matrix is equal to zero (e.g. (15)), and improves the one presented in [GRR17b].

For a fixed $a \in \mathcal{D}_0^{\perp}$ (i.e. $a_{i,i} = 0$ for $i = 1, 2, 3$), consider a set of plaintexts of the form:

$$V_\delta \equiv \left\{ a \oplus \begin{bmatrix} x & 0 & 0 & 0 \\ 0 & x \oplus \delta_{1,1} & 0 & 0 \\ 0 & 0 & x \oplus \delta_{2,2} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \Big| \, \forall x \in \mathbb{F}_{2^8} \right\} \tag{27}$$

and let $\delta \equiv (\delta_{1,1}, \delta_{2,2})$. Since

$$M^{MC}_{r,1} \oplus M^{MC}_{r,2} \oplus M^{MC}_{r,3} = 0 \qquad \text{for } r = 0, 1,$$

it follows by Prop. 2 that the set V_δ is mapped into a coset of $\mathcal{C}_0 \cap \mathcal{D}_{2,3}$ with probability 1 after one round if $\delta_{1,1} = k_{1,1} \oplus k_{0,0}$ *and* $\delta_{2,2} = k_{2,2} \oplus k_{0,0}$. In the other cases, that is if $\delta_{1,1} \neq k_{1,1} \oplus k_{0,0}$ and/or $\delta_{2,2} \neq k_{2,2} \oplus k_{0,0}$ the set V_δ is mapped into a coset of $\mathcal{C}_0$ with probability 1, and into a coset of $\mathcal{C}_0 \cap \mathcal{D}_I \subseteq \mathcal{D}_I$ for a certain I with $|I| = 2$ with probability $6 \cdot 2^{-16} = 3 \cdot 2^{-15}$.

Since $Prob(R^4(x) \oplus R^4(y) \in \mathcal{M}_J \,|\, x \oplus y \in \mathcal{D}_I) = 0$ for $|I| + |J| \leq 4$ - see Prop. 1, if $\delta_{1,1} = k_{1,1} \oplus k_{0,0}$ *and* $\delta_{2,2} = k_{2,2} \oplus k_{0,0}$, it follows that given two plaintexts in the same coset of V_δ, then the corresponding ciphertexts after five rounds can not belong to the same coset of $\mathcal{M}_J$ for $|J| = 2$:

$$Prob(R^5(x) \oplus R^5(y) \in \mathcal{M}_J \,|\, x, y \in V_\delta \quad \text{and} \quad \delta_{i,i} = k_{i,i} \oplus k_{0,0} \text{ for } i = 1, 2) = 0.$$

In the other cases - if $\delta_{1,1} \neq k_{1,1} \oplus k_{0,0}$ and/or $\delta_{2,2} \neq k_{2,2} \oplus k_{0,0}$, given two plaintexts in the same coset of V_δ, then the corresponding ciphertexts after 5-round belong to the same coset of $\mathcal{M}_J$ for $|J| = 2$ with prob. $6 \cdot 2^{-64} = 3 \cdot 2^{-63}$. The idea is to exploit this difference

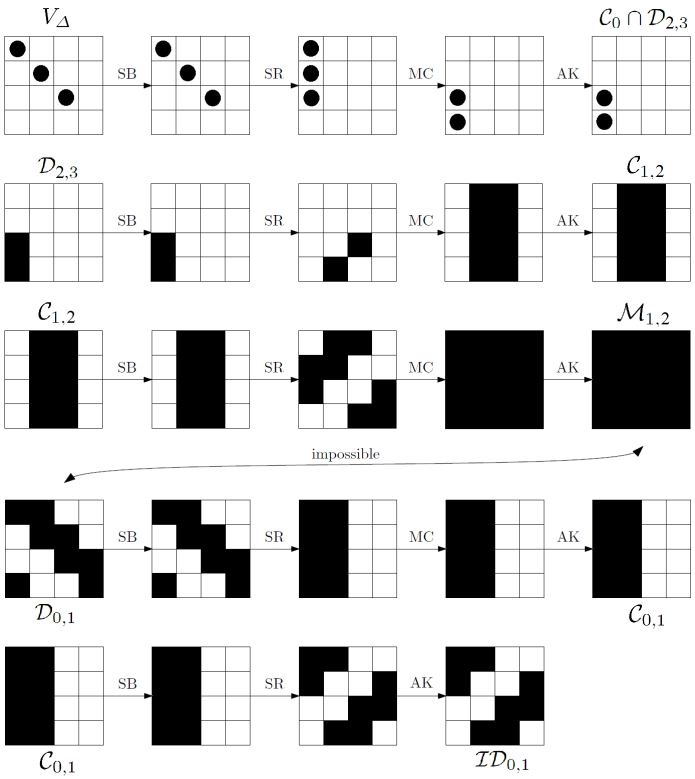

Figure 5: *5-Round secret-key distinguisher for AES with a single secret S-Box* with data complexity $2^{76.4}$ based on a 4-round Impossible Subspace Trail. The choice of the plaintexts (i.e. $p_{0,0} \oplus p_{i,i} = k_{0,0} \oplus k_{i,i}$ for $i = 1, 2$) guarantees that after one round there are only two bytes with non-zero difference instead of four, that is the plaintexts belong to the same coset of $\mathcal{C}_0 \cap \mathcal{D}_{2,3}$. Thus, the probability the two ciphertexts belong to the same coset of $\mathcal{M}_K$ for $|K| = 2$ is zero. White box denotes denotes a byte with a zero-difference, while a black box denotes a byte with non-zero difference.

in the probabilities to recover the secret key.

Data and Computational Costs. The data and the computational costs analysis are similar to the ones proposed in [GRR17b]. Consider the attack on 2 bytes of the secret key. In order to discard a wrong candidate δ of the key, it is sufficient that at least one set V_δ for which a pair of ciphertexts belong to the same coset of $\mathcal{M}_J$ with $|J| = 2$ exists (note that this can never happen for the right value of δ - the secret key). Since there are $2^{16} - 1$ wrong candidates, in order to have a total probability of success of 95%, such a set must exist for each δ with probability higher than $(0.95)^{2^{-16}} \simeq 99.999922\%$.

Given a set V_δ, it is possible to construct approximately $2^7 \cdot (2^8 - 1) = 2^{15}$ different pairs of ciphertexts. Since each pair can belong to the same coset of $\mathcal{M}_J$ with a probability of $3 \cdot 2^{-63}$, given n different pairs, the probability that at least one of them belong to the same coset of $\mathcal{M}_J$ is $1 - (1 - 3 \cdot 2^{-63})^n$. By simple computation, the condition $1 - (1 - 3 \cdot 2^{-63})^n > 0.99999922$ is satisfied for $n > 2^{65.23}$. Since each set V_δ is composed of 2^{15} pairs and since one has to repeat the attack for each possible value of δ, the attacker needs approximately $2^{65.23} \cdot 2^{-7} \cdot 2^{16} = 2^{74.23}$ chosen plaintexts to find two bytes of the secret key (note that each set V_δ contains 2^8 texts, so $2^{-15} \cdot 2^8 = 2^{-7}$).

The idea is to repeat this attack 4 times in order to find 8 bytes of the key (i.e. 2

Data: $2^{74.4}$ different sets V_δ defined as in (27) - $2^{58.4}$ for each $\delta \equiv (\delta_{1,1}, \delta_{2,2})$ - and corresponding ciphertexts after 5 rounds
Result: $k_{0,0} \oplus k_{1,1}$ and $k_{0,0} \oplus k_{2,2}$
for *each* $\delta_{1,1}$ *from 0 to* $2^8 - 1$ *and each* $\delta_{2,2}$ *from 0 to* $2^8 - 1$ **do**
  $flag \leftarrow 0$;
  for *each set* V_δ **do**
    for *each* $I \subseteq \{0, 1, 2, 3\}$ *with* $|I| = 2$ **do**
      let (p^i, c^i) for $i = 0, ..., 2^8 - 1$ be the 2^8 (plaintexts, ciphertexts) of V_δ;
      re-order this set of elements w.r.t. the partial order $\preceq$ defined in analogous way of Def. 11 s.t. $c^i \preceq c^{i+1}\ \forall i$; `// ⪯ depends on` I
      for i *from 0 to* $2^8 - 2$ **do**
        if $c^i \oplus c^{i+1} \in \mathcal{M}_I$ **then**
          $flag \leftarrow 1$;
          next δ;
        end
      end
    end
  end
  if $flag = 0$ **then**
    identify $\delta_{1,1}$ as candidate for $k_{0,0} \oplus k_{1,1}$ and $\delta_{2,2}$ as candidate for $k_{0,0} \oplus k_{2,2}$;
  end
end
return *Candidates for* $k_{0,0} \oplus k_{1,1}$ *and* $k_{0,0} \oplus k_{2,2}$. `// Only one candidate with Prob. 95%`

Algorithm 7: *Impossible Differential Attack on 5 rounds of AES with a single secret S-Box.* For simplicity, the goal of the attack is to find two bytes of the key - $k_{0,0} \oplus k_{1,1}$ and $k_{0,0} \oplus k_{2,2}$. The same attack on the other diagonals can be used to recover the entire key up to 2^{32} variants.

for column). In this case, for each candidate δ of the key at least one set V_δ with the previous property must exist with probability higher $(0.95)^{2^{-18}} \simeq 99.99998\%$. Using the same calculation as before, one needs approximately $n > 2^{65.37}$ pairs of ciphertexts for each δ, that is approximately $2^{50.37}$ different sets V_δ.

Finally, in order to find the final 4 bytes of the key (remember that we are to find it up to 2^{32} variants), the idea is to repeat again the previous attack. However, note that in this case the attacker must guess only one byte of the key for each diagonal instead of two (since two of three differences are already known). Thus, for each wrong δ, at least one set for which two ciphertexts belong to the same coset of $\mathcal{M}_J$ with $|J| = 2$ must exist with probability higher $(0.95)^{2^{-10}} \simeq 99.995\%$. Using the same calculation as before, one needs approximately $n > 2^{64.73}$ pairs of ciphertexts for each δ, that is approximately $2^{57.73}$ different sets V_δ. It follows that the total data complexity is approximately of $4 \cdot 2^{58.37} \cdot 2^{16} + 4 \cdot 2^{57.73} \cdot 2^8 = 2^{76.374}$ chosen plaintexts.

As for the impossible differential attack on 5-round AES with a single secret S-Box presented in [GRR17b], the computational cost is well approximated by the re-ordering algorithm, which can be approximated by $4 \cdot 4 \cdot 2^{58.37} \cdot 2^{16} \cdot (\log 2^8 + 1) = 2^{81.54}$ table look-ups, or approximately $2^{74.9}$ five-round encryptions.

G.2 Computational Cost of Key-Recovery Attacks on 5-round AES of Sect. 6.2

In this section, we give all the details of the 5-round key-recovery attacks for AES presented in Sect. 6.2 about computational costs. We refer to those sections for all the details about the attacks.

G.2.1 Attack of Sect. 6.2 - Computational Cost

In order to count the number of collisions, one can use the same procedure of the attack described in Sect. 5, i.e. one can re-oder the texts with respect to a particular partial order $\preceq$ as defined in 11. Here we propose an alternative strategy, which exploits *data structure*.

Assume $I \subseteq \{0,1,2,3\}$ is fixed with $|I| = 1$, and that the final MixColumns operation is not omitted. The goal is to count the number of pairs of ciphertexts (c^1, c^2) such that $c^1 \oplus c^2 \in \mathcal{M}_I$, or equivalently

$$MC^{-1}(c^1)_{i,j-i} = MC^{-1}(c^2)_{i,j-i} \qquad \forall i = 0,1,2,3 \tag{28}$$

where $j = \{0,1,2,3\} \setminus I$, and the index is computed modulo 4. To do this, consider an array W of 2^{32} elements completely initialized to zero. The element of W in position x for $0 \leq x \leq 2^{32} - 1$ - denoted by $W[x]$ - 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 (28), then they increment the same element x of the array W. It follows that given $r \geq 0$ texts that increment the same element x of the array W, then it is possible to construct $\binom{r}{2} = \frac{r \cdot (r-1)}{2}$ different pairs of texts that satisfy (28). The complete pseudo-code of such an algorithm is given in Algorithm 3.

What is the total computational cost of this procedure? Given a set of 2^{40} (plaintexts, ciphertexts) pairs, one has first to fill the array W 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 $2^{40} + 2 \cdot 2^{32} = 2^{40.01}$ table look-ups - these operations require 2^{32} table look-ups (for the W case) or 2^{40} table look-ups (for the A_δ case). Since one has to repeat this procedure 16 times for each candidate of δ, and 12 times in order to find the key up to 2^{32} variants, the total cost of this attack is well approximated by $12 \cdot 2^8 \cdot 16 \cdot 2^{40.01} \simeq 2^{55.6}$ table look-ups or approximately $2^{48.96}$ five-rounds encryptions.

For comparison, the computational cost using the re-ordering algorithm is well approximated by $12 \cdot 2^8 \cdot 2^{40} \cdot (\log 2^{40} + 1) \cdot 16 \simeq 2^{60.9}$ table look-ups, that is approximately $2^{54.25}$ five-round encryptions.

G.3 Attack on 5-round AES with single secret S-Box - MixColumns Matrix with Zero-Sum of Coefficients

In this section, we show how to adapt the attack just presented in order to exploit e.g. condition (15), i.e.the fact that a sum of elements that lie on the same row of the MixColumns matrix are equal to zero.

Similar to before, the idea is to consider a set of plaintexts $\mathcal{A}'_\delta$ which depends on the guessed value of the key of the form:

$$\mathcal{A}'_\delta \equiv \left\{ a \oplus \begin{bmatrix} 0 & y_0 & 0 & 0 \\ 0 & x & y_1 & 0 \\ 0 & 0 & x \oplus \delta_{2,2} & y_2 \\ y_3 & 0 & 0 & x \oplus \delta_{3,3} \end{bmatrix} \;\middle|\; \forall x, y_0, ..., y_3 \in \mathbb{F}_{2^8} \right\} \tag{29}$$

where $\delta = (\delta_{2,2}, \delta_{3,3})$ and $a \in \mathcal{D}_0^{\perp}$ (i.e. $a_{i,i} = 0$ for $i = 1, 2, 3$) is a constant. Given a set $\mathcal{A}'_\delta$, we claim that if $\delta_{i,i} = k_{1,1} \oplus k_{i,i}$ for $i = 2, 3$ then the number of collisions among the ciphertexts after 5 rounds in the same coset of $\mathcal{M}_I$ for a fixed $I \subseteq \{0, 1, 2, 3\}$ with $|I| = 3$ is a multiple of 4. More formally:

Proposition 9. *Consider a set of plaintexts $\mathcal{A}'_\delta$ defined as in (29), and the corresponding ciphertexts after 5 rounds. If $\delta_{i,i} = k_{1,1} \oplus k_{i,i}$ for $i = 2, 3$, then the number of different pairs of ciphertexts that belong to the same coset of $\mathcal{M}_I$ for a fixed $I \subseteq \{0, 1, 2, 3\}$ with $|I| = 3$ is a multiple of 4.*

Proof. Let $\delta_{2,2} = k_{1,1} \oplus k_{2,2}$ and $\delta_{3,3} = k_{1,1} \oplus k_{3,3}$. By simple computation, there exists b such that the set $\mathcal{A}'_\delta$ is mapped after one round in

$$R(\mathcal{A}'_\delta) \equiv \left\{ b \oplus \begin{bmatrix} 0\text{x}03 \cdot w & z_0 & 0 & 0 \\ 0 & z_1 & 0 & 0 \\ 0 & z_2 & 0 & 0 \\ 0\text{x}02 \cdot w & z_3 & 0 & 0 \end{bmatrix} \;\middle|\; \forall w, z_0, ..., z_3 \in \mathbb{F}_{2^8} \right\}.$$

Consider two elements $z, z' \in R(\mathcal{A}'_\delta)$ generated respectively by $z \equiv (z_0, z_1, z_2, z_3, w)$ and $z' \equiv (z'_0, z'_1, z'_2, z'_3, w)$. The idea is to consider separately the cases (1) $z_2 \neq z'_2$ and $z_3 \neq z'_3$, (2) $z_2 = z'_2$ and $z_3 = z'_3$ and (3) $z_2 = z'_2$ and $z_3 \neq z'_3$ (or viceversa), and to show that in the first case the number of collisions is a multiple of 4, while in the second case it is a multiple of 2^{16} and in the third case it is a multiple of 2^9. It follows that there exist $n', n'', n''' \in \mathbb{N}$ such that the total number of collisions n can be written as $n = 4 \cdot n' + 2^{16} \cdot n'' + 2^9 \cdot n''' = 4 \cdot (n' + 2^{14} \cdot n'' + 2^7 \cdot n''')$. In other words, the total number of collisions is a multiple of 4.

The details of the proof can be found in App. H. □

Note that the previous result doesn't hold for the cases $\delta_{2,2} \neq k_{1,1} \oplus k_{2,2}$ and/or $\delta_{3,3} \neq k_{1,1} \oplus k_{3,3}$. In these cases, the number of collisions for $\delta_{i,i} \neq k_{1,1} \oplus k_{i,i}$ is a multiple of 4 only with probability $1/4 = 25\%$.

Since the procedure of the attack is completely equivalent to the one just described in App. 6.2, we limit here to give the details of the data and of the computational costs of the attack.

Working in the same way just described for the attack of App. 6.2, an attacker can recover the secret key up to 2^{32} variants. Note that in this case for each set $\mathcal{A}'_\delta$, the attacker has to test 2^{16} different keys, i.e. she has to test 2 bytes of the key (instead of 1 as before). Due to similar argumentation as before, for each possible wrong candidate of the key δ, at least one set $\mathcal{A}'_\delta$ must exist for which the number of collisions is not a multiple of 4 with a probability higher than $(0.95)^{2^{-16}} \simeq 99.999922\%$. Since given n sets $\mathcal{A}'_\delta$ the probability that a set with the required property exists is $1 - 2^{-2n}$, one needs approximately $n \geq 11$ different tests (i.e. 3 different sets $\mathcal{A}'_\delta$ - remember that there are 4 different subspace $\mathcal{M}_I$ with $|I| = 3$) for each δ in order to find the right key.

The idea is to use the same procedure to find the rest of the key. In particular, one repeats the same procedure for each one of the four columns in order to recover 8 bytes of the key (2 for each column). It follows that a set $\mathcal{A}'_\delta$ must exist for each wrong guessed δ with probability higher than $(0.95)^{2^{-18}} \simeq 99.99998\%$, that is one needs approximately $n \geq 12$ different tests (i.e. 3 different sets $\mathcal{A}'_\delta$) for each δ in order to find the right key. To find the final 4 bytes of the key, the attacker repeats the previous procedure, noting that in this case one has to guess only one byte of difference of the key instead of two, since the other one is already known. Thus, for each one of the $4 \cdot 2^8$ possible candidates of the key, one needs that at least a set $\mathcal{A}'_\delta$ for which the number of collisions is not a multiple of 4 exists with probability higher than $(0.95)^{2^{-10}} \simeq 99.995\%$, that is approximately $n \geq 8$

Data: $3 \cdot 2^{16}$ different sets $\mathcal{A}'_\delta$ defined as in (29) - 3 different sets for each $\delta \equiv (\delta_{2,2}, \delta_{3,3})$ - and corresponding ciphertexts after 5 rounds
Result: $k_{2,2} \oplus k_{1,1}$ and $k_{3,3} \oplus k_{1,1}$
for *each* $\delta_{2,2}$ *from 0 to* $2^8 - 1$ *and each* $\delta_{3,3}$ *from 0 to* $2^8 - 1$ **do**
  $flag \leftarrow 0$;
  for *each set* $\mathcal{A}'_\delta$ **do**
    let (p^i, c^i) for $i = 0, ..., 2^{40} - 1$ be the 2^{40} (plaintexts, ciphertexts) of $\mathcal{A}'_\delta$;
    for *all* $j \in \{0, 1, 2, 3\}$ **do**
      Let $W[0, ..., 2^{32} - 1]$ be an array initialized to zero;
      for i from *0 to* $2^{40} - 1$ **do**
        $x \leftarrow \sum_{k=0}^{3} MC^{-1}(c^i)_{k,j-k} \cdot 256^k$; // $MC^{-1}(c^i)_{k,j-k}$ denotes the byte of $MC^{-1}(c^i)$ in row k and column $j-k \mod 4$
        $W[x] \leftarrow W[x] + 1$; // $W[x]$ denotes the value stored in the x-th address of the array W
      end
      $n \leftarrow 0$;
      for i from *0 to* $2^{32} - 1$ **do**
        $n \leftarrow n + W[i] \cdot (W[i] - 1)/2$;
      end
      if $(n \mod 4) \neq 0$ **then**
        $flag \leftarrow 1$;
        next δ;
      end
    end
  end
  if $flag = 0$ **then**
    identify $\delta_{2,2}$ as candidate for $k_{2,2} \oplus k_{1,1}$ and $\delta_{3,3}$ as candidate for $k_{3,3} \oplus k_{1,1}$;
  end
end
return *Candidates for* $k_{2,2} \oplus k_{1,1}$ *and* $k_{3,3} \oplus k_{1,1}$. // Only one candidate with Prob. 95%

Algorithm 8: *Key-Recovery Attack on 5 rounds of AES with a single secret S-Box.* For simplicity, the goal of the attack is to find two bytes of the key - $k_{2,2} \oplus k_{1,1}$ and $k_{3,3} \oplus k_{1,1}$. The same attack can be used to recover the entire key up to 2^{32} variants.

different tests (i.e. 2 different sets $\mathcal{A}'_\delta$) for each δ are sufficient in order to find the right key.

In conclusion, the data cost of the attack is well approximated by 4 (columns) $\cdot 3$ (cosets) $\cdot 2^{40}$ (number of texts in $\mathcal{A}'_\delta$) $\cdot 2^{16}$ (candidates of the key) $+4 \cdot 2 \cdot 2^{40} \cdot 2^8 = 2^{59.6}$ chosen plaintexts. Using the same strategy proposed in Sect. 6.2 and described in details in Algorithm 8, the computational cost using data-structure is well approximated by $4 \cdot 4 \cdot 3 \cdot (2^{40} + 2 \cdot 2^{32}) \cdot 2^{16} \simeq 2^{61.6}$ table look-ups, that is approximately $2^{54.96}$ five-round encryptions. For comparison, the computational cost using a re-ordering algorithm is well approximated by $4 \cdot 4 \cdot 3 \cdot 2^{40} \cdot (\log 2^{40} + 1) \cdot 2^{16} \simeq 2^{66.9}$ table look-ups, that is approximately $2^{60.26}$ five-round encryptions.

Practical Verification

Using a C/C++ implementation[31], we have practically verified the attack just described on a small-scale variant of AES, as presented in [CMR05] - not on real AES due to the

[31] The source codes of the attacks on AES with a secret S-Box are available at https://github.com/Krypto-iaik/Attacks_AES_SecretSBox2

large computational cost of the attack. We emphasize that Prop. 9 is independent of the fact that each word is composed of 8 or 4 bits. Thus, our verification on small-scale variant of AES is strong evidence for it to hold for the real AES.

For simplicity, we limit here to report the result for the attack on two bytes of the key, e.g. $k_{1,1} \oplus k_{2,2}$ and $k_{1,1} \oplus k_{3,3}$. For small-scale AES, since there are only $(2^4)^2 = 2^8$ possible candidates, it is sufficient that a set $\mathcal{A}_\delta$ for which the number of collisions is odd exists for each wrong candidate of $(k_{1,1} \oplus k_{2,2}, k_{1,1} \oplus k_{3,3})$ with probability higher than $(0.95)^{2^{-8}} = 99.98\%$. It follows that 7 tests (that is 2 different sets $\mathcal{A}_\delta$) for each candidate of $(k_{1,1} \oplus k_{2,2}, k_{1,1} \oplus k_{3,3})$ are sufficient to find the right value. Re-ordering the texts as described previously, the theoretical computational cost is well approximated by $4 \cdot 2 \cdot 2^8 \cdot 2^{20} \cdot (\log 2^{20} + 1) \simeq 2^{35.32}$ table look-ups, while using data-structure is well approximated by $4 \cdot 2 \cdot 2^8 \cdot (2^{20} + 2 \cdot 2^{16}) \simeq 2^{31.17}$ table look-ups.

Our tests confirm that 2 different sets $\mathcal{A}_\delta$ are largely sufficient to find the key. The average practical computational cost is of $2^{33.6}$ table look-ups using the re-ordering algorithm and 2^{30} table look-ups using data-structure. As before, the difference with the theoretical value is justified by the fact that the this last one is computed in the worst case.

H Proof of Sect. 6.2 - 6.3 and App. G.3

H.1 Proof of Sect. 6.2

For a fixed a, consider a set of plaintexts $\mathcal{A}_\delta$ of the form (16):

$$\mathcal{A}_\delta \equiv \left\{ a \oplus \begin{bmatrix} y_0 & x & 0 & 0 \\ 0 & y_1 & x \oplus \delta & 0 \\ 0 & 0 & y_2 & 0 \\ 0 & 0 & 0 & y_3 \end{bmatrix} \;\middle|\; \forall x, y_0, ..., y_3 \in \mathbb{F}_{2^8} \right\}.$$

Proposition 10. *Consider a set of plaintexts $\mathcal{A}_\delta$ defined as in (16), and the corresponding ciphertexts after 5 rounds. If $\delta = k_{0,1} \oplus k_{1,2}$, then the number of different pairs of ciphertexts that belong to the same coset of $\mathcal{M}_I$ for a fixed $I \subseteq \{0,1,2,3\}$ with $|I| = 3$ is a multiple of 2.*

Proof. Let $\delta = k_{0,1} \oplus k_{1,2}$. By simple computation, there exists b such that the set $\mathcal{A}_\delta$ is mapped after one round into

$$R(\mathcal{A}_\delta) \equiv \left\{ b \oplus \begin{bmatrix} z_0 & w & 0 & 0 \\ z_1 & 0\text{x}03 \cdot w & 0 & 0 \\ z_2 & 0 & 0 & 0 \\ z_3 & 0\text{x}02 \cdot w & 0 & 0 \end{bmatrix} \;\middle|\; \forall w, z_0, ..., z_3 \in \mathbb{F}_{2^8} \right\}.$$

Consider two elements $z, z' \in R(\mathcal{A}_\delta)$ generated respectively by $z \equiv (z_0, z_1, z_2, z_3, w)$ and $z' \equiv (z'_0, z'_1, z'_2, z'_3, w)$. In the following, we consider separately the two cases $z_1 \neq z'_1$ and $z_1 = z'_1$. We show that in the first case (i.e. the set of all different pairs of elements with $z_{1,1} \neq z'_{1,1}$) the number of collisions is a multiple of 2, while in the second case (i.e. the set of all different pairs of elements with $z_1 = z'_{1,1}$) the number of collisions is a multiple of 256. It follows that there exist $n', n'' \in \mathbb{N}$ such that the total number of collisions n can be written as $n = 2 \cdot n' + 256 \cdot n'' = 2 \cdot (n' + 128 \cdot n'')$. In other words, the total number of collisions is a multiple of 2.

Case: $\mathbf{z_1 \neq z'_1}$. Consider two elements $z, z' \in R(\mathcal{A}_\delta)$ generated respectively by $z \equiv (z_0, z_1, z_2, z_3, w)$ and $z' \equiv (z'_0, z'_1, z'_2, z'_3, w)$ with $z_1 \neq z'_1$. For a fixed $I \in \{0,1,2,3\}$ with

$|I| = 3$, the idea is to show that $R^4(z) \oplus R^4(z') \in \mathcal{M}_I$ *if and only if* $R^4(v) \oplus R^4(v') \in \mathcal{M}_I$ where the texts $v, v' \in R(\mathcal{A}_\delta)$ are generated respectively by $v \equiv (z_0, z'_1, z_2, z_3, w)$ and $v' \equiv (z'_0, z_1, z'_2, z'_3, w)$. This follows by Theorem 2 of [GRR17a] - recalled in Sect. 4, and implies that the number of collision must be a multiple of 2 for this case.

For more details, let v and v' defined as before. The idea is to prove (1) that $R^2(z) \oplus R^2(z') = R^2(v) \oplus R^2(v')$ and (2) that $z, z' \in R(\mathcal{A}_\delta)$ can exist such that $R^4(z) \oplus R^4(z') \in \mathcal{M}_I$.

First of all, note that if $R^2(z) \oplus R^2(z') = R^2(v) \oplus R^2(v')$ and if $R^4(z) \oplus R^4(z') \in \mathcal{M}_I$, then also $R^4(v) \oplus R^4(v') \in \mathcal{M}_I$. Indeed, if $R^4(z) \oplus R^4(z') \in \mathcal{M}_I$ (i.e. $R^4(z)$ and $R^4(z')$ belong to the same coset of $\mathcal{M}_I$), then $R^2(z) \oplus R^2(z') \in \mathcal{D}_I$ by Theorem. 1. By $R^2(z) \oplus R^2(z') = R^2(v) \oplus R^2(v')$, it follow that $R^2(v) \oplus R^2(v') \in \mathcal{D}_I$ and so $R^4(v) \oplus R^4(v') \in \mathcal{M}_I$.

Secondly, one has to prove $[R^2(z) \oplus R^2(z')]_{i,j} = [R^2(v) \oplus R^2(v')]_{i,j}$ for each i, j. For simplicity, we limit to prove that $[R^2(z) \oplus R^2(z')]_{0,0} = [R^2(v) \oplus R^2(v')]_{0,0}$, i.e. we focus on the byte in position (0,0) - the proof for the other bytes is analogous. By simple computation, there exist constants c_i, d_i and e_i for $i = 0, ..., 3$ - which depend only on the secret key and by the constant b which defines $R(\mathcal{A}_\delta)$ - such that :

$$
\begin{aligned}
&[R^2(z) \oplus R^2(z')]_{0,0} = \\
&= \text{0x02} \cdot \text{S-Box}\big(\text{0x02} \cdot \text{S-Box}(z_0 \oplus d_0) \oplus \text{0x03} \cdot \text{S-Box}(\text{0x03} \cdot w \oplus e_0) \oplus c_0\big) \oplus \\
&\oplus \text{0x02} \cdot \text{S-Box}\big(\text{0x02} \cdot \text{S-Box}(z'_0 \oplus d_0) \oplus \text{0x03} \cdot \text{S-Box}(\text{0x03} \cdot w' \oplus e_0) \oplus c_0\big) \oplus \\
&\oplus \text{0x03} \cdot \text{S-Box}\big(\text{S-Box}(z_3 \oplus d_3) \oplus \text{0x02} \cdot \text{S-Box}(w \oplus e_1) \oplus c_1\big) \oplus \\
&\oplus \text{0x03} \cdot \text{S-Box}\big(\text{S-Box}(z'_3 \oplus d_3) \oplus \text{0x02} \cdot \text{S-Box}(w' \oplus e_1) \oplus c_1\big) \oplus \\
&\oplus \text{S-Box}\big(\text{0x02} \cdot \text{S-Box}(z_2 \oplus d_2) \oplus \text{0x03} \cdot \text{S-Box}(\text{0x02} \cdot w \oplus e_2) \oplus c_2\big) \oplus \\
&\oplus \text{S-Box}\big(\text{0x02} \cdot \text{S-Box}(z'_2 \oplus d_2) \oplus \text{0x03} \cdot \text{S-Box}(\text{0x02} \cdot w' \oplus e_2) \oplus c_2\big) \oplus \\
&\oplus \text{S-Box}\big(\text{S-Box}(z_1 \oplus d_1) \oplus c_3)\big) \oplus \text{S-Box}\big(\text{S-Box}(z'_1 \oplus d_1) \oplus c_3)\big) = \\
&= [R^2(v) \oplus R^2(v')]_{0,0}.
\end{aligned}
$$

More generally, there exist some constants $A, B, C \in \mathbb{F}_{2^8}$ such that each byte of $[R^2(z) \oplus R^2(z')]_{i,j} = [R^2(w) \oplus R^2(w')]_{i,j}$ for $i, j = 0, ..., 3$ can be written as:

$$
\begin{aligned}
&[R^2(z) \oplus R^2(z')]_{i,j} = [R^2(v) \oplus R^2(v')]_{i,j} = F(z_0, z'_0, z_2, z'_2, z_3, z'_3, w, w') \oplus \\
&\oplus A \cdot \text{S-Box}\big(B \cdot \text{S-Box}(z_1 \oplus k_{1,0}) \oplus C)\big) \oplus A \cdot \text{S-Box}\big(B \cdot \text{S-Box}(z'_1 \oplus k_{1,0}) \oplus C)\big).
\end{aligned}
\tag{30}
$$

Thirdly, consider $z, z' \in R(\mathcal{A}_\delta)$ generated respectively by $z \equiv (z_0, z_1, z_2, z_3, w)$ and $z' \equiv (z'_0, z'_1, z'_2, z'_3, w)$. The two texts satisfy $R^2(z) \oplus R^2(z') \in \mathcal{D}_I$ for $|I| = 3$ if four (particular) bytes (one per column) of $R^2(z) \oplus R^2(z')$ are equal to zero (remember that the bytes of $R^2(z) \oplus R^2(z')$ don't depend on z_1, z'_1). Since the two elements depend on $10 - 2 = 8$ variables and only 4 conditions must be satisfied, such elements z, z' can exist. A similar argumentation holds also for the case in which $z_1 = z'_1$. As a result, it follows that the number of collisions for the case $z_1 \neq z'_1$ is a multiple of 2.

Case: $\mathbf{z_1 = z'_1}$. As second case, we consider two elements $z, z' \in R(\mathcal{A}_\delta)$ generated respectively by $z \equiv (z_0, z_1, z_2, z_3, w)$ and $z' \equiv (z'_0, z'_1, z'_2, z'_3, w)$ with $z_1 = z'_1$.

First of all, note that if $z_{1,1} = z'_{1,1}$, then $z \oplus z' \in \mathcal{D}_{0,2,3}$. By Prop. 4, note that $R^4(z) \oplus R^4(z') \notin \mathcal{M}_I$ for all $I \in \{0, 1, 2, 3\}$ with $|I| = 1$. However, for the case $|I| = 3$ the idea is to prove that if $z, z' \in R(\mathcal{A}_\delta)$ satisfy the condition $R^2(z) \oplus R^2(z') \in \mathcal{D}_I$, then each pair of elements $v, v' \in R(\mathcal{A}_\delta)$ generated respectively by $v \equiv (z_0, v_1, z_2, z_3, w)$ and $v' \equiv (z'_0, v_1, z'_2, z'_3, w)$ for each $v_1 \in \mathbb{F}_{2^8}$ have the same property, that is $R^2(v) \oplus R^2(v') \in \mathcal{D}_I$. Since there are $2^8 = 256$ different values for v_1, then the number of collisions must be a multiple of 256.

This follows immediately by the fact that each byte of $R^2(z) \oplus R^2(z')$ doesn't depend on $z_1 = z'_1$. Indeed, if $z_1 = z'_1$, then each byte of $R^2(z) \oplus R^2(z')$ doesn't depend on $z_1 = z'_1$, i.e. by (30) it can be re-written as

$$[R^2(z) \oplus R^2(z')]_{i,j} = F(z_0, z'_0, z_2, z'_2, z_3, z'_3, w, w')$$

for a particular function $F(\cdot)$. For each pair of elements $v, v' \in R(\mathcal{A}_\delta)$ generated respectively by $v \equiv (z_0, v_1, z_2, z_3, w)$ and $v' \equiv (z'_0, v_1, z'_2, z'_3, w)$ follows immediately that $R^2(v) \oplus R^2(v') = R^2(z) \oplus R^2(z')$ for all v_1. That is, $R^2(v) \oplus R^2(v') \in \mathcal{D}_I$ if and only if $R^2(z) \oplus R^2(z') \in \mathcal{D}_I$ for all v_1. □

H.2 Proof of App. G.3

For a fixed a, consider a set of plaintexts $\mathcal{A}'_\delta$ of the form (29)

$$\mathcal{A}'_\delta \equiv \left\{ a \oplus \begin{bmatrix} 0 & y_0 & 0 & 0 \\ 0 & x & y_1 & 0 \\ 0 & 0 & x \oplus \delta_{2,2} & y_2 \\ y_3 & 0 & 0 & x \oplus \delta_{3,3} \end{bmatrix} \;\middle|\; \forall x, y_0, ..., y_3 \in \mathbb{F}_{2^8} \right\}$$

where $\delta = (\delta_{2,2}, \delta_{3,3})$.

Proposition 11. *Consider a set of plaintexts $\mathcal{A}'$ defined as in (29), and the corresponding ciphertexts after 5 rounds. If $\delta_{i,i} = k_{1,1} \oplus k_{i,i}$ for $i = 2, 3$, then the number of different pairs of ciphertexts that belong to the same coset of $\mathcal{M}_I$ for a fixed $I \subseteq \{0, 1, 2, 3\}$ with $|I| = 3$ is a multiple of 4.*

Proof. Let $\delta_{2,2} = k_{1,1} \oplus k_{2,2}$ and $\delta_{3,3} = k_{1,1} \oplus k_{3,3}$. By simple computation, there exists b such that the set $\mathcal{A}'_\delta$ is mapped after one round into

$$R(\mathcal{A}'_\delta) \equiv \left\{ b \oplus \begin{bmatrix} 0\text{x}03 \cdot w & z_0 & 0 & 0 \\ 0 & z_1 & 0 & 0 \\ 0 & z_2 & 0 & 0 \\ 0\text{x}02 \cdot w & z_3 & 0 & 0 \end{bmatrix} \;\middle|\; \forall w, z_0, ..., z_3 \in \mathbb{F}_{2^8} \right\}.$$

Consider two elements $z, z' \in R(\mathcal{A}'_\delta)$ generated respectively by $z \equiv (z_0, z_1, z_2, z_3, w)$ and $z' \equiv (z'_0, z'_1, z'_2, z'_3, w)$. In the following, we consider separately the cases (1) $z_2 \neq z'_2$ and $z_3 \neq z'_3$, (2) $z_2 = z'_2$ and $z_3 = z'_3$ and (3) $z_2 = z'_2$ and $z_3 \neq z'_3$ (or viceversa). We show that in the first case the number of collisions is a multiple of 4, in the second case it is a multiple of 2^{16} and in the third case it is a multiple of 2^9. It follows that there exist $n', n'', n''' \in \mathbb{N}$ such that the total number of collisions n can be written as $n = 4 \cdot n' + 2^{16} \cdot n'' + 2^{10} \cdot n''' = 4 \cdot (n' + 2^{14} \cdot n'' + 2^8 \cdot n''')$. In other words, the total number of collisions is a multiple of 4.

Case: $\mathbf{z_2 \neq z'_2}$ and $\mathbf{z_3 \neq z'_3}$. Consider two elements $z, z' \in R(\mathcal{A}_\delta)$ generated respectively by $z \equiv (z_0, z_1, z_2, z_3, w)$ and $z' \equiv (z'_0, z'_1, z'_2, z'_3, w)$ with $z_2 \neq z'_2$ and $z_3 \neq z'_3$. For a fixed $I \in \{0, 1, 2, 3\}$ with $|I| = 3$, the idea is to show that $R^4(z) \oplus R^4(z') \in \mathcal{M}_I$ *if and only if* $R^4(v) \oplus R^4(v') \in \mathcal{M}_I$ where the texts $v, v' \in R(\mathcal{A}_\delta)$ are generated respectively by the following combinations:

- $v \equiv (z_0, z_1, z'_2, z_3, w)$ and $v' \equiv (z'_0, z'_1, z_2, z'_3, w)$;
- $v \equiv (z_0, z_1, z_2, z'_3, w)$ and $v' \equiv (z'_0, z'_1, z'_2, z_3, w)$;
- $v \equiv (z_0, z_1, z'_2, z'_3, w)$ and $v' \equiv (z'_0, z'_1, z_2, z_3, w)$.

This follows by Theorem 2 of [GRR17a] - recalled in Sect. 4, and implies that the number of collision must a multiple of 4 for this case.

For more details, Let v and v' defined as before. As before, it is sufficient to prove that (1) $R^2(z) \oplus R^2(z') = R^2(v) \oplus R^2(v')$ and (2) that $z, z' \in R(\mathcal{A}_\delta)$ can exist such that $R^4(z) \oplus R^4(z') \in \mathcal{M}_I$. Since the proof of these two facts is equivalent to that given in App. H.1, we refer to that section for more details and we limit here to highlight the major differences.

By simple computation, the first point is due to the fact that there exist some constants $A, B, C, D, E, F \in \mathbb{F}_{2^8}$ such that each byte of $[R^2(z) \oplus R^2(z')]_{i,j} = [R^2(v) \oplus R^2(v')]_{i,j}$ for $i, j = 0, ..., 3$ can be written as:

$$\begin{aligned}
&[R^2(z) \oplus R^2(z')]_{i,j} = [R^2(v) \oplus R^2(v')]_{i,j} = F(z_0, z'_0, z_1, z'_1, w, w') \oplus \\
&\oplus A \cdot \text{S-Box}\big(B \cdot \text{S-Box}(z_2 \oplus k_{2,1}) \oplus C\big)\big) \oplus A \cdot \text{S-Box}\big(B \cdot \text{S-Box}(z'_2 \oplus k_{2,1}) \oplus C\big)\big) \oplus \\
&\oplus D \cdot \text{S-Box}\big(E \cdot \text{S-Box}(z_3 \oplus k_{3,1}) \oplus F\big)\big) \oplus D \cdot \text{S-Box}\big(E \cdot \text{S-Box}(z'_3 \oplus k_{3,1}) \oplus F\big)\big).
\end{aligned} \tag{31}$$

As an example, the first byte of $[R^2(z) \oplus R^2(z')]_{0,0}$ (analogous for the others):

$$\begin{aligned}
&[R^2(z) \oplus R^2(z')]_{0,0} = \\
&= 0\text{x}02 \cdot \text{S-Box}\big(0\text{x}03 \cdot \text{S-Box}(z_1 \oplus d_1) \oplus 0\text{x}02 \cdot \text{S-Box}(0\text{x}02 \cdot w \oplus e_0) \oplus c_0\big) \oplus \\
&\oplus 0\text{x}02 \cdot \text{S-Box}\big(0\text{x}03 \cdot \text{S-Box}(z'_1 \oplus d_1) \oplus 0\text{x}02 \cdot \text{S-Box}(0\text{x}02 \cdot w' \oplus e_0) \oplus c_0\big) \oplus \\
&\oplus 0\text{x}03 \cdot \text{S-Box}\big(0\text{x}03 \cdot \text{S-Box}(z_0 \oplus d_0) \oplus 0\text{x}02 \cdot \text{S-Box}(0\text{x}02 \cdot w \oplus e_1) \oplus c_1\big) \oplus \\
&\oplus 0\text{x}03 \cdot \text{S-Box}\big(0\text{x}03 \cdot \text{S-Box}(z'_0 \oplus d_0) \oplus 0\text{x}02 \cdot \text{S-Box}(0\text{x}02 \cdot w' \oplus e_1) \oplus c_1\big) \oplus \\
&\oplus \text{S-Box}\big(0\text{x}02 \cdot \text{S-Box}(z_2 \oplus d_2) \oplus c_2\big) \oplus \text{S-Box}\big(0\text{x}02 \cdot \text{S-Box}(z'_2 \oplus d_2) \oplus c_2\big) \oplus \\
&\oplus \text{S-Box}\big(0\text{x}02 \cdot \text{S-Box}(z_3 \oplus d_3) \oplus c_3\big) \oplus \text{S-Box}\big(0\text{x}02 \cdot \text{S-Box}(z'_3 \oplus d_3) \oplus c_3\big) = \\
&= [R^2(v) \oplus R^2(v')]_{0,0} =
\end{aligned}$$

where the constants c_i, d_i and e_i depend only on the secret key and by the constant b which defines $R(\mathcal{A}'_\delta)$.

Secondly, consider $z, z' \in R(\mathcal{A}_\delta)$ generated respectively by $z \equiv (z_0, z_1, z_2, z_3, w)$ and $z' \equiv (z'_0, z'_1, z'_2, z'_3, w)$. The two elements satisfy $R^2(z) \oplus R^2(z') \in \mathcal{D}_I$ for $|I| = 3$ if four (particular) bytes (one per column) of $R^2(z) \oplus R^2(z')$ are equal to zero (remember that the bytes of $R^2(z) \oplus R^2(z')$ don't depend on z_i, z'_i for $i = 2, 3$). Since the two elements depend on $10 - 4 = 6$ variables and only 4 conditions must be satisfied, such elements z, z' can exist. A similar argumentation holds also for the other cases.

Case: $\mathbf{z_2 = z'_2}$ and $\mathbf{z_3 = z'_3}$. As second case, we consider two elements in $z, z' \in R(\mathcal{A}_\delta)$ generated respectively by $z \equiv (z_0, z_1, z_2, z_3, w)$ and $z' \equiv (z'_0, z'_1, z'_2, z'_3, w)$ with $z_2 = z'_2$ and $z_3 = z'_3$.

In this case, the idea is to prove that if $z, z' \in R(\mathcal{A}_\delta)$ satisfy the condition $R^2(z) \oplus R^2(z') \in \mathcal{D}_I$, then each pair of texts $v, v' \in R(\mathcal{A}_\delta)$ generated respectively by $v \equiv (z_0, z_1, v_2, v_3, w)$ and $v' \equiv (z'_0, z'_1, v_2, v_3, w)$ for all $v_2, v_3 \in \mathbb{F}_{2^8}$ have the same property, that is $R^2(v) \oplus R^2(v') \in \mathcal{D}_I$. Since there are $2^8 \cdot 2^8 = 2^{16}$ different values for v_2, v_3, then the number of collisions must be a multiple of 2^{16}.

As for the proof given in App. H.1, this follows by the fact that each byte of $R^2(z) \oplus R^2(z')$ doesn't depend on $z_2 = z'_2$ and $z_3 = z'_3$. Indeed, if for $z_2 = z'_2$ and $z_3 = z'_3$ and by (31), each byte of $R^2(z) \oplus R^2(z')$ depends on the following variables

$$[R^2(z) \oplus R^2(z')]_{i,j} = F(z_0, z'_0, z_1, z'_1, w, w')$$

for a particular function $F(\cdot)$. For each pair of elements $v, v' \in R(\mathcal{A}_\delta)$ generated respectively by $v \equiv (z_0, z_1, v_2, v_3, w)$ and $v' \equiv (z'_0, z'_1, v_2, v_3, w)$ follows immediately that

$R^2(v) \oplus R^2(v') = R^2(z) \oplus R^2(z')$ for all v_1. That is, $R^2(v) \oplus R^2(v') \in \mathcal{D}_I$ if and only if $R^2(z) \oplus R^2(z') \in \mathcal{D}_I$ for all v_1.

Case: $\mathbf{z_2 \neq z'_2}$ and $\mathbf{z_3 = z'_3}$. As final case, we consider two elements $z, z' \in R(\mathcal{A}_\delta)$ generated respectively by $z \equiv (z_0, z_1, z_2, z_3, w)$ and $z' \equiv (z'_0, z'_1, z'_2, z'_3, w)$ with $z_2 \neq z'_2$ and $z_3 = z'_3$ - analogous for $z_2 = z'_2$ and $z_3 \neq z'_3$.

Using similar argumentations as before, in this case the idea is to prove that if $z, z' \in R(\mathcal{A}_\delta)$ satisfy the condition $R^2(z) \oplus R^2(z') \in \mathcal{D}_I$, then each pair of elements $v, v' \in R(\mathcal{A}_\delta)$ generated respectively by

- $v \equiv (z_0, z_1, z_2, v_3, w)$ and $v' \equiv (z'_0, z'_1, z'_2, v_3, w)$;
- $v \equiv (z_0, z_1, z'_2, v_3, w)$ and $v' \equiv (z'_0, z'_1, z_2, v_3, w)$;

for all $v_3 \in \mathbb{F}_{2^8}$ have the same property. Since there are 2^8 different values for v_3, then the number of collisions must be a multiple of $2 \cdot 2^8 = 512$. □

H.3 Proof of Sect. 6.3

For a fixed a, consider a set of plaintexts $\mathcal{A}''_\delta$ of the form (17):

$$\mathcal{A}''_\delta \equiv \left\{ a \oplus \begin{bmatrix} 0 & y & 0 & 0 \\ 0 & x & y \oplus \delta_{1,2} & 0 \\ 0 & 0 & x \oplus \delta_{2,2} & w \oplus \delta_{2,3} \\ 0 & 0 & 0 & x \oplus \delta_{3,3} \end{bmatrix} \;\middle|\; \forall x, y \in \mathbb{F}_{2^8} \right\}$$

where $\delta \equiv (\delta_{1,2}, \delta_{2,2}, \delta_{2,3}, \delta_{3,3})$.

Proposition 12. *Consider a set of plaintexts $\mathcal{A}''_\delta$ defined as in (17), and the corresponding ciphertexts after 5 rounds. If $\delta_{i,i} = k_{1,1} \oplus k_{i,i}$ and $\delta_{j,j+1} = k_{0,1} \oplus k_{j,j+1}$ for $i = 2, 3$ and $j = 1, 2$ (where the indexes are taken modulo 4), then the number of different pairs of ciphertexts that belong to the same coset of $\mathcal{M}_I$ for a fixed $I \subseteq \{0, 1, 2, 3\}$ with $|I| = 3$ is a multiple of 2.*

Proof. Let $\delta_{i,i} = k_{i,i} \oplus k_{1,1}$ for $i = 2, 3$ and $\delta_{j,j+1} = k_{j,j+1} \oplus k_{0,1}$ for $j = 1, 2$. By simple computation, there exists a constant b such that $\mathcal{A}''_\delta$ is mapped into

$$R(\mathcal{A}''_\delta) \equiv \left\{ b \oplus \begin{bmatrix} 0\text{x}03 \cdot z & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0\text{x}02 \cdot w & 0 & 0 \\ 0\text{x}02 \cdot z & 0\text{x}03 \cdot w & 0 & 0 \end{bmatrix} \;\middle|\; \forall z, w \in \mathbb{F}_{2^8} \right\}.$$

Consider a pair of texts $t^1, t^2 \in R(\mathcal{A}''_\delta)$ generated respectively by $t^1 = (z, w)$ and $t^2 = (z', w')$. We consider the following two cases separately: (1) $z = z'$ and $w \neq w'$ (or viceversa) and (2) $z \neq z'$ and $w \neq w'$. We show that in the first case (1) the number of collisions is a multiple of 256, while in the second case (2) the number of collisions is a multiple of 2. Thus, there exist $n', n'' \in \mathbb{N}$ such that the total number of collisions n can be written as $n = 2 \cdot n' + 256 \cdot n'' = 2 \cdot (n' + 128 \cdot n'')$, that is n is a multiple of 2.

Case: $\mathbf{z \neq z'}$ and $\mathbf{w \neq w'}$. Consider a pair of texts $t^1, t^2 \in R(\mathcal{A}''_\delta)$ generated respectively by $t^1 = (z, w)$ and $t^2 = (z', w')$ with $z \neq z'$ and $w \neq w'$.

Similar to the previous proofs, the idea is to show that $R^4(t^1) \oplus R^4(t^2) \in \mathcal{M}_I$ *if and only if* $R^4(s^1) \oplus R^4(s^2) \in \mathcal{M}_I$ for $|I| = 3$, where the texts $s^1, s^2 \in R(\mathcal{A}''_\delta)$ are generated respectively by $s^1 = (z, w')$ and $s^2 = (z', w)$. Since each coset of $\mathcal{M}_I$ is mapped two round before into a coset of $\mathcal{D}_I$ (i.e. for each $a \in \mathcal{M}_I^\perp$ there exists unique $b \in \mathcal{D}_I^\perp$ such that

$R^{-2}(\mathcal{M}_I \oplus a) = \mathcal{D}_I \oplus b)$, it is sufficient to prove that $R^2(t^1) \oplus R^2(t^2) \in \mathcal{D}_I$ for $|I| = 3$ if and only if $R^2(s^1) \oplus R^2(s^2) \in \mathcal{D}_I$ in order to guarantee that $R^4(s^1) \oplus R^4(s^2) \in \mathcal{M}_I$. To do this, we show that each byte of $R^2(t^1) \oplus R^2(t^2)$ is equal to each byte of $R^2(s^1) \oplus R^2(s^2)$, that is:

$$[R^2(t^1) \oplus R^2(t^2)]_{i,j} = [R^2(s^1) \oplus R^2(s^2)]_{i,j}$$

for $i, j = 0, ..., 3$. By simple computation, there exist constants c, d - that depend only on the secret key and on b which defined $R(\mathcal{A}_\delta^{''})$ - such that:

$$R^2(\mathcal{A}_\delta^{''}) \equiv c \oplus M^{MC} \times \begin{bmatrix} \text{S-Box}(z_0) & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \text{S-Box}(w_1) \\ 0 & \text{S-Box}(z_1) & \text{S-Box}(w_0) & 0 \end{bmatrix}$$

where

$$\begin{aligned} z_0 &= \text{0x03} \cdot z \oplus d_{0,0}, & z_1 &= \text{0x02} \cdot z \oplus d_{3,0}, \\ w_0 &= \text{0x03} \cdot w \oplus d_{3,1} & w_1 &= \text{0x02} \cdot w \oplus d_{2,1} \end{aligned}$$

for all $z, w \in \mathbb{F}_{2^8}$. It follows that each byte of $[R^2(t^1) \oplus R^2(t^2)]_{i,j} = [R^2(s^1) \oplus R^2(s^2)]_{i,j}$ for $i, j = 0, ..., 3$ can be re-written as:

$$\begin{aligned} &[R^2(t^1) \oplus R^2(t^2)]_{i,j} = \\ =&A_0 \cdot \text{S-Box}(B_0 \cdot \text{S-Box}(z_0) \oplus C_0) \oplus A_0 \cdot \text{S-Box}(B_0 \cdot \text{S-Box}(z_0') \oplus C_0) \oplus \\ \oplus&A_1 \cdot \text{S-Box}(B_1 \cdot \text{S-Box}(z_1) \oplus C_1) \oplus A_1 \cdot \text{S-Box}(B_1 \cdot \text{S-Box}(z_1') \oplus C_1) \oplus \\ \oplus&A_2 \cdot \text{S-Box}(B_2 \cdot \text{S-Box}(w_0) \oplus C_2) \oplus A_2 \cdot \text{S-Box}(B_2 \cdot \text{S-Box}(w_0') \oplus C_2) \oplus \\ \oplus&A_3 \cdot \text{S-Box}(B_3 \cdot \text{S-Box}(w_1) \oplus C_3) \oplus A_3 \cdot \text{S-Box}(B_3 \cdot \text{S-Box}(w_1') \oplus C_3) = \\ =&[R^2(s^1) \oplus R^2(s^2)]_{i,j} \end{aligned} \tag{32}$$

for some constants A_i, B_i, C_i that depend only on the secret key and on c, d which define $R^2(\mathcal{A}_\delta^{''})$, that is the thesis.

Case: $\mathbf{z \neq z'}$ and $\mathbf{w = w'}$. Consider a pair of texts $t^1, t^2 \in R(\mathcal{A}_\delta^{''})$ generated respectively by $t^1 = (z, w)$ and $t^2 = (z', w')$, with the condition $z \neq z'$ and $w = w'$ (or viceversa). By definition of $\mathcal{D}_J$, the two elements belong to the same coset of $\mathcal{D}_{0,3}$ (or more generally of $\mathcal{D}_J$ for $|J| = 2$). By Prop. 1, it follows that the two texts can not belong to the same coset of $\mathcal{M}_I$ for $|I| \leq 2$, but no restriction holds for the case $\mathcal{M}_I$ for $|I| = 3$.

Using similar argumentations of before, the idea is to prove that if $t^1, t^2 \in R(\mathcal{A}_\delta^{''})$ satisfy the condition $R^4(t^1) \oplus R^4(t^2) \in \mathcal{M}_I$ for $|I| = 3$, then all the pairs of texts $s^1, s^2 \in R(\mathcal{A}_\delta^{''})$ generated respectively by $t^1 = (z, s)$ and $t^2 = (z', s)$ for all $s \in \mathbb{F}_{2^8}$ have the same property. To do this, it is sufficient to show that $[R^2(t^1) \oplus R^2(t^2)]_{i,j} = [R^2(s^1) \oplus R^2(s^2)]_{i,j}$ for $i, j = 0, ..., 3$. By previous considerations - see (32), it follows that if $w = w'$ then $[R^2(t^1) \oplus R^2(t^2)]_{i,j}$ depends only on z and z', that is it is independent of w, w'. This implies the thesis, that is the number of collisions for this case must be a multiple of 256. □

H.4 Final Considerations of App. 6.1 - Details

As last thing, one may ask what is the probability that a random matrix M^{MC} satisfies one of the two following requirements:

- for each row, at least two elements are equal;
- for each row, the XOR-sum of at least two elements is equal to zero.

Since designers usually choose an MDS (Maximal Distance Separable) circulant[32] matrices, we limit to consider such kind of $n \times n$ matrix with elements in $GF(2^m)$ for our analysis. In particular, since the elements of the rows of a circulant matrix are identical, we focus on a single generic row.

We emphasize that our goal is only to give a (rough) estimation of this ratio, and not to give the exact number. Thus, we simply consider the number of *all* the matrices with two identical elements for each row and for which the sum of some elements is zero, without worrying about the condition that the matrix is invertible and about the MDS property.

First of all, note that if $n > 2^m$, then at least two elements must be equal. Thus, for the following we limit to consider the case $n \leq 2^m$. By simple computation, the number of circulant matrices with (at least) two identical elements is given by

$$(2^m)^n - \frac{2^m!}{(2^m - n)!}$$

that is the total number of matrices minus the number of matrices with all different elements. Note that

$$(2^m)^n - \frac{2^m!}{(2^m - n)!} = 2^{m\cdot(n-1)} \sum_{i=1}^{n+1} i - 2^{m\cdot(n-2)} \cdot \sum_{i=1}^{n+1} \sum_{j=1,\, j\neq i}^{n+1} i \cdot j + ... \approx 2^{m\cdot(n-1)} \cdot \frac{n^2}{2}$$

where the approximation[33] holds if $2^{m+1} \gg n^2 + 5 \cdot n$.

Note that a similar result can be obtained in a different way. In particular, the number of n sets of elements in $\{0, 1, ..., 2^m - 1\}$ for which two elements is well approximated by

$$(2^m)^{n-1} \times \binom{n}{2} = 2^{m\cdot(n-1)} \times \frac{n \cdot (n-1)}{2}.$$

To give some concrete numbers, in the AES case (that is, $n = 4$ and $m = 8$), the first number is equal to $(2^m)^n - \frac{2^m!}{(2^m-n)!} = 99\,943\,936 \simeq 2^{26.575}$ while the second one is equal to $2^{m\cdot(n-1)} \times \frac{n\cdot(n-1)}{2} = 100\,663\,296 \simeq 2^{26.585}$.

In a similar way, the number of n sets of elements in $\{0, 1, ..., 2^m - 1\}$ for which the sum of two or more elements is equal to zero is well approximated by[34]

$$(2^m)^{n-1} \times \sum_{i=2}^{n} \binom{n}{i} = 2^{m\cdot(n-1)} \times \big(2^n - n - 1\big).$$

It follows that the ratio between the number of matrices for which the sum of some elements is equal to zero with respect to the ones for which (at least) two elements are equal is well approximated by

$$\frac{2^{n+1} - 2 \cdot n - 2}{n^2 - n} \approx \frac{2^{n+1}}{n^2}.$$

Note that this ratio increases with n and it is independent of m. In order to give an example, for the AES case (that is $n = 4$ and $m = 8$ - note that the condition $2^{m+1} = 512 \gg 36 = n^2 + 5 \cdot n$ is satisfied) this ratio is approximately equal to $11/6 \approx 2$.

[32] A circulant matrix is a matrix where each row vector is rotated one element to the right relative to the preceding row vector matrices.

[33] By computation:

$$\sum_{i=1}^{n+1} i = \frac{(n+1)\cdot(n+2)}{2} \quad \text{and} \quad \sum_{i=1}^{n+1} \sum_{j=1,\, j\neq i}^{n+1} i \cdot j = \frac{(n+1)^2 \cdot (n+2)^2}{4} - \frac{(n+1)\cdot(n+2)\cdot(2n+3)}{6}.$$

Thus $2^{m\cdot(n-1)} \cdot \frac{n^2}{2} \gg 2^{m\cdot(n-2)} \cdot \frac{3n^4+14n^3}{6}$ if $2^{m+1} \gg n^2 + 5 \cdot n$.

[34] Remember that $\sum_{i=0}^{n} \binom{n}{i} = 2^n$.

For completeness, a rough approximation of the same ratio for *generic* matrix is given by $\left(\frac{2^{n+1}}{n^2}\right)^n$ under the same assumption of the previous case. This rough result can be simply obtained by assuming that the n rows are independent. For the AES case, this ration is approximately equal to $2^4 = 16$.