Subspace Trail Cryptanalysis and its Applications to AES

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

[1] IAIK, Graz University of Technology, Austria
[2] Nasjonal sikkerhetsmyndighet, Norway
[3] DTU Compute, DTU, Denmark
{firstname.lastname}@iaik.tugraz.at, sondrer@gmail.com

Abstract. We introduce subspace trail cryptanalysis, a generalization of invariant subspace cryptanalysis. With this more generic treatment of subspaces we do no longer rely on specific choices of round constants or subkeys, and the resulting method is as such a potentially more powerful attack vector.
We provide a general framework for subspace trail cryptanalysis of AES-like Substitution-Permutation Network (SPN) constructions. Interestingly, subspace trail cryptanalysis in fact includes earlier techniques based on impossible or truncated differential cryptanalysis and the integral property as special cases.
Choosing AES-128 as the perhaps most studied cipher, we describe distinguishers up to 5 round-reduced AES with a single unknown key. As the perhaps most interesting concrete result, we are able to describe *the first 5-round distinguisher for* (all versions of) *AES* that does not require any knowledge about subkeys and needs much less than the full codebook.

Keywords: Block cipher - AES - Invariant Subspace Attack - Subspace Trail Cryptanalysis - Secret-Key Distinguisher - Truncated Differential Cryptanalysis - Zero-Sum - Impossible Differential Cryptanalysis

1 Introduction

In this paper we present a new cryptanalysis technique that adds to the toolbox of techniques at the disposal for cryptanalysts to evaluate the security of designs in symmetric cryptography. Our main contribution is the analysis of subspaces in SPNs (substitution-permutation networks) constructions, which can be seen as a generalization of the invariant subspace attack [15, 16]. While invariant subspace cryptanalysis relies on iterative subspace structures, our analysis focuses on *trails* of different subspaces. To clarify the presentation, we focus on the well-known block cipher AES-128. In particular, we study the propagation of subspaces trough various building blocks like S-Box and linear layers. In that sense it has similarities with SASAS cryptanalysis [6], but also with Evertse's linear structures [12].

If a cryptographic primitive succumbs to particular non-random behavior, it might be possible to distinguish it from what one would expect from sufficiently generic behavior. Invariant subspace cryptanalysis is a cryptanalytic technique that is extremely powerful for certain block ciphers. If there exists invariant subspace for the round function and for the key schedule, then this technique can be used to mount fast distinguishers and key recovery. This technique was introduced in [15] at CRYPTO 2011 for the cryptanalysis of PRINTcipher. Its efficiency has also been demonstrated on on the CAESAR candidate iSCREAM, on the LS-design Robin and on the lightweight cipher Zorro in [16], and on the block cipher Midori64 [13]. However, if such symmetries do not exist or are not found, invariant subspace cryptanalysis is not applicable.

In this paper we investigate the behavior of subspaces in keyed permutations. At a high level, we fix subspaces of the plaintext that maintain predictable properties after repeated applications of a key-dependent round function. First we identify what we call *subspace trails* which is essentially a coset of a plaintext subspace that encrypts to proper subspaces of the state space over several rounds. The trails are formed by the affine hulls of the intermediate ciphertexts. Subspace trails typically consist of subspaces that increase in dimension for each round, meaning that if the plaintext subspace has low dimension in comparison to the block length, the subsequent subspaces dimension increases for each round. For byte-based ciphers (like AES), a quick and dirty test for subspaces is to compute the affine hulls of a n-round encryption (for a certain $n \geq 1$) of all values for each byte and then identify these subspaces. For bit-based ciphers, it is more important to determine what was coined a *nucleon* in [16], that is candidate plaintext subspaces that seem to fit symmetries in the round function. Trails of affine hulls of the intermediate ciphertexts that grow slowly in dimension for each round, typically reflect slow diffusion in the round function. This is often the case for ciphers that iterate simple round functions many times. In this paper we will focus on what we call *constant dimensional subspace trails*, which are trails of cosets that preserve dimension over several rounds. We show how to connect two or more trails and form longer trails that preserve predictable structure. In particular, when we connect two trails we typically seek to describe an output coset of a first trail in terms of cosets of the input coset for the second trail.

We consider the introduction of the generalization of the known distinguishers from 1 up to 4 rounds using the subspace trails, and the 5-rounds distinguisher of AES in the secret-key setting as the most important contributions of the paper. It is important to note that well known techniques such as impossible or truncated differentials as well as integral properties can be seen as special cases of subspace trails. We discuss these links in Section 5. The approach to the generalization from invariant subspace cryptanalysis to subspace trail is outlined in Sect. 2. In Sect. 3 we give technical preliminaries with respect to AES-like permutations, and in Sect. 4 we state central theorems related to subspace trails and their intersections.

When concretely applying it to AES, the perhaps most widely used and analyzed cipher, we describe in Sect. 5 distinguishers of round-reduced AES with a single unknown key up to 4 rounds, which correspond to truncated differential, impossible differential, and integral distinguishers. Finally, in Sect. 6, we are able to present *the first 5-rounds secret key distinguisher of* AES which needs much less than the full codebooks (it has a data complexity of $2^{98.2}$ texts). However, before we start with these sections, we discuss our concrete results about the distinguishers in the unknown (secret)-key model.

1.1 Secret-Key Distinguishers for AES

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

In Table 1.1 we summarize the secret-key distinguishers for 1 up to 5 rounds. Such results often serve as a basis for key recovery attacks in the most relevant single-key setting. The subspace trail cryptanalysis includes as special cases and can be viewed as a generalization of differential cryptanalysis techniques (like truncated or impossible differentials) and integral cryptanalysis.

About the 5-round secret key distinguishers for AES, there exist some distinguishers for AES-192 and AES-256 [11], while the first distinguisher for AES-128 has been proposed recently in CRYPTO 2016. However, it requires the *whole* input-output space to work. Thus, *our proposed secret key distinguisher of* AES *is the first one that requires (much) less than the whole input-output space.*

Relation to Differential and Integral Distinguishers. The 1-, 2- and 3-round distinguishers exploit the same well-known structural properties that also truncated differentials exhibit. Using a different notation (namely the AES "Super S-Box"), 2-rounds subspace trails were already discovered and investigated in [9] and [10], with the objective to understand how the components of the AES interact. In these papers, authors study the probability of differentials and characteristics over 2 rounds of AES, giving bound on the maximum differential probability (which can be used to derive bounds on the expected differential probability of four-round differentials). Starting from such a 2-rounds subspace trail, in the paper we present competitive key-recovery attacks on 2-, 3- and 4-rounds of AES.

The first key-recovery attacks on round-reduced AES were obtained by introducing an attack vector that uses a 3-round distinguisher to attack up to 6 rounds of the cipher that goes back to the block cipher Square [7] and later became known as integral attacks.

In the meanwhile the most recent attacks achieve 7 rounds (using either impossible differentials or meet-in-the-middle techniques) with complexity significantly faster than brute-force search [18, 11]. The impossible differential distinguishers used for those attacks are up to 4 rounds. Our 4-round subspace

Table 1. *AES secret-key distinguishers, independent of key schedule.* Data Complexity is measured in minimum number of chosen plaintexts CP or/and chosen ciphertexts CC (which is equal for the random and the subspace case) which are needed to distinguish the two cases with high probability (usually higher than 95%). The case in which the MixColumns operation is omitted in the last round is denoted by "r.5 rounds", that is r full rounds and the final round.

Rounds	**Data**	CP	CC	**Property**	**Reference**
1 - 1.5 - 2	2	×	×	Subspace Trail	Sect. 5.1
1 - 1.5 - 2	2	×	×	Truncated Differential	[9]
2.5 - 3	$20 \simeq 2^{4.3}$	×	×	Subspace Trail	Sect. 5.2
2.5 - 3	$20 \simeq 2^{4.3}$	×	×	Truncated Differential	[5]
2.5	2^8	×	×	Integral	[7]
3	2^8		×	Subspace Trail	Sect. 5.3
3.5 - 4	$2^{16.25}$	×	×	Impossible Differential	[3]
3.5 - 4	$2^{16.25}$	×	×	Subspace Trail	Sect. 5.3
3.5	2^{32}	×	×	Integral	[7]
4	2^{32}		×	Subspace Trail	Sect. 5.3
4.5 - 5	$2^{98.2}$	×		Subspace Trail	Sect. 6
5	2^{128}		×	Integral	[19]

trail distinguisher uses the same structural properties exploited by impossible differential distinguishers.

In [19], authors present the first 5-rounds secret key distinguisher for AES-128. First they construct several types of 5-rounds zero-correlation linear hulls for AES-like ciphers, and then, using the link between integrals and zero correlation linear hulls [20], they are able to construct an integral distinguisher on 5 rounds. However, this distinguisher requires all the input-output space to work, that is the data complexity is of 2^{128} texts, or alternatively some knowleduge about subkey bits. Moreover, this distinguisher is constructed in the chosen-ciphertext mode, and only in the case in which MixColumns in the last round is not omitted. For this reason, authors claim that "*since the 5-round distinguisher for AES can only be constructed in the chosen-ciphertexts mode, the security margin for the round-reduced AES under the chosen-plaintext attack may be different from that under the chosen-ciphertext attack*".

Our 5-rounds secret key distinguisher presented in Sect. 6 is constructed in the chosen-plaintexts setting, extending the impossible 4-rounds distinguisher presented in Sect. 5.3 at the beginning. Our distinguisher works independent of the presence of the last MixColumns operation. Hence it provides a counterexample to the conjecture made in [19], i.e. it seems there is no clear evidence that chosen-ciphertext security is less than chosen-plaintext security in AES. Moreover, the data complexity is only $2^{98.2}$ chosen plaintexts instead of 2^{128}.

The subspace trail approach is mostly providing an alternative description of known properties under the umbrella of a single framework. However, there are other recent techniques that this approach does *not* seem to include. Recently integral distinguishers have been generalized by Todo [23] and in there also applied to AES-like primitives. Distinguishers for AES itself were not improved, but clear progress e.g. with MISTY cryptanalysis was demonstrated [22]. Todo's generalization can take S-Box properties into account, on the other hand the property exploited is still a type of zero-sum. Thus it complements our approach which is independent of the S-Box, but exploits properties more subtle than zero-sums.

Polytopic cryptanalysis, introduced by Tiessen in [21], is a generalization of differential cryptanalysis, and provides another type of distinguisher. While standard differential cryptanalysis uses statistical dependencies between the difference of two plaintexts and the difference of the respective two ciphertexts to attack a cipher, polytopic cryptanalysis considers interdependencies between larger sets of texts as they traverse through the cipher. Subspace trails do not seem to capture this type of distinguisher.

1.2 Practical Results

We practically verified the secret-key distinguishers using a C implementation[1] for 1 up to 4 rounds, and we have found that the practical results are consistent with our theory.

2 Subspace Trails and Distinguishers

In this section, we recall the invariant subspace cryptanalysis of [15, 16] (depicted in Fig. 1), and then we introduce the concept of subspace trails (Fig. 2).

Invariant subspace cryptanalysis can be a powerful cryptanalytic tool. Let F denote a round function in an iterative block cipher and assume there exists a coset[2] $V \oplus a$ such that $F(V \oplus a) = V \oplus a'$. Then if the round key K resides in $V \oplus (a \oplus a')$, it follows that $F(V \oplus a) \oplus K = V \oplus a$ and we get an iterative invariant subspace.

A slightly more powerful property can occur if for each a, there exists unique b such that $F_K(V \oplus a) := F(V \oplus a) \oplus K = V \oplus b$ meaning that the subspace property is invariant, but not the initial coset. That is, for each initial coset $V \oplus a$, its image under the application of F_K is another coset of V, in general different from the initial one. Equivalently, the initial coset $V \oplus a$ is mapped into another coset $V \oplus b$, where b depends on a and on the round key. In this paper,

[1] The source code of the distinguishers is available on `https://github.com/Krypto-iaik/Distinguishers_AES`.

[2] For completeness, we recall the definition of coset, largely used in the paper. Let W a vector space and V a subspace of W. A *coset* of V in W is a subset of the form $V \oplus a = \{v \oplus a \,|\, \forall v \in V\}$.

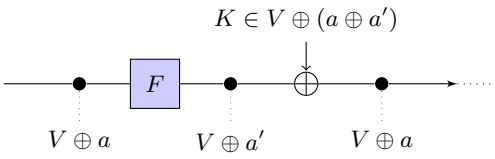

Fig. 1. Invariant subspaces.

we generalize this concept and search for trails of subspaces. In the simplest case we look for pairs of subspaces V_1 and V_2 such that

$$F(V_1 \oplus a) \oplus K = V_2 \oplus b$$

holds for any constant a, that is for each a there exists unique b for which the previous equivalence is satisfied.

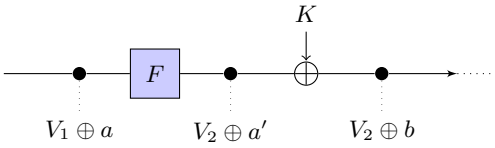

Fig. 2. Trail of subspaces.

A *subspace trail* of length r is then simply a set of $r+1$ subspaces $(V_1, V_2, \ldots, V_{r+1})$ that satisfy

$$F(V_i \oplus a_i) \oplus K \subseteq V_{i+1} \oplus a_{i+1}.$$

When the relation holds with equality, the trail is called a *constant-dimensional* subspace trail. In this case, if we let F_K^t denote the application of t rounds with fixed keys, it means that

$$F_K^t(V_1 \oplus a_1) = V_{t+1} \oplus a_{t+1}.$$

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*

$$F_K(V_i \oplus a_i) \subseteq V_{i+1} \oplus a_{i+1},$$

then $(V_1, V_2, ..., V_{r+1})$ is subspace trail of length r for the function F_K. If all the previous relations hold with equality, the trail is called a constant-dimensional subspace trail.

Note that a_{i+1} depends on a_i and on the secret round key. With the aim to simplify the notation, we use simply a_{i+1} instead of $a_{i+1}(a_i, k)$.

With subspace structures at hand, we might ask questions about the probability that ciphertexts or sums of ciphertexts reside in certain subspaces, given that the plaintexts obey certain subspace structure (e.g. their sum is also in a fixed subspace). For AES-type block ciphers, we are typically not able to construct very long trails. In this case we can connect trails together and depending on the intersection properties of the endpoints of the trails, get predictable subspace properties for longer trails. However, in general these are not necessarily simple constant dimensional trails. In the following we describe subspace trail cryptanalysis and later on distinguishers based on it. For sake of concreteness and better exposition we focus on the case of AES. We'd like to emphasize that the properties described here extend almost immediately to any AES-like cipher with little modifications.

Before to continue, we give the following definition of *equivalence cosets* of a generic subspace X:

Definition 2. *Let X a generic subspace, and let $X \oplus a$ and $X \oplus b$ two different cosets of X (that is $a \neq b$). We say that they are equivalent under an "equivalence relationship" (that is $X \oplus a \sim X \oplus b$) if and only if $a \oplus b \in X$:*

$$X \oplus a \sim X \oplus b \qquad \textit{if and only if} \qquad a \oplus b \in X.$$

3 Preliminaries - Description of AES

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

- *SubBytes* (S-Box) - applying the same 8-bit to 8-bit invertible S-Box 16 times in parallel on each byte of the state (it provides the non-linearity in the cipher);
- *ShiftRows* (SR) - cyclic shift of each row (i-th row is shifted by i bytes to the left);
- *MixColumns* (MC) - multiplication of each column by a constant 4×4 invertible matrix over the field $GF(2^8)$ (it and ShiftRows provide diffusion in the cipher[3]);
- *AddRoundKey* (ARK) - XORing the state with a 128-bit subkey.

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

[3] ShiftRows makes sure column values are spread and MixColumns makes sure each column is mixed.

As we consider only AES with 128-bit key, we shall describe only its key schedule algorithm. The key schedule of AES-128 takes the user key and transforms it into 11 subkeys of 128 bits each. The subkey array is denoted by $W[0, ..., 43]$, where each word of $W[\cdot]$ consists of 32 bits and where the first 4 words of $W[\cdot]$ are loaded with the user secret key. The remaining words of $W[\cdot]$ are updated according to the following rule:

- if $i \equiv 0 \mod 4$, then $W[i] = W[i-4] \oplus \mathit{RotByte}(\text{S-Box}(W[i-1])) \oplus \mathit{RCON}[i/4]$,
- otherwise, $W[i] = W[i-1] \oplus W[i-4]$,

where $i = 4, ..., 43$, $\mathit{RotByte}$ rotates the word by 8 bits to the left and $\mathit{RCON}[\cdot]$ is an array of predetermined constant.

The Notation Used in the Paper Let x denote a plaintext, a ciphertext, an intermediate state or a key. Then $x_{i,j}$ with $i, j \in \{0, ..., 3\}$ denotes the byte in the row i and in the column j. We denote by k^r the key of the r-th round, where k^0 is the secret key. If only the key of the final round is used, then we denote it by k to simplify the notation. Finally, we denote by R one round of AES[4], while we denote i rounds of AES by $R^{(i)}$. If the MixColumns operation is omitted in the last round, then we denote it by R_f. As last thing, in the paper we often use the term "*collision*" when two texts belong to the same coset of a given subspace X.

3.1 Subspaces through 1-Round of AES

For a vector space V and a function F on $\mathbb{F}_{2^8}^{4\times4}$, let $F(V) = \{F(v) \,|\, v \in V\}$ (as usual). For a subset $I \subseteq \{1, 2, \ldots, n\}$ and a subset of vector spaces $\{G_1, G_2, \ldots, G_n\}$, we define G_I as $G_I := \bigoplus_{i \in I} G_i$.

In the following we define three families of subspaces essential to AES; the diagonal spaces $\mathcal{D}_I$, the column spaces $\mathcal{C}_I$ and the mixed spaces $\mathcal{M}_I$. Since AES operates on 4×4 matrices over $\mathbb{F}_{2^8}$, then we work with vectors and vector spaces over $\mathbb{F}_{2^8}^{4\times4}$ (that is, all the subspaces considered in the paper are subspace over $\mathbb{F}_{2^8}^{4\times4}$). Moreover, we denote with $E = \{e_{0,0}, ..., e_{3,3}\}$ the unit vectors of $\mathbb{F}_{2^8}^{4\times4}$ (e.g. $e_{i,j}$ has a single 1 in row i and column j).

Definition 3. ***(Diagonal spaces)*** *The diagonal spaces $\mathcal{D}_i$ are defined as*

$$\mathcal{D}_i = < e_{0,i}, e_{1,i+1}, e_{2,i+2}, e_{3,i+3} >$$

where the index $i + j$ is computed modulo 4. For instance, the diagonal space $\mathcal{D}_0$ corresponds to the symbolic matrix

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

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

Definition 4. *(**Column spaces**) The column spaces $\mathcal{C}_i$ are defined as*

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

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

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

The last type of subspaces we define are called mixed subspaces.

Definition 5. *(**Mixed spaces**) The i-th mixed subspace $\mathcal{M}_i$ is defined as*

$$\mathcal{M}_i = MC \circ SR(\mathcal{C}_i).$$

These subspaces are formed by applying ShiftRows and then MixColumns to a column space. For instance, $\mathcal{M}_0$ corresponds to symbolic matrix

$$\mathcal{M}_0 = \left\{ \begin{bmatrix} \alpha \cdot x_1 & x_4 & x_3 & (\alpha+1) \cdot x_2 \\ x_1 & x_4 & (\alpha+1) \cdot x_3 & \alpha \cdot x_2 \\ x_1 & (\alpha+1) \cdot x_4 & \alpha \cdot x_3 & x_2 \\ (\alpha+1) \cdot x_1 & \alpha \cdot x_4 & x_3 & x_2 \end{bmatrix} \;\middle|\; \forall x_1, x_2, x_3, x_4 \in \mathbb{F}_{2^8} \right\}.$$

Definition 6. *Given $I \subseteq \{0,1,2,3\}$ where $0 < |I| \leq 3$, we define:*

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

The dimension of any of the spaces $\mathcal{D}_I, \mathcal{C}_I$ and $\mathcal{M}_I$ is $4 \cdot |I|$. The essential subspaces in AES are built from diagonal spaces $\mathcal{D}_i$, column spaces $\mathcal{C}_j$ and mixed spaces $\mathcal{M}_k$. There are four of each of these spaces, and direct sums of subsets of these result in higher-dimensional diagonal, column and mixed spaces.

It is easy to see that SubBytes maps cosets of diagonal and column spaces to cosets of diagonal and column spaces. Since SubBytes operates on each byte individually and it is bijective, and since the bytes of column and diagonal spaces are independent, its only effect is to change the coset. It is also easy to see that ShiftRows maps a coset of a diagonal space to a coset of a column space, since diagonals are mapped to columns. The effect of MixColumns to a columns space $\mathcal{C}_I \oplus a$ is simply to change the coset, since applying the MixColumns matrix to a column space $\mathcal{C}_i$ has no effect.

Lemma 1. *Let $I \subseteq \{0,1,2,3\}$ where $0 < |I| \leq 3$ and $a \in \mathcal{D}_I^{\perp}$. There exists unique $b \in \mathcal{C}_I^{\perp}$ such that*

$$R_K(\mathcal{D}_I \oplus a) = \mathcal{C}_I \oplus b.$$

Note that b is unique with respect to the equivalence relationship defined before (analogous in the following).

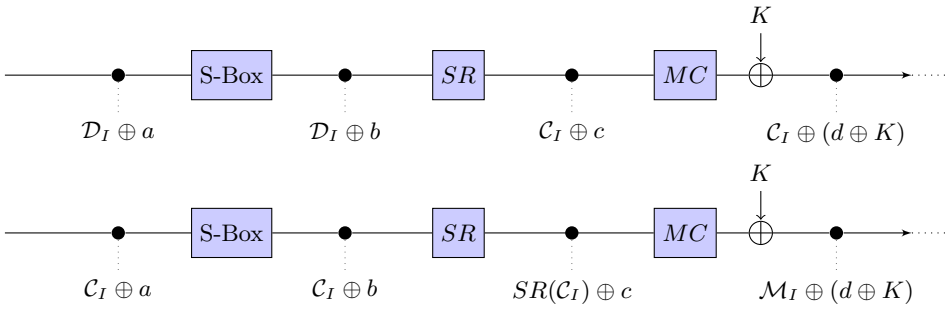

Fig. 3. The essential subspaces in the AES round.

Proof. As we have just seen, since SubBytes is bijective and operates on each byte independently, it simply changes the coset $\mathcal{D}_I \oplus a$ to $\mathcal{D}_I \oplus a'$, where $a'_{i,j} =$ S-Box$(a_{i,j})$ for each $i, j = 0, ..., 3$. ShiftRows simply moves the bytes of $\mathcal{D}_I \oplus a'$ to a column space $\mathcal{C}_I \oplus b^{'}$, where $b' = SR(a')$. MixColumns affects only the constant columns, thus $MC(\mathcal{C}_I \oplus b^{'}) = \mathcal{C}_I \oplus MC(b') = \mathcal{C}_I \oplus b^{''}$. Key addition then changes the coset to $\mathcal{C}_I \oplus b$. □

Lemma 2. *Let $I \subseteq \{0, 1, 2, 3\}$ where $0 < |I| \leq 3$ and $a \in \mathcal{C}_I^{\perp}$. There exists unique $b \in \mathcal{M}_I^{\perp}$ such that*

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

Proof. By definition 5, the mixed spaces $\mathcal{M}_I$ are defined as the application of the linear layer in AES to column spaces $\mathcal{C}_I$. Since the SubBytes layer only moves a coset $\mathcal{C}_I \oplus a$ to a coset $\mathcal{C}_I \oplus a'$, it follows that for any fixed coset $\mathcal{C}_I \oplus a$, there exists $b \in \mathcal{M}_I^{\perp}$ such that $MC \circ SR \circ SB(\mathcal{C}_I \oplus a) \oplus K = \mathcal{M}_I \oplus b$, where $b = MC \circ SR(a') \oplus K$ and $a'_{i,j} =$ S-Box$(a_{i,j})$ for each $i, j = 0, ..., 3$. □

This simply states that a coset of a sum of diagonal spaces $\mathcal{D}_I$ encrypt to a coset of a corresponding sum of column spaces $\mathcal{C}_I$ through one round. Similarly, a coset of a sum of column spaces $\mathcal{C}_I$ encrypts to a coset of the corresponding sum of mixed spaces $\mathcal{M}_I$ over one round.

4 Intersecting AES Subspaces

We continue with useful properties of AES subspaces. In this section we show the following: diagonal spaces and column spaces have non-trivial intersection, column spaces and mixed spaces have non-trivial intersection, but diagonal spaces and mixed spaces have only trivial intersection. This will be useful for creating subspace trails covering a higher number of rounds.

Lemma 3. $\mathcal{D}_i \cap \mathcal{C}_j =< e_{j+i,j} >$.

Proof. $\mathcal{D}_i$ space corresponds to a symbolic matrix with variables along the i-th diagonal, while $\mathcal{C}_j$ has variables in the j-th column. Any diagonal and column meets in exactly one byte, precisely in row $j+i$ and column j. □

It follows that $\mathcal{D}_I \cap \mathcal{C}_J =< e_{j+i,j} \,|\, i \in I, j \in J >$ where $j+i$ is taken modulo 4. In particular, the intersection has dimension $|I| \cdot |J|$.

Lemma 4. $\mathcal{C}_i \cap \mathcal{M}_j = MC \circ SR(\mathcal{D}_i \cap \mathcal{C}_j) =< MC(e_{j+i,i}) >$.

Proof. We have that $MC \circ SR(\mathcal{D}_i) = \mathcal{C}_i$ and by definition 5, $\mathcal{M}_i = MC \circ SR(\mathcal{C}_i)$. By Lemma 3, $\mathcal{D}_i \cap \mathcal{C}_j =< e_{j+i,j} >$. Thus it follows that $< MC(e_{j+i,j}) >= MC \circ SR(\mathcal{D}_i) \cap MC \circ SR(\mathcal{C}_j) = \mathcal{D}_i \cap \mathcal{M}_j$. Finally, since $SR(e_{r,c}) = e_{r,c-r}$, we obtain that $< MC \circ SR(e_{j+i,j}) >=< MC(e_{j+i,i}) >$. □

Thus, for two subspaces $\mathcal{C}_I$ and $\mathcal{M}_J$ for non-empty subsets I and J of $\{0,1,2,3\}$, it follows that $\mathcal{C}_I \cap \mathcal{M}_J =< MC(e_{j+i,i}) \,|\, i \in I, j \in J >$ (where $i+j$ is taken modulo 4) which has dimension $|I| \cdot |J|$. While the spaces $\mathcal{D}_I$ and $\mathcal{C}_J$, and $\mathcal{C}_I$ and $\mathcal{M}_J$ intersect non-trivially, the spaces $\mathcal{D}_I$ and $\mathcal{M}_J$ intersect trivially.

Lemma 5. $\mathcal{D}_i \cap \mathcal{M}_j = \{0\}$ *for all* i *and* j.

Proof. A basis for $\mathcal{M}_j$ is given by:

$$\mathcal{M}_j =< MC(e_{0,j}), MC(e_{1,j-1}), MC(e_{2,j-2}), MC(e_{3,j-3}) >,$$

while a basis for $\mathcal{D}_i$ is given by $\mathcal{D}_i =< e_{0,i}, e_{1,i+1}, e_{2,i+2}, e_{3,i+3} >$, where in both cases the indexes are taken modulo 4.

Suppose by contradiction that $\mathcal{D}_i$ and $\mathcal{M}_j$ has a nonzero intersection. This implies that there exist x_k and y_k for $k = 0, ..., 3$ such that

$$\begin{aligned} &\bigoplus_{k=0}^{3} x_k \cdot < e_{k,i+k} > \oplus \bigoplus_{k=0}^{3} y_k \cdot < MC(e_{k,j-k}) >= \\ &= \bigoplus_{k=0}^{3} \Big[x_{k-i} \cdot < e_{k-i,k} > \oplus y_{k+j} \cdot < MC(e_{k+j,k}) > \Big] = 0. \end{aligned} \tag{1}$$

has a nontrivial solution. This is clearly impossible since $< e_{k-i,k} >$ and $< MC(e_{k+j,k}) >$ are linearly independent for each $k = 0, ..., 3$. Thus, $\mathcal{D}_i$ and $\mathcal{M}_j$ intersect only in zero. □

As long as $|I| + |J| \leq 4$, we have that any combinations of subspaces $\mathcal{D}_I$ and $\mathcal{M}_J$ only intersect in the zero vector. Indeed, consider the sum over k defined in eq. (1). If $|I| + |J| \leq 4$, then for each k (i.e. for each column) there are at most four terms. Among them, there is at least one term of the form $< e_{\cdot,k} >$ and at least one of the form $< MC(e_{\cdot,k}) >$. Thus, equation (1) has only trivial solutions. Instead, note that this is not true if $|I| + |J| > 4$. Indeed, in this case

for each k (i.e. for each column), the equation (1) has at least 5 terms. Since there are only 4 rows, it is always possible to find non trivial solutions[5].

Lemma 6. *$\mathcal{D}_I \cap \mathcal{M}_J = \{0\}$ for all I and J such that $|I| + |J| \leq 4$.*

5 Subspace Distinguishers for AES with Secret Round-Keys

In this section we describe a series of subspace trails for AES. Additionally we also describe how these trails can be used to formulate ways to detect non-randomness, often colloquially referred to a distinguishers. All distinguishers in this section, ranging from two up to four rounds, are independent of the round keys and are formulated without the knowledge of the key. From now on, we assume that any subspaces $\mathcal{D}_I$, $\mathcal{C}_I$ or $\mathcal{M}_I$ has nonzero dimension (that is, $I \subseteq \{0,1,2,3\}$ is not empty). Moreover, when we intersect two subspaces $\mathcal{D}_I$ and $\mathcal{M}_J$, where both I and J are assumed non-empty, we always assume that the sum of their dimensions is not larger than 16. Typically, the sum of their dimensions will be exactly 16.

5.1 2-Rounds Subspace Distinguisher for AES

It follows directly from Section 3.1 that plaintexts from diagonal spaces are encrypted over two rounds to ciphertexts in mixed subspaces. Let $R^{(2)}$ denote two AES rounds with fixed random round keys $K = K_1, K_2$. Let $I \subseteq \{1,2,3,4\}$ nonzero and fixed. By Lemma 1, a coset $\mathcal{D}_I \oplus a$ of dimension $4 \cdot |I|$ encrypts to a coset $R_{K_1}(\mathcal{D}_I \oplus a) = \mathcal{C}_I \oplus a'$ over one round. By Lemma 2, there exists unique b (relative to the round keys and the constant a') such that $R_{K_2}(\mathcal{C}_I \oplus a') = \mathcal{M}_I \oplus b$. By combining the two rounds, we get that for each $a \in \mathcal{D}_I^{\perp}$, there exists unique $b \in \mathcal{M}_I^{\perp}$ such that $R^{(2)}(\mathcal{D}_I \oplus a) = \mathcal{M}_I \oplus b$.

Consequently, we get the following properties. If two plaintexts belong to the same coset of a diagonal space $\mathcal{D}_I$, then their encryption belongs to the same coset of a mixed space $\mathcal{M}_I$. In particular, for a two round encryption R^2 with fixed keys, we have that

$$Pr(R^{(2)}(u) \oplus R^{(2)}(v) \in \mathcal{M}_I \,|\, u \oplus v \in \mathcal{D}_I) = 1 \tag{2}$$

for nonzero set I of $\{0,1,2,3\}$ (i.e. $|I| \neq 0$) and where $u \neq v$. The opposite follows directly; if two plaintexts belong to different cosets of a diagonal space

[5] For example, the first column (i.e. $k = 0$) of the intersection $\mathcal{D}_{0,1,2} \cap \mathcal{M}_{0,1}$ is equal to:

$$(\mathcal{D}_{0,1,2} \cap \mathcal{M}_{0,1})_{col(0)} \equiv MC\left(\begin{bmatrix} x \\ (\alpha+1)\cdot x \\ 0 \\ 0 \end{bmatrix}\right) \equiv \begin{bmatrix} (\alpha^2+\alpha+1)\cdot x \\ (\alpha^2+\alpha+1)\cdot x \\ \alpha \cdot x \\ 0 \end{bmatrix} \qquad \forall x \in \mathbb{F}_{2^8}.$$

$\mathcal{D}_I$, then their encryption belongs to different cosets of a mixed space $\mathcal{W}_I$. In other words,

$$Pr(R^{(2)}(u) \oplus R^{(2)}(v) \in \mathcal{M}_I \,|\, u \oplus v \notin \mathcal{D}_I) = 0.$$

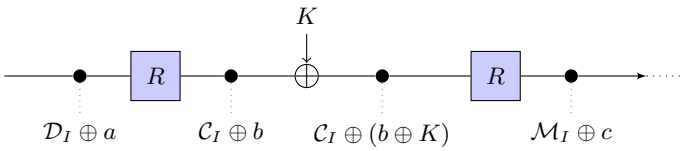

Fig. 4. Subspaces over two rounds.

These properties are used to set up the distinguisher for two rounds. However, other interesting properties hold when one considers two rounds of encryption. In particular, by Lemma 6, the intersection between a mixed space $\mathcal{M}_I$ space and a diagonal space $\mathcal{D}_J$ space contains only zero, if $|I| + |J|$ is less than 4. Thus, if two plaintexts are in the same coset of $\mathcal{M}_I$, they must belong to different cosets of $\mathcal{D}_J$. In other words, for $\mathcal{D}_I$ and $\mathcal{D}_J$ such that $\dim(\mathcal{D}_I) + \dim(\mathcal{D}_J) \leq 16$ (and $|I|, |J| \neq 0$)

$$Pr(R^{(2)}(u) \oplus R^{(2)}(v) \in \mathcal{D}_J \,|\, u \oplus v \in \mathcal{D}_I) = 0 \tag{3}$$

where $u \neq v$, since $R^{(2)}(u)$ and $R^{(2)}(v)$ are both in the same coset of $\mathcal{M}_I$ and thus are always in different cosets of $\mathcal{D}_J$. We can get similar results for the mixed spaces $\mathcal{M}_I$. In particular, if two plaintexts belong to the same coset of a mixed space $\mathcal{M}_I$, then their two round encryptions belong to different cosets of any mixed space $\mathcal{M}_J$. Indeed, two (different) elements of $\mathcal{M}_I$ belong to different cosets of $\mathcal{D}_J$ (since $\mathcal{M}_I \cap \mathcal{D}_J = \{0\}$). Since $R^{(2)}(u) \oplus R^{(2)}(v) \in \mathcal{M}_J$ if and only if $u \oplus v \in \mathcal{D}_J$, we obtain the desired result. Thus, for $\mathcal{M}_I$ and $\mathcal{M}_J$ such that $0 < \dim(\mathcal{M}_I) + \dim(\mathcal{M}_J) \leq 16$, we have that

$$Pr(R^{(2)}(u) \oplus R^{(2)}(v) \in \mathcal{M}_J \,|\, u \oplus v \in \mathcal{M}_I) = 0 \tag{4}$$

if $u \neq v$. We'll use these probabilities to set up an efficient 4 rounds distinguisher.

A Concrete Distinguisher for 2 Rounds. As we have seen, if two plaintexts belong to the same coset of $\mathcal{D}_I$, then they belong to the same coset of $\mathcal{M}_I$ with probability 1 after two rounds - for each I. Consider instead two random texts. By simple computation, the probability that there exists I such that they belong to the same cosets of $\mathcal{M}_I$ is $\binom{4}{|I|} \cdot (2^8)^{-16+4\cdot|I|}$ (note that there are $\binom{4}{|I|}$ different subspaces $\mathcal{M}_I$). Setting $|I| = 1$, this probability is equal to 2^{-94}.

Thus, one pair of plaintexts (that is 2 texts) is sufficient to distinguish the random case from the other one. Indeed, on average in the random case we expect $2^{-94} \cdot 2 = 2^{-93} \simeq 0$ *collisions* (when two elements belong to the same coset of

$\mathcal{M}_I$, we say that there is a "collision"), while this number is equal to 1 (with probability 1) in the other case. The cost of this distinguisher is hence 2 texts. An equivalent distinguisher over 2 rounds was already introduced in [10], where authors investigated how the components of the AES interact over 2 rounds.

Data: Pair of texts c^1 and c^2.
Result: i such that $c^1 \oplus c^2 \in \mathcal{M}_i$, -1 otherwise.
$c \leftarrow MC^{-1}(c^1 \oplus c^2)$;
for i from *0 to* 3 **do**
  if $c_{(i+1)\%4,0} = 0$ *AND* $c_{(i+2)\%4,0} = 0$ *AND* $c_{(i+3)\%4,0} = 0$
   AND $c_{i,1} = 0$ *AND* $c_{(i+1)\%4,1} = 0$ *AND* $c_{(i+2)\%4,1} = 0$
   AND $c_{i,2} = 0$ *AND* $c_{(i+1)\%4,2} = 0$ *AND* $c_{(i+3)\%4,2} = 0$
   AND $c_{i,3} = 0$ *AND* $c_{(i+2)\%4,3} = 0$ *AND* $c_{(i+3)\%4,3} = 0$ **then**
    return i;
  end
end
return -1.

Algorithm 1: Distinguisher for 2-rounds of AES - Pseudo-code.

Finally, note that a similar distinguisher can be used for the 1 round case. Indeed, note that if two plaintexts belong to the same coset of $\mathcal{D}_I$ (equivalently $\mathcal{C}_I$), then they belong to the same coset of $\mathcal{C}_I$ (equivalently $\mathcal{M}_I$) with probability 1 for each I after 1 round. Moreover, observe that it also is possible to set up a 2 rounds distinguisher using the impossible differential properties defined in (3) or (4).

5.2 3-Round Subspace Distinguisher for AES

To form a three round distinguisher, we extend a two round distinguisher to three rounds. The following theorem describes the essential step for the extension.

Theorem 1. *For any $\mathcal{M}_I$ and $\mathcal{M}_J$, we have that*

$$Pr(R(u) \oplus R(v) \in \mathcal{M}_J \,|\, u \oplus v \in \mathcal{M}_I) = (2^8)^{-4|I|+|I|\cdot|J|}.$$

Proof. In the previous section, we have seen that $R(x) \oplus R(y) \in \mathcal{M}_J$ if and only if $x \oplus y \in \mathcal{C}_J$. This implies that the probability given in the Theorem is equivalent to the following:

$$Pr(R(x) \oplus R(y) \in \mathcal{M}_J \,|\, x \oplus y \in \mathcal{M}_I) = Pr(x \oplus y \in \mathcal{C}_J \,|\, x \oplus y \in \mathcal{M}_I) = Pr(z \in \mathcal{C}_J \,|\, z \in \mathcal{M}_I).$$

Let $\mathcal{Z} = \mathcal{M}_I \cap \mathcal{C}_J$. In Section 4, it is shown that $\dim(\mathcal{Z}) = \dim(\mathcal{M}_I \cap \mathcal{C}_J) = |I|\cdot|J|$. Let $\mathcal{Y}$ the subspace of dimension $4 \cdot |I| - |I| \cdot |J|$ such that $\mathcal{M}_I = \mathcal{Y} \oplus \mathcal{Z}$, and let $\pi_{\mathcal{Y}}$ and $\pi_{\mathcal{Z}}$ the projection of $\mathcal{M}_I$ on $\mathcal{Y}$ and $\mathcal{Z}$ respectively:

$$\pi_{\mathcal{Y}} : \mathcal{M}_I \to \mathcal{Y}, \qquad \pi_{\mathcal{Y}}(x) = x_y,$$
$$\pi_{\mathcal{Z}} : \mathcal{M}_I \to \mathcal{Z}, \qquad \pi_{\mathcal{Z}}(x) = x_z.$$

That is, $\forall x \in \mathcal{M}_I$, there exists unique $x_y \in \mathcal{Y}$ and $x_z \in \mathcal{Z}$ such that $x = x_z \oplus x_y$.

It follows that:

$$Pr(x \in \mathcal{C}_J \,|\, x \in \mathcal{M}_I) = Pr(\pi_{\mathcal{Y}}(z) = 0 \,|\, z \in \mathcal{M}_I).$$

Since $\mathcal{Y}$ has dimension $4 \cdot |I| - |I| \cdot |J|$, we obtain:

$$Pr(R(x) \oplus R(y) \in \mathcal{M}_J \,|\, x \oplus y \in \mathcal{M}_I) = Pr(\pi_{\mathcal{Y}}(z) = 0 \,|\, z \in \mathcal{M}_I) = (2^8)^{-4\cdot|I|+|I|\cdot|J|}.$$

□

Note that if $|J| = 4$ (i.e. if $\mathcal{M}_J$ is all the space), then the probability is equal to 1.

Let $c \in \mathcal{W}_I^{\perp}$. Given $\mathcal{Z} := \mathcal{M}_I \cap \mathcal{C}_J$ and $\mathcal{Y} := \mathcal{M}_I \setminus \mathcal{Z}$, then

$$\mathcal{M}_I = \mathcal{Z} \oplus \mathcal{Y} = \bigcup_{a \in \mathcal{Y}} \mathcal{Z} \oplus a \subseteq \bigcup_{a \in \mathcal{Y}} \mathcal{C}_J \oplus a = \mathcal{C}_J \oplus \mathcal{Y}, \qquad \text{and}$$

$$\mathcal{M}_I \oplus c = \bigcup_{a'_i \in \mathcal{C}_J \setminus \mathcal{Z}} \mathcal{Z} \oplus (a'_i \oplus c) = \bigcup_{i=1}^{(2^8)^{4\cdot|I|-|I|\cdot|J|}} \mathcal{Z} \oplus a_i,$$

where $\mathcal{Z} \oplus a_i = (\mathcal{M}_I \cap \mathcal{C}_J) \oplus a_i$ are cosets of dimension $|I| \cdot |J|$.

Let $A_i := \mathcal{Z} \oplus a_i$. Since cosets of $\mathcal{C}_J$ spaces encrypts to cosets of $\mathcal{M}_J$ spaces, we get that

$$B_i := R(A_i) \subseteq R(\mathcal{C}_J \oplus a_i) = \mathcal{M}_J \oplus b_i,$$

since $A_i = \mathcal{Z} \oplus a_i \subseteq \mathcal{C}_J \oplus a_i$ by definition of $\mathcal{Z}$. As a consequence:

$$R^{(3)}(\mathcal{D}_I \oplus a) = R(\mathcal{M}_I \oplus c) = R\left(\bigcup_{i=1}^{n} A_i\right) = \bigcup_{i=1}^{n} R(A_i) = \bigcup_{i=1}^{n} B_i \subseteq \bigcup_{i=1}^{n} \mathcal{M}_J \oplus b_i,$$

where $n := (2^8)^{4\cdot|I|-|I|\cdot|J|}$, R is the application of one round with a fixed key and $R^{(3)}$ the application of three rounds.

For instance, consider a coset of $\mathcal{D}_I$. After two rounds, each element belongs to a coset of $\mathcal{M}_I$. Equivalently, for a given J, after two rounds the texts are *uniform* distributed in $(2^8)^{4\cdot|I|-|I|\cdot|J|}$ cosets of $\mathcal{C}_J$. That is, after two rounds, there exist $(2^8)^{4\cdot|I|-|I|\cdot|J|}$ cosets of $\mathcal{C}_J$ such that each one of these cosets contains exactly $(2^8)^{|I|\cdot|J|}$ elements. Since each coset of $\mathcal{C}_J$ encrypts to a unique coset of $\mathcal{M}_J$ (that is, two elements that belong to different coset of $\mathcal{C}_J$ can not belong to the same coset of $\mathcal{M}_J$), if we start with a coset of $\mathcal{D}_I$, after three rounds the texts are *uniform* distributed in $(2^8)^{4\cdot|I|-|I|\cdot|J|}$ cosets of $\mathcal{M}_J$.

In order to better understand it, we give an example for the particular case in which $\mathcal{D}_I = \mathcal{D}_i$ is of dimension 1 and $\mathcal{M}_J = \mathcal{M}_{i_1} \oplus \mathcal{M}_{i_2} \oplus \mathcal{M}_{i_3}$ of dimension 12. So, after 2 rounds, the texts are uniform distributed in 2^8 cosets of $\mathcal{C}_J$, i.e. there exist 2^8 cosets of $\mathcal{C}_J$ such that each one of them contains exactly 2^{24} texts. Note that the remaining $2^{32} - 2^8$ cosets of $\mathcal{C}_J$ don't contain any texts. Thus, after 3 rounds the texts are uniform distributed in 2^8 cosets of $\mathcal{M}_J$, i.e. there

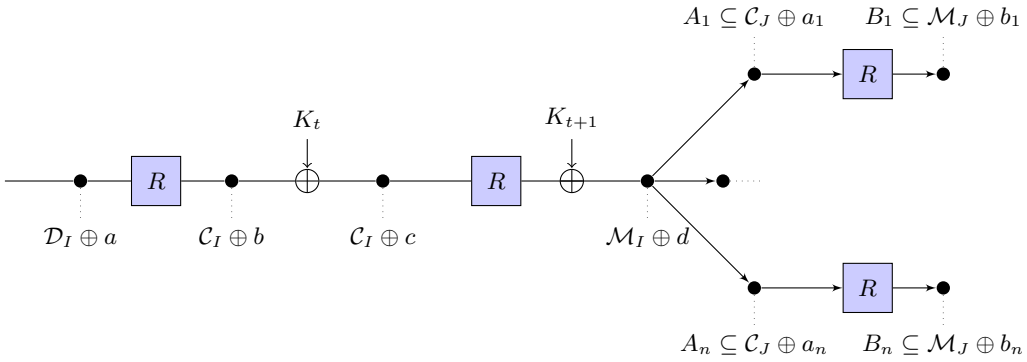

Fig. 5. 3-round distinguishers for AES (the index n is defined as $n := (2^8)^{4\cdot|I|-|I|\cdot|J|}$).

exist 2^8 cosets of $\mathcal{M}_J$ such that each one of them contains exactly 2^{24} texts, since (as we have seen) each coset of $\mathcal{C}_J$ is mapped in exactly one coset of $\mathcal{M}_J$, while the remaining $2^{32} - 2^8$ cosets of $\mathcal{C}_J$ don't contain any texts.

A Concrete Distinguisher for 3 Rounds. In order to set up the distinguisher, we exploit the difference of probability to have a collision in the ciphertexts set between the case in which two plaintexts are taken in a random way and the case in which two plaintexts belong to the same coset of $\mathcal{D}_I$.

The probabilities that two elements drawn randomly from $\mathbb{F}_{2^8}^{4\times 4}$ (denoted by p_1) and that two plaintexts drawn from a coset of $\mathcal{D}_I$ (denoted by p_2) belong to the same coset of $\mathcal{M}_J$ are respectively:

$$p_1 = \binom{4}{|J|} \cdot (2^8)^{-16+4|J|}, \qquad p_2 = \binom{4}{|J|} \cdot (2^8)^{-4|I|+|I||J|}.$$

It is very easy to observe that the probability to have a collision in the second case is higher than in the random case. In particular, for $|J| = 3$ and $|I| = 1$, we obtain that $p_2 = 2^{-6}$ while $p_1 = 2^{-30}$. Thus, the idea is to look for the minimum number of texts m in order to guarantee at least one collision in the "subspace case" and zero in the random case (with high probability).

To do this, we recall the *birthday paradox*. Given d (equally likely) values and n variables, the probability that at least two of them have the same value is given by:

$$p = 1 - \frac{n!}{(n-d)! \cdot n^d} = 1 - \frac{(d)!}{n^d} \cdot \binom{n}{d} \simeq 1 - e^{\frac{-d(d-1)}{2n}},$$

where the last one is an useful approximation.

Since if we encrypt two plaintexts from a coset of $\mathcal{D}_I$, each of them can only belong to one of the 2^8 cosets of $\mathcal{M}_J$ defined as before, the probability that there is at least one collision in a coset is equal to the probability that two elements belong to the same cosets of $\mathcal{M}_J$, that is $p = 1 - e^{-m(m-1)/(2\cdot 2^8)}$. However,

this property holds if we choose any of the four 12-dimensional space $\mathcal{M}_J$ as a target distinguisher space, each yielding an independent experiment. Since this experiments are independent, we have that the probability to have at least one collision in the subspace case given m texts is:

$$p = 1 - \left(\frac{2^8!}{(2^8 - m)! \cdot (2^8)^d} \right)^4 \simeq 1 - \left(e^{\frac{-m(m-1)}{2 \cdot 2^8}} \right)^4 = 1 - e^{\frac{-m(m-1)}{2 \cdot 2^6}}.$$

Thus, if we set $m = 20$, we get that the probability to have at least one collision in one of the four different $\mathcal{M}_J$ spaces (with $|J| = 1$) is 95.251% (14 texts are sufficient to have at least one collision with probability greater than 75%). In order to distinguish the two sets (that is, the random one and the "subspace" one), the verifier has to construct all the possible pairs of texts and to count the number of collisions, for each of them. In particular, given 20 texts (that is, 190 different pairs), we expect $190 \cdot 2^{-6} \simeq 3$ collisions in the subspace case and $190 \cdot 2^{-30} = 2^{-22.4} \simeq 0$ in the random case.

Observe that the distinguisher works in similar way in the decryption direction, with the same complexity.

Data: 20 texts c^i (for $i = 1, ..., 20$).
Result: number of collisions.
$n \leftarrow 0$;
for each pair (c^i, c^j) *with* $i \neq j$ **do**
  $c \leftarrow MC^{-1}(c^i \oplus c^j)$;
  for k from *0 to 3* **do**
    if $c_{k,0} = 0$ *AND* $c_{(3+k)\%4,1} = 0$ *AND* $c_{(2+k)\%4,2} = 0$ *AND* $c_{(1+k)\%4,3} = 0$ **then**
      $n \leftarrow n + 1$;
      next pair
    end
  end
end
return n.

Algorithm 2: Distinguisher for 3-rounds of AES - Pseudo-code.

Finally, in a very different scenario, an analogous distinguisher (based on truncated differential) was introduced in [5], and showed in Fig. 6. Consider a pair of plaintexts that belong to the same coset of $\mathcal{D}_0$. With probability $2^{-8} \cdot 4 = 2^{-6}$, after one round they belong to the same coset of $\mathcal{C}_0 \cap \mathcal{D}_I$, for a certain I with $|I| = 3$. That is, with probability 2^{-6}, after one round only three bytes are active instead of four. Thus, since $\mathcal{C}_0 \cap \mathcal{D}_I \subseteq \mathcal{D}_I$ and since for each $a \in \mathcal{D}_I^{\perp}$ there exists unique $b \in \mathcal{M}_I^{\perp}$ such that $R^{(2)}(\mathcal{D}_I \oplus a) = \mathcal{M}_I \oplus b$, if two texts belong to the same coset of $\mathcal{D}_i$, then they belong to the same coset of $\mathcal{M}_I$ with $|I| = 3$ after three rounds with probability 2^{-6}.

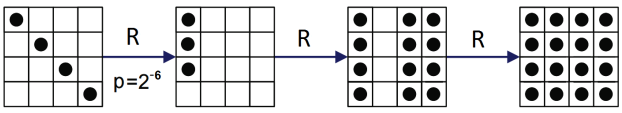

Fig. 6. Truncated differential characteristic over 3-rounds AES. White box denotes a byte with a zero difference, while black box denotes a byte with a non-zero difference.

5.3 4-Rounds Subspace Distinguisher for AES

From now on, we assume that I and J satisfy the condition $0 < |I| + |J| \leq 4$ (in order to use Lemma 6). To set up the 4-rounds distinguisher, we start from the 2-rounds one. Fix $\mathcal{D}_I$ and $\mathcal{D}_J$ such that $0 < \dim(\mathcal{D}_I) + \dim(\mathcal{D}_J) \leq 16$. We can construct a four round trail by simply combining two-round subspaces properties. Indeed, we have seen that

$$Pr(R^{(2)}(u) \oplus R^{(2)}(v) \in \mathcal{M}_I \,|\, u \oplus v \in \mathcal{D}_I) = 1$$
$$Pr(R^{(2)}(u) \oplus R^{(2)}(v) \in \mathcal{M}_J \,|\, u \oplus v \in \mathcal{M}_I) = 0$$

if $u \neq v$. Combining these two probabilities for two-rounds yields a four round probability

$$Pr(R^{(4)}(u) \oplus R^{(4)}(v) \in \mathcal{M}_J \,|\, u \oplus v \in \mathcal{D}_I) = 0 \tag{5}$$

where $u \neq v$. This means that the adversary can pick any coset of a non-zero plaintext space $\mathcal{D}_I$ and a non-zero ciphertext space $\mathcal{M}_J$, as long as $0 < \dim(\mathcal{D}_I) + \dim(\mathcal{M}_J) \leq 16$, and distinguish on the fact that the probability that two plaintexts encrypt to the same coset of the ciphertext space is zero over four rounds.

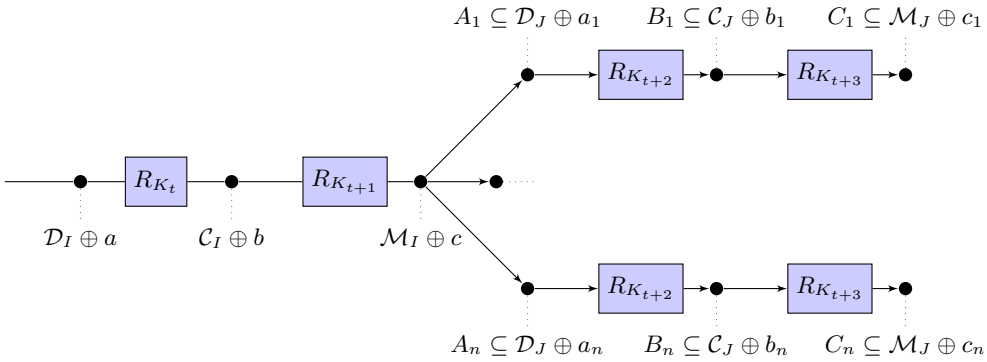

Fig. 7. 4-round distinguishers for AES (where the index n is defined as $n := (2^8)^{4|I|}$ and the indexes I and J satisfy the condition $0 < |I| + |J| \leq 4$).

A Concrete Distinguisher for 4 Rounds. The idea is pick parameters that maximize probability in the random case. The best minimal data complexity is found if we choose $|J| = 3$. This implies that $|I| = 1$, since we have the condition that $|I| + |J| \leq 4$. In this case, the probability that two random elements belong to the same coset of $\mathcal{M}_J$ for a certain J with $|J| = 3$ is 2^{-30} (as we have already seen). Instead, the probability that two elements, that belong to the same coset of $\mathcal{D}_I$, belong to the same coset of $\mathcal{M}_J$ after four rounds is 0.

Exactly as before, the idea is to look for the minimum number of texts m in order to guarantee at least one collision in the random case with high probability. Since there are four 12-dimensional space $\mathcal{M}_J$ and using the birthday paradox, the probability to have at least one collision in the random case given m texts is well approximated by $p = 1 - e^{-m(m-1)/(2 \cdot 2^{30})}$. Thus, $m \simeq 2^{16.25}$ texts are sufficient to set up a 4-Rounds distinguisher (in this case, the probability to have a collision in the random case is approximately 95% - note that $2^{15.75}$ texts are sufficient to have at least one collision with probability of 75%). Indeed, given $2^{16.25}$ texts (that is about $2^{31.5}$ pairs), the number of collision in the random case is on average $2^{31.5} \cdot 2^{-30} = 2^{1.5} \approx 3$, while the number of collision in the other case is $2^{31.5} \cdot 0 = 0$. That is, $2^{16.25}$ chosen plaintexts are sufficient for this distinguisher.

Data: $2^{16.25}$ texts c^i (for $i = 1, ..., 2^{16.25}$).
Result: 1 if there is at least one collision, 0 otherwise.
for each pair (c^i, c^j) *with* $i \neq j$ **do**
  $c \leftarrow MC^{-1}(c^i \oplus c^j)$;
  for k from *0 to 3* **do**
    if $c_{k,0} = 0$ *AND* $c_{(3+k)\%4,1} = 0$ *AND* $c_{(2+k)\%4,2} = 0$ *AND* $c_{(1+k)\%4,3} = 0$ **then**
      return 1;
    end
  end
end
return 0.

Algorithm 3: Distinguisher for 4-rounds of AES - Pseudo-code.

Note that this distinguisher exploits the Impossible Differential property presented in [3]. Thus, it is not a surprise that the computational complexity of these two distinguishers is the same. Only for completeness, note that it is possible to set up a 0-probability distinguishers also for the 3-rounds case:

$$Pr(R^{(3)}(x) \oplus R^{(3)}(y) \in \mathcal{M}_I \,|\, x \oplus y \in \mathcal{C}_J) =$$
$$= Pr(R^{(3)}(x) \oplus R^{(3)}(y) \in \mathcal{C}_I \,|\, x \oplus y \in \mathcal{D}_J) = 0$$

where $0 < |I| + |J| \leq 4$. Since in the random case, the probability that two elements belong to the same coset of $\mathcal{C}_I$ or $\mathcal{M}_I$ is upper bounded by 2^{-30} for each I and J, one needs at least $2^{15.75}$ chosen plaintexts to set up this distinguisher.

That is, in the case of 3-rounds of AES, the 0-probability distinguisher is worse than the one described in the previous section

Finally, note that also the 4-rounds distinguisher (as the 3-rounds one) works also in the decryption direction. In this case, using the same argumentation as before, if we two texts belong to the same coset of $\mathcal{M}_I$, then they belong to two different cosets of $\mathcal{D}_J$ four rounds before for $|I| + |J| \leq 4$.

Relationship between 4-Rounds Subspace Trail and Impossible Differential Cryptanalysis. We would like to highlight the relationship between the 4-rounds subspace trails found in Sect. 5.3 and the impossible differential cryptanalysis. As we have seen, if $0 < \dim(\mathcal{D}_I) + \dim(\mathcal{M}_J) \leq 16$ then $Pr(R^{(4)}(x) \oplus R^{(4)}(y) \in \mathcal{M}_J \,|\, x \oplus y \in \mathcal{D}_I) = 0$. We define this subspace trail as a "*0-Probability Subspace Trail*" or "*Impossible subspace trail*". In the following, we'd like to show the relationship between (5) and *Impossible Differential Analysis* [3], [2], which is a generalization of Differential Analysis [4]. Differential cryptanalysis traditionally considers characteristics or differentials with relatively high probabilities and uses them to distinguish the correct unknown keys from the wrong keys. The idea is that the difference predicted by the differential appears frequently only when the correct key is used to decrypt the last few rounds of many pairs of ciphertexts. Impossible differential analysis exploits instead the differences which should not occur (i.e., that have probability exactly zero). In this case, a key that decrypts a pair of ciphertexts to that difference is certainly wrong.

Definition 7. *(**Inverse-diagonal spaces**) The inverse-diagonal spaces $\mathcal{ID}_i$ are defined as*

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

If $I \subseteq \{0,1,2,3\}$, the subspace $\mathcal{ID}_I$ is defined as $\mathcal{ID}_I = \bigoplus_{i \in I} \mathcal{ID}_i$.

For instance, $\mathcal{ID}_0 = SR(\mathcal{C}_0)$ corresponds to the symbolic matrix

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

Using similar argumentations as before, if $|I| + |J| \leq 4$ and if the final MixColumns operation is omitted, then $Pr(R_f \circ R^{(3)}(x) \oplus R_f \circ R^{(3)}(y) \in \mathcal{ID}_J \,|\, x \oplus y \in \mathcal{D}_I) = 0$. Thus, consider 5 rounds of AES:

$$p^h \xrightarrow{R(\cdot)} s^h \xrightarrow{R_f \circ R^{(3)}(\cdot)} c^h$$

for $h = 1, 2$. If there exists a pair of ciphertexts c^1 and c^2 that belong to the same coset of $\mathcal{ID}_J$ (that is $c^1 \oplus c^2 \in \mathcal{ID}_J$), then all the keys of the first round such that $s^1 \oplus s^2 = R(p^1) \oplus R(p^2) \in \mathcal{D}_I$ for $0 < \dim(\mathcal{D}_I) + \dim(\mathcal{ID}_J) \leq 16$ are certainly wrong.

To exploit this fact in order to discover the key, the idea is to choose plaintexts with a particular shape. For simplicity, let $I = \{0\}$ fixed. Suppose to considers pair of plaintexts p^1, p^2 such that $p^1_{i,j} = p^2_{i,j}$ for each $i, j = 0, ...3$ with $(i,j) \neq \{(0,0),(1,3),(2,2),(3,1)\}$ (that is $SR^{-1}(p^1)_{col(i)} = SR^{-1}(p^2)_{col(i)}$ for $i = 1,2,3$). This choice implies that for each key K:

$$R(p^1)_{col(i)} = R(p^2)_{col(i)} \qquad \forall i = 1,2,3,$$

that is the second, the third and the fourth columns of the two texts are equal after one round[6]. Given c^1 and c^2 such that $c^1 \oplus c^2 \in \mathcal{ID}_J$ (with $\dim(Y_J) \geq 12$), in order to guarantee that $R(p^1) \oplus R(p^2) \in \mathcal{D}_I$, the attacker has to work only on the first column of $R(p^1)$ and $R(p^2)$, that is only on the first column of $SR^{-1}(k)$ (for the other columns, all the values are fine). Thus, all the keys such that $R(p^1) \oplus R(p^2) \in \mathcal{D}_I$ are certainly wrong.

There are three possibilities that can be exploited for an impossible differential attack, which are $\dim(\mathcal{D}_I) = 4$ and $\dim(\mathcal{ID}_J) = 12$, $\dim(\mathcal{D}_I) = 12$ and $\dim(\mathcal{ID}_J) = 4$, and finally $\dim(\mathcal{D}_I) = \dim(\mathcal{ID}_J) = 8$. For each of these combinations, using the definitions of $\mathcal{D}_I$ and $\mathcal{ID}_J$ it is possible to obtain and to list all the impossible input/output combinations of difference that can be exploited to set up the attack. In particular, the first combination is exploited for example in [17] and in [1], while the second one is exploited in [18]. Interestingly, in literature there isn't any attack that exploits the last (impossible) input/output combination of differences. A possible reason of this fact is that using this combination it is not possible to attack 7 rounds of AES-128 as for the other combinations. Moreover, even if it is possible to attack 7 rounds of AES-192 and 8 rounds of AES-256 using it, our results (omitted due to page limit) show that in this case the data and the computational complexity is not better than the other attacks already present in literature that exploit the first and the second impossible combinations.

Relationship between 3- and 4-Rounds Subspace Trail and Integral Attack. For comparison, another four round (*without* the final MixColumns operation) integral distinguisher for AES uses the fact that summing over all 2^{32} ciphertexts (formed by encrypting a coset of a diagonal space $\mathcal{D}_i$ four rounds without the final MixColumns operation) is zero. In terms of subspaces, this has a different interpretation.

[6] For completeness, to show this fact we compute the i-th column of $SR^{-1}(p^1)$ and $SR^{-1}(p^2)$ after one round for $i = 1,2,3$. By simple computation, we have that for each $j = 1,2$:

$$R(SR^{-1}(p^j)_{col(i)} = [k^1 \oplus MC \circ \text{S-Box}(p^j \oplus k^0)]_{col(i)},$$

where we use the fact that she ShiftRows, the SubBytes and the AddRoundKey operations can be switched positions. Thus, since the MixColumns operation works on each column independently by the others and since $SR \circ SR^{-1}(p^1)_{col(i)} = p^1_{col(i)} = p^2_{col(i)} = SR \circ SR^{-1}(p^1)_{col(i)}$ for each $i = 1,2,3$, it follows that the second, the third and the fourth columns of the two texts are equal after one round.

First of all, note that the entire space $\mathbb{F}_{2^8}^{4\times 4}$ can be decomposed as $\mathbb{F}_{2^8}^{4\times 4} = \mathcal{ID}_0 \oplus \mathcal{ID}_1 \oplus \mathcal{ID}_2 \oplus \mathcal{ID}_3$, where $\mathcal{ID}_j$ is the j-th inverse diagonal space defined above. Let $\mathcal{ID}_I = \mathcal{ID}_0 \oplus \mathcal{ID}_1 \oplus \mathcal{ID}_2$. If we encrypt the 2^{32} plaintexts of a coset of $\mathcal{D}_i$ (for four round without the final MixColumns), we get a set of 2^{32} ciphertexts $C = \{c_1, c_2, \ldots, c_{2^{32}}\}$, where each c_i belongs to a different coset of $\mathcal{ID}_I$. If we decompose these vectors with respect to the subspaces $\mathcal{ID}_i$, each c_i can be written as $c_i = c_{i,0} \oplus c_{i,1} \oplus c_{i,2} \oplus c_{i,3}$ where $c_{i,j} \in \mathcal{ID}_j$. Since each c_i belongs to a different coset of $\mathcal{ID}_I$, it means that the components $c_{i,3}$ are all different; thus their sum must be zero since it amounts to summing over all vectors in $\mathcal{ID}_3$. Since this property holds for all four choices of $\mathcal{ID}_I$, it means that all of the components $c_{i,j}$ must be different with respect to the same subspace $\mathcal{ID}_j$, thus the sum over all the vectors in C is zero. In comparison to integrals we have more structure that allows for distinguisher with lower data-complexity.

Finally, note that the integral attack on 4 rounds works in a similar way also in the decryption direction, where in this case it is not necessary to omit the final MixColumns operation. Indeed, given 2^{32} ciphertexts that belong to the same coset of $\mathcal{M}_i$ with $|i| = 1$, then the sum of the corresponding plaintexts is equal to zero. In fact, after one decryption round all these texts belong to the same coset of $\mathcal{C}_i$. Thus, it is a well-known fact that their sum is equal to zero three rounds before. For example, this property is exploited by Knudsen and Rijmen in [14] to construct the first 7-rounds known-key distinguisher for AES. Instead, if the final MixColumns operation is omitted, then one has to take the 2^{32} ciphertexts in the same coset of $\mathcal{ID}_i$ with $|i| = 1$. Since the decryption algorithm on 3.5 rounds is equivalent to the encryption one, it follows that the sum of the corresponding plaintexts is equal to zero.

Note that the same holds also for the 3 rounds case. In this case, if the final MixColumns is not omitted, given 2^8 ciphertexts that belong to the same coset of $\mathcal{C}_i \cap \mathcal{M}_j$ for $|j| = |i| = 1$, then they belong to the same coset of $\mathcal{D}_i \cap \mathcal{C}_j$ one round before (that is, only one byte is active one round before), and the sum of the corresponding plaintexts is equal to zero. Instead, if the final MixColumns operation is omitted, then one has to take the 2^8 ciphertexts in the same coset of $\mathcal{C}_i \cap \mathcal{ID}_j$ for $|j| = |i| = 1$. For the same reason of before, the sum of the corresponding plaintexts is equal to zero.

6 A 5-Rounds Secret Key Distinguisher for AES

In CRYPTO 2016, new 5-rounds secret key distinguishers of AES-128 have been presented [19]. In this paper, authors construct a 5-rounds *zero-correlation linear hulls* for AES, and then use it to construct 5-rounds integral distinguisher for AES. First, they present them in the case in which the difference of two sub-key bytes is known (in this case, the distinguisher requires 2^{120} texts). Then they extend it to the general case, i.e. they prove that it is always possible to distinguish 5 rounds of AES from random permutations even when the difference of the sub-keys is unknown (in this case, the distinguisher requires 2^{128} texts, i.e. the entire input-output space). We refer to [19] for more details.

Following a similar procedure, we present a new 5-rounds secret key distinguisher for AES-128 in the chosen-plaintexts mode, which needs much less than the full codebook. Our idea is basically to extend the impossible subspace trail distinguisher on 4-rounds presented in Sect. 5.3 at the beginning. As a result, this distinguisher works both in the case in which the last MixColumns operation is omitted or not. As a result, our secret key distinguisher on 5 rounds of AES needs only $2^{98.2}$ chosen plaintexts instead of 2^{128}.

In order to set up it, first we consider the case in which the difference two sub-key bytes is known. Then we show how to generalize it in the case in which no information about the secret key is known. Finally, we evaluate the possibility to exploit more than a single sub-key bytes difference. However, as we show in the following, the best distinguisher (from the point of view of the data and cost complexity) is the one in which only a single difference of two sub-key bytes is considered. The distinguisher is shown in Fig. 8.

The Difference of Two Sub-Key Bytes is Known. Suppose for the moment to know the difference of two sub-key bytes, that is $\Delta := k_{0,0} \oplus k_{1,1}$. As we've already said, the idea is to extend at the beginning the 4-rounds distinguisher based on impossible differential presented in Sect. 5.3. To do this, the idea is to choose plaintexts that belong to the same coset of $\mathcal{D}_I$ for a certain I after one round.

Thus, consider a set of plaintexts-ciphertexts V_Δ of the form[7]:

$$\begin{aligned} V_\Delta = &\{(p^i, c^i) \text{ for } i = 0, ..., 2^8 - 1 \,|\, p^i_{0,0} \oplus p^i_{1,1} = \Delta \quad \forall i \quad \text{and} \\ &\text{and} \quad p^i_{k,l} = p^j_{k,l} \, \forall (k,l) \neq \{(0,0),(1,1)\} \text{ and } i \neq j\}, \end{aligned} \tag{6}$$

that is plaintexts with 14 constants bytes and where $\Delta := k_{0,0} \oplus k_{1,1}$ and that $|V_\Delta| = 2^8$. It is easy to prove that this choice of plaintexts guarantees that after one round they belong to the same coset of $\mathcal{D}_I$ where $I = \{0, 1, 3\}$ (see for example Footnote 6). That is, there exists unique (unknown) $\tilde{a} \in \mathcal{D}_I^\perp$ such that for each $p \in V_\Delta$, then $R(p) \in \mathcal{D}_I \oplus \tilde{a}$ for $I = \{0, 1, 3\}$. Equivalently, if $p, q \in \mathcal{V}_\Delta$, then $R(p) \oplus R(q) \in \mathcal{D}_I$. More in details, there exists unique (unknown) $a \in (\mathcal{D}_I \cap \mathcal{C}_0)^\perp$ such that $R(V_\Delta) \subseteq (\mathcal{C}_0 \cap \mathcal{D}_{0,1,3}) \oplus a$ (note that $|\mathcal{C}_3 \cap \mathcal{D}_{0,1,3} \oplus a| = 2^{24}$).

Proposition 1. *Let V_Δ defined as in (6) and let $I = \{0, 1, 3\}$. There exists $a \in (\mathcal{D}_I \cap \mathcal{C}_0)^\perp$ such that $R(V_\Delta) \subseteq (\mathcal{C}_0 \cap \mathcal{D}_{0,1,3}) \oplus a$.*

Proof. First of all, note that given two arbitrary elements p and q in V_Δ, then after one round their second, third and fourth columns are equal, that is $R(p)_{i,j} = R(q)_{i,j}$ $\forall i = 0, ..., 3$ and $\forall j \neq 0$. Thus, in order to prove that $R(V_\Delta) \subseteq (\mathcal{C}_0 \cap \mathcal{D}_{0,1,3}) \oplus a$, it is sufficient to prove that given two arbitrary elements p and q in

[7] In [19], authors consider a set of plaintexts-ciphertexts $\tilde{V}_{\tilde{\Delta}}$ of the form $\tilde{V}_{\tilde{\Delta}} = \{(p, c) \,|\, c_{0,0} \oplus c_{1,3} = \tilde{\Delta}\}$ where $\tilde{\Delta} = k_{0,0} \oplus k_{1,3}$ and with anyone assumptions on the other bytes. Note that $|\tilde{V}_{\tilde{\Delta}}| = 2^{120}$.

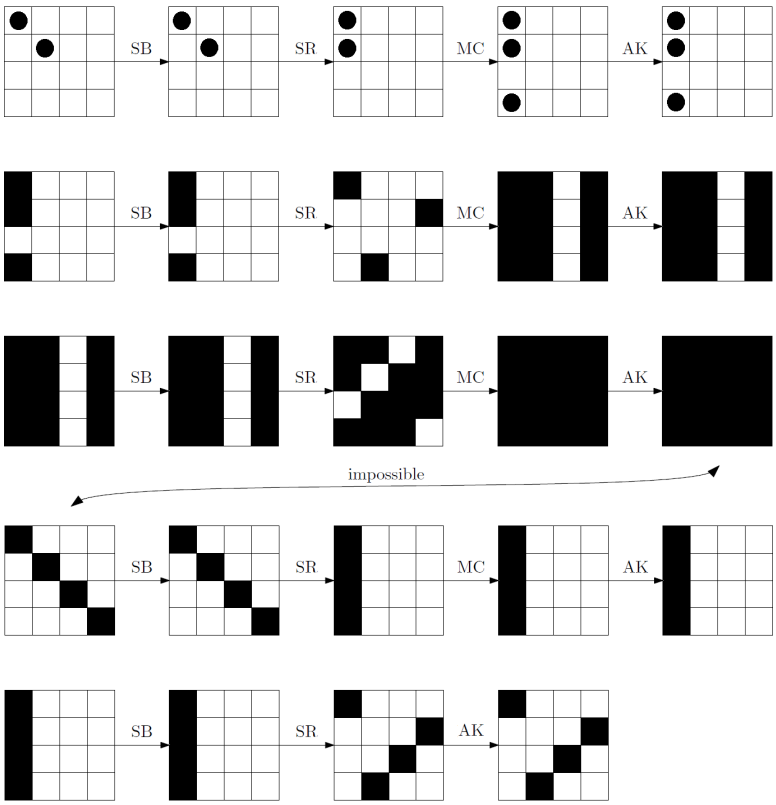

Fig. 8. 5-Rounds Secret Key Distinguisher based on the Impossible Subspace Trail on 4-Rounds (from Sect. 5.3). The choice of the plaintexts (i.e. $p_{0,0} \oplus p_{1,1} = k_{0,0} \oplus k_{1,1}$) guarantees that after one round there are only three bytes with non-zero difference instead of four. White box denotes denotes a byte with a zero-difference, while a black box denotes a byte with non-zero difference.

V_Δ, then $R(p)_{2,0} \oplus R(q)_{2,0} = 0$. By simple computation:

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

First of all observe that $\text{S-Box}(p_{0,0} \oplus k^0_{0,0}) \oplus \text{S-Box}(p_{1,1} \oplus k^0_{1,1}) = 0$. Indeed, since $p_{0,0} \oplus p_{1,1} = k_{0,0} \oplus k_{1,1}$ by definition, then $p_{0,0} \oplus k^0_{0,0} = p_{1,1} \oplus k^0_{1,1}$, that is $\text{S-Box}(p_{0,0} \oplus k^0_{0,0}) = \text{S-Box}(p_{1,1} \oplus k^0_{1,1})$ and so $\text{S-Box}(p_{0,0} \oplus k^0_{0,0}) \oplus \text{S-Box}(p_{1,1} \oplus k^0_{1,1}) = 0$. Thus:

$$R(p)_{2,0} = \alpha \cdot\ \text{S-Box}(p_{2,2} \oplus k_{2,2}) \oplus (\alpha + 1) \cdot\ \text{S-Box}(p_{3,3} \oplus k_{3,3})$$

and in a similar way:

$$R(q)_{2,0} = \alpha \cdot\ \text{S-Box}(q_{2,2} \oplus k_{2,2}) \oplus (\alpha + 1) \cdot\ \text{S-Box}(q_{3,3} \oplus k_{3,3}).$$

Since $p_{2,2} = q_{2,2}$ and $p_{3,3} = q_{3,3}$ by definition, it follows that $R(p)_{2,0} = R(q)_{2,0}$, and so the thesis. □

Note that the assumption $p^i_{k,l} = p^j_{k,l}$ for each $(k,l) \neq \{(0,0),(1,1)\}$ and $i \neq j$ is necessary. Indeed, without this assumption, it is not true that all the plaintexts belong to the same coset of $\mathcal{D}_I$ for $I = \{0,1,3\}$ after one round.

Since $R(p) \oplus R(q) \in \mathcal{D}_{\{0,1,3\}}$ for each pair of plaintexts p and q in V_Δ, then $R^{(4)} \circ R(p) \oplus R^{(4)} \circ R(q) = R^{(5)}(p) \oplus R^{(5)}(q) \notin \mathcal{M}_J$ for $|I| + |J| \leq 4$ with probability 1 due to the 4-rounds impossible differential distinguisher of Sect. 5.3. That is:

$$Pr(R^{(5)}(x) \oplus R^{(5)}(y) \in \mathcal{M}_J \,|\, x_{0,0} \oplus x_{1,1} = y_{0,0} \oplus y_{1,1} = \Delta \text{ and} \\ \text{and } x_{i,j} = y_{i,j} \quad \forall (i,j) \neq \{(0,0),(1,1)\}) = 0,$$

for each J with $|J| = 1$ and where $\Delta := k_{0,0} \oplus k_{1,1}$ is known. Thus, if the difference of two sub-key bytes ($\Delta := k_{0,0} \oplus k_{1,1}$) is known, it is possible to construct an impossible differential distinguisher over 5 rounds. Only for completeness, in the case in which the final MixColumns operation is omitted, the previous probability becomes

$$Pr(R_f^{(5)}(x) \oplus R_f^{(5)}(y) \in \mathcal{ID}_J \,|\, x_{0,0} \oplus x_{1,1} = y_{0,0} \oplus y_{1,1} = \Delta \text{ and} \\ \text{and } x_{i,j} = y_{i,j} \quad \forall (i,j) \neq \{(0,0),(1,1)\}) = 0,$$

for each J with $|J| = 1$, where $R_f^{(5)}(\cdot) := R_f \circ R^{(4)}(\cdot)$ and $\mathcal{ID}_j$ is the inverse-diagonal space (defined as $\mathcal{ID}_j = SR(\mathcal{C}_j)$).

In order to set up the distinguisher, we look for the minimum number of texts necessary to have a collision in the random case with high probability. Since $|J| = 1$ and since there are four different J such that $|J| = 1$, the probability that two texts belong to the same coset of $\mathcal{M}_J$ is $4 \cdot (2^8)^{-16+4} = 2^{-94}$ (analogous for $\mathcal{ID}_J$). Thus, given n pairs, the probability to have at least one collision in the same coset of $\mathcal{M}_J$ for $|J| = 1$ is given by

$$p = 1 - \left(e^{-n/2^{96}}\right)^4 = 1 - e^{-n/2^{94}}.$$

If the number of pairs n is approximately $2^{95.6}$, then p is greater than 95%. Given a single set V_Δ, it is possible to construct $2^7 \cdot (2^8 - 1) \simeq 2^{15}$ different pairs. Thus, for the distinguisher we need approximately $2^{95.6} \cdot 2^{-15} = 2^{80.6}$ different sets V_Δ. Since each of this set contains 2^8 texts, the data complexity of the distinguisher is of $2^{80.6} \cdot 2^8 = 2^{88.6}$ text. We'd like to emphasize that for each difference Δ fixed, there are $2^{128} \cdot 2^{-8} = 2^{120}$ different sets of V_Δ.

The 5-Round Secret Key Distinguisher for AES. Next we choose how to extend the previous distinguisher in the case in which the difference $\Delta := k_{0,0} \oplus k_{1,3}$ is not known.

First of all, note that Δ can only assume 2^8 values. The idea is simply to "repeat" the previous distinguisher for each possible values of Δ, i.e. the idea is

to construct a sufficient number of different sets V_Δ for each possible values of Δ [8]. That is, the idea is to construct 2^8 collections of sets, one for each possible value of Δ. For each one of these 2^8 collections, one has to count the number of collisions, i.e. the number of pairs of texts that belong to the same coset of $\mathcal{M}_J$ for $|J| = 1$. For a random permutation, the goal is to have at least one collision for each one of the 2^8 collections, i.e. for each value of Δ. Instead, for the AES permutation, note that there exists one collection in which there is no collisions with probability 1. This collection corresponds to the one for which $\Delta := k_{0,0} \oplus k_{1,1}$. For the other values of Δ, the behavior is similar to that of the random case. Thus, it is not difficult to distinguish the two cases: the random case is the one for which all the collections have at least one collision, while the AES case is the one for which there is one collection with no collisions.

To set up the distinguisher, we are interested to compute the number of sets of the form V_Δ for each one of the 2^8 collection. If each collection has $2^{80.6}$ sets (as before), then for each fixed collection the probability to have one collision is 95%. Since all the 2^8 collections are independent, the probability that there is at least one collision for each one of the 2^8 collections is $0.95^{256} \simeq 2 \cdot 10^{-6}$. In order to have a *total* probability of about 95%, the probability to have at least one collision in each fixed collection has to be approximately $(0.95)^{1/2^8} = 0.9998$. In this way, the total probability is given by $0.9998^{256} = 0.95$. Thus, for each one of the 2^8 collections (i.e. for each Δ), we need at least $2^{97.2}$ pairs to have at least one collision with probability 0.9998 (analogous computation as before). Since each set V_Δ has about 2^{15} different pairs, then we need about $2^{97.2} \cdot 2^{-15} = 2^{82.2}$ different sets for each Δ (instead of $2^{80.6}$ as before), that is $2^{90.2}$ texts for each Δ. Since each set V_Δ has 2^8 texts, the total number of texts required for this distinguisher is of $2^8 \cdot 2^{90.2} = 2^{98.2}$ texts, which is lower than the total input-output space.

To summarize, suppose to have 2^8 collections (one for each Δ), each one with $2^{82.2}$ different sets V_Δ, where each of this set contains 2^8 texts, for a total of $2^{98.2}$ texts. In the random case and with probability 95%, we expect that in each one of these 2^8 collections there is at least one collision. Note that the average number of collisions for each collection (i.e. for each Δ) is about $2^{-94} \cdot 2^{97.2} = 2^{3.2} \simeq 9$. For the AES permutation, we expect that there exists one Δ for which there is no collision with probability 1 in the corresponding collection of sets. For all the other collections, we expect to have at least one collision with probability 95%. We'd like to highlight that given the $2^{98.2}$ texts defined as before, it is always possible to divide them in 2^8 collections (one for each Δ), and that each collection can be divided in a very simple way in $2^{82.2}$ different sets V_Δ (simply using the definition of V_Δ). For example, given a fixed Δ, the corresponding collection is composed of all the texts p such that $p_{0,0} \oplus p_{1,1} = \Delta$.

[8] In [19], in order to construct the secret key distinguisher, authors simply consider all the input-output space, and divide it in the 2^8 subsets defined by $\tilde{V}_{\tilde{\Delta}}$. Then they argue that there exists $\tilde{\Delta}$ such that after 5 rounds the zero-sum property holds (and which corresponds to $\tilde{\Delta} := k_{0,0} \oplus k_{1,3}$). For random permutation, this happens with probability 2^{-120}.

In order to compare this distinguisher with the one presented in [19], we analyze the data and the computational cost of our distinguisher. First of all, to construct all the plaintext-ciphertext pairs, the cost is of $2^{98.2}$ encryptions or oracle queries. For comparison, in [19], since the entire input-output space is required, the cost is of 2^{128} Encryptions or Oracle Queries. Consider instead the cost to check that there exists at least one collision. To do this check, one has to construct all the possible pairs and to check that there is at least one pair that collides in the same coset of $\mathcal{M}_J$ for $|J| = 1$ (note that when the first collision is found, one can consider the next collection). First of all, given a pair of texts, the cost to verify that they collide in the same coset of $\mathcal{M}_I$ is approximate the cost of 1 bit-XOR operation and the cost of an inverse MixColumns. Since the costs of these two operations is negligible compared to a table look-ups, the total cost can be approximated by the cost to construct all the pairs. Note that one has to construct only the pairs of texts that belong to the same coset of V_{Δ}. Thus, the cost for this step can be approximated by 2^8 (number of Δ) $\cdot 2^{82.2}$ (number of sets)$\cdot 2^7 \cdot (2^8 - 1)$ (number of pairs) $\simeq 2^{105.2}$ table look-ups. For the distinguisher presented in [19], the cost to do the verification operation can be approximated to 2^{128} bit-XOR operations. In conclusion, our distinguisher of 5 rounds of AES with a secret key requires $2^{98.2}$ texts, the cost to construct them is of $2^{98.2}$ encryptions and the verification cost is of $2^{105.2}$ table look-ups. The distinguisher presented in [19] requires 2^{128} texts, the cost to construct them is of 2^{128} encryptions or oracle queries and the verification cost is of 2^{132} XOR operations. We'd like to emphasize that *our distinguisher on 5-rounds AES with secret key is the first one that doesn't require all the entire input-output space, but only about* $2^{98.2}$ *texts*. To the best of our knowledge, this is the best distinguisher for the 5-rounds reduced AES in the secret-key setting.

Finally, note that besides the possibility to distinguish between a random permutation and an AES one using less than the entire input-output space, we can also recover some information on the secret key (that is, the difference $k_{0,0} \oplus k_{1,1}$), with a computational cost that is lower than a brute force attack.

Other Distinguishers. In order to construct the previous distinguisher, we focus only on the difference of two sub-key bytes that belong to the same column after the first ShiftRows operation. However, since there are two 1's in each column of the MixColumns matrix, we exploit the possibility to construct distinguishers using the differences of two sub-key bytes (that belong to the same column after the first ShiftRows operation) for more than one column. However, we found that the best distinguisher (from the point of view of the data complexity) is obtained when only one difference of two sub-key bytes is considered. In the following, we present as example the distinguisher in which all the four differences (one for each column) of two sub-key bytes are exploited - the other cases are similar.

Data: 2^8 collections (one for each possible value of Δ. Each collection contains $2^{82.2}$ different sets V_Δ defined as in (6).
Result: Δ if the permutation is an AES permutation (where $\Delta = k_{0,0} \oplus k_{1,1}$); -1 if the permutation is a Random one.

```
for Δ from 0 to 2^8 − 1 do
    flag = 0;
    for each one of the 2^82.2 different sets V_Δ do
        for each pair (c^i, c^j) ∈ V_Δ do          // about 2^15 different pairs
            if c^i ⊕ c^j ∈ M_k for |k| = 1 then  // for details, see Algorithm 1
                flag = 1;
                next collection Δ;
            end
        end
    end
    if flag = 0 then                                  // AES permutation
        return Δ;
    end
end
return −1.                                          // Random permutation
```

Algorithm 4: Distinguisher for 5-rounds of AES - Pseudo-code. The $2^{98.2}$ input texts are already divided in 2^8 collections (one for each Δ), and for each collection the texts are already divided in the sets V_Δ. Given a pair (c^i, c^j), to check that $c^i \oplus c^j$ belongs in $\mathcal{M}_k$ for a $|k| = 1$ or not see for example Algorithm 1.

In the same way as before, let the set of plaintext-ciphertext W_Δ defined as follows

$$W_\Delta = \{(p^i, c^i) \text{ for } i = 0, ..., 2^{32} - 1 \,|\, p^i_{0,0} \oplus p^i_{1,1} = \Delta_0,\ p^i_{1,2} \oplus p^i_{2,3} = \Delta_1,$$
$$p^i_{2,0} \oplus p^i_{3,1} = \Delta_2,\ p^i_{0,3} \oplus p^i_{3,2} = \Delta_3 \quad \forall i \quad \text{and} \quad p^i_{k,l} = p^j_{k,l} \text{ otherwise}\},$$

where $\Delta_0 = k_{0,0} \oplus k_{1,1}$, $\Delta_1 = k_{1,2} \oplus k_{2,3}$, $\Delta_2 = k_{2,0} \oplus k_{3,1}$ and $\Delta_3 = k_{0,3} \oplus k_{3,2}$. Note that $|W_\Delta| = 2^{32}$, thus it is possible to construct $2^{31} \cdot (2^{32} - 1) = 2^{63}$ different pairs.

Proposition 2. *Let W_Δ defined as before and let $I = \{0, 1, 3\}$. There exists $a \in \mathcal{D}_I^\perp$ such that $R(W_\Delta) \subseteq \mathcal{D}_{0,1,3} \oplus a$.*

The proof of this proposition is analogous to that given for Prop. 1. As a consequence, given two elements p and q in W_Δ, then $R(p) \oplus R(q) \in \mathcal{D}_{0,1,3}$, and so as before $R^{(5)}(p) \oplus R^{(5)}(q) \notin \mathcal{M}_J$ with probability 1 for each J with $|J| = 1$.

Suppose to know all the difference Δ_i for each $i = 0, ..., 3$. As we've already seen, to distinguish between a random permutations and the AES one in this case, we need about $2^{95.6}$ pairs. Since each W_Δ contains about $2^{31} \cdot (2^{32} - 1) = 2^{63}$ pairs, we need approximately $2^{95.6} \cdot 2^{-63} = 2^{32.6}$ different sets of W_Δ, that is about $2^{32.6} \cdot 2^{32} = 2^{64.6}$ texts (each set of W_Δ contains 2^{32} texts).

Suppose instead that all the difference Δ_i are unknown. As before, the idea is to construct 2^{32} collections (one for each possible combination of values of Δ_i for $i = 0, ..., 3$), each one with a certain number of sets W_Δ. To compute this number, our goal is to guarantee that in the random case, there is at least one collision for each possible combination of values of Δ_i for $i = 0, ..., 3$ with probability 95%. Using the same computation as before, for each one of the 2^{32} collections (i.e. for each combination of values of Δ_i for $i = 0, ..., 3$), we need at least one collision with probability higher than $0.95^{1/2^{32}} \simeq 1 - 1.1 \cdot 10^{-11}$. Thus, to have at least one collision with this probability for each one of the 2^{32} collections, each collection has to be composed of $2^{100.1}$ pairs (instead of $2^{95.6}$ pairs). Equivalently, this means that for each one of the 2^{32} collections we need about $2^{100.1} \cdot 2^{-63} = 2^{37.1}$ different sets W_Δ. Since each one of these sets contains 2^{32} texts (equivalently 2^{63} pairs) and since there are 2^{32} possible Δ, the total number of texts is $2^{32} \cdot 2^{37.1} \cdot 2^{32} = 2^{101.1}$, which is higher than before.

Observations. We'd like to conclude with some observations regarding our distinguishers. To present them, we focus on the AES case. However, it is simply to generalize them for other encryption scheme design *if* some important assumptions hold. These assumptions (equivalent to the ones for the distinguisher proposed in [19]) are:

- the encryption scheme has to adopt *identical S-Boxes*;
- *at least one column of the MixColumns matrix MC* (or its inverse) *has to contain* (at least) *two identical elements*.

If one of these two assumptions is missing, the above distinguishers don't work. Actually, the first one can be relaxed. Indeed, it is sufficient that only the two S-Boxes that are in the positions in which the MixColumns matrix has identical elements are equal. Note that both the assumptions are necessary to construct V_Δ (for example, the second one is necessary to prove Prop. 1). However, note that *the above distinguishers don't depend on the particular choice of which S-Box and MixColumns matrix MC are used in the cipher*. That is, it works for each S-Box and for each MC matrix for which the previous assumptions hold.

Finally, we'd like to emphasize that these assumptions are quite common for the construction of AES-like ciphers (or more in general, for Substitution-Permutation Network (SPN) ciphers). Indeed, symmetric encryption schemes are usually a trade-off between the security and computational efficiency. Thus, to enhance the performance of an encryption scheme (especially for lightweight cryptography), designers usually use identical S-Box and a diffusion layer which maximize the number of 1's (or elements with relatively low hamming weights). However, these choices can cause some weakness, as we have shown.

7 Conclusion

We have proposed a generalization of invariant subspace cryptanalysis. Compared to other attack vectors and the state-of-the-art, it's application to 1-4

rounds of AES lead to similar or identical distinguishers, for 5-round of AES however we reported the best distinguisher known so far. Future work includes using this approach to devise key-recovery attacks, and apply it to other schemes.

References

1. B. Bahrak and M. R. Aref, "Impossible differential attack on seven-round AES-128." *IET Information Security*, vol. 2, no. 2, pp. 28–32, 2008.
2. E. Biham, A. Biryukov, and A. Shamir, "Cryptanalysis of Skipjack Reduced to 31 Rounds Using Impossible Differentials," in *Advances in Cryptology — EUROCRYPT 1999: International Conference on the Theory and Application of Cryptographic Techniques, Czech Republic. Proceedings*, J. Stern, Ed., 1999, pp. 12–23.
3. E. Biham and N. Keller, "Cryptanalysis of Reduced Variants of Rijndael," unpublished, 2001, http://csrc.nist.gov/archive/aes/round2/conf3/papers/35-ebiham.pdf.
4. E. Biham and A. Shamir, "Differential cryptanalysis of DES-like cryptosystems," *Journal of Cryptology*, vol. 4, no. 1, pp. 3–72, 1991.
5. A. Biryukov and D. Khovratovich, "Two New Techniques of Side-Channel Cryptanalysis," in *Cryptographic Hardware and Embedded Systems - CHES 2007: 9th International Workshop, Austria. Proceedings*, 2007, pp. 195–208.
6. A. Biryukov and A. Shamir, "Structural Cryptanalysis of SASAS," *Journal of Cryptology*, vol. 23, no. 4, pp. 505–518, 2010.
7. J. Daemen, L. R. Knudsen, and V. Rijmen, "The Block Cipher Square," in *Fast Software Encryption - FSE 1997: 4th International Workshop, Israel. Proceedings*, 1997, pp. 149–165.
8. J. Daemen and V. Rijmen, *The Design of Rijndael: AES - The Advanced Encryption Standard*, ser. Information Security and Cryptography. Springer, 2002.
9. ——, "Two-Round AES Differentials," Cryptology ePrint Archive, Report 2006/039, 2006.
10. ——, "Understanding Two-Round Differentials in AES," in *Security and Cryptography for Networks 2006*, vol. 4116, 2006, pp. 78 – 94.
11. P. Derbez, P. Fouque, and J. Jean, "Improved key recovery attacks on reduced-round AES in the single-key setting," in *Advances in Cryptology - EUROCRYPT 2013: 32nd Annual International Conference on the Theory and Applications of Cryptographic Techniques, Greece. Proceedings*, 2013, pp. 371–387.
12. J. Evertse, "Linear Structures in Blockciphers," in *Advances in Cryptology - EUROCRYPT 1987: Workshop on the Theory and Application of of Cryptographic Techniques, Netherlands. Proceedings*, 1987, pp. 249–266.
13. J. Guo, J. Jean, I. Nikolic, K. Qiao, Y. Sasaki, and S. M. Sim, "Invariant Subspace Attack Against Full Midori64," Cryptology ePrint Archive, Report 2015/1189, 2015.
14. L. R. Knudsen and V. Rijmen, "Known-Key Distinguishers for Some Block Ciphers," in *Advances in Cryptology – ASIACRYPT 2007: 13th International Conference on the Theory and Application of Cryptology and Information Security, Malaysia, 2007. Proceedings*, 2007, pp. 315–324.
15. G. Leander, M. A. Abdelraheem, H. AlKhzaimi, and E. Zenner, "A Cryptanalysis of PRINTcipher: The Invariant Subspace Attack," in *Advances in Cryptology – CRYPTO 2011: 31st Annual Cryptology Conference, Santa Barbara, CA, USA, 2011. Proceedings*, 2011, pp. 206–221.

16. G. Leander, B. Minaud, and S. Rønjom, "A Generic Approach to Invariant Subspace Attacks: Cryptanalysis of Robin, iSCREAM and Zorro," in *Advances in Cryptology - EUROCRYPT 2015: 34th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Bulgaria. Proceedings, Part I*, 2015, pp. 254–283.
17. J. Lu, O. Dunkelman, N. Keller, and J. Kim, *Progress in Cryptology - INDOCRYPT 2008: 9th International Conference on Cryptology in India, India. Proceedings*, 2008, ch. New Impossible Differential Attacks on AES, pp. 279–293.
18. H. Mala, M. Dakhilalian, V. Rijmen, and M. Modarres-Hashemi, "Improved impossible differential cryptanalysis of 7-round AES-128," in *Progress in Cryptology - INDOCRYPT 2010: 11th International Conference on Cryptology in India, India. Proceedings*, 2010, pp. 282–291.
19. B. Sun, M. Liu, J. Guo, L. Qu, and V. Rijmen, "New Insights on AES-like SPN Ciphers," Cryptology ePrint Archive, Report 2016/533, 2016.
20. B. Sun, Z. Liu, V. Rijmen, R. Li, L. Cheng, Q. Wang, H. Alkhzaimi, and C. Li, "Links among Impossible Differential, Integral and Zero Correlation Linear Cryptanalysis," in *Advances in Cryptology – CRYPTO 2015: 35th Annual Cryptology Conference, Santa Barbara, CA, USA. Proceedings*, 2015, pp. 95–115.
21. T. Tiessen, "Polytopic Cryptanalysis," in *Advances in Cryptology - EUROCRYPT 2016 - 35th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Austria. Proceedings, Part I*, 2016, pp. 214–239.
22. Y. Todo, "Integral cryptanalysis on full MISTY1," in *Advances in Cryptology - CRYPTO 2015: 35th Annual Cryptology Conference, Santa Barbara, CA, USA. Proceedings, Part I*, 2015, pp. 413–432.
23. ——, "Structural evaluation by generalized integral property," in *Advances in Cryptology - EUROCRYPT 2015: 34th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Bulgaria. Proceedings, Part I*, 2015, pp. 287–314.