MixColumns Properties and Attacks on (round-reduced) AES with a Single Secret S-Box

Lorenzo Grassi

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

Abstract. In this paper, we present new key-recovery attacks on AES with a single secret S-Box. Several attacks for this model have been proposed in literature, the most recent ones at Crypto'16 and FSE'17. Both these attacks exploit a particular property of the MixColumns matrix to recover the secret-key.
In this work, we show that the same attacks work exploiting a weaker property of the MixColumns matrix. As first result, this allows to (largely) increase the number of MixColumns matrices for which it is possible to set up all these attacks. As a second result, we present new attacks on 5-round AES with a single secret S-Box that exploit the new multiple-of-n property recently proposed at Eurocrypt'17. This property is based on the fact that choosing a particular set of plaintexts, the number of pairs of ciphertexts that lie in a particular subspace is a multiple of n.

Keywords: AES, MixColumns, key-recovery attack, secret S-Box

1 Introduction

A key-recovery attack is any adversary's attempt to recover the cryptographic key of an encryption scheme. As stated by the Kerckhoffs Principle, one common assumption is that the security of a cryptosystem must lie in the choice of its keys only: everything else (including the algorithm itself) should be considered public knowledge. *What happens if part of the crypto-system is instead kept secret?*

This problem has been first introduced by Biryukov and Shamir [6], where authors studied the security of AES-like ciphers which contain alternate (secret) layers of invertible S-Boxes and (secret) affine mappings. In particular, they analyzed an AES-like cipher with 128-bit blocks using eight-bit S-Boxes. An attack was presented on five layers (SASAS, where S stands for substitution and A stands for affine mapping) of this construction which finds all secret components (up to an equivalence). Using the terminology of "rounds" as in the AES, this version consists of two and a half rounds.

After this first work, several other results regarding cryptanalysis of ciphers with secret S-Boxes have been presented in literature. To cite some examples, Gilbert and Chauvaud [14] presented a differential attack on the cipher Khufu (an unbalanced Feistel cipher), while Vaudenay provided cryptanalysis of reduced-round variants of Blowfish [22]. Most recently, the lightweight cipher PRESENT (standardized ISO) was cryptanalyzed by Borghof *et al.* [9] also in

the (extreme) case in which the S-Boxes are chosen uniformly at random for each round. In [5], authors considered the ASASA scheme in order to design public key or white-box constructions using symmetric cipher components.

Focusing on AES, several works considered the security of this cipher in the case in which the S-Box is replaced by a secret S-Box, about which the adversary has no knowledge. At FSE 2015 Tiessen *et al.* [21] presented the first attack up to 6-round AES with a single secret S-Box, based on the integral technique [11]. At Crypto 2016, Sun *et al.* [20] proposed the first key-dependent distinguisher on 5-round AES with a single secret S-Box, based on zero-correlation linear hulls [8]. Such distinguisher has been then improved by Grassi *et al.* at FSE 2017 [16], using a technique based on impossible differential cryptanalysis [2, 17, 3].

State of the Art and Our Contributions

Background. The Advanced Encryption Standard (AES) [12] is an iterated block cipher using 10, 12, or 14 rounds depending on the key size of 128, 192, or 256 bits. These variants are named AES-128, AES-192, and AES-256. In this paper we focus on the cipher that is derived from the AES by replacing the S-Box with a secret 8-bit S-Box while keeping everything else unchanged. If the choice of S-Box is made uniformly at random from all 8-bit S-Boxes[1], the size of the secret information increases from 128 - 256 bits (the key size in the AES) to $128 + \log_2(2^8!) = 1812$ and $256 + \log_2(2^8!) = 1940$ bits respectively.

To better understand the attacks on AES with a single secret S-Box, we briefly recall few details of AES. Without going into the details here, AES is a key-iterated block cipher that consists of the repeated application of a round transformation on the state (called intermediate result). Each round transformation is a sequence of four steps, an S-Box (the only non-linear operation), a ShiftRows (a permutation on the byte positions), a MixColumns matrix (a linear operation) and the AddRoundKey.

The attacks on AES with a single secret S-Box present in literature can be divided in *two categories*:

1. in the first case (e.g. [6] and [21]), the attacker first determines the secret S-Box up to additive constants (that is, $\text{S-Box}(x \oplus a) \oplus b$ for unknown a and b), and then she uses this knowledge and applies attacks present in literature (e.g. the integral one) to derive the whitening key;
2. in the second case (e.g. [20] and [16]), the attacker exploits a particular property of the MixColumns matrix (i.e. the fact that two elements for each row of the matrix are equal) in order to find *directly* the secret key.

In this second strategy, *no information of the secret S-Box is derived and/or exploited to find the key.* This second strategy is so generic that can be applied to integral, truncated differential and impossible differential attack. In this case, the idea of the attack is to choose a set of plaintexts that depends on some guessed bytes of the key. Exploiting the fact that particular properties holds

[1] For completeness, we mention that a randomly chosen S-Box is likely to have good properties against differential and linear cryptanalysis, as shown in [21].

Table 1. *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 (20 table look-ups $\approx$ 1-round encryption). 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. Symbol * denotes an attack of the $1st$ category (as defined in the main text).

Attack	Rounds	Data	Computation	Memory	Reference
I*	4.5 - 5	2^{40} CC	$2^{38.7}$ E	2^{40}	[21]
I*	4.5 - 5	2^{40} CP	$2^{54.7}$ E	2^{40}	[21, Sect. 3.5]
Mult-of-n	**4.5 – 5**	$\mathbf{2^{53.25}}$ **CP**	$\mathbf{2^{59.25}}$ **M** $\approx \mathbf{2^{52.6}}$ **E**	$\mathbf{2^{16}}$	**Sect. 5.2**
Mult-of-n	**4.5 – 5**	$\mathbf{2^{53.6}}$ **CP**	$\mathbf{2^{55.6}}$ **M** $\approx \mathbf{2^{48.96}}$ **E**	$\mathbf{2^{40}}$	**Sect. 3.1**
ImD	**4.5 – 5**	$\mathbf{2^{76.37}}$ **CP**	$\mathbf{2^{81.54}}$ **M** $\approx \mathbf{2^{74.9}}$ **E**	$\mathbf{2^{8}}$	**Sect. 5.1**
ImD	4.5 - 5	2^{102} CP	2^{107} M $\approx 2^{100.4}$ E	2^8	[16]
I	5	2^{128} CC	$2^{129.6}$ XOR	small	[20]

I: Integral, ImD: Impossible Differential, Mult-of-n: Multiple-of-n

with higher probability for the right key than for the wrongly guessed one, it is possible to find the secret key.

Our Contributions. In this paper, we focus only on this second strategy, and we propose the following contributions.

First Contribution. As first contribution, in Sect. 4 we generalize the strategy proposed in [20] and in [16]. While attacks proposed in these papers 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 the XOR-sum of more than two coefficients of each row of the MixColumns matrix is equal to zero. As main result, the strategies proposed in [20] and in [16] work for a bigger class of MixColumns matrices. Moreover, in some cases this allows to improve the data and/or the computational costs of some attacks proposed in [20] and in [16], as the the impossible differential attack on 5-round AES with a single secret S-Box (see Sect. 5.1 for details).

Second Contribution. Recently, Grassi *et al.* [15] present the first secret-key distinguisher on 5-round AES which is *independent* of the secret key. 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[2] is always a multiple of 8. In Sect. 5, we show how to exploit an equivalent property to set up new (competitive) key-recovery attacks on 5-round AES with a single secret S-Box. In particular, by appropriate choice of a set of

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

plaintexts (that depends on the guessed key), 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 for the rightly guessed key, while this happens only with probability strictly less than 1 for wrongly guessed keys.

Potential Impact of Our Results

Round-Reduced AES as Part of New Designs. Many constructions employ reduced round AES as part of their design. Reduced versions of AES have nice and well-studied properties that can be favorable as components of larger designs. Only to cite some of them, in the on-going "Competition for Authenticated Encryption: Security, Applicability, and Robustness" (CAESAR) [1] which is currently at its third round, among many others, AEGIS [23] uses five AES round-functions in the state update functions, while ELmD v1.0 [13] recommends to use round-reduced AES including 5-round AES to partially encrypt the data[3]. In a very different context, Mennink and Neves [19] propose a method for transforming a dedicated block-cipher design into a dedicated PRF design. The main proposal AES-PRF-128 is defined to be AES xored with the internal state after 5 rounds, that is $AES\text{-}PRF(\cdot) = AES_{10}(\cdot) \oplus AES_5(\cdot)$.

Since the security level of AES-like cipher with a single secret S-Box could be very high (e.g. 1812-1940 bits) and since many constructions employ reduced round AES as part of their design, a natural question arises: *Could the number of rounds of AES-like cipher be reduced to fewer than 10 rounds (as in AES-128) in the case of secret S-Box?* The answer seems to be negative, since our results - together with the ones already present in literature - show that, despite the increased size of the secret information in the cipher, key-recovery attacks on round-reduced AES with a single secret S-Box are still possible.

MixColumns Matrix Design. The security of a block cipher depends on the details of the S-Box function and of the mixing linear transformation. If one chooses such functions carefully, the dedicated cipher based on the AES-like structure can be resilient to both differential [4] and linear cryptanalysis [18]. For example, based on the fact that the branch number of the AES MixColumns is 5, it is proved in [12] that the number of active S-boxes of 4-round AES is at least 25. Since the maximal differential probability of the S-Box is 2^{-6}, there does not exist any differential characteristic[4] of 4-round AES with probability larger than 2^{-150}.

Focusing only on the mixing linear transformation, in order to increase the performance of a block cipher, designers usually use a *circulant matrix* whose elements are restricted to low hamming weights in order to reduce the workload of the multiplications over finite fields. Furthermore, not only the matrix are always circulant, but also there are usually identical elements in each row.

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

[4] For completeness, we remark that bounding characteristic probability is not enough to prove resistance against other kinds of differential and linear attacks.

Most known cryptanalysis techniques don't make use of these observations, and there is little literature concentrating on the choices of these matrices in constructing distinguishers of round-reduced AES. On the other hand, our results - together with the ones already present in literature - show that some properties of the MixColumns matrix can be exploited to set up key-recovery attacks on AES-like cipher with a single secret S-Box. Thus, when designing an AES-like cipher, it seems better to choose those MDS matrices MC s.t. no XOR-sum of two or more coefficients of each row of both MC and MC^{-1} is equal to zero.

2 Preliminary

2.1 Description of AES

The Advanced Encryption Standard [12] 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 ;
- *MixColumns* (MC) -multiplication of each column by a constant 4×4 invertible matrix M^{MC} (MC and SR provide diffusion in the cipher[5]);
- *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. The secret key is usually denoted by k. We denote by R one round[6] 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 $\mathcal{X}$.

2.2 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* [16], recently introduced at FSE 2017.

[5] SR makes sure column values are spread, MC makes sure each column is mixed.
[6] Sometimes we use the notation R_k instead of R to highlight the round key k.

Definition 1. *Let* $(V_1, V_2, ..., V_{r+1})$ *denote a set of* $r+1$ *subspaces with* $\dim(V_i) \leq \dim(V_{i+1})$*. If for each* $i = 1, ..., r$ *and for each* $a_i \in V_i^{\perp}$*, there exist (unique)* $a_{i+1} \in V_{i+1}^{\perp}$ *such that* $F(V_i \oplus a_i) \subseteq V_{i+1} \oplus a_{i+1}$*, then* $(V_1, V_2, ..., V_{r+1})$ *is subspace trail of length* r *for the function* F*.*

This means that if F^t denotes the application of t rounds with fixed keys, then $F^t(V_1 \oplus a_1) = V_{t+1} \oplus a_{t+1}$. We refer to [16] for more details about the concept of subspace trails. Our treatment here is however meant to be self-contained.

Subspace Trails of AES. Here we recall the subspace trails of AES presented in [16], 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 [16], for any coset $\mathcal{D}_I \oplus a$ there exists unique $b \in \mathcal{C}_I^\perp$ such that $R(\mathcal{D}_I \oplus a) = \mathcal{C}_I \oplus b$. Similarly, for any coset $\mathcal{C}_I \oplus a$ there exists unique $b \in \mathcal{M}_I^\perp$ such that $R(\mathcal{C}_I \oplus a) = \mathcal{M}_I \oplus b$.

Theorem 1 ([16]). *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 emphasize that b depends on the initial constant a and on the secret key k.

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 $I, J \subseteq \{0,1,2,3\}$, then $\mathcal{M}_I \cap \mathcal{D}_J = \{0\}$ if and only if $|I| + |J| \leq 4$, as demonstrated in [16]. It follows that:

Proposition 1 ([16]). *Let $I, J \subseteq \{0,1,2,3\}$ such that $|I| + |J| \leq 4$. For all x, y with $x \neq y$:*

$$Prob(R^4(x) \oplus R^4(y) \in \mathcal{M}_I \,|\, x \oplus y \in \mathcal{D}_J) = 0. \tag{3}$$

We remark that all these results can be re-described using a more "classical" - but equivalent - truncated differential notation, as formally pointed out in [7]. To be more concrete, if two texts t^1 and t^2 are equal expect for the bytes in the i-th diagonal[7] for each $i \notin I$, 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 $\mathcal{ID}_I$.

5-round Secret-Key Distinguisher proposed in [15]. For the following, we briefly recall the property exploited in [15] to set up the first 5-round secret-key distinguisher of AES (independent of the secret key).

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, that is $(p^i, c^i \equiv R^5(p^i))$ for $i = 0, ..., 2^{32 \cdot |I|} - 1$ such that $p^i \in \mathcal{D}_I \oplus a$ for all i. The 5-round AES distinguisher proposed in [15] exploits the fact that the number of different pairs[8] of ciphertexts (c^i, c^j) that belong to the same coset of $\mathcal{M}_J$ for a fixed $J \subset \{0,1,2,3\}$ (that is $c^i \oplus c^j \in \mathcal{M}_J$) has the special property to be a multiple of 8 with prob. 1 independently of the secret key, of the details of the S-Box and of the MixColumns matrix (assuming branch number equal to 5).

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

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

The proof of this property is based on the following argumentation. Given two different texts $t^1, t^2 \in \mathcal{D}_I \oplus a$, it is possible to prove that there exist other two texts $s^1, s^2 \in \mathcal{D}_I \oplus a$ (*related to* t^1 and t^2) such that

$$R^5(t^1) \oplus R^5(t^2) \in \mathcal{M}_J \qquad \text{if and only if} \qquad R^5(s^1) \oplus R^5(s^2) \in \mathcal{M}_J.$$

3 MixColumns Property and Key-Recovery Attacks 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 [20] and in [16]. Instead of finding the secret S-Box up to additive constants (as in [21]), 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. For the following, we recall the details of such strategy, and we show how to combine it with the new multiple-of-n property proposed in [15] just recalled.

MixColumns Matrix with Two Equal Coefficients: Strategy of the Attack. The strategy proposed in [20] and [16] 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 on the 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 wrong.

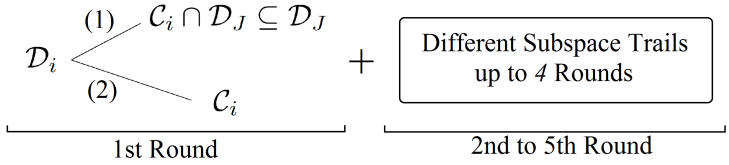

Fig. 1. *Strategy of the attacks on AES with a secret S-Box* proposed in [16]. A subset of a coset of $\mathcal{D}_i$ (which depends on the guessed values of the secret key) is mapped after one round into a subset of a coset of $\mathcal{D}_J$ if the guessed values is correct - *(1st)* case, or into a subset of a coset of $\mathcal{C}_i$ if the guessed values is wrong - *(2nd)* case. 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.

We limit here to recall a concrete example, and we refer to [16] 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 prob. 1 - case (1) of Fig. 1, otherwise they belong to the same coset of $\mathcal{D}_{1,2,3}$ only with prob. 2^{-8} - case (2) of Fig. 1. 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. 1. This allows to recover the key. We emphasize that *no information on the S-Box is recovered or used.*

As concrete example, consider the set of plaintexts-ciphertexts V_δ defined as

$$\begin{aligned} V_\delta = &\{(p^i, c^i) \text{ for } i = 0, ..., 2^8 - 1 \,|\, p^i_{2,2} \oplus p^i_{3,3} = \delta \quad \forall i \quad \text{and} \\ &\text{and} \quad p^i_{k,l} = p^j_{k,l} \quad \forall (k,l) \neq \{(2,2),(3,3)\} \text{ and } i \neq j\}, \end{aligned} \tag{4}$$

that is 2^8 plaintexts with 14 constants bytes and for which the difference on the other two bytes is fixed and equal to a guessed value of the key. If the guessed key is the 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. Moreover, if the guessed key is the correct one, then after 5 rounds the previous texts belong to the same coset of $\mathcal{M}_I$ for each $I \subseteq \{0,1,2,3\}$ for $|I| = 1$ with probability 0, while this happens with probability 2^{-94} for a wrongly guessed key. If the final MixColumns is omitted, it is sufficient to replace $\mathcal{M}_I$ with $\mathcal{ID}_I$.

3.1 Multiple-of-n Attack on 5-round AES with a secret S-Box

As first thing, we show how to adapt the previous strategy to set up an attack on 5-round AES with a single secret S-Box which exploits the multiple-of-n property proposed in [15]. 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 $a \in \mathcal{D}_1^\perp$ (i.e. $a_{0,1} = a_{1,2} = 0$), let $\mathcal{A}_\delta$ be the set of plaintexts of the form:

$$\mathcal{A}_\delta \equiv \left\{ a \oplus \begin{bmatrix} y_0 & x & 0 & 0 \\ 0 & y_1 & x \oplus \delta & 0 \\ 0 & 0 & y_2 & 0 \\ 0 & 0 & 0 & y_3 \end{bmatrix} \;\middle|\; \forall x, y_0, ..., y_3 \in \mathbb{F}_{2^8} \right\}. \tag{5}$$

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 2. *Consider a set of plaintexts $\mathcal{A}_\delta$ defined as in (5), 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. E. □

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 a constant 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 Cost. 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 an 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 $(0.95)^{1/255} = 99.98\%$ (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[9], it follows that 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. In order to count the number of collisions, one can exploit *data structure* - the complete pseudo-code of such an algorithm is given in Algorithm 1. 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[10], we have practically verified the attack just described on a small-scale variant of AES, as presented in [10] - not on real AES due to the large computational cost of the attack. We emphasize that Prop. 2 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.

[9] In this case, one needs that for each one of the $2^8 - 1$ wrong possible values for δ, at least one set $\mathcal{A}_\delta$ for which the number of collision is odd exists with probability higher than $(0.9998)^{1/12} = 99.99835\%$.

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

Data: 2^{10} different sets $\mathcal{A}_\delta$ defined as in (5) - 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 1: *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 is used to recover the entire key up to 2^{32} variants.

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.

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. To explain the (small) difference with the theoretical value, note that the theoretical value is computed in the worst case. As 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.

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

As we have just recalled, the strategy proposed in [20] and in [16] exploits the fact that two coefficients of each row of the MixColumns matrix are equal. Here we show how 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

$$\bigoplus_{j \in J_r} M^{MC}_{r,j} = 0 \tag{6}$$

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

$$M^{MC}_{0,0} \oplus M^{MC}_{0,1} \oplus M^{MC}_{0,2} = \text{0x02} \oplus \text{0x03} \oplus \text{0x01} = 0, \quad M^{MC}_{0,i} \neq M^{MC}_{0,j} \ \forall i,j \in \{0,1,2\}.$$

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 [20] and in [16] as a particular case.

To explain how to exploit property (6), we show how to adapt the attacks described in [16] (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 on 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 3. *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

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

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 or four coefficients are equal to zero. Moreover, given J fixed, if the sum $\bigoplus_{j \in J} M^{MC}_{r,j}$ is equal to zero for more than a single row r, the following Lemma follows immediately.

Lemma 2. *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 $\mathcal{C}_k \cap \mathcal{D}_{\{0,1,2,3\} \setminus \{r,w\}}$.

To give a concrete example of this strategy, in App. B we show how to adapt the attack presented in Sect. 3.1 in order to exploit the new property (6). In the following - Sect. 5.2, we present another (a little more complicated) variant of such attack which is more completive both for the data and computational cost.

What is the number of matrices that satisfy condition (6) with respect to the number of matrices with two equal coefficients in each row? Since we consider AES-like ciphers, we limit to practical count[11] both these numbers for the cases of *circulant* matrices in $\mathbb{F}_{2^m}^{4 \times 4}$ for $m = 4, 8$. We remember that the strategy just proposed works in the encryption direction if the MixColumns matrix satisfies one of the two previous properties and/or in

[11] The source codes are available at `https://github.com/Krypto-iaik/Attacks_AES_SecretSBox2`

Table 2. *Practical Numbers for the case of Circulant Invertible Matrices.* The second column gives the number of invertible matrices MC for which MC or MC^{-1} has two equal coefficients in each row, while the third one gives the number of invertible matrices for which the sum of ≥ 2 the same row of MC or MC^{-1} is equal to zero.

$\mathbb{F}_{2^m}^{4\times 4}$	**Number Invertible Matrices**	**Two Equal Coeff.**	**Zero-Sum of ≥ 2 Coeff.**
$m = 4$	61 440	32 640 (53.125%)	45 600 (74.22%)
$m = 8$	4 278 190 080	165 550 080 (3.87%)	293 556 000 (6.87%)

Table 3. *Practical Numbers for the case of Circulant MDS Matrices.* The second column gives the number of MDS matrices MC for which MC or MC^{-1} has two equal coefficients in each row, while the third one gives the number of MDS matrices for which the sum of ≥ 2 elements in the same row of MC or MC^{-1} is equal to zero.

$\mathbb{F}_{2^m}^{4\times 4}$	**Number MDS Matrices**	**Two Equal Coeff.**	**Zero-Sum of ≥ 2 Coeff.**
$m = 4$	16 560	10 080 (60.87%)	12 480 (75.36%)
$m = 8$	4 015 735 920	126 977 760 (3.16%)	249 418 560 (6.21%)

the decryption direction if the inverse MixColumns matrix satisfies them. For this reason, we compute the number of MixColumns matrices for which one of the two previous properties is satisfied in the encryption direction (i.e. by MC) *or* in the decryption direction (i.e. by MC^{-1}). For completeness, in App. A we list similar numbers in the case in which only one of the two directions (e.g. encryption - MC) is considered.

In Table 2 we list our results limiting to consider invertible matrices, while in Table 3 we list our results limiting to consider MDS (Maximal Distance Separable)[12] matrices. Observing the numbers in the tables, both for these two cases and both for $m = 4$ and $m = 8$, the number of matrices that satisfy condition (6) is (largely) higher than the number of matrices with two equal coefficients in each row. E.g. for the case $m = 8$, this number increases of 77.32% (e.g. $2^{27.3}$ vs $2^{28.13}$) for the invertible matrices case, and of 96.42% (e.g. $2^{26.92}$ vs $2^{27.89}$) for the MDS matrices case (that is, the number has doubled).

5 New Attacks on 5-round AES with a secret S-Box

In this section, we propose two attacks on AES with a single secret S-Box that exploit the fact that the sum of some coefficients of the MixColumns matrix is equal to zero. In particular, we show how to set up an impossible differential attack up to 5 rounds of AES that exploits (6), which improves the impossible differential attack presented in [16]. Then, we show how to adapt the attack presented in Sect. 3.1 in order to exploit the new property just presented.

[12] A matrix $M \in \mathbb{F}_{2^m}^{n\times n}$ is called *Maximum Distance Separable* (MDS) matrix if and only if it has branch number $\mathcal{B}(M)$ equal to $B(M) = n + 1$. Equivalently, a matrix M is MDS if and only if all square sub-matrices of M are of full rank. It follows immediately that if a matrix is not invertible, it can not be MDS.

5.1 Impossible Diff. Attack on 5-round AES with a secret S-Box

Here 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. (6)), and improves the one presented in [16].

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:

$$V_\delta \equiv \left\{ a \oplus \begin{bmatrix} x & 0 & 0 & 0 \\ 0 & x \oplus \delta_{1,1} & 0 & 0 \\ 0 & 0 & x \oplus \delta_{2,2} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \;\middle|\; \forall x \in \mathbb{F}_{2^8} \right\} \tag{7}$$

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

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

it follows by Prop. 3 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 $|I| + |J| \leq 4$ (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 in the probabilities to recover the secret key.

Comparison with the Impossible-Differential Attack of [16]. For completeness, we briefly discuss the difference with the attack proposed in [16]. In this last case, a similar set V_δ is defined, and the idea is to exploit the fact two elements of each row of the MixColumns matrix are equal. As before, for the right guessed key and given two plaintexts in the same coset of V_δ, then the corresponding ciphertexts after 5-round can not belong to the same coset of $\mathcal{M}_J$ for $|J| = 1$ The main difference regards the case of a wrong guessed key, for which the previous event happens with prob. 2^{-94}. As a result, one needs more texts to detect the wrong guessed keys.

Data and Computational Costs. The data and the computational costs analysis are similar to the ones proposed in [16]. For this reason, we limit here to report the data and computational costs of the attack, and we refer to App. C for all the details. 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, while - using the re-ordering

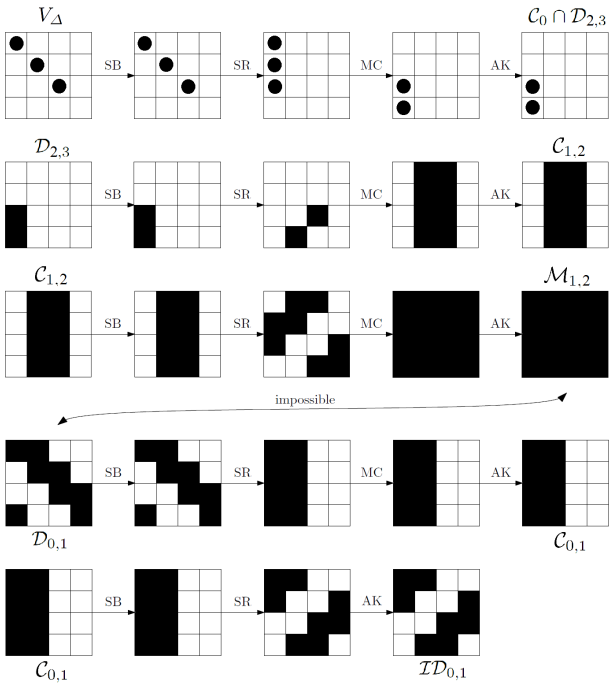

Fig. 2. *5-Round secret-key distinguisher for AES with a single secret S-Box.* The choice of the plaintexts (i.e. $p_{0,0} \oplus p_{i,i} = k_{0,0} \oplus k_{i,i}$ for $i = 1, 2$) guarantees that after one round there are only two bytes with non-zero difference instead of four, that is the plaintexts 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.

algorithm proposed in Algorithm 3 (see App. C) - the computational cost is well 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. For comparison, the attack proposed in [16] requires 2^{102} chosen plaintexts and computational cost is of $2^{100.4}$ five-round encryptions.

5.2 Improved Multiple-of-n Attack on 5-round AES with a secret S-Box

Here we show how to adapt the attack proposed in Sect. 3.1 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 the key δ of the form:

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

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 (8), 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

$$R(\mathcal{A}_\delta^{''}) \equiv \left\{ b \oplus \begin{bmatrix} \text{0x03} \cdot z & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & \text{0x02} \cdot w & 0 & 0 \\ \text{0x02} \cdot z & \text{0x03} \cdot w & 0 & 0 \end{bmatrix} \;\middle|\; \forall z, w \in \mathbb{F}_{2^8} \right\}.$$

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 vice-versa) 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 vice-versa). 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. G. □

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. 3.1, we limit here to report the data and computational costs of the attack and we refer to App. D for all the details. The total data complexity is approximately of $2 \cdot 2^{52.248} + 12 \cdot 2^{16} \cdot 2^{16} = 2^{53.25}$ chosen plaintexts, while - using the re-ordering algorithm proposed in Algorithm 4 (see App. D) - the computational cost 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.

Practical Verification Using a C/C++ implementation, we have practically verified the attack just described on a small-scale variant of AES [10] - 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 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. As before, the difference is explained by the fact that in general it is possible to discard wrong candidates without considering all the corresponding 11 sets $\mathcal{A}_\delta^{''}$.

6 Summary and Open Problems

In this work, we studied the impact of replacing the S-Box in the AES by a secret S-Box unknown to the adversary. Despite the expected increase in difficulty of recovering the secret information, we are able to mount (efficient) attacks based on a new propriety of the MixColumns matrix combined with dedicated techniques. It is an open problem if a weaker property of the MixColumns matrix can be exploited to set up similar attacks.

Cryptanalysis of cipher derived from the AES (with known S-Box) by replacing the ShiftRows and the MixColumns operation with a secret linear (or - more generally - affine) mixing transformation is still an open problem. In this setting, is it possible to set up attacks on more than 6-round AES with a single secret linear mixing transformation? What is the gap between the data/time complexities of such attacks with respect to the cases of standard AES or/and

AES with a single secret S-Box?

Acknowledgements. The author thanks Christian Rechberger for fruitful discussions and comments that helped to improve the quality of the paper.

References

1. "CAESAR: Competition for Authenticated Encryption: Security, Applicability, and Robustness," http://competitions.cr.yp.to/caesar.html.
2. E. Biham, A. Biryukov, and A. Shamir, "Cryptanalysis of Skipjack Reduced to 31 Rounds Using Impossible Differentials," in *Advances in Cryptology - EUROCRYPT 1999*, 1999, pp. 12–23.
3. E. Biham and N. Keller, "Cryptanalysis of Reduced Variants of Rijndael," 2001, unpublished, http://csrc.nist.gov/archive/aes/round2/conf3/papers/35-ebiham.pdf.
4. E. Biham and A. Shamir, *Differential Cryptanalysis of the Data Encryption Standard*. Springer, 1993.
5. A. Biryukov, C. Bouillaguet, and D. Khovratovich, "Cryptographic Schemes Based on the ASASA Structure:Black-Box,White-Box, andPublic-Key," in *Advances in Cryptology - ASIACRYPT 2014*, ser. LNCS, vol. 8873, 2014, pp. 63–84.
6. A. Biryukov and A. Shamir, "Structural Cryptanalysis of SASAS," *Journal of Cryptology*, vol. 23, no. 4, pp. 505–518, 2010.
7. C. Blondeau, G. Leander, and K. Nyberg, "Differential-Linear Cryptanalysis Revisited," *Journal of Cryptology*, vol. 30, no. 3, pp. 859–888, 2017.
8. A. Bogdanov and V. Rijmen, "Linear hulls with correlation zero and linear cryptanalysis of block ciphers," *Designs, Codes and Cryptography*, vol. 70, no. 3, pp. 369–383, 2014.
9. J. Borghoff, L. R. Knudsen, G. Leander, and S. S. Thomsen, "Cryptanalysis of PRESENT-Like Ciphers with Secret S-Boxes," in *Fast Software Encryption - FSE 2011*, ser. LNCS, vol. 6733, 2011, pp. 270–289.
10. C. Cid, S. Murphy, and M. J. B. Robshaw, "Small Scale Variants of the AES," in *Fast Software Encryption - FSE 2005*, ser. LNCS, vol. 3557, 2005, pp. 145–162.
11. J. Daemen, L. R. Knudsen, and V. Rijmen, "The Block Cipher Square," in *Fast Software Encryption - FSE 1997*, ser. LNCS, vol. 1267, 1997, pp. 149–165.
12. J. Daemen and V. Rijmen, *The Design of Rijndael: AES - The Advanced Encryption Standard*, ser. Information Security and Cryptography. Springer, 2002.
13. N. Datta and M. Nandi, "ELmD," https://competitions.cr.yp.to/round1/elmdv10.pdf.
14. H. Gilbert and P. Chauvaud, "A Chosen Plaintext Attack of the 16-round Khufu Cryptosystem," in *Advances in Cryptology - CRYPTO 1994*, ser. LNCS, vol. 839, 1994, pp. 359–368.
15. L. Grassi, C. Rechberger, and S. Rønjom, "A New Structural-Differential Property of 5-Round AES," in *Advances in Cryptology - EUROCRYPT 2017*, ser. LNCS, vol. 10211, 2017, pp. 289–317.
16. ——, "Subspace Trail Cryptanalysis and its Applications to AES," *IACR Transactions on Symmetric Cryptology*, vol. 2016, no. 2, pp. 192–225, 2017. [Online]. Available: http://ojs.ub.rub.de/index.php/ToSC/article/view/571
17. L. R. Knudsen, "DEAL - a 128-bit block cipher," Technical Report 151, Department of Informatics, University of Bergen, Norway, Feb. 1998.

18. M. Matsui, "Linear Cryptanalysis Method for DES Cipher," in *Advances in Cryptology — EUROCRYPT 1993*, ser. LNCS, no. 765, 1994, pp. 386–397.
19. B. Mennink and S. Neves, "Optimal PRFs from Blockcipher Designs," *IACR Transactions on Symmetric Cryptology*, vol. 2017, no. 3, pp. 228–252, 2017.
20. B. Sun, M. Liu, J. Guo, L. Qu, and V. Rijmen, "New Insights on AES-Like SPN Ciphers," in *Advances in Cryptology - CRYPTO 2016*, ser. LNCS, vol. 9814, 2016, pp. 605–624.
21. T. Tiessen, L. R. Knudsen, S. Kölbl, and M. M. Lauridsen, "Security of the AES with a Secret S-Box," in *Fast Software Encryption - FSE 2015*, ser. LNCS, vol. 9054, 2015, pp. 175–189.
22. S. Vaudenay, "On the weak keys of blowfish," in *Fast Software Encryption - FSE 1996*, ser. LNCS, vol. 1039, 1996, pp. 27–32.
23. H. Wu and B. Preneel, "A Fast Authenticated Encryption Algorithm," http://competitions.cr.yp.to/round1/aegisv11.pdf.

A Number of Matrices with Particular Properties

What is the number of matrices that satisfy condition (6) with respect to the number of matrices with two equal coefficients in each row? Since we consider AES-like ciphers, we limit to practical compute both these numbers for the cases of *circulant* matrices in $\mathbb{F}_{2^m}^{4\times4}$ for $m = 4, 8$. We remember that the strategy just proposed works in the encryption direction if the MixColumns matrix satisfies one of the two previous property and/or in the decryption direction if the inverse MixColumns matrix satisfies them. For this reason, in Sect. 4 we compute the number of MixColumns matrices for which one of the two previous properties is satisfied in the encryption direction (i.e. by MC) or in the decryption direction (i.e. by MC^{-1}). *Here we list similar numbers in the case in which only one of the two directions (e.g. encryption - MC) is considered.*

Table 4. *Practical Numbers for the case of Circulant Invertible Matrices.* The second column gives the number of invertible matrices for which there are two equal coefficients in each row, while the third one gives the number of invertible matrices for which the sum of two or more elements in the same row is equal to zero.

$\mathbb{F}_{2^m}^{4\times4}$	**Number Invertible Matrices**	**Two Equal Coeff.**	**Zero-Sum of ≥ 2 Coeff.**
$m = 4$	61 440	21 120 (34.38%)	31 200 (50.78%)
$m = 8$	4 278 190 080	99 747 840 (2.33%)	165 036 000 (3.86%)

In Table 4 we list our results limiting to consider invertible matrices, while in Table 5 we list our results limiting to consider MDS (Maximal Distance Separable) matrices. Observing the numbers in the tables, both for these two cases and both for $m = 4$ and $m = 8$, the number of matrices that satisfy condition (6) is largely higher than the number of matrices with two equal coefficients in each row. E.g. for the case $m = 8$, this number increases of 65.45% (e.g. $2^{26.571}$ vs $2^{27.298}$) for the invertible matrices case, and of 98.01% (e.g. $2^{25.925}$ vs $2^{26.911}$) for the MDS matrices case.

Table 5. *Practical Numbers for the case of Circulant MDS Matrices.* The second column gives the number of MDS matrices for which there are two equal coefficients in each row, while the third one gives the number of MDS matrices for which the sum of two or more elements in the same row is equal to zero.

$\mathbb{F}_{2^m}^{4\times 4}$	**Number MDS Matrices**	**Two Equal Coeff.**	**Zero-Sum of ≥ 2 Coeff.**
$m = 4$	16 560	5 760 (34.78%)	8 640 (52.18%)
$m = 8$	4 015 735 920	63 745 920 (1.59 %)	126 218 880 (3.15%)

B Second Version of the Multiple-of-n Attack on 5-round AES with a secret S-Box

In this section, we show how to adapt the attack of Sect. 3.1 in order to exploit e.g. condition (6), 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:

$$\mathcal{A}'_\delta \equiv \left\{ a \oplus \begin{bmatrix} 0 & y_0 & 0 & 0 \\ 0 & x & y_1 & 0 \\ 0 & 0 & x \oplus \delta_{2,2} & y_2 \\ y_3 & 0 & 0 & x \oplus \delta_{3,3} \end{bmatrix} \;\middle|\; \forall x, y_0, ..., y_3 \in \mathbb{F}_{2^8} \right\} \tag{9}$$

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 5. *Consider a set of plaintexts $\mathcal{A}'_\delta$ defined as in (9), 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

$$R(\mathcal{A}'_\delta) \equiv \left\{ b \oplus \begin{bmatrix} \text{0x03} \cdot w & z_0 & 0 & 0 \\ 0 & z_1 & 0 & 0 \\ 0 & z_2 & 0 & 0 \\ \text{0x02} \cdot w & z_3 & 0 & 0 \end{bmatrix} \;\middle|\; \forall w, z_0, ..., z_3 \in \mathbb{F}_{2^8} \right\}.$$

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 vice-versa), 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

Data: $3 \cdot 2^{16}$ different sets $\mathcal{A}'_\delta$ defined as in (9) - 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 $(n \bmod 4) \neq 0$ **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 2: *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.

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. F. □

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. 3.1, 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. 3.1, 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 $n \geq 8$ 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. 3.1 and described in details in Algorithm 2, 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, we have practically verified the attack just described on a small-scale variant of AES, as presented in [10] - not on real AES due to the large computational cost of the attack. We emphasize that Prop. 5 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.

C Impossible Differential Attack of Sect. 5.1 - Details

In Sect. 5.1 we show how to set up an impossible differential attack on 5-round AES with a single secret S-Box that exploits the fact that a sum of coefficients of the MixColumns matrix is equal to zero (e.g. (6)). We refer to that section for all the details, and we limit here to describe the data and the computational costs.

Data Cost. First of all, 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 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

Data: $2^{74.4}$ different sets V_δ defined as in (7) - $2^{58.4}$ 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 $0 \leq i \leq 2^8 - 1$ 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. 6 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 3: *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.

$(0.95)^{2^{-10}} \simeq 99.995\%$. 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.

Computational Cost. As for the impossible differential attack on 5-round AES with a single secret S-Box presented in [16], using the re-ordering algorithm proposed in Algorithm 3, the computational cost is well 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 (20 table look-ups $\approx$ 1-round of encryption[13]).

Such re-ordering algorithm exploits the following partial-order:

[13] *This approximation is not formally correct*, since the size of the table of an S-Box look-up is lower than the size of the table used for our proposed distinguisher. However, it allows to give a comparison between our distinguishers and the others currently present in the literature. Moreover, it is largely used in literature.

Definition 6. *Let $I \subset \{0,1,2,3\}$ with $|I| = 3$ and let $l \in \{0,1,2,3\} \setminus I$. Let $t^1, t^2 \in \mathbb{F}_{2^8}^{4\times 4}$ with $t^1 \neq t^2$. The text t^1 is less or equal than the text t^2 with respect to the partial order $\preceq$ (i.e. $t^1 \preceq t^2$) if and only if one of the two following conditions is satisfied (the indexes are taken modulo 4):*

- *there exists $j \in \{0,1,2,3\}$ 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 there exists $i, j \in \{0,1,2,3\}$ such that (1) $MC^{-1}(t^1)_{k,l} = MC^{-1}(t^2)_{k,l}$ for all $k, l \in \{0,1,2,3\}$ with $k + 4 \cdot l < i + 4 \cdot j$ and (2) $MC^{-1}(t^1)_{i,j} < MC^{-1}(t^2)_{i,j}$.*

D Multiple-of-n Attack of Sect. 5.2 - Details

In Sect. 5.2 we show how to set up an attack on 5-round AES with a single secret S-Box that exploits the multiple-of-n property and the fact that a sum of coefficients of the MixColumns matrix is equal to zero (e.g. (6)). We refer to that section for all the details, and we limit here to describe the data and the computational costs.

Data Costs. Since the procedure of the attack is completely equivalent to the one described in Sect. 3.1, we refer to that section for all the details and we limit here to report the main differences.

First of all, 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 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

Data: $19 \cdot 2^{32}$ different sets $\mathcal{A}_\delta^{'''}$ defined as in (8) - 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. 6 s.t. $c^i \preceq c^{i+1}$ for each i; `// ⪯ 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**
        $r \leftarrow 1$ and $j \leftarrow i$;
        while $c^j \oplus c^{j+1} \in \mathcal{M}_I$ **do**
          $r \leftarrow r + 1$ and $j \leftarrow j + 1$;
        end
        $i \leftarrow j + 1$ and $n \leftarrow n + r \cdot (r-1)/2$;
      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 4-bytes of the 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.

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[14], 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.

[14] 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 $(0.95)^{2^{-18}}$.

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.

Computational Costs. Using a re-ordering algorithm proposed in Algorithm 4, the computational cost 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. 3.1 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 Algorithm 1 - is (much) larger than the size of the sets $\mathcal{A}_\delta^{''}$ (i.e. 2^{32} versus 2^{16}).

E Proof of Sect. 3.1

For a fixed a, consider a set of plaintexts $\mathcal{A}_\delta$ of the form (5):

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

Proposition 6. *Consider a set of plaintexts $\mathcal{A}_\delta$ defined as in (5), 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

$$R(\mathcal{A}_\delta) \equiv \left\{ b \oplus \begin{bmatrix} z_0 & w & 0 & 0 \\ z_1 & \text{0x03} \cdot w & 0 & 0 \\ z_2 & 0 & 0 & 0 \\ z_3 & \text{0x02} \cdot w & 0 & 0 \end{bmatrix} \;\middle|\; \forall w, z_0, ..., z_3 \in \mathbb{F}_{2^8} \right\}.$$

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

$$R^4(z) \oplus R^4(z') \in \mathcal{M}_I \textit{ if and only if} R^4(v) \oplus R^4(v') \in \mathcal{M}_I$$

where the texts $v, v' \in R(\mathcal{A}_\delta)$ are generated respectively by

$$v \equiv (z_0, z_1', z_2, z_3, w) \quad \text{and} \quad v' \equiv (z_0', z_1, z_2', z_3', w).$$

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

Step (1). 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 $R^4(v) \oplus R^4(v') \in \mathcal{M}_I$.

Step (2). 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:

$$\begin{aligned}
&[R^2(z) \oplus R^2(z')]_{i,j} = [R^2(v) \oplus R^2(v')]_{i,j} = F(z_0, z_0', z_2, z_2', z_3, z_3', w, w') \oplus \\
&\oplus A \cdot \text{S-Box}\big(B \cdot \text{S-Box}(z_1 \oplus k_{1,0}) \oplus C)\big) \oplus A \cdot \text{S-Box}\big(B \cdot \text{S-Box}(z_1' \oplus k_{1,0}) \oplus C)\big).
\end{aligned} \tag{10}$$

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. 3, 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 (10) 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. □

F Proof of App. B

For a fixed a, consider a set of plaintexts $\mathcal{A}'_\delta$ of the form (9)

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

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

Proposition 7. *Consider a set of plaintexts $\mathcal{A}'$ defined as in (9), 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

$$R(\mathcal{A}'_\delta) \equiv \left\{ b \oplus \begin{bmatrix} \text{0x03} \cdot w & z_0 & 0 & 0 \\ 0 & z_1 & 0 & 0 \\ 0 & z_2 & 0 & 0 \\ \text{0x02} \cdot w & z_3 & 0 & 0 \end{bmatrix} \;\middle|\; \forall w, z_0, ..., z_3 \in \mathbb{F}_{2^8} \right\}.$$

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 vice-versa). 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 $|I| = 3$, as before the idea is to show that

$$R^4(z) \oplus R^4(z') \in \mathcal{M}_I \quad \text{if and only if} \quad R^4(v) \oplus R^4(v') \in \mathcal{M}_I$$

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 $v' \equiv (z'_0, z'_1, z_2, z_3, w)$.

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. E, 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:

$$\begin{aligned}
&[R^2(z) \oplus R^2(z')]_{i,j} = [R^2(v) \oplus R^2(v')]_{i,j} = F(z_0, z'_0, z_1, z'_1, w, w') \oplus \\
&\oplus A \cdot \text{S-Box}\big(B \cdot \text{S-Box}(z_2 \oplus k_{2,1}) \oplus C)\big) \oplus A \cdot \text{S-Box}\big(B \cdot \text{S-Box}(z'_2 \oplus k_{2,1}) \oplus C)\big) \oplus \\
&\oplus D \cdot \text{S-Box}\big(E \cdot \text{S-Box}(z_3 \oplus k_{3,1}) \oplus F)\big) \oplus D \cdot \text{S-Box}\big(E \cdot \text{S-Box}(z'_3 \oplus k_{3,1}) \oplus F)\big).
\end{aligned} \tag{11}$$

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. E, 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 (11), 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 $2 \cdot 2^8 = 512$. □

G Proof of Sect. 5.2

For a fixed a, consider a set of plaintexts $\mathcal{A}_\delta^{''}$ of the form (8):

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

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

Proposition 8. *Consider a set of plaintexts $\mathcal{A}_\delta^{''}$ defined as in (8), 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

$$R(\mathcal{A}_\delta^{''}) \equiv \left\{ b \oplus \begin{bmatrix} \text{0x03} \cdot z & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & \text{0x02} \cdot w & 0 & 0 \\ \text{0x02} \cdot z & \text{0x03} \cdot w & 0 & 0 \end{bmatrix} \;\middle|\; \forall z, w \in \mathbb{F}_{2^8} \right\}.$$

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 vice-versa) 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

$$R^4(t^1) \oplus R^4(t^2) \in \mathcal{M}_I \quad \text{if and only if} \quad R^4(s^1) \oplus R^4(s^2) \in \mathcal{M}_I$$

for $|I| = 3$, where the texts $s^1, s^2 \in R(\mathcal{A}_\delta^{''})$ are generated respectively by

$$s^1 = (z, w') \quad \text{and} \quad s^2 = (z', w).$$

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:

$$\begin{aligned} &[R^2(t^1) \oplus R^2(t^2)]_{i,j} = \\ =&A_0 \cdot \text{S-Box}(B_0 \cdot \text{S-Box}(z_0) \oplus C_0) \oplus A_0 \cdot \text{S-Box}(B_0 \cdot \text{S-Box}(z_0') \oplus C_0) \oplus \\ \oplus&A_1 \cdot \text{S-Box}(B_1 \cdot \text{S-Box}(z_1) \oplus C_1) \oplus A_1 \cdot \text{S-Box}(B_1 \cdot \text{S-Box}(z_1') \oplus C_1) \oplus \\ \oplus&A_2 \cdot \text{S-Box}(B_2 \cdot \text{S-Box}(w_0) \oplus C_2) \oplus A_2 \cdot \text{S-Box}(B_2 \cdot \text{S-Box}(w_0') \oplus C_2) \oplus \\ \oplus&A_3 \cdot \text{S-Box}(B_3 \cdot \text{S-Box}(w_1) \oplus C_3) \oplus A_3 \cdot \text{S-Box}(B_3 \cdot \text{S-Box}(w_1') \oplus C_3) = \\ =&[R^2(s^1) \oplus R^2(s^2)]_{i,j} \end{aligned} \tag{12}$$

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 vice-versa). 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 (12), 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. □