Extended Expectation Cryptanalysis on Round-reduced AES

Zhenzhen Bao[1,2], Jian Guo[1] and Eik List[3(✉)]

[1] Division of Mathematical Sciences, School of Physical and Mathematical Sciences,
Nanyang Technological University, Singapore
{zzbao,guojian}(at)ntu.edu.sg
[2] Strategic Centre for Research in Privacy-Preserving Technologies and Systems,
Nanyang Technological University, Singapore
[3] Bauhaus-Universität Weimar, Weimar, Germany
<firstname>.<lastname>(at)uni-weimar.de

Abstract. Distinguishers on round-reduced AES have attracted considerable attention in the recent years. Although the number of rounds covered in key-recovery attacks has not been increased since, subspace, yoyo, and multiple-of-n cryptanalysis advanced the understanding of properties of the cipher.

Expectation cryptanalysis is an umbrella term for all forms of statistical analysis that try to identify properties whose expectation differs from that of an ideal primitive. For substitution-permutation networks, integral attacks seem a suitable target for extension since they usually end after a linear layer sums several subcomponents. Based on results by Patarin, Chen et al. already observed that the expected number of collisions differs slightly for a sum of permutations from the ideal. Though, their target remained lightweight primitives.

The present work applies expectation-based distinguishers from a sum of PRPs to round-reduced AES. We show how to extend the well-known 3-round integral distinguisher to expectation distinguishers over 4 and 5 rounds. In contrast to previous expectation distinguishers by Grassi et al., our approach allows to prepend a round that starts from a diagonal subspace. We demonstrate how the prepended round can be used for key recovery. Moreover, we show how the prepended round can be integrated to form a six-round distinguisher. For all distinguishers, our results are supported by their implementations with Cid et al.'s established Small-AES version.

Keywords: Cryptanalysis · block cipher · AES

1 Introduction

During the previous two decades, the Advanced Encryption Standard (AES) [Nat01] has withstood vast amounts of cryptanalysis. Besides the biclique-based accelerated exhaustive search by Bogdanov et al. [BKR11][1], the best known attacks in the secret-key model cover seven rounds of the AES-128, as had been the state close after its announcement [FKL+00]. However, the community's efforts led to attacks with considerably reduced resources. Among the best attacks in number of rounds [MDRM10, FKL+00], the Demirci-Selçuk-based meet-in-the-middle attacks by Derbez et al. possess the lowest time and data complexities for more than half a decade [DFJ13].

[1] Instead of exploiting dedicated properties of a given cipher, biclique-based attacks represent rather a general approach to speed up brute force since their outer loop iterates over all possible keys, cf. [DDKS13].

1.1 Distinguishers on Round-reduced AES

Although the number of rounds covered in key-recovery attacks did not increase since, the recent years were filled with research on the AES that significantly raised the understanding of the cipher's components. This direction appears promising – metaphorically, it is comparable to heuristics that sometimes also have to leave a local optimum to improve upon existing results in the long run. Until a few years ago, the best key-independent distinguishers on the AES had covered at most four rounds. Those distinguishers, be they differential, impossible differential, zero-correlation, or integrals, serve as base for the construction of longer key-recovery attacks. It is well-known [SLR+15] that there exists a dual relationship between the existence of impossible-differential, zero-correlation, and integral distinguishers. This means, the existence of an impossible differential on a cipher $\mathcal{E}$ implies the existence of affine layers $\mathcal{A}_1$ and $\mathcal{A}_2$ such that there is an integral and a zero-correlation distinguisher on $\mathcal{A}_2 \circ \mathcal{E} \circ \mathcal{A}_1$. Thus, results on one kind of distinguisher are also applicable to others.

Negative Results paved a rocky start for the search for new distinguishers. Sun et al. [SLG+16b] proved the absence of impossible differentials over more than four rounds for the AES structure, which was tantamount with the absence of integrals or zero-correlation attacks over more than four rounds. Interestingly, [SLG+16b] targeted a generalized structure instead of a concrete cipher; since it ignored the details of the AES S-box and its key schedule, there remained a spark of hope for longer distinguishers of those kinds. The work by Wang and Jin [WJ18] extinguished this spark. Under the Markov-cipher assumption, they showed the absence of any truncated impossible differentials over more than four-round of the AES even if the details of the S-box are taken into account. For the AES-256 in particular, they showed the absence even if the key schedule is also considered.

Key-dependent Distinguishers. Despite the negative results, an active series of works has been focusing on novel properties for distinguishers on less rounds than the best known attacks. First, a number of key-dependent distinguishers were crafted, e.g., the chosen-ciphertext zero-correlation hull on five rounds by Sun et al. [SLG+16a] exploited a known difference in two key bytes to produce a five-round distinguisher. While their distinguisher required the full codebook when converted to the single-key model, it reignited the community's efforts on analyzing round-reduced AES.
Subsequent works improved on their result and proposed further key-dependent distinguishers. In [GRR16], Grassi et al. reconsidered the distinguisher by Sun et al. and derived key-dependent chosen-plaintext distinguisher with lower complexity. Their more relevant contribution was their formalism: Differential, linear, and integral distinguishers already implicitly exploited that texts or tuples thereof had a larger probability to lie in cosets of certain subspaces, which were more than often enough formulated in a complicated manner. Grassi et al. unified them in their notion of subspace trails. In particular for the AES, subspace trails allow to describe diagonal, column, inverse-diagonal, and mixed (results of MixColumns) subspaces quite elegantly.
Grassi proposed further key-dependent distinguishers in [Gra18a]; starting from the distinguisher by Sun et al., Hu et al. [HCGW18] improved the complexity and transformed it into a chosen-plaintext attack, and derived impossible-differential attacks from it. Cui et al. [CCM+18] later reduced the complexity. Those considered attacks on the AES with a secret S-box [TKKL15]. Among those attacks, the yoyo distinguisher on five rounds by Bardeh and Rønjom [BR19] possesses the lowest complexity, with less than 2^{30} chosen plaintexts and 2^{32} adaptively chosen ciphertexts.

Key-independent Distinguishers. Besides the key-dependent results, several powerful key-independent distinguishers have been proposed recently. In a series of works [GRR16, GRR17, GR18, Gra18b], Grassi et al. proposed several novel observations and distinguishers on five-round AES. At their core, [GRR17] proposed a strong property dubbed multiple-of-n: starting from a structure of a diagonal space, the number of different ciphertext pairs that belonged to the same coset of a mixed space was always a multiple of eight after five rounds.

Boura et al. [BCC19] revisited the multiple-of-n property and derived similar distinguishers for further AES-like primitives. They traced the property back to a combination of (1) an equivalence relation between the differences after one round that holds for n pairs at a time, and (2) probability-one subspace trails that wrap the relation. Moreover, they showed that the property is independent from the details of the MixColumns matrix.

Mixture-differential Distinguishers. In [Gra18b], Grassi considered mixture-differential differentials. At their core, they exploited the following property: let (P, P') denote a pair of plaintexts that differ in two bytes (x, y) and (x', y') in the same column, and whose corresponding ciphertexts (C, C') have a difference in a certain (mixed) subspace after four rounds. Then, the difference of the ciphertext pair (C'', C''') that corresponds to the plaintexts (P'', P''') with byte values (x, y') and (x', y) (i.e., which "mixes" the original pair of plaintexts), respectively, will also lie in a coset of the same subspace. This pair of plaintext-ciphertext pairs is called a couple. In [Gra18b], Grassi proposed this and its generic variant (whose plaintexts differ in four bytes in one column) as efficient four-round distinguishers and a five-round key-recovery attack. Soon upon, Bar-On et al. [BODK+18] improved the key-recovery to the attack on five-round AES with the lowest data and time complexity known so far.

In [Gra17], the corresponding full version [Gra18b], Grassi further proposed probabilistic, threshold, as well as impossible mixture-differential distinguishers. The former distinguisher exploited a tiny difference between two expectations: the expected number of sets with at least one couple whose both ciphertexts belong to the same coset of a subspace is a little lower for five-round AES than for a random permutation. Grassi's threshold distinguisher exploited different expectations between sets of couples. Each set was formed by couples whose both plaintexts mixed the byte values in two diagonals, and differed only in those two diagonals. This distinguisher exploited a complex fact: Grassi counted the expected number of sets whose number of couples where both ciphertexts belong to the same coset of a subspace is higher than a threshold. This number is higher for five-round AES than it is for a random permutation. Finally, he considered an impossible mixture-differential distinguisher. This attack exploited that it is impossible for five-round AES, for specifically defined sets, they all have at least one couple whose both ciphertexts belong to the same coset of subspace (i.e., for five-round AES, at least one set does not have the corresponding property). Again, this attack considered sets formed by couples whose plaintexts mixed the byte values of the first two diagonals.

The Best Previously Published Distinguishers. To the best of our knowledge, the best previously published distinguishers on round-reduced AES-128 in terms of minimal complexity up to now are the yoyo-based proposals by Rønjom et al. [RBH17]. They proposed a five-round distinguisher with minimal time complexity among the known results, and a result on six rounds, at the cost of $2^{122.3}$ time and data requirements. Recently, Bardeh and Rønjom et al. proposed a similar key-independent attack, also in [BR19]. Those share the need of adaptive chosen ciphertexts. Table 1 provides a summary of existing distinguishers on five or more rounds of AES-128; note that many results hold for diverse versions of the AES; the focus of this work resides on the 128-bit variant.

Table 1: Existing secret-key distinguishers on 5+ rounds of the AES-128, ordered by rounds (descending) then time (ascending). MAs = memory accesses; CP = chosen plaintexts; (A)CC = (adaptive) chosen ciphertexts; MD = mixture differential; TD = truncated differential.

Attack Type	Time		Data		Ref.
5 Rounds					
Integral	2^{128}	XORs	2^{128}	CC	[SLG+16a]
Impossible Differential	2^{107}	MAs	$2^{98.2}$	CP	[GRR16]
Threshold MD	$2^{98.1}$	MAs	2^{89}	CP	[Gra17]
Impossible MD	$2^{97.8}$	MAs	2^{82}	CP	[Gra17]
Probabilistic MD	$2^{71.5}$	MAs	2^{52}	CP	[Gra17]
Expectation of TD	$\mathbf{2^{70.2}}$	**MAs**	$\mathbf{2^{65}}$	**CP**	**Sect. 4.2**
Expectation of TD	2^{51}	MAs	$2^{47.4}$	CP	[GR18]
Variance of TD	$2^{41.6}$	MAs	2^{38}	CP	[GR18]
Multiple-of-8	$2^{35.6}$	MAs	2^{32}	CP	[GRR17]
Yoyo	$2^{26.2}$	XORs	$2^{27.2}$	ACC	[BR19]
Yoyo	$2^{25.8}$	XORs	$2^{26.8}$	ACC	[RBH17]
6 Rounds					
Impossible Yoyo	$2^{121.83}$	XORs	$2^{122.83}$	ACC	[RBH17]
Expectation of TD	$\mathbf{2^{96.52}}$	**MAs**	$\mathbf{2^{89.43}}$	**CP**	**Sect. 6**

1.2 From Integral To Expectation Distinguishers

Integral Distinguishers map multi-sets of inputs that iterate over all values to multi-sets of outputs that are balanced. Traditionally, the properties of bits or bytes are — in order of their strength — either constant ($\mathcal{C}$), iterate over all values ($\mathcal{A}$), are balanced ($\mathcal{B}$), or unknown ($\mathcal{U}$). A traditional integral distinguisher usually ends directly before all parts of the state become unknown. For the AES, the three-round integral distinguisher [DKR97] that maps sets of a single active byte to a set of states where each byte is balanced after three rounds is well-understood, and so is the extension to a distinguisher on four rounds [DKR97, FKL+00] that prepends a round and starts from an active diagonal.

Probabilistic Integrals. Wang et al. [WCC+16] proposed so-called statistical integrals. Assuming that integral structures map structures of 2^s inputs that iterate over all values to b pairwise disjoint sets of t output bits each that are uniformly distributed. Statistical integrals exploit that the results are hypergeometrically distributed in contrast to a close-to-normally distribution in random permutations. Therefore, they do not reconstruct the full integral of 2^s texts, but can reduce the data requirements to $2^{(s-t)/2}$. Cui et al. [CSCW17, CCM+18] transferred the approach to the AES and exploited multiple integral structures. If N_s structures are necessary, their approach reduced the data complexity further to $O(\sqrt{N_s/n} \cdot 2^{(s-t)/2})$, where n is the state size. Though, this approach does not directly aim at extending distinguishers, but represents a data-reduction technique instead. Moreover, the attacks by Cui et al. were in the secret-s-box setting. For Skipjack [WCC+16], it nevertheless allowed to extend previous attacks by reducing the number of texts to trace through the cipher.

Expectation and Standard-deviation Cryptanalysis aim to distinguish distributions from the means and standard deviations of certain properties. In [Pat08, Pat13], Patarin studied the number of pairwise collisions for the sums of multiple permutations in comparison with those from random functions. Follow-up works by Nachef, Marrière or Patarin, and Volte [NMV16, NPV18, NPV14, VNM16a, VNM16b] also employed the standard deviation as distinguishing property. Though, their focus resided mainly on Feistel networks.

Extending Integrals To Expectation Distinguishers. The core observation at the beginning of this work was that an integral distinguisher usually ends with a linear operation. In many SPNs, the linear layer often consists of a sum of multiple words. At the end of an integral, such a sum is equivalent to the sum of values that iterate over all values in the subspace—hence, a sum of permutations. The sum still has a Balanced (i.e., zero-sum) property, which is usually destroyed by the subsequent non-linear layer. As illustrated in Patarin's works, the number of collisions induced by a sum of permutations differs slightly from that of an ideal function. The collisions due to the linear sum will be preserved by the subsequent non-linear S-box operation. Therefore, an integral distinguisher can be extended through the subsequent non-linear operation.
We point out that Chen et al. [CMSZ15] had already considered this approach of Patarin's analysis [Pat08, Pat13] for extending integrals of SPNs. Chen et al. considered Type-II and Nyberg-type Feistel networks and conducted experiments on lightweight ciphers for which they could confirm that this strategy can lead to extended attacks. This work, however, focuses on the AES.

Previous Expectation Distinguishers on the AES. In [GR18], Grassi and Rechberger also considered truncated differentials and exploited smaller statistical differences. The core results are five-round distinguishers that exploit the following property of the AES: A structure of 2^{32} plaintexts that differ only in a single diagonal leads to a mixed space after two rounds with probability one; this trail can be used twice and connected with a probabilistic trail over the middle. Grassi and Rechberger observed that the expectation of this probabilistic event is slightly higher for five-round AES than for a random permutation. Furthermore, they show how to exploit the considerably different variances. Their results are similar to what we aim to in this work. Though, we start and view from a different path (from integrals), whereas their work did not depend on Patarin's result. In contrast, the expectation distinguishers in [Gra17] considered the expectation of couples, whereas we will consider pairs.

Contribution. This work tries to extend the known integral distinguisher to expectation distinguishers. As a result, it describes a five-round distinguisher from a single byte to a mixed space. Since inputs start from single-byte differences, plaintext structures can form less pairs than in e.g., the structures from diagonals as in [GR18]. As a consequence, the data and computational complexity of the five-round distinguisher here is higher than the probabilistic distinguishers in [GR18]. However, our proposal allows a straightforward extension to a six-round key-recovery attack by prepending a round. We report on the results of a practical implementation of the five-round distinguisher and the six-round key-recovery attacks with a small-scale variant of the AES. Finally, we propose a possible extension to a six-round expectation distinguisher and report on our results of its implementation for the small variant. [2]

Outline. The remainder is structured as follows. Next, Section 2 will revisit the necessary preliminaries, as well as the known results from Patarin on sums of independent permutations and subspaces of the AES by Grassi et al. [GRR16]. Thereupon, Section 4 will develop our five-round distinguisher and report on a verification. Section 5 describes a key-recovery attack on six rounds. Section 6 derives a six-round distinguisher and again reports on a verification.

[2] Our implementations can be found freely available at
`https://github.com/medsec/expectation-cryptanalysis-on-round-reduced-aes`.

Table 2: Expected numbers of collisions after q queries for the sums of k permutations and distinguishing complexity for $q \simeq 2^n$ from [Pat08].

k	2	3	4
$\mathbb{E}\left[N_k\right]$	$\frac{g\binom{q}{2}}{2^n}\left(1+\frac{1}{2^n-1}\right)$	$\frac{g\binom{q}{2}}{2^n}\left(1-\frac{1}{(2^n-1)^2}\right)$	$\frac{g\binom{q}{2}}{2^n}\left(1+\frac{1}{(2^n-1)^3}\right)$
Complexity	$O(2^{2n})$	$O(2^{4n})$	$O(2^{6n})$

2 Preliminaries

General Notations. We denote by $\mathbb{F}_2$ the finite field of characteristic two. We represent functions and variables by upper case letters and indices by lowercase letters, sets by calligraphic letters. We employ typewriter font for hexadecimal values. Let $X, Y \in \mathbb{F}_2^n$ for some positive integer n in the following. Then, we denote by $X \parallel Y$ the concatenation of X and Y, by $X \oplus Y$ their bitwise XOR. For all $X \in \mathbb{F}_2^n$, we index the bits $X = (X_{n-1} \ldots X_1 X_0)$ where X_{n-1} is the most significant and X_0 the least significant bit of X. For integers $x \geq y$, we write $X_{x..y}$ as short form of $(X_x X_{x-1} \ldots X_y)$.
We denote by $\mathbb{E}\left[X\right]$ the expectation of a random variable X and by σ_X its standard deviation. We denote by μ and σ^2 the mean and the variance of a distribution.
We use the binomial distribution $\mathcal{B}(n, p)$, which yields the number of successes in the a sequence of n independent Boolean experiments, each of which is successful with probability p. The values μ and variance σ^2 are given by $\mu = n \cdot p$ and $\sigma^2 = n \cdot p \cdot (1-p)$.
We approximate the binomial distribution with a normal distribution and can consider the difference again as a normal distribution. Let $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ follow a normal distribution with mean μ_1 and variance σ_1^2, let $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$. Then, $X_1 - X_2 \sim \mathcal{N}(\mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2)$.

Functions and Permutations. For sets $\mathcal{X}$ and $\mathcal{Y}$, we write $\mathsf{Func}(\mathcal{X}, \mathcal{Y}) =^{\text{def}} \{F | F : \mathcal{X} \to \mathcal{Y}\}$ as the set of all functions with domain $\mathcal{X}$ and range $\mathcal{Y}$. Let further $\mathsf{Perm}(\mathcal{X})$ be the set of all permutations over $\mathcal{X}$. We define as short forms $\mathsf{Func}_n =^{\text{def}} \mathsf{Func}(\mathbb{F}_2^n, \mathbb{F}_2^n)$ and Perm_n define the set of permutations over $\mathbb{F}_2^n$. Clearly, it holds that $|\mathsf{Func}_n| = (2^n)^{2^n}$ and $|\mathsf{Perm}_n| = (2^n)!$. We call π an ideal permutation (over $\mathbb{F}_2^n$) if $\pi \twoheadleftarrow \mathsf{Perm}_n$, i.e., if it is sampled uniformly at random from Perm_n. Similarly, we call ρ an ideal function (over $\mathbb{F}_2^n$) if $\rho \twoheadleftarrow \mathsf{Func}_n$. For integers $m \leq n$ and arbitrary $X \in \mathbb{F}_2^n$, we define $\mathsf{trunc}_m(X) =^{\text{def}} \mathsf{msb}_m(X) =^{\text{def}} X_{(n-1)..(n-1-m)}$ to truncate the input x and return only the most-significant (i.e., leftmost) m bits of x.

2.1 Distinguishers for Sums of Permutations

In the following, we recall briefly the results by [Pat08]. Given a function set $\mathcal{F}$, we define by $\mathsf{Gen}(\mathcal{F})$ a function generator that gives access to multiple pairwise independent instances from $\mathcal{F}$. Let $\pi_1, \ldots, \pi_k \twoheadleftarrow \mathsf{Perm}_n$ be independent ideal random permutations, and let $\rho \twoheadleftarrow \mathsf{Func}_n$. We define a k-sum of permutations as $\Sigma_k[\pi_1, \ldots, \pi_k](x) =^{\text{def}} \bigoplus_{i=1}^{k} \pi_i(x)$ and write Σ_k as short form.
The goal of a k-sum-distinguisher $\mathbf{A}$ is to distinguish ρ from Σ_k: we write $\Delta_{\mathbf{A}}(\Sigma_k; \rho)$ for its advantage. W.l.o.g., we assume that $\mathbf{A}$ is deterministic, information-theoretic, and does not ask queries to which it already knows the answer. The advantages considered here may be irrelevant if $\mathbf{A}$ has access to only a single instance of ρ or Σ_k since it usually exceeds the codebook. Instead, Patarin considered already multiple such independent functions in the generator $\mathsf{Gen}(\mathcal{F})$. We denote by g the number of available functions from the generator, and by q the number of queries x_i that $\mathbf{A}$ can ask to each of the functions. The queries by $\mathbf{A}$ are collected together with the responses of the oracle in a transcript $\tau = \{(x_i^j, y_i^j)\}_{1 \leq i \leq q, 1 \leq j \leq g}$.

Let N be a random variable for the number of collisions between outputs y_i, i.e., $N = |\{i, j \in [q] : i \neq j \wedge y_i = y_j\}|$. In general, for functions F, we define random variables N_F for the number of collisions of F after q queries. So, let N_ρ be the number of collisions for ρ. Since all responses y_i are sampled independently uniformly at random, it holds that [Pat08]

$$\mathbb{E}\left[N_\rho\right] = \frac{g\binom{q}{2}}{2^n} \quad \text{and} \quad \sigma(N_\rho) = O\left(\frac{\sqrt{g}q}{\sqrt{2^n}}\right) .$$

Let N_k denote a random variable for the number of collisions of Σ_k (N_k is a short form of N_{Σ_k}). Scenario 2 in [Pat08] gives

$$\mathbb{E}\left[N_k\right] = \frac{g\binom{q}{2}}{2^n} \cdot \left(1 + \frac{(-1)^k}{(2^n - 1)^{k-1}}\right) \quad \text{and} \quad \sigma(N_k) = O\left(\frac{\sqrt{g}q}{\sqrt{2^n}}\right) .$$

Patarin argues with the Tchebichev theorem $\Pr\left[|X - \mu|\right] \geq c\sigma] \leq 1/c^2$, that the distinguishing advantage between the collision distributions of two functions F and G becomes non-negligible if

$$\sigma(N_F) \ll \left|\mathbb{E}\left[N_F\right] - \mathbb{E}\left[N_G\right]\right| \quad \text{and} \quad \sigma(N_G) \ll \left|\mathbb{E}\left[N_F\right] - \mathbb{E}\left[N_G\right]\right| ,$$

hold. For example, the sum of $k = 2$ permutations has

$$\mathbb{E}\left[N_2\right] = \frac{g\binom{q}{2}}{2^n}\left(1 + \frac{1}{2^n - 1}\right) .$$

Thus, if

$$\frac{\sqrt{g}q}{\sqrt{2^n}} \ll \frac{g\binom{q}{2}}{2 \cdot 2^{2n}} ,$$

both distributions can be distinguished with non-negligible advantage. So, for $q \simeq 2^n$, the adversary can distinguish both settings in $g \geq 2^n$, i.e., the adversary needs $O(2^{2n})$ queries. For k permutations, Patarin showed that the advantage is non-negligible when

$$\frac{\sqrt{g}q}{\sqrt{2^n}} \ll \frac{gq^2}{2^{kn}}.$$

For $q \simeq 2^n$, this yields that $g \geq 2^{(2k-3)n}$ functions are necessary. Table 2 illustrates the expected number of collisions and distinguishing efforts for $k \in \{2, 3, 4\}$.

2.2 The AES-128 and Subspaces

Brief Definition of The AES-128. We assume, the reader is familiar with the details of the AES and provide only a very brief summary here. Details can be found in, e.g., [DR02, Nat01]. The AES-128 is a substitution-permutation network that transforms 128-bit inputs through ten rounds, consisting of SubBytes (SB), ShiftRows (SR), MixColumns (MC), and a round-key addition with a round key K^i. Before the first round, a whitening key K^0 is XORed to the state; the final round omits the MixColumns operation. We write S^i for the state after Round i, and $S^i[j]$ for the j-th byte, for $0 \leq i \leq 10$ and $0 \leq j \leq 15$. Though, we will interchangeably also use the indices for a 4×4-byte matrix, i.e., $0, 0$ for Byte 0, and $3, 3$ for Byte 15. So, the byte ordering is given by either

$$\begin{bmatrix} 0 & 4 & 8 & 12 \\ 1 & 5 & 9 & 13 \\ 2 & 6 & 10 & 14 \\ 3 & 7 & 11 & 15 \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} 0,0 & 0,1 & 0,2 & 0,3 \\ 1,0 & 1,1 & 1,2 & 1,3 \\ 2,0 & 2,1 & 2,2 & 2,3 \\ 3,0 & 3,1 & 3,2 & 3,3 \end{bmatrix} .$$

Table 3: The 4-bit S-box of small-scale AES from [CMR05].

X	0	1	2	3	4	5	6	7	8	9	a	b	c	d	e	f
$S(X)$	6	B	5	4	2	E	7	A	9	D	F	C	3	1	0	8

When using two-dimensional indices, we will assume that all indices are taken modulo four, which will make our life easier in sums. $\mathsf{R}[K^i] =^{\text{def}} \mathsf{AK}[K^i] \circ \mathsf{MC} \circ \mathsf{SR} \circ \mathsf{SB}$ denotes one application of the round function and denote by $S^{r,\mathsf{SB}}$, $S^{r,\mathsf{SR}}$, and $S^{r,\mathsf{MC}}$ the states in the r-th round directly after the application of SubBytes, ShiftRows, and MixColumns, respectively. Moreover, we will use $\mathbf{M}$ to denote the MixColumns matrix.

Subspaces of The AES. We adopt the notation of subspaces for the AES from Grassi et al. [GRR16]. Let $\mathcal{W}$ denote a vector space and $\mathcal{V} \subseteq \mathcal{W}$ be a subspace. If a is an element of $\mathcal{W}$, then a coset $\mathcal{V} \oplus a$ of $\mathcal{V}$ in $\mathcal{W}$ is a subset $\mathcal{V} \oplus a = \{v \oplus a | \forall v \in \mathcal{V}\}$. We consider vectors and vector spaces over $\mathbb{F}_{2^8}^{4\times4}$, and denote by $\{e_{0,0}, \ldots, e_{3,3}\}$ the unit vectors of $\mathbb{F}_{2^8}^{4\times4}$, i.e., $e_{i,j}$ has a single 1 in the i-th row and j-th column. For a vector space $\mathcal{V}$ and a function $F : \mathbb{F}_{2^8}^{4\times4} \to \mathbb{F}_{2^8}^{4\times4}$, we let $F(\mathcal{V}) =^{\text{def}} \{F(v) | v \in \mathcal{V}\}$. For a subset $\mathcal{I} \subseteq \{1, 2, \ldots, n\}$ and a subset of vector spaces $\{\mathcal{V}_1, \mathcal{V}_2, \ldots, \mathcal{V}_n\}$, we define $\mathcal{V}_\mathcal{I} =^{\text{def}} \bigoplus_{i\in\mathcal{I}} \mathcal{V}_i$. We adopt the definitions by Grassi et al. of four families of subspaces for the AES for each $i \in \{0, 1, 2, 3\}$:

- the column spaces $\mathcal{C}_i = \langle e_{0,i}, e_{1,i}, e_{2,i}, e_{3,i} \rangle$,
- the diagonal spaces $\mathcal{D}_i = \mathsf{SR}^{-1}(\mathcal{C}_i)$,
- the inverse-diagonal spaces $\mathcal{ID}_i = \mathsf{SR}(\mathcal{C}_i)$, and
- the mixed spaces $\mathcal{M}_i = \mathsf{MC}(\mathcal{ID}_i)$.

For $\mathcal{I} \subseteq \{0, 1, 2, 3\}$, the spaces $\mathcal{C}_\mathcal{I}$, $\mathcal{D}_\mathcal{I}$, $\mathcal{ID}_\mathcal{I}$, and $\mathcal{M}_\mathcal{I}$ are defined as

$$\mathcal{C}_\mathcal{I} \stackrel{\text{def}}{=} \bigoplus_{i\in\mathcal{I}} \mathcal{C}_i\,, \qquad \mathcal{D}_\mathcal{I} \stackrel{\text{def}}{=} \bigoplus_{i\in\mathcal{I}} \mathcal{D}_i\,, \qquad \mathcal{ID}_\mathcal{I} \stackrel{\text{def}}{=} \bigoplus_{i\in\mathcal{I}} \mathcal{ID}_i\,, \quad \text{and} \quad \mathcal{M}_\mathcal{I} \stackrel{\text{def}}{=} \bigoplus_{i\in\mathcal{I}} \mathcal{M}_i\,.$$

Small-AES. Cid et al. [CMR05] proposed small-scale variants of the AES to help cryptanalysts study attacks whose complexity were impractical on the full-fledged cipher. We employ the four-bit variant with a 4 × 4-nibble matrix in the following.
Small-AES differs from the AES only in the following aspects:

- It operates on a 16-nibble state of 64 bits, i.e., states and keys are elements of $\mathbb{F}_{2^4}^{4\times4}$.
- The S-box operates on nibbles; it is given in Table 3 for completeness.
- The MixColumns multiplications operate in $\mathbb{F}_{2^4}$ modulo $p(\mathtt{x}) = \mathtt{x}^4 + \mathtt{x} + \mathtt{1}$. The values in the MixColumns matrix are equal to those of $\mathbf{M}$ in the original AES.
- The round constants to derive the round key K^i are $\mathtt{x}^{i-1}$ in $\mathbb{F}_{2^4}/p(\mathtt{x})$.

3 Integral Properties

Three-round Integral Distinguisher. We briefly recap the well-known three-round integral distinguisher for the AES [DKR97]. Let $\mathcal{I}, \mathcal{I}', \mathcal{J}, \mathcal{J}' \subseteq \{0, 1, 2, 3\}$. Let X^i denote the i-th element in $\mathcal{S}$ and define $S^{r,i} = R^r(X^i)$ denote the encryption of X^i through r consecutive rounds of AES. Let $r, c \in \{0, 1, 2, 3\}$. We denote by $\mathcal{S} = a \oplus \left(\mathcal{D}_{\{r\}} \cap \mathcal{C}_{\{c\}}\right)$ a δ-set, that is a set of 2^8 texts that iterate over all values $x_{c-r \bmod 4,c}$ (the byte at Row $(c-r) \bmod 4$ and Column c is indicated as $\mathbf{A}$) and are constant at all other bytes ($\mathbf{C}$). Then, the texts in $\mathcal{S}$ iterate over all 2^8 values in each byte of S^2 after two rounds of AES.

$$\begin{bmatrix} \mathbf{A} & \mathbf{C} & \mathbf{C} & \mathbf{C} \\ \mathbf{C} & \mathbf{C} & \mathbf{C} & \mathbf{C} \\ \mathbf{C} & \mathbf{C} & \mathbf{C} & \mathbf{C} \\ \mathbf{C} & \mathbf{C} & \mathbf{C} & \mathbf{C} \end{bmatrix} \xrightarrow{R^2} \begin{bmatrix} \mathbf{A} & \mathbf{A} & \mathbf{A} & \mathbf{A} \\ \mathbf{A} & \mathbf{A} & \mathbf{A} & \mathbf{A} \\ \mathbf{A} & \mathbf{A} & \mathbf{A} & \mathbf{A} \\ \mathbf{A} & \mathbf{A} & \mathbf{A} & \mathbf{A} \end{bmatrix} \xrightarrow{\mathsf{SR} \circ \mathsf{SB}} \begin{bmatrix} \mathbf{A} & \mathbf{A} & \mathbf{A} & \mathbf{A} \\ \mathbf{A} & \mathbf{A} & \mathbf{A} & \mathbf{A} \\ \mathbf{A} & \mathbf{A} & \mathbf{A} & \mathbf{A} \\ \mathbf{A} & \mathbf{A} & \mathbf{A} & \mathbf{A} \end{bmatrix} \xrightarrow{\mathsf{MC}} \begin{bmatrix} \mathbf{B} & \mathbf{B} & \mathbf{B} & \mathbf{B} \\ \mathbf{B} & \mathbf{B} & \mathbf{B} & \mathbf{B} \\ \mathbf{B} & \mathbf{B} & \mathbf{B} & \mathbf{B} \\ \mathbf{B} & \mathbf{B} & \mathbf{B} & \mathbf{B} \end{bmatrix}.$$

This property is preserved through SubBytes and ShiftRows, but is not guaranteed by the MixColumns operation at the end of Round 3. Since MixColumns is linear, it preserves balanced input sets, i.e., the sum of all 2^8 states $\bigoplus_{i=1}^{2^8} S^{3,i} = 0$. This is indicated by $\mathbf{B}$. The subsequent SubBytes operation in Round 4 destroys this Balanced property.

Three-round Integral Distinguisher. Though, there exists a well-known further property.

Property 1. For all $i, j \in \{1, \ldots, 2^8\}$, $i \neq j$, there is no all-zero column in $S^{3,i} \oplus S^{3,j}$. Hence, there is no all-zero anti-diagonal in the difference $\mathsf{SR}(\mathsf{SB}(S^{3,i})) \oplus \mathsf{SR}(\mathsf{SB}(S^{3,j}))$. Let $\mathcal{I} \subset \{0, 1, 2, 3\}$ with $|\mathcal{I}| < 4$. Then, $\mathcal{D}_i \cap \mathcal{C}_j$ for any $i, j \in \{0, 1, 2, 3\}$ maps to $\mathcal{M}_{j-1}$ after two rounds with probability one, and cannot map to $\mathcal{M}_{\mathcal{I}}$ after two further rounds:

$$\mathcal{D}_i \cap \mathcal{C}_j \xrightarrow{R^2} \mathcal{M}_{j-i} \not\xrightarrow{R^2} \mathcal{M}_{\mathcal{I}}.$$

Four-round Integral Distinguisher. The integral distinguisher is well-known to be extendable to four rounds [DKR97]. Let $\mathcal{D}_{\mathcal{I}}$ for $|\mathcal{I}| = 1$ be a diagonal space. Then, the texts in $\mathcal{D}_{\mathcal{I}}$ iterate over all 2^{32} values in each column of S^3 after three rounds of AES.

$$\begin{bmatrix} \mathbf{A} & \mathbf{C} & \mathbf{C} & \mathbf{C} \\ \mathbf{C} & \mathbf{A} & \mathbf{C} & \mathbf{C} \\ \mathbf{C} & \mathbf{C} & \mathbf{A} & \mathbf{C} \\ \mathbf{C} & \mathbf{C} & \mathbf{C} & \mathbf{A} \end{bmatrix} \xrightarrow{R^3} \begin{bmatrix} \mathbf{A} & \mathbf{A} & \mathbf{A} & \mathbf{A} \\ \mathbf{A} & \mathbf{A} & \mathbf{A} & \mathbf{A} \\ \mathbf{A} & \mathbf{A} & \mathbf{A} & \mathbf{A} \\ \mathbf{A} & \mathbf{A} & \mathbf{A} & \mathbf{A} \end{bmatrix} \xrightarrow{\mathsf{SR} \circ \mathsf{SB}} \begin{bmatrix} \mathbf{A} & \mathbf{A} & \mathbf{A} & \mathbf{A} \\ \mathbf{A} & \mathbf{A} & \mathbf{A} & \mathbf{A} \\ \mathbf{A} & \mathbf{A} & \mathbf{A} & \mathbf{A} \\ \mathbf{A} & \mathbf{A} & \mathbf{A} & \mathbf{A} \end{bmatrix} \xrightarrow{\mathsf{MC}} \begin{bmatrix} \mathbf{B} & \mathbf{B} & \mathbf{B} & \mathbf{B} \\ \mathbf{B} & \mathbf{B} & \mathbf{B} & \mathbf{B} \\ \mathbf{B} & \mathbf{B} & \mathbf{B} & \mathbf{B} \\ \mathbf{B} & \mathbf{B} & \mathbf{B} & \mathbf{B} \end{bmatrix}.$$

This property is preserved through SubBytes and ShiftRows, but is not guaranteed by the MixColumns operation at the end of Round 3. Since MixColumns is linear, it preserves the Balanced property, i.e., the sum of all 2^{32} states $\bigoplus_{i=1}^{2^{32}} S^{4,i} = 0$. The subsequent SubBytes operation in Round 5 destroys the Balanced property.

4 Distinguishers

4.1 Four-round Expectation Distinguisher

We can extend the deterministic three-round integral distinguisher to a probabilistic four-round expectation distinguisher in the following. Prior, we articulate an easy observation: Let $\mathcal{I} \subseteq \{0, 1, 2, 3\}$. If there exists an all-zero anti-diagonal $\mathcal{ID}_{\mathcal{I}}$ in the difference of two texts after r rounds without the final MixColumns operation, then $\mathcal{C}_{\mathcal{I}}$ after $r-1$ rounds must have been an all-zero column, and $\mathcal{D}_{\mathcal{I}}$ after $r-2$ rounds must have been an all-zero diagonal:

$$\Pr\left[\mathcal{ID}_{\mathcal{I}} \xrightarrow{\mathsf{SR}^{-1} \circ \mathsf{SB}^{-1} \circ \mathsf{AK}^{-1}} \mathcal{C}_{\mathcal{I}} \xrightarrow{R^{-1}} \mathcal{D}_{\mathcal{I}} \right] = 1.$$

Let $R^r_{i,j} : \mathbb{F}_{2^8}^{4 \times 4} \to \mathbb{F}_{2^8}$ denote the mapping of a 16-byte state through r consecutive rounds of AES and output only the Byte at Index (i, j). Let $\widetilde{R}^r_{i,j}$ denote the similar mapping through almost r rounds, without the final key addition. Moreover, let $\widehat{R}^r_{i,j}$ denote the similar mapping through almost r rounds, without the final key addition and $\mathsf{MixColumns}$ operation. We use the matrix indexing analogously as before.

Theorem 1. Then, for an input $X^r \in \mathbb{F}_{2^8}^{4\times 4}$, we can rewrite $\widehat{R}^3_{i,j}(X^r)$ as

$$\widehat{R}^3_{i,j}(X^r) \overset{\text{def}}{=} \pi_{i,j}(\widetilde{R}^2_{i,j}(X^r)),$$

where the $\pi_{i,j} =^{\text{def}} \mathsf{SR} \circ \mathsf{SB} \circ \mathsf{AK}[K^{r+2}_{i,j}]$ and $\pi_{i,j} \in \mathsf{Perm}(\mathbb{F}_{2^8})$ are independent, for all $\{i, j\} \in \{0, 1, 2, 3\}$, and $\widetilde{R}^2_{i,j}$ preserves the ALL property.

Proof. The proof is very brief. Let SBox denote the AES S-box and $\mathbf{M}$ the $\mathsf{MixColumns}$ matrix. After one round, each byte is the sum of four transformed input bytes:

$$X^{r+1}_{i,j} = K^{r+1}_{i,j} \oplus \bigoplus_{k=0}^{3} \left(\mathbf{M}_{i,k} \cdot \text{SBox}\left(X^r_{j+k,k}\right)\right).$$

After two rounds, each byte is the sum of all transformed input bytes:

$$\begin{aligned} X^{r+2}_{i,j} &= K^{r+2}_{i,j} \oplus \bigoplus_{k=0}^{3} \left(\mathbf{M}_{i,k} \cdot \text{SBox}\left(X^{r+1}_{j+k,k}\right)\right) \\ &= K^{r+2}_{i,j} \oplus \bigoplus_{k=0}^{3} \left(\mathbf{M}_{i,k} \cdot \text{SBox}\left(K^{r+1}_{j+k,k} \oplus \bigoplus_{\ell=0}^{3} \left(\mathbf{M}_{j+k,\ell} \cdot \text{SBox}\left(X^r_{k+\ell,\ell}\right)\right)\right)\right). \end{aligned}$$

So, each input byte index (i, j) occurs exactly once in the equation for each output byte index (i', j'). For a δ-set wherein one input byte $X^r_{i,j}$ iterates over all values, this means that R^2 without the final key addition yields the state where each byte iterates over all values. Under the (simplifying) assumption that the round keys are independent and uniformly random, the remaining operations $\pi_{i,j} =^{\text{def}} \mathsf{SR}\circ\mathsf{SB}\circ\mathsf{AK}[K^{r+2}_{i,j}]$ form an independent permutation each, preserving the ALL property for all output bytes of $\widehat{R}^3$. □

For a δ-set, all bytes iterate over all values after $\widehat{R}^3$. So, for each column, the $\mathsf{MixColumns}$ operation in Round 3 can be viewed as the sum of the results of four independent permutations where the inputs iterate over all values. Hence, we approximate the expected probability that a fixed byte of interest $S^{3,i}_{r,c} = S^{3,j}_{r,c}$ collides after $\mathsf{MixColumns}$ by $\mathbb{E}[N_4]$:

$$\Pr_{\text{AES}}\left[S^{3,i}_{r,c} = S^{3,j}_{r,c}\right] \simeq \frac{1}{2^8} + \frac{1}{2^8(2^8-1)^3} \simeq 2^{-8} + 2^{-31.983}.$$

For a random permutation, the probability can be approximated by

$$\Pr_{\mathsf{rand}}\left[S^{3,i}_{r,c} = S^{3,j}_{r,c}\right] = \frac{2^{120}-1}{2^{128}-1} \simeq 2^{-8}.$$

The difference between those two probabilities thus can be exploited to build an expectation distinguisher on four-round AES, because $S^{3,i}_{r,c} = S^{3,j}_{r,c}$ directly implies a collision between the corresponding bytes in $S^{4,\mathsf{SB}}$, which can be computed from the ciphertext by inverting the final $\mathsf{ShiftRows}$ operation. The final round-key addition can be converted into an addition of an equivalent key, such that it does not influence detecting the collision.

Statistical Framework. The results by Patarin yield a good intuition for the number of necessary queries of distinguishers. To obtain more precise success probabilities, we further consider the updated framework from [Gra18b], which is close to that used by

Chen et al. [CMSZ15]. We consider two distributions, where we approximate that their difference is normally distributed with $\mathcal{N}(\mu, \sigma^2)$, with

$$\mu = n \cdot |p_{\mathsf{AES}} - p_{\mathsf{rand}}| \quad \text{and} \quad \sigma^2 = n \cdot (p_{\mathsf{rand}} \cdot (1 - p_{\mathsf{rand}}) + p_{\mathsf{AES}} \cdot (1 - p_{\mathsf{AES}}))\,.$$

Since the probability density of the normal distribution is

$$F\left(x|\mu, \sigma^2\right) \stackrel{\text{def}}{=} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \cdot \frac{1}{\sigma\sqrt{2\pi}},$$

it follows that

$$p = \int_0^{+\infty} \frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}}\, \mathrm{d}x = \int_{-\frac{\mu}{\sigma}}^{+\infty} \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}\, \mathrm{d}x = \frac{1}{2}\left(1 + \mathsf{erf}\left(\frac{-\mu}{\sigma\sqrt{2}}\right)\right),$$

where $\mathsf{erf}(x)$ is the error function, i.e., the probability that a normally distributed random variable $X \sim \mathcal{N}(0, 0.5)$ falls into the interval $[-x, x]$. To obtain a success probability of at least p, the number of experiments n has to satisfy

$$n \geq \frac{2\left(p_{\mathsf{AES}}(1 - p_{\mathsf{AES}}) + p_{\mathsf{rand}}(1 - p_{\mathsf{rand}})\right)}{(p_{\mathsf{AES}} - p_{\mathsf{rand}})^2} \cdot \left(\mathsf{erfinv}(2 \cdot p - 1)^2\right), \tag{1}$$

where $\mathsf{erfinv}(x)$ is the inverse error function. Since $p_{\mathsf{AES}}, p_{\mathsf{rand}} \ll 1$, a good approximation of n is given by

$$n \geq \frac{4 \max(p_{\mathsf{AES}}, p_{\mathsf{rand}})}{(p_{\mathsf{AES}} - p_{\mathsf{rand}})^2} \cdot \left(\mathsf{erfinv}(2 \cdot p - 1)^2\right). \tag{2}$$

Complexity. The complexity of our four-round expectation distinguisher is approximated by Equation (2) (plugging $\Pr_{\text{AES}}\left[S^{3,i}_{r,c} = S^{3,j}_{r,c}\right]$ into p_{AES} and $\Pr_{\mathsf{rand}}\left[S^{3,i}_{r,c} = S^{3,j}_{r,c}\right]$ into p_{rand}). For a success probability of $p = 0.95$, we obtain $n > 2^{58.402}$ pairs. In total, a δ-set contains $\binom{2^8}{2} \simeq 2^{15}$ pairs. So, we need about $2^{43.402}$ δ-sets with $2^{51.402}$ chosen plaintexts.

Reduced Variant. In order to allow a practical verification, we derive the corresponding probabilities for four rounds of Small-AES with four-bit S-boxes. We start again by viewing the downscaled variant of MixColumns as a sum of four independent permutations on $\mathbb{F}_{2^4}$. Then, the probability that a nibble after three rounds leads to a zero difference is approximately

$$\Pr_{\text{Small-AES}}\left[S^{3,i}_{r,c} = S^{3,j}_{r,c}\right] \simeq \frac{1}{2^4} + \frac{1}{2^4(2^4-1)^3} \simeq 2^{-4} + 2^{-15.721}.$$

For a random permutation, the probability can be approximated by

$$\Pr_{\mathsf{rand}}\left[S^{3,i}_{r,c} = S^{3,j}_{r,c}\right] = \frac{2^{60}-1}{2^{64}-1} \simeq 2^{-4} - 2^{-64.093}.$$

Applying Equation (2) yields $n > 2^{29.878}$ for $p = 0.95$. Since a δ-set yields $\binom{2^4}{2} = 120$ pairs, we need about 2^{23} δ-sets that consist of 2^{27} chosen plaintexts.

Experimental Verification. We verified the distinguisher experimentally with 100 random keys per experiment and 2^s δ-sets, for $s \in \{20, \ldots, 23\}$, on Small-AES. In each experiment, we evaluated the number of collisions of the first nibble of the output. As approximation of a random permutation, we employed full-round Speck-64-96 with 100 random keys. Our results are listed in Table 4. The values μ denote the obtained means of the number of pairs that collide in at least on inactive inverse diagonal, over all experiments, e.g. for 2^{20} δ-sets, one could expect $2^{20} \cdot \binom{2^4}{2} \cdot (2^{-4} + 2^{-15.721}) \approx 7\,866\,650$ colliding pairs per experiment.

Table 4: Means and standard deviations for the number of pairs that collided in at least one inverse diagonal for our four-round expectation distinguisher with Small-AES. Each experiment employed 100 random independent keys and 2^s random δ-sets. Experimental values are rounded. Full-round Speck-64-96 was used as pseudorandom permutation π.

	Theory		Experiments			
#δ-sets	Small-AES	π	Small-AES		π	
($\log_2$)	μ	μ	μ	σ	μ	σ
20	7 866 650	7 863 200	7 870 789.	2 918.	7 864 396.	2 566.
21	15 733 300	15 728 600	15 742 188.	3 809.	15 728 650.	3 957.
22	31 466 600	31 457 300	31 484 544.	6 007.	31 457 205.	5 096.
23	62 933 200	62 914 600	62 967 244.	7 030.	62 915 004.	7 820.

4.2 Five-round Expectation Distinguisher

We can extend the four-round distinguisher to five rounds. Consider some diagonal space $\mathcal{D}_{\{c\}}$ for some index $c \in \{0, 1, 2, 3\}$. Then, the expected probability that all four bytes in that diagonal space collide for two texts in a δ-set can be approximated by:

$$\Pr_{\text{AES}}\left[S^3 \in \mathcal{D}_{\{c\}}\right] \simeq \left(2^{-8} + \frac{1}{2^8(2^8-1)^3}\right)^4 \simeq \left(2^{-8} + 2^{-31.983}\right)^4,$$

Naively, we can approximate the same probability for a random permutation by

$$\Pr_{\mathsf{rand}}\left[S^3 \in \mathcal{D}_{\{c\}}\right] = \frac{2^{96}-1}{2^{128}-1} \simeq 2^{-32} - 2^{-128}.$$

This approximation is naive since the texts are not independent. In [GR18, Appendix C], Grassi and Rechberger provide arguments for the number of collisions for 2^{32} input texts that can be combined to pairs with each other to a mixed space $\mathcal{M}_{\mathcal{I}}$ for $|\mathcal{I}| = 3$. They consider the dependency between multiple texts. Still, they show that the collision probability can be well approximated by 2^{-32}.

Say, we observe the ciphertexts after five rounds that correspond to a δ-set of 2^8 plaintexts. Again, the final MixColumns operation can be simply inverted to obtain the differences before it. Then, the probability to have at least one all-zero anti-diagonal in the difference is given for five-round AES by

$$p_{\mathsf{AES}} \simeq 1 - \left(1 - \Pr_{\text{AES}}\left[S^3 \in \mathcal{D}_{\{c\}}\right]\right)^4 \simeq 2^{-30} + 2^{-51.985},$$

whereas for a random permutation, it is approximately

$$p_{\mathsf{rand}} \simeq 1 - \left(1 - \Pr_{\mathsf{rand}}\left[S^3 \in \mathcal{D}_{\{c\}}\right]\right)^4 \simeq 2^{-30} - 2^{-61.415}.$$

Steps. The steps of our distinguisher are as follows:

1. Initialize a collision counter.
2. For $i = 1..2^s$, collect a structure $\mathcal{S}$ of 2^{64} texts that iterate over all values in any eight bytes and leave the remaining bytes constants. Query the plaintexts of a structure and ask for their ciphertexts. Invert the final ShiftRows operations to get the states $S^{5,\mathsf{SB}}$ and store them in some list $\mathcal{Q}$.

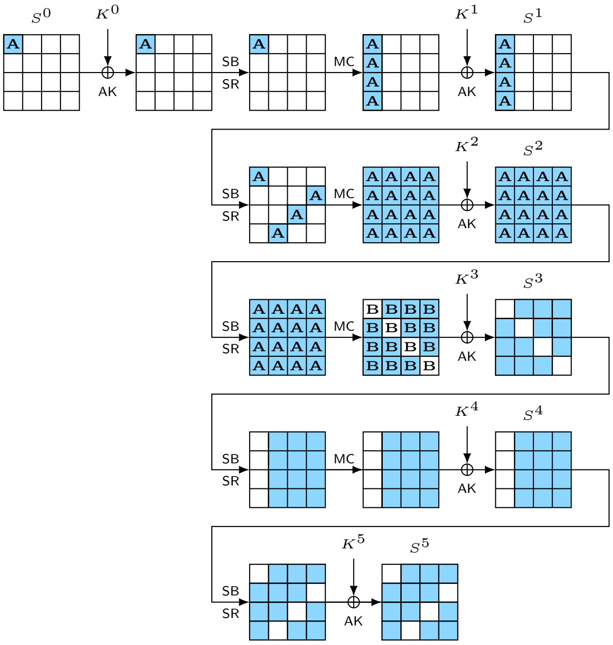


Figure 1: Five-round distinguisher. Darkened cells represent bytes with active (non-zero) difference, white cells represent bytes with zero difference.

3. Form $8 \cdot 2^{56}$ δ-sets from a structure, i.e., the texts in each structure iterate over the 2^8 values in one byte and are constant in the 15 remaining bytes. For each δ-set:
   3.1 Initialize four lists $\mathcal{L}_i$, for $i = 0, 1, 2, 3$ of 2^{32} elements.
   3.2 For each column i of $S^{5,\mathsf{SB}}$, interpret the column as 32-bit integer and append the text to $\mathcal{L}_i$ the index corresponding to the column value for each list.
4. For each of the lists $\mathcal{L}_i$:
   4.1 Look for collisions, e.g., multiple values at the same index.
   4.2 For each collision, look those pairs up in other columns if they have already been counted. If not, increment the counter.
5. If the counter is higher than the threshold θ, output real; otherwise, output random.

An approximation for θ can be

$$2 \cdot 8 \cdot 2^{56} \cdot \binom{2^8}{2} \cdot \frac{p_{\mathsf{AES}} + p_{\mathsf{rand}}}{2} \simeq 35\,046\,937\,350\,825.148\,.$$

Complexity. For a success probability of approximately $p = 0.95$, we obtain $n > 2^{74.436}$. By fixing any set of eight distinct bytes in the plaintexts, we can form 2^{64} texts by iterating over all values of those eight bytes and leaving the remaining eight bytes constant. For each byte, the texts can be partitioned into 2^{56} sets. Each set consists of $\binom{2^8}{2}$ pairs of two texts that differ in a single byte only. In total, the plaintext structure consists of $8 \cdot 2^{56}$ such sets for each byte of interest. So, one structure provides

$$8 \cdot 2^{56} \cdot \binom{2^8}{2} \simeq 2^{73.994} \simeq 2^{74} \text{ pairs.}$$

Two structures from 2^{65} chosen plaintexts with approximately 2^{75} pairs suffice for a success probability of more than 0.95 on average. The memory complexity is given by storing 2^{64} states in $\mathcal{Q}$ and four lists $\mathcal{L}_i$ of $4 \cdot 2^{32}$ columns at a time. The time complexity consists of

- 2^{65} encryptions and 1/5 partial decryptions or $2^{65.3}$ encryption equivalents.
- $2^{65} + 2^{65} \cdot 4 \simeq 2^{67.4}$ memory accesses for inserting the texts. Depending on the data structure, insertions or lookups need sorting, which requires $N \log_2(N)$ operations on average with e.g., quicksort. Given four lists of 2^8 elements, sorting needs approximately $2 \cdot 2^{56} \cdot 4 \cdot 2^8 \cdot 8 \simeq 2^{70}$ memory accesses.
- At most $2 \cdot 2^{56} \cdot 8 \cdot \binom{2^8}{2} \cdot 2^{-32} \cdot 3 \simeq 2^{44.6}$ additional memory accesses when collisions occur to look up if other columns collide in the other lists.
- Approximately $2 \cdot 2^{56} \cdot 8 \cdot \binom{2^8}{2} \cdot 2^{-32} \simeq 2^{43}$ memory accesses to increment the counter.

So, the computational complexity can be approximated by $2^{67.4} + 2^{70} + 2^{44.6} + 2^{43} \simeq 2^{70.2}$ memory accesses and $2^{65.3}$ encryptions.

Reduced Variant. In order to allow a practical verification, we derive the corresponding probabilities for a downscaled variant of five-round AES with four-bit S-boxes.
Again, we view the downscaled variant of MixColumns as a sum of four independent permutations on $\mathbb{F}_{2^4}$. Then, the probability that a nibble after three rounds leads to a zero difference is approximately

$$\Pr_{\text{Small-AES}}\left[S^{3,i}_{r,c} = S^{3,j}_{r,c}\right] \simeq \frac{1}{2^4} + \frac{1}{2^4(2^4-1)^3} \simeq 2^{-4} + 2^{-15.721}.$$

For a random permutation, the probability can be approximated by

$$\Pr_{\mathsf{rand}}\left[S^{3,i}_{r,c} = S^{3,j}_{r,c}\right] = \frac{2^{60}-1}{2^{64}-1} \simeq 2^{-4} - 2^{-64.093}.$$

The expected probability that all four bytes in a diagonal space $\mathcal{D}_{\{c\}}$ collide for two texts in a δ-set can be approximated by $\mathbb{E}\left[N_4\right]^4$:

$$\Pr_{\text{Small-AES}}\left[S^3 \in \mathcal{D}_{\{c\}}\right] = \left(2^{-4} + \frac{1}{2^4(2^4-1)^3}\right)^4 \simeq \left(2^{-4} + 2^{-15.721}\right)^4,$$

and that for a random permutation by

$$\Pr_{\mathsf{rand}}\left[S^3 \in \mathcal{D}_{\{c\}}\right] = \frac{2^{48}-1}{2^{64}-1} \simeq 2^{-16} - 2^{-64}.$$

Then, the probability to have at least one all-zero anti-diagonal in the difference is given for five-round Small-AES by

$$p_{\mathsf{SmallAES}} \simeq 1 - \left(1 - \Pr_{\text{Small-AES}}\left[S^3 \in \mathcal{D}_{\{c\}}\right]\right)^4 \simeq 2^{-14} + 2^{-23.748},$$

whereas for a random permutation, it is approximately

$$p_{\mathsf{rand}} \simeq 1 - \left(1 - \Pr_{\mathsf{rand}}\left[S^3 \in \mathcal{D}_{\{c\}}\right]\right)^4 \simeq 2^{-14} - 2^{-29.415}.$$

Feeding those figures into Equation (2) yields, for a success probability of approximately $p = \{0.95, 0.99\}$, that the number of pairs has to exceed $n > \{2^{35.878}, 2^{36.878}\}$. A structure of $(2^4)^8 = 2^{32}$ texts that fixes eight cells can form $8 \cdot 2^{4 \cdot 7}$ sets of $\binom{2^4}{2}$ pairs each, which corresponds to

$$8 \cdot 2^{28} \cdot \binom{2^4}{2} \simeq 2^{37.907}$$

pairs. So, one structure may suffice to have a distinguishing advantage of ≥ 0.99.

Table 5: Means and standard deviations for the number of pairs that collided in at least one inverse diagonal for our five-round expectation distinguisher with Small-AES. Each experiment employed 100 random independent keys and 2^s random δ-sets. Experimental values are rounded. Full-round Speck-64-96 was used as pseudorandom permutation π.

	Theory		Experiments			
$\#\delta$-sets	Small-AES	π	Small-AES		π	
($\log_2$)	μ	μ	μ	σ	μ	σ
23	61 512	61 439	61 518.80	237.14	61 416.79	246.80
24	123 023	122 877	123 042.21	345.99	122 833.97	329.42
25	246 046	245 754	246 039.13	485.50	245 778.46	543.67
26	492 092	491 509	492 213.04	644.16	491 464.80	724.31

Experimental Verification. Again, we tried to verify our claims experimentally. Table 5 depicts our results with 100 random keys and 2^s random δ-sets of 2^4 texts each, for $s \in \{23, \ldots, 26\}$. All values μ denote the means for the number of pairs in δ-sets that collided in at least one inverse diagonal after five rounds, over all δ sets, e.g., $\mu = \binom{2^4}{2} \cdot 2^{23} \cdot p_{\mathsf{SmallAES}}$ for Small-AES for 2^{23} δ-sets.

5 Six-round Key-recovery Attack

From the five-round distinguisher from Section 4.2, we can mount a key-recovery attack on six rounds of AES that recovers 32 key bits. We apply the attack twice in shifted form to recover 64 key bits and the remaining key bits exhaustively.

Precomputation. We construct four hash tables $\mathcal{H}_i : \mathbb{F}_{2^8}^4 \to (\mathbb{F}_{2^8}^4)^*$, for $0 \le i \le 3$. For a column difference $\Delta X = (\Delta X_0, \Delta X_1, \Delta X_2, \Delta X_3)$, $\mathcal{H}_i(\Delta X)$ contains a set of all four-byte pairs $(X, X') = ((X_0, X_1, X_2, X_3), (X'_0, X'_1, X'_2, X'_3))$ to SubBytes such that $\Delta S^1 = \mathsf{MC}(\mathsf{SB}(X)) \oplus \mathsf{MC}(\mathsf{SB}(X \oplus \Delta X))$ is active in exactly the i-th byte. For each $\mathcal{H}_i$, there exist 255 differences ΔS^1. Since the AES S-box and its inverse map an input difference to 127 output differences of 126 pairs and one 4-tuple, each mapping $\Delta X \xleftarrow{\mathsf{SB}^{-1} \circ \mathsf{MC}^{-1}} \Delta S^1$ contains one pair on average. So, each entry in $\mathcal{H}_i$ consists of 255 pairs on average. Since we have four tables with 255 entries each, they need $4 \cdot 255 \cdot 255 \cdot 8/16$ Bytes, which corresponds to approximately 2^{17} state equivalents.

Steps. The steps are as follows:

1. Initialize a zeroed list $\mathcal{K}$ for the 2^{32} key candidates for $K^0[0, 5, 10, 15]$.
2. Precompute the tables $\mathcal{H}_i$, for $0 \le i \le 3$.
3. For $i = 1..2^s$, collect a structure $\mathcal{S}$ of 2^{32} texts in a coset of $\mathcal{D}_{\{0\}}$.
   - 3.1 So, the texts in each structure iterate over the 2^{32} values in Bytes $P[0, 5, 10, 15]$ and are constant in the 12 remaining bytes.
   - 3.2 Initialize four lists $\mathcal{L}_i$, for $i = 0, 1, 2, 3$. Query the plaintexts of a structure to an encryption oracle to obtain the corresponding ciphertexts C. Invert the final MixColumns and ShiftRows operations to get the states $S^{6,\mathsf{SB}}$ and store the tuples of $(P[0, 5, 10, 15], S^{6,\mathsf{SB}})$ into the four lists $\mathcal{L}_i$, where the texts are indexed by the i-th column of $S^{6,\mathsf{SB}}$.

3.3 For $j = 0..3$, consider the lists $\mathcal{L}_j$:

   i. Look up collisions of $(S^{6,\mathsf{SB}}, S'^{6,\mathsf{SB}})$. For each collision, look up in the lists with lower indices j if the same pair collided already in a different anti-diagonal to prevent double counting.

   ii. For each collision, consider their corresponding plaintexts P and P' and derive their difference $\Delta P = P \oplus P'$. From $\mathcal{H}_0$, $\mathcal{H}_1$, $\mathcal{H}_2$, and $\mathcal{H}_3$, look up all possible states X that $P \oplus K^0[0, 5, 10, 15]$ could take such that $P \oplus P'$ is in a δ-set. For each value X, derive the candidate $K^0[0, 5, 10, 15] = P \oplus X$ and increment its counter in $\mathcal{K}$.

4. Sort the list of key candidates and output the sorted list.

5. Apply the attack another time from a shifted diagonal, e.g., $\mathcal{D}_{\{1\}}$, and with another 2^s structures to recover another 32 bits of key material.

6. Test the keys in descending order of their counters to recover the remaining 64 key bits.

Complexity. Each structure yields $4 \cdot \binom{2^8}{2} \cdot 2^{24}$ δ-sets, which corresponds to approximately 2^{41} pairs. We choose $2^s = 2^{33.5}$ structures, which corresponds to $2^{65.5}$ chosen plaintexts. The time complexity is given by the following:

- The precomputations cost $4 \cdot 255 \cdot 255 \cdot 8/16 \cdot 1/6 \simeq 2^{14.5}$ encryption equivalents.
- Encrypting the data takes $2^{65.5}$ encryptions and $1/6$ partial decryptions, or at most $2^{65.8}$ encryption equivalents.
- We need $2^{67.5}$ memory accesses for storing the $2^{65.5}$ texts into four lists. Sorting the four lists requires $2^{33.5} \cdot 4 \cdot 2^{32} \cdot 32 \simeq 2^{72.5}$ memory accesses.
- Each structure yields about 2^{63} pairs, and each pair has a probability of approximately 2^{-30} to collide on a column. So, we expect approximately $2^{s+63} \cdot 2^{-30} \simeq 2^{s+33} = 2^{66.5}$ collisions. For each collision that occurs in a table $\mathcal{L}_i$, we need at most three memory accesses for the three other tables in $\mathcal{L}_j$, $j \neq i$, to prevent double counting. Thus, we need $2^{66.5} \cdot 3 \simeq 2^{68.1}$ MAs at that point.
- For each collision, we expect on average $4 \cdot 255$ suggestions from the hash tables $\mathcal{H}_i$. So, we need about $2^{66.5} \cdot 4 \cdot 255$ memory accesses plus the same amount in $\mathcal{K}$, plus the same amount of XORs.

Since the attack is performed twice with the same plaintext material but on 32 different key bits, e.g., $K[3, 4, 9, 14]$, we need the same number of encryptions (alternatively, they could be stored) and lookups, plus 2^{64} encryptions for the key search. In total, the time complexity consists of

$$2^{14.5} + 2 \cdot 2^{65.8} + 2^{64} \simeq 2^{67} \quad \text{encs. and} \quad 2 \cdot (2^{67.5} + 2^{72.5} + 2^{68.1} + 2^{76.5}) \simeq 2^{77.6} \text{ MAs.}$$

The attack stores 2^{32} tuples of $4 + 16$ bytes at a time for the states plus 2^{32} byte counters candidates plus 2^{17} states for the hash tables $\mathcal{H}_i$. So, the memory complexity is circa $2^{32.4}$ states.

Experimental Verification. We verified our attack with the small-scale variant. The four-bit S-box had been constructed to have similar properties as its bigger brother in the AES. It is 4-differential-uniform; for each input difference α, there exist six output differences β with $\Pr[S(x) \oplus S(x \oplus \alpha) = \beta] = 2/16$, and one difference β for which $\Pr[S(x) \oplus S(x \oplus \alpha) = \beta] = 4/16$. Similar properties hold for the inverse S-box.

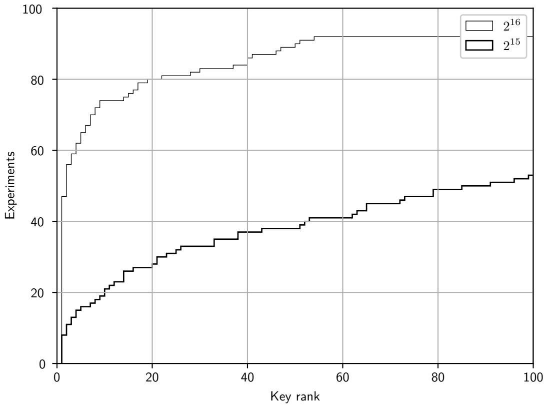


Figure 2: Rank distribution for the correct key among 2^{16} candidates from 100 runs of our six-round attack with Small-AES, with random keys and 2^{15} or 2^{16} structures of 2^{16} texts each.

So, each hash table $\mathcal{H}_i$, for $0 \leq i \leq 3$ contains on average $15 \cdot 2^{16}$ 16-bit entries for each input difference ΔP, or 2^{20} bytes, which corresponds to 2^{17} 64-bit states for Small-AES. We choose one structure of 2^{32} texts which yields $8 \cdot 2^{28}\binom{2^4}{2}$ δ-sets, which corresponds to approximately $2^{37.907}$ pairs. The time complexity is given by

- $4 \cdot 15 \cdot 15 \cdot 2^{16} \cdot 8/16 \cdot 1/6 \simeq 2^{22.3}$ encryptions for the hash tables.
- 2^{32} encryptions and 1/6 partial decryptions, or at most $2^{32.22}$ encryption equivalents.
- 2^{34} memory accesses for storing the texts and $4 \cdot 2^{16} \cdot 16 \cdot 2^{16} = 2^{38}$ memory accesses for sorting the lists.
- It yields about $8 \cdot \binom{2^4}{2} \cdot 2^{28} \cdot 2^{-14} \simeq 2^{23.91}$ collisions. Given that we need at most three accesses to the lists $\mathcal{L}$ per collision as in our attack on the original five-round AES, this corresponds to at most $2^{25.5}$ memory accesses.
- Per collision, we expect on average $4 \cdot 15$ suggestions from the hash tables. So, we require $2 \cdot (2^{23.91} \cdot 4 \cdot 15) \simeq 2^{30.82}$ XORs for deriving the keys in $\mathcal{K}$ and the same maximum number of MAs to increment the counters in $\mathcal{K}$.

We omit the second execution; the attack requires

$$2^{22.3} + 2^{32.22} \simeq 2^{32.23} \text{ encryptions and } 2^{34} + 2^{38} + 2^{25.5} + 2^{30.82} \simeq 2^{38.1} \text{ MAs}$$

to reduce the key space for $K^0[0, 5, 10, 15]$. The attack stores 2^{16} tuples of $4 + 16$ nibbles at a time for the states plus 2^{16} four-nibble key candidates, plus 2^{17} states for the hash tables. So, less than $2^{17.9}$ states.
Figure 2 illustrates the results of 100 experiments that employed independent random keys and 2^{15} and 2^{16} structures of 2^{16} plaintexts each. Our experiments aimed at recovering the 16 key bits $K^0[0, 5, 10, 15]$ from the first diagonal. In 53 runs, the correct key was among the top 100 keys, among which it was in the top 16 keys in 27 runs. This yields a conservative advantage of approximately one (1.089) bit over all runs, and an advantage of about ten (10.117) bits in more than half of the experiments. The results improved for 2^{16} structures, where the correct key ranked top for 47 experiments, and 92 times among the top 100 keys, and at worst at rank 313. Over the runs, this yielded an advantage of 10.55 bits over all experiments.

6 Six-round Expectation Distinguisher

We can derive a six-round distinguisher that uses the same higher-level concept as our six-round key-recovery attack.

Core Idea. In our previous five-round distinguisher, we exploited a non-negligible but very small difference between the expectation for a certain event of two distributions D_0 and D_1^5 of a random ideal permutation and a real construction on five rounds, respectively. In theory, we can go one step further.

Let us focus on a set of inputs $\mathcal{X}$, and on any event E with non-zero probability in both worlds. Let p_0 and p_1^5 denote two expected probabilities of a certain event in the respective distributions. We exploited that $|p_0 - p_1^5|$ was not too small. For simplicity, assume for the moment that the probability of an event for a given input is independent of previous inputs and the ideal distribution D_0 behaves equally on all inputs. So

$$\Pr_{X \twoheadleftarrow \mathcal{X}}[\mathsf{D}_0 = E] = p_0.$$

In the following, consider p_1^6 and D_1^6 representing the expected probability of a certain event and the distribution for six rounds, respectively. Assume that we sample over the set of inputs that contains our previous input set – allowed us to determine p_1^5– but that also contains many further inputs, without partitioning them into sets. Clearly, the distance between both distributions reduces considerably. More formally, the set of inputs $\mathcal{X}$ can be partitioned into two disjoint sets $\mathcal{X} = \mathcal{X}_a \cup \mathcal{X}_b$, i.e., $\mathcal{X}_a \cap \mathcal{X}_b = \emptyset$. Assume, the distribution of the real construction has an expectation of p_0 for E over the inputs from $\mathcal{X}_a$, but and expectation of q for E to occur over inputs from $\mathcal{X}_b$. So

$$\Pr_{X \twoheadleftarrow \mathcal{X}}[\mathsf{D}_1^6(X) = E] = p_1^6 = \frac{|\mathcal{X}_a| \cdot p_0 + |\mathcal{X}_b| \cdot q}{|\mathcal{X}|}.$$

Application to The AES. Compared to the five-round distinguisher from Grassi et al. [Gra18b], our approach from the previous section starts from a single active byte. Thus, it can benefit from the fact that one can easily prepend one round for an attack and start from a diagonal structure of 2^{32} texts. We know that a diagonal structure contains $4 \cdot 2^{24}$ δ-sets that had been used in the distinguisher. Clearly, the event E was to find at least one inactive inverse diagonal after five rounds. Since we aimed at finding the probability for such collisions for p_0 and p_{AES}^5. For this purpose, we had to guess the initial subkeys used in the initial diagonal.

The following distinguisher avoids to guess key bits, but considers instead the probability of collisions over all possible pairs of the diagonal. Clearly, among the $\binom{2^{32}}{2}$ pairs, only $4 \cdot 2^{24} \cdot \binom{2^8}{2}$ pairs have a probability of p_{AES}^5 to yield the inactive inverse diagonal after six rounds. Naively, one could assume that all other pairs would behave pseudorandomly, i.e., their probability for at least one inactive inverse diagonal is also p_0. Under those naive assumptions, we could approximate p_{AES}^6 by

$$\begin{aligned} &\simeq \frac{4 \cdot 2^{24} \cdot \binom{2^8}{2} \cdot \left(2^{-30} + 2^{-51.985}\right) + \left(\binom{2^{32}}{2} - 4 \cdot 2^{24} \cdot \binom{2^8}{2}\right) \cdot \left(2^{-30} - 2^{-61.415}\right)}{\binom{2^{32}}{2}} \\ &\simeq 2^{-30} - 2^{-61.415} + 2^{-73.995}. \end{aligned} \tag{3}$$

This would imply that $|p_{\mathsf{rand}} - p_{\mathsf{AES}}^6| \simeq 2^{-73.995}$. This statistical difference is very small, but may still be high enough to distinguish between the distributions.

Equation (2) yields, again for a success probability of 0.95, that one would need approximately $n > 2^{120.43}$ pairs. Assuming that a diagonal structure of 2^{32} texts contains $\binom{2^{32}}{2}$

pairs, this amount of pairs can be obtained by querying $2^{57.43}$ structures or $2^{89.43}$ chosen plaintexts.

Steps. The steps are as follows:

1. Initialize a collision counter.
2. For $i = 1..2^s$, collect a structure $\mathcal{S}$ of 2^{32} texts that iterate over all values in $\mathcal{D}_{\{0\}}$ and leave the remaining bytes constant. Query the plaintexts of a structure and ask for their corresponding ciphertexts after six rounds. Invert the final MixColumns and ShiftRows operations to get the states $S^{6,\mathsf{SB}}$ and store them in a list $\mathcal{Q}$.
3. For each structure and each state $S^{6,\mathsf{SB}}$ therein:
   - 3.1 Initialize four lists $\mathcal{L}_i$, for $i = 0, 1, 2, 3$ of 2^{32} elements.
   - 3.2 For each column $i \in \{0, 1, 2, 3\}$ of $S^{6,\mathsf{SB}}$, interpret the i-th column as 32-bit integer $\mathbf{c}$ and append the text to $\mathcal{L}_i[\mathbf{c}]$ at the index corresponding to the column value $\mathbf{c}$, e.g., $\mathbf{c} = S^{6,\mathsf{SB}}[0, 1, 2, 3]$ for Column 0.
4. For each list $\mathcal{L}_i$:
   - 4.1 Look for collisions, e.g., multiple values at the same index.
   - 4.2 For each collision, look up those pairs in other lists $\mathcal{L}_j$, for $i \neq j$, if they have already been counted. Otherwise, increment the corresponding key counter.
5. If the counter exceeds a given threshold θ, output real; otherwise, output random.

An approximation for θ can be $2^{57.43} \cdot \binom{2^{32}}{2} \cdot (p^6_{\mathsf{AES}} + p_{\mathsf{rand}})/2$.

Complexity. The attack employs $2^{57.43}$ structures or $2^{89.43}$ chosen plaintexts for a success probability of about 0.95. The memory complexity is given by storing 2^{32} states in $\mathcal{Q}$ and four lists $\mathcal{L}_i$ of $4 \cdot 2^{32}$ columns at a time. The time complexity consists of

- $2^{89.43}$ encryptions and 1/6 partial decryptions or $2^{89.7}$ encryption equivalents.
- $2^{89.43} \cdot 4 \simeq 2^{91.43}$ memory accesses for inserting the texts. Sorting requires approximately $2^{57.43} \cdot 4 \cdot 32 \cdot 2^{32} \simeq 2^{96.43}$ memory accesses.
- If a collision occurs when inserting for a column, at most $2 \cdot 2^{57.43} \cdot \binom{2^{32}}{2} \cdot 2^{-32} \cdot 3 \simeq 2^{91.02}$ additional memory accesses are needed to look up if other columns collide in the other lists.
- Approximately $2 \cdot 2^{57.43} \cdot \binom{2^{32}}{2} \cdot 2^{-32} \simeq 2^{89.43}$ memory accesses to increment the counter.

So, the computational complexity is given by approximately $2^{89.7}$ encryptions and $2^{91.43} + 2^{96.43} + 2^{91.02} + 2^{89.43} \simeq 2^{96.52}$ memory accesses.

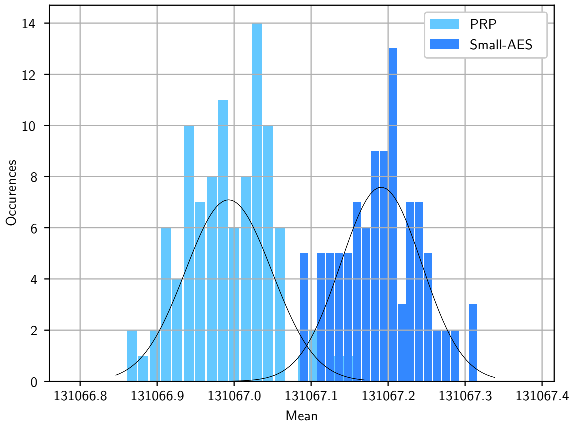
Small Variant. For the small-scaled variant of AES with four-bit S-boxes, we obtain from Equation (3)

$$p^6_{\mathsf{SmallAES}} \simeq 2^{-14} - 2^{-29.415} + 2^{-33.869},$$

which yields a mean of $p^6_{\mathsf{SmallAES}} \cdot \binom{2^{16}}{2} \simeq 131\,067.137$ colliding pairs per structure. For a random permutation, the probability of a pair to collide is approximately

$$p_{\mathsf{rand}} \simeq 2^{-14} - 2^{-29.415},$$

which gives a mean of $p_{\mathsf{rand}} \cdot \binom{2^{16}}{2} \simeq 131\,067.000$ colliding pairs per structure. So, a similar distinguisher on the small-scale version of the AES would need $2^{56.18}$ experiments, which corresponds to $2^{25.18}$ structures of 2^{16} texts each, or $2^{41.18}$ chosen plaintexts.



Small-AES	π
Theoretic	
131 067.137	131 067.000
Experimental	
131 067.191	131 066.993

Figure 3: Distribution of the means from 100 experiments that counted the number of collisions in at least one ciphertext column per structure of 2^{16} texts of our 6-round distinguisher on Small-AES and full Speck-64 as pseudorandom permutation.

Experimental Verification. We implemented the distinguisher for Small-AES. For each experiment, we encrypted $37 \cdot 2^{10} \simeq 2^{25.21}$ structures of 2^{16} texts each. For Small-AES, we used five full rounds plus SubBytes and AddRoundKeys in the last round since MixColumns and ShiftRows are easily invertible. As a pseudorandom permutation π, we employed full Speck-64-96. For each primitive, the 100 experiments required about three CPU years of computations.
As random independent keys, we employed the highest eight bytes from the first 100 output values from the NIST random Beacon service[3]. For each structure, we counted the number of pairs that collided in at least one column. For Small-AES, we observed a mean of $\mu = 131\,067.191$ pairs per structure, and $\mu = 131\,066.993$ for our pseudorandom permutation. Figure 3 illustrates the results of our experiments. Both illustrate that the difference in the distributions is even slightly higher in our experiments than in the theory. Most importantly, they show that both distributions can be distinguished well.

7 Conclusion

This work extends the well-known integral distinguisher on three-round AES to expectation distinguishers on four and five rounds. At the core, our attacks exploit the small bias in the number of byte collisions between the sum of four permutations—which MixColumns approximates after three rounds—and the number of byte collisions of a truncated random permutation. Thus, we could extend the integral attack to four rounds. By extending this approach to collisions in the four bytes of an inverse diagonal, we proposed a novel five-round distinguisher for the AES. Since those results start from a single byte, they could be extended easily by prepending a key-recovery round. Last but not least, we showed that even this prepended round could still be included into a secret-key distinguisher, with considerably lower but still distinguishable bias. We note that our distinguishers are infeasible to verify directly for the AES. Therefore, we implemented them with Cid et al.'s defacto standard of Small-AES with four-bit cells.

[3] See `https://beacon.nist.gov/beacon/2.0/chain/1/pulse/<i>` for $1 \leq i \leq 100$.

Acknoledgments. The stay by Eik List at Nanyang Technical University Singapore was kindly made possible by the DAAD IPID4all programme through the Bauhaus Research School at Weimar. Parts of the research leading to these results was made possible by DFG Grant LU 608/9-1. Thanks go to Johannes Kiesel and Michael Völske for their kind support for working on the Webis betaweb cluster. This research is supported by the National Research Foundation, Prime Minister's Office, Singapore under its Strategic Capability Research Centres Funding Initiative, Nanyang Technological University under research grant M4082123, and Singapore's Ministry of Education under grant M4012049.

References

[BCC19] Christina Boura, Anne Canteaut, and Daniel Coggia. A General Proof Framework for Recent AES Distinguishers. *IACR Trans. Symmetric Cryptol.*, 2019(1):170–191, 2019.

[BKR11] Andrey Bogdanov, Dmitry Khovratovich, and Christian Rechberger. Biclique Cryptanalysis of the Full AES. In Dong Hoon Lee and Xiaoyun Wang, editors, *ASIACRYPT*, volume 7073 of *LNCS*, pages 344–371. Springer, 2011.

[BODK$^+$18] Achiya Bar-On, Orr Dunkelman, Nathan Keller, Eyal Ronen, and Adi Shamir. Improved Key Recovery Attacks on Reduced-Round AES with Practical Data and Memory Complexities. In Hovav Shacham and Alexandra Boldyreva, editors, *CRYPTO II*, volume 10992 of *LNCS*, pages 185–212. Springer, 2018.

[BR19] Navid Ghaedi Bardeh and Sondre Rønjom. Practical Attacks on Reduced-Round AES. pages 1–13, 2019. To appear in the proceedings of Africacrypt 2019.

[CCM$^+$18] Tingting Cui, Huaifeng Chen, Sihem Mesnager, Ling Sun, and Meiqin Wang. Statistical integral distinguisher with multi-structure and its application on AES-like ciphers. *Cryptography and Communications*, 10(5):755–776, 2018.

[CMR05] Carlos Cid, Sean Murphy, and Matthew J. B. Robshaw. Small Scale Variants of the AES. In Henri Gilbert and Helena Handschuh, editors, *FSE*, volume 3557 of *LNCS*, pages 145–162. Springer, 2005.

[CMSZ15] Jiageng Chen, Atsuko Miyaji, Chunhua Su, and Liang Zhao. A New Statistical Approach for Integral Attack. In Meikang Qiu, Shouhuai Xu, Moti Yung, and Haibo Zhang, editors, *NSS*, volume 9408 of *LNCS*, pages 345–356. Springer, 2015.

[CSCW17] Tingting Cui, Ling Sun, Huaifeng Chen, and Meiqin Wang. Statistical Integral Distinguisher with Multi-structure and Its Application on AES. In Josef Pieprzyk and Suriadi Suriadi, editors, *ACISP I*, volume 10342 of *LNCS*, pages 402–420. Springer, 2017.

[DDKS13] Itai Dinur, Orr Dunkelman, Nathan Keller, and Adi Shamir. Key Recovery Attacks on 3-round Even-Mansour, 8-step LED-128, and Full AES^2. In Kazue Sako and Palash Sarkar, editors, *ASIACRYPT I*, volume 8269 of *LNCS*, pages 337–356. Springer, 2013.

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

[DKR97] Joan Daemen, Lars R. Knudsen, and Vincent Rijmen. The Block Cipher Square. In Eli Biham, editor, *FSE*, volume 1267 of *LNCS*, pages 149–165. Springer, 1997.

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

[FKL$^+$00] Niels Ferguson, John Kelsey, Stefan Lucks, Bruce Schneier, Michael Stay, David A. Wagner, and Doug Whiting. Improved Cryptanalysis of Rijndael. In Bruce Schneier, editor, *FSE*, volume 1978 of *LNCS*, pages 213–230. Springer, 2000.

[GR18] Lorenzo Grassi and Christian Rechberger. New Rigorous Analysis of Truncated Differentials for 5-round AES. *IACR Cryptology ePrint Archive*, 2018:182, 2018.

[Gra17] Lorenzo Grassi. Mixture Differential Cryptanalysis: a New Approach to Distinguishers and Attacks on round-reduced AES. *IACR Cryptology ePrint Archive*, 2017:832, 2017.

[Gra18a] Lorenzo Grassi. MixColumns Properties and Attacks on (Round-Reduced) AES with a Single Secret S-Box. In Nigel P. Smart, editor, *CT-RSA*, volume 10808 of *LNCS*, pages 243–263. Springer, 2018.

[Gra18b] Lorenzo Grassi. Mixture Differential Cryptanalysis: a New Approach to Distinguishers and Attacks on round-reduced AES. *IACR Transactions on Symmetric Cryptology*, 2018(2):133–160, 2018.

[GRR16] Lorenzo Grassi, Christian Rechberger, and Sondre Rønjom. Subspace Trail Cryptanalysis and its Applications to AES. *IACR Trans. Symmetric Cryptol.*, 2016(2):192–225, 2016.

[GRR17] Lorenzo Grassi, Christian Rechberger, and Sondre Rønjom. A New Structural-Differential Property of 5-Round AES. In Jean-Sébastien Coron and Jesper Buus Nielsen, editors, *EUROCRYPT II*, volume 10211 of *LNCS*, pages 289–317, 2017.

[HCGW18] Kai Hu, Tingting Cui, Chao Gao, and Meiqin Wang. Towards Key-Dependent Integral and Impossible Differential Distinguishers on 5-Round AES. In Carlos Cid and Michael J. Jacobson Jr., editors, *SAC*, volume 11349 of *LNCS*, pages 139–162. Springer, 2018.

[MDRM10] Hamid Mala, Mohammad Dakhilalian, Vincent Rijmen, and Mahmoud Modarres-Hashemi. Improved Impossible Differential Cryptanalysis of 7-Round AES-128. In Guang Gong and Kishan Chand Gupta, editors, *INDOCRYPT*, volume 6498 of *LNCS*, pages 282–291. Springer, 2010.

[Nat01] National Institute of Standards and Technology. FIPS 197. *National Institute of Standards and Technology, November*, pages 1–51, 2001.

[NMV16] Valérie Nachef, Nicolas Marrière, and Emmanuel Volte. Improved Attacks on Extended Generalized Feistel Networks. In Sara Foresti and Giuseppe Persiano, editors, *CANS*, volume 10052 of *LNCS*, pages 562–572, 2016.

[NPV14] Valérie Nachef, Jacques Patarin, and Emmanuel Volte. 4-point Attacks with Standard Deviation Analysis on A-Feistel Schemes. *IACR Cryptology ePrint Archive*, 2014:446, 2014.

[NPV18] Valérie Nachef, Jacques Patarin, and Emmanuel Volte. Generic attacks with standard deviation analysis on a-feistel schemes. *Cryptography and Communications*, 10(1):59–77, 2018.

[Pat08] Jacques Patarin. Generic Attacks for the Xor of k random permutations. *IACR Cryptology ePrint Archive*, 2008:9, 2008.

[Pat13] Jacques Patarin. Generic Attacks for the Xor of k Random Permutations. In Michael J. Jacobson Jr., Michael E. Locasto, Payman Mohassel, and Reihaneh Safavi-Naini, editors, *ACNS*, volume 7954 of *LNCS*, pages 154–169. Springer, 2013.

[RBH17] Sondre Rønjom, Navid Ghaedi Bardeh, and Tor Helleseth. Yoyo Tricks with AES. In Tsuyoshi Takagi and Thomas Peyrin, editors, *ASIACRYPT I*, volume 10624 of *LNCS*, pages 217–243. Springer, 2017.

[SLG$^+$16a] Bing Sun, Meicheng Liu, Jian Guo, Longjiang Qu, and Vincent Rijmen. New Insights on AES-Like SPN Ciphers. In Matthew Robshaw and Jonathan Katz, editors, *CRYPTO I*, volume 9814 of *LNCS*, pages 605–624. Springer, 2016.

[SLG$^+$16b] Bing Sun, Meicheng Liu, Jian Guo, Vincent Rijmen, and Ruilin Li. Provable Security Evaluation of Structures Against Impossible Differential and Zero Correlation Linear Cryptanalysis. In Marc Fischlin and Jean-Sébastien Coron, editors, *EUROCRYPT I*, volume 9665 of *LNCS*, pages 196–213. Springer, 2016.

[SLR$^+$15] Bing Sun, Zhiqiang Liu, Vincent Rijmen, Ruilin Li, Lei Cheng, Qingju Wang, Hoda AlKhzaimi, and Chao Li. Links Among Impossible Differential, Integral and Zero Correlation Linear Cryptanalysis. In Rosario Gennaro and Matthew Robshaw, editors, *CRYPTO I*, volume 9215 of *LNCS*, pages 95–115. Springer, 2015.

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

[VNM16a] Emmanuel Volte, Valérie Nachef, and Nicolas Marrière. Automatic Expectation and Variance Computing for Attacks on Feistel Schemes. *IACR Cryptology ePrint Archive*, 2016:136, 2016.

[VNM16b] Emmanuel Volte, Valérie Nachef, and Nicolas Marrière. Improvements of Attacks on Various Feistel Schemes. In Raphael C.-W. Phan and Moti Yung, editors, *Mycrypt*, volume 10311 of *LNCS*, pages 321–344. Springer, 2016.

[WCC$^+$16] Meiqin Wang, Tingting Cui, Huaifeng Chen, Ling Sun, Long Wen, and Andrey Bogdanov. Integrals Go Statistical: Cryptanalysis of Full Skipjack Variants. In Thomas Peyrin, editor, *FSE*, volume 9783 of *LNCS*, pages 399–415. Springer, 2016.

[WJ18] Qian Wang and Chenhui Jin. Upper bound of the length of truncated impossible differentials for AES. *Des. Codes Cryptography*, 86(7):1541–1552, 2018.