Two-Round AES Differentials

Joan Daemen[1] and Vincent Rijmen[2,3]

[1] STMicroelectronics Belgium
joan.daemen@st.com
[2] IAIK, Graz University of Technology
vincent.rijmen@iaik.tugraz.at
[3] Cryptomathic A/S

Abstract. In this paper we study the probability of differentials and characteristics over 2 rounds of the AES with the objective to understand how the components of the AES round transformation interact. We extend and correct the analysis of the differential properties of the multiplicative inverse in GF(2^n) given in [17]. We show that AES has characteristics with a fixed-key probability that is many times larger than the EDP. For instance, in the case of 2-round AES, we measured factors up to 2^{100}. We study the number of characteristics with EDP > 0 whose probability adds up to the probability of a differential and derive formulas that allow to produce a close estimate of this number for any differential. We show how the properties discovered in our study can be used to explain the values of the differentials with the largest EDP values and to construct new distinguishers using truncated differentials.

1 Introduction

In this paper we study the probability of differentials and characteristics [3, 13] over 2 rounds of the AES where the difference is the bitwise XOR. Bounds on the expected differential probability (EDP) of characteristics were proven in the design documentation of Rijndael [7]. Bounds on the EDP of differentials have been investigated in [10, 18, 19]. In [11] it was shown that the exact maximum value of the EDP of a differential over 2 rounds is equal to $53/2^{34}$.

We investigated differential propagation in AES, with the objective to understand how the components of the AES interact. We explain observed difference propagation probability values, including the maximum EDP value over 2 rounds. This paper contains the results of this study.

In Section 3, we extend and correct the analysis of the differential properties of the multiplicative inverse in GF(2^n) given in [17]. In Section 4 we derive Theorem 1, which leads to a surprising result on the fixed-key probability of two-round characteristics. Except for a side result in [12], we are not aware of previous work explaining how a characteristic's fixed key probability relates to its EDP. We discovered that several ciphers have characteristics with a fixed-key probability that is many times larger than the EDP. For instance, in the case of 2-round AES, we measured factors up to 2^{100}. For this kind of characteristics, the widely made assumption that the EDP of a characteristic is a good approximation for its key-dependent probability for most keys is not justified. In Section 5, we study the number of characteristics with EDP > 0 whose probability adds up to the probability of a differential. We derive formulas that allow to produce a close estimate of this number for any differential. Finally, in Section 6, we give some applications of our results on the distribution of the number of trails. We discuss the maximum EDP value of [11] in the light of our results and present a new truncated differential distinguisher. But first we briefly introduce some new terminology and define notations.

2 AES and Differential Cryptanalysis Basics

2.1 Basics of differential probability

We denote a differential over an arbitrary map by (a, b) and assume that it is clear from the context which map we mean. We call a the input difference and b the output difference. The probability of a differential is denoted by $\mathrm{DP}(a, b)$. The weight of a possible differential is defined as minus the binary logarithm of its DP.

$$\mathrm{weight}(a, b) = -\log_2 \mathrm{DP}(a, b) \ . \tag{1}$$

If the mapping is keyed, we can define a differential probability $\mathrm{DP}[k](a, b)$ for each value k of the key. We define the expected differential probability (EDP) of a differential as the average of the differential probability $\mathrm{DP}(a, b)$ over all keys.

2.2 The AES Super box

The AES S-box operates on $\mathrm{GF}(2^8)$ and can be described as

$$S[x] = L(x^{-1}) + q, \tag{2}$$

Here x^{-1} denotes the multiplicative inverse of x in $\mathrm{GF}(2^8)$, extended with 0 being mapped to 0. L a linear transformation over $GF(2)$ and q a constant. Note that L is not linear over $\mathrm{GF}(2^8)$ and can be expressed as a so-called *linearized polynomial* [14]. The additive group of the finite field $\mathrm{GF}(2^8)$ forms a vector space. In the remainder of this paper, we will sometimes tacitly switch from one representation to another.

For reasons of clarity, we introduce the structure of the (*AES*) *super box* (our notation). The differential probabilities over this structure are equivalent to those over 2 AES rounds. The AES super box maps a 4-byte array $a = (a_0, a_1, a_2, a_3)$ to a 4-byte array e and takes a 4-byte key k. It consists of the sequence of four transformations:

- $b_i = S[a_i]$: the AES S-box
- $c = \mathrm{M_c} b$: the MixColumns matrix multiplication
- $d = c \oplus k$: key addition
- $e_i = d_i^{-1}$: the multiplicative inverse in $\mathrm{GF}(2^8)$ being the AES S-box with L and q removed.

We denote the inverse of $\mathrm{M_c}$ by $\mathrm{M_{ci}}$. If we consider two AES rounds, swap the steps ShiftRows and SubBytes in the first round, and remove all linear operations before the first multiplicative inverse map and after the second multiplicative inverse map, then we obtain a map that can also be described as 4 parallel instances of the AES super box.

We can partition input differences, where all differences in a given subset are zero in a number of byte positions. Such a subset is characterized by an *activity pattern.* The activity pattern has a single bit for each byte position indicating whether its value must be 0 (passive) or not (active). The activity pattern of a differential (a, e) is the couple of the activity patterns of a and e.

Definition 1. *A* differential trail *through an iterative function is a sequence of differences a, b, . . . such that there are pairs $\{x, x \oplus a\}$ that exhibit the sequence of differences through the iterative function for some keys.*

Hence, a differential trail is a characteristic with EDP > 0. A differential trail through the AES super box consists of a sequence of 5 differences: a, b, c, d and e. In a trail through the AES super box, we always have $c = d$ (so we omit c) and $d = \mathrm{M_c} b$. We denote these trails by (a, b, d, e).

Given a key, a trail has a differential probability $\mathrm{DP}[k](a, b, d, e)$. The EDP of a trail is the average of its $\mathrm{DP}[k]$ over all keys. For Markov ciphers, the EDP of a trail Q is the product of the DP of its S-boxes and $\mathrm{weight}(Q) = -\log_2 \mathrm{EDP}(Q)$ is the sum of the weights of its S-boxes [13]. A trail (a, b, d, e) *contributes to* a differential (f, g) if $a = f$ and $e = g$. We denote the number of trails that contribute to a differential (a, e) by $\mathrm{N_t}(a, e)$.

We define an *inner trail* in the AES super box as a couple $|b, d|$ such that $d = \mathrm{M_c} b$. We say an inner trail $|b, d|$ is *compatible with* a differential (a, e) if (a, b, d, e) is a trail. The activity pattern of an inner trail $|b, d|$ is the couple of the activity patterns of b and d. An inner trail $|b, d|$ can only be compatible with a differential (a, e) if they have matching activity patterns. Due to the diffusion properties of $\mathrm{M_c}$, activity patterns of differentials must have a minimum of 5 active positions. In total there are 93 such activity patterns.

3 The multiplicative inverse in GF(2^n)

In this section we discuss the differential properties of the single component in AES that is non-linear over GF(2): the multiplicative inverse in GF(2^n), extended with 0 being mapped to 0. In fact this is the operation of raising to the power $2^n - 2$. For readability we use the notation x^{-1} rather than x^{2^n-2}. Differential properties of this map were previously already studied in [17]. In the following, a and b denote arbitrary non-zero differences. We need the *trace map* defined over a finite field GF(p^n), denoted by $\mathrm{Tr}(x)$:

$$\mathrm{Tr}(x) = \sum_{i=0}^{n-1} x^{p^i} \tag{3}$$

Note that the trace map is linear over GF(p) and that $\mathrm{Tr}(x^{p^i}) = \mathrm{Tr}(x)$ for any value of i. The differential (a, b) over the multiplicative inverse map has $\mathrm{DP}(a, b) > 0$ if and only if the equation

$$(x + a)^{-1} + x^{-1} = b \tag{4}$$

has solutions. If $x = a$ or $x = 0$ is a solution of (4), we have $b = a^{-1}$ and both are solutions. Otherwise, $x = a$ or $x = 0$ is not a solution, we can transform (4) by multiplying with $x(x + a)$ yielding the following equivalent equation:

$$bx^2 + abx + a = 0, \tag{5}$$

To investigate the condition for this equation to have solutions we have the following lemma:

Lemma 1 ([14, Theorem 2.25]). *$z^2 + z + t$ is irreducible over GF(2) iff Tr(t) $\neq 0$.*

From this it follows easily that:

Lemma 2 ([9]). *For $b \neq a^{-1}$, equation (4) has 2 solutions if Tr($(ab)^{-1}$) $= 0$, and zero solutions otherwise.*

Consider now the case $b = a^{-1}$. We present now the following new result:

Lemma 3. *For even n, the solutions of*

$$(x + a)^{-1} + x^{-1} = a^{-1} \tag{6}$$

form the set $T_a = \{0, a, ea, e^2a\}$ where e and e^2 are the two elements of GF(2^n) of order 3. [4]

Proof. $x = a$ and $x = 0$ are solutions of (6). Assume there are other solutions. We can write such a solution as a product of a with an element z different from 0 or 1. We have

$$(za + a)^{-1} + (za)^{-1} = a^{-1} \ . \tag{7}$$

Or, equivalently,

$$(z + 1)^{-1} + z^{-1} = 1 \ . \tag{8}$$

Multiplication with $z(z + 1)$ yields:

$$z^2 + z + 1 = 0 \ . \tag{9}$$

Applying Lemma 1, Equation (9) has two solutions iff $\mathrm{Tr}(1) = 0$ and none otherwise. $\mathrm{Tr}(1) = 0$ iff n is even. As for a solution of (9) we have $z^3 = 1$, its solutions are the two elements of $\mathrm{GF}(2^n)$ of order three. □

Note that $\{0, 1, e, e^2\}$ is $\mathrm{GF}(2^2)$ and that $e^2 + e = 1$. From these lemmas follow several corollaries.

Corollary 1 ([17]). *For odd n,*

$$(x + a)^{-1} + x^{-1} = a^{-1} \tag{10}$$

has two solutions: 0 and a.

Corollary 2. *For even n, the possible output differences b for a given input difference a are those with Tr$((ab)^{-1}) = 0$ except $b = 0$. For odd n, the possible output differences b for a given input difference a are those with Tr$((ab)^{-1}) = 0$ except $b = 0$ and extended with $b = a^{-1}$.*

Together with the fact that (4) has 4 solutions only if $b = a^{-1}$, this leads to the following corollary:

Corollary 3. *For all non-zero $k \in GF(2^n)$ and for all positive integers t:*

$$DP(a, b) = DP(b, a) = DP(ka, bk^{-1}) = DP(a^{2^t}, b^{2^t}), \tag{11}$$

The following observation will be used in Section 5 to describe the distribution of the number of trails in a 2-round differential.

Observation 1. Since the trace map is linear over $\mathrm{GF}(2)$, the solution space of $\mathrm{Tr}(ax) = 0$ is a vector space of dimension $n - 1$ over $\mathrm{GF}(2)$. The intersection of $\mathrm{Tr}(ax) = 0$ and $\mathrm{Tr}(bx) = 0$ is a vector space of dimension $n - 2$ or $n - 1$. If the dimension is $n - 1$, this implies $a = b$. In general, the dimension of the intersection of a set $\mathrm{Tr}(v_j x) = 0$ is equal to n minus the dimension of the vector space generated by the elements v_j. For example, the solution space of $\mathrm{Tr}(ax) = \mathrm{Tr}(bx) = \mathrm{Tr}(cx) = 0$ with $a \neq b \neq c \neq a$ has dimension $n - 2$ if $c = a + b$ and dimension $n - 3$ otherwise.

It follows that for even n, for a given input difference a, the multiplicative inverses of the possible output differences b^{-1} form the vector space given by $\mathrm{Tr}(a^{-1}x)$ minus the origin.

[4] Note that the description of the solutions given in [17]: $T_a = \{0, a, a^{1+d}, a^{1+2d}\}$ with $d = (2^n - 1)/3$ is only correct if $a^d \neq 1$, i.e. if the order of a does not divide $(2^n - 1)/3$.

4 Plateau trails

In this section we show that for a large class of trails the DP$[k]$ can take only a small number of values.

4.1 Planar differentials and mappings

Let $F_{(a,b)}$ denote the set containing the inputs x for which the pair $\{x, x+a\}$ follows the differential (a, b). Let $G_{(a,b)}$ denote the set containing the corresponding outputs.

We introduce the concept of planar differentials:

Definition 2. *A differential (a, b) is* planar *if $F_{(a,b)}$ and $G_{(a,b)}$ form affine variants:*

$$F_{(a,b)} = u \oplus U_{(a,b)} \tag{12}$$

$$G_{(a,b)} = v \oplus V_{(a,b)} \tag{13}$$

with $U_{(a,b)}$ and $V_{(a,b)}$ vector spaces, u any element in $F_{(a,b)}$ and v any element in $G_{(a,b)}$.

Since $F_{(a,b)}$ contains the elements x and $x \oplus a$, it follows that $a \in U_{(a,b)}$. The number of elements in $F_{(a,b)}$ is $2^{\dim(U_{(a,b)})}$, so $\dim(U_{(a,b)}) = \mathrm{n_b} - \mathrm{weight}(a, b)$, with $\mathrm{n_b}$ the input size of the mapping. Similarly, we have $b \in V_{(a,b)}$ and $\dim(V_{(a,b)}) = \mathrm{n_b} - \mathrm{weight}(a, b)$. We can now prove the following lemmas.

Lemma 4. *A differential (a, b) which is followed by exactly two pairs,*[5] *is planar.*

Proof. Denote the pairs by $\{p, p \oplus a\}$, $\{p \oplus a, p\}$. The elements p and $p \oplus a$ form an affine variant of dimension 1 with offset $u = p$ and the basis of $U_{(a,b)}$ equal to (a). A similar argument is valid for the elements of the pairs at the output. □

Lemma 5. *A differential (a, b) which is followed by exactly four pairs, is planar.*

Proof. Denote the inputs of the pairs by $p, p \oplus a, q$ and $q \oplus a$. These 4 elements lie in an affine variant of dimension 2 with offset $u = p$ and basis of $U_{(a,b)}$ equal to $(a, p \oplus q)$. This yields $U_{(a,b)} = \{0, a, p \oplus q, a \oplus p \oplus q\}$ and hence $F_{(a,b)} = \{p, p \oplus a, q, a \oplus q\}$. A similar argument is valid for the elements of the pairs at the output. □

Lemma 6. *Any differential with $DP = 1$ is a planar differential.*

Proof. $F_{(a,b)}$ and $G_{(a,b)}$ form the complete input space and output space respectively. □

Examples of differentials with DP $= 1$ are the trivial differential $(0, 0)$ and differentials over linear mappings. If $\mathrm{DP}(a, b) = 2^{t-\mathrm{n_b}}$, with $t \notin \{1, 2, \mathrm{n_b}\}$, the differential may or may not be planar.

Definition 3. *A mapping is planar if all differentials over it are planar.*

Any mapping for which all non-trivial differentials have $\mathrm{DP}(a, b) \leq 2^{2-\mathrm{n_b}}$ is planar. Such mappings are sometimes called differentially 4-uniform. Now we give a number of properties on planar differentials over composed mappings.

Property 1. Let $y = \alpha(x)$ be a mapping consisting of a set of parallel mappings $y_i = \alpha_i(x_i)$ with $x = (x_0, x_1, \ldots, x_t)$ and $y = (y_0, y_1, \ldots, y_t)$. A differential (a, b) for which the differentials (a_i, b_i) are planar, is planar.

[5] We adopt the convention that $\{p, q\}$ and $\{q, p\}$ are two different pairs.

We have

$$U_{(a,b)} = U_{(a_0,b_0)} \times U_{(a_1,b_1)} \times \cdots \times U_{(a_t,b_t)}$$
$$V_{(a,b)} = V_{(a_0,b_0)} \times V_{(a_1,b_1)} \times \cdots \times V_{(a_t,b_t)}$$

with × denoting the direct product [8].

Property 2. If (a, b) is a planar differential of α, then for any pair of affine mappings L_1 and L_2 with L_1 invertible, the differential $(L_1(a), L_2(b))$ is planar over $L_2 \circ \alpha \circ L_1^{-1}$.

Examples of ciphers in which single-round differentials are planar are AES, but also 3-Way [4], SHARK [20], Square [5], Camellia [2], Serpent [1] and Noekeon [6].

4.2 Two-round plateau trails

We now introduce two related concepts: a *plateau trail* and its *height*.

Definition 4. *A trail Q is called a* plateau trail *with height* height(Q) *if and only if the following holds:*

1. *For a fraction $2^{\mathrm{n_b}-\mathrm{weight}(Q)-\mathrm{height}(Q)}$ of the keys $DP[k](Q) = 2^{\mathrm{height}(Q)-\mathrm{n_b}}$, and*
2. *For all other keys $DP[k](Q) = 0$.*

The height of a plateau trail is a nonzero positive integer. It can be bounded as follows. height(Q) is maximal when all but one key have DP equal to zero. In that case EDP $= 2^{-\mathrm{n_b}}$DP. Taking the logarithm, we obtain $-\mathrm{weight}(Q) = -\mathrm{n_b} + \mathrm{height}(Q) - \mathrm{n_b}$. Hence, $\mathrm{height}(Q) \leq 2\mathrm{n_b} - \mathrm{weight}(Q)$. We can now prove the following:

Theorem 1 (Two-Round Plateau Trail Theorem). *A trail $Q = (a, b, c)$ over an iterative mapping consisting of two steps with a key addition in between, in which the differentials (a, b) and (b, c) are planar, is a plateau trail with* $\mathrm{height}(Q) = \dim(V_{(a,b)} \cap U_{(b,c)})$.

Proof. The proof is based on geometrical arguments [8]. For pairs following the trail, the values at the output of the first step are in $G_{(a,b)}$. The values at the input of the second step are in $F_{(b,c)}$, or equivalently, the values at the output of the first step are in $k \oplus F_{(b,c)}$. It follows that the values at the output of the first step are in:

$$H = G_{(a,b)} \cap (k \oplus F_{(b,c)}) \ .$$

Since both differentials are planar, there exist offsets u, v such that

$$H = (v \oplus V_{(a,b)}) \cap (k \oplus u \oplus U_{(b,c)}) \ ,$$

with $V_{(a,b)}$ and $U_{(b,c)}$ vector spaces. Let W denote the vector space $V_{(a,b)} \cap U_{(b,c)}$. Suppose now that H contains at least one element, denoted by h. Hence,

$$h \oplus v \in V_{(a,b)} \text{ and } h \oplus k \oplus u \in U_{(b,c)}.$$

Since W is a subset of $V_{(a,b)}$ and $U_{(b,c)}$, for all $w \in W$:

$$h \oplus v \oplus w \in V_{(a,b)} \text{ and } h \oplus k \oplus u \oplus w \in U_{(b,c)},$$

which implies $w \oplus h \in H$. Hence, if H contains an element h, then $W \oplus h \subseteq H$.

Secondly, if $h \in H$, then there are vectors $v_i \in V_{(a,b)}$, $u_j \in U_{(b,c)}$ such that

$$v \oplus v_i = k \oplus u \oplus u_j.$$

If we consider this as an equation in k, then it follows that if there is one solution for k, then there are 2^z solutions for k, with $z = \dim(V_{(a,b)} \oplus U_{(b,c)})$.

We have proven so far that for 2^z different keys, there are at least $2^{\dim(W)}$ pairs following the differential. Invoking the subspace dimension theorem, we obtain:

$$\begin{aligned} z + \dim(W) &= \dim(V_{(a,b)} \oplus U_{(b,c)}) + \dim(V_{(a,b)} \cap U_{(b,c)}) \\ &= \dim(V_{(a,b)}) + \dim(U_{(b,c)}) \\ &= \mathrm{n_b} - \mathrm{weight}(a,b) + \mathrm{n_b} - \mathrm{weight}(b,c) = 2\mathrm{n_b} - \mathrm{weight}(Q) \\ 2^z 2^{\dim(W)} &= 2^{\mathrm{n_b}}\mathrm{EDP}(Q). \end{aligned}$$

Hence, we have accounted for all the pairs, and denoting by $K(Q)$ the affine variant $v \oplus u \oplus V_{(a,b)} \oplus U_{(b,c)}$, we can summarize as follows. If $k \notin K(Q)$, then H is empty. If $k \in K(Q)$, then $H = W \oplus h$.

It follows that the trail is a plateau trail with $\mathrm{height}(Q) = \dim(V_{(a,b)} \cap U_{(b,c)})$. □

This theorem is valid for all ciphers in which single-round differentials are planar and round keys are applied with XOR. This includes all ciphers mentioned in Section 4.1.

Like any other trail, a plateau trail has $\mathrm{EDP}(Q) = \mathrm{DP}(a,b)\mathrm{DP}(b,c) = 2^{-\mathrm{weight}(Q)}$. Only if $\mathrm{height}(Q) = \mathrm{n_b} - \mathrm{weight}(Q)$, it holds that $\mathrm{DP}[k](Q) = \mathrm{EDP}(Q)$ for all keys. This can only be the case for trails with $\mathrm{weight}(Q) < \mathrm{n_b}$.

4.3 Trails in the AES super box

The AES super box satisfies the criteria of Theorem 1 and hence all trails Q in the AES super box are plateau trails. $\mathrm{DP}[k](Q)$ can be described by defining $W = V_{(a,b)} \cap U_{(b,e)}$ and $U_{(b,e)} = \mathrm{M_{ci}}(U_{(d,e)})$, where $\mathrm{M_{ci}}(U) = \{\mathrm{M_{ci}}u | u \in U\}$.

We have determined for all trails over the AES super box their weight and height. The weight ranges from 30 to 56, the height from 1 to 5 and the ratio $\mathrm{DP}[k](Q)/\mathrm{EDP}(Q)$ from 1 to 2^{25}.

We call trails for which the ratio is 1 *flat trails* because for these the equality $\mathrm{DP}[k](Q) = \mathrm{EDP}(Q)$ holds for all keys. The majority of the trails has ratio 2^{25}. Since the AES round transformation can be described as the parallel application of 4 super boxes, it follows that for most trails over two rounds of AES, there are keys with $\mathrm{DP}[k](Q) = 2^{100}\mathrm{EDP}(Q)$. A summary of the results is given in Table 1. Note that the portion of trails with height larger than 1 is only 1/8000. For more detailed results, we refer to Appendix C, Table 4.

Table 1. Number of trails (binary logarithm) and their height for the AES Super box.

Height	1	2	3	4	5
Number of trails	87.9	74.9	60.5	45.4	31.7

4.4 Impact on the DP of differentials

The dependence of the DP of trails on the key value makes that the DP of differentials also depends on the key. If we denote the trails that contribute to a differential (a, b) by Q_i we have:

$$\mathrm{DP}[k](a, b) = 2^{-32} \sum_{i|k \in K(Q_i)} 2^{\mathrm{height}(Q_i)} \ . \tag{14}$$

Hence this value varies per key k depending on the number of affine subspaces $K(Q_i)$ it is in. Trails with larger height contribute more to the variability. Flat trails add a constant term and do not contribute to the variability. There is at most one flat trail per differential with 5 active S-boxes and none for differentials with more than 5 active S-boxes. The exact distribution of $\mathrm{DP}[k](a, b)$ depends on the relative positions of the affine subspaces $K(Q_i)$ and the height of the trails.

5 Distribution of the number of trails

In this section we treat the distribution of the number of trails in differentials. This is valuable in understanding the distribution of the EDP of differentials. We first derive the distribution of the number of trails in differentials with 5 active S-boxes. This is followed by a treatment of the distribution of number of trails in differentials with more than 5 active S-boxes.

5.1 Inner trail bundles

Assume we want to count the number of trails in a differential (a, e) with activity pattern (1000; 1111). Thanks to MixColumns we have $d_0 = 2b_0$, $d_1 = b_0$, $d_2 = b_0$ and $d_3 = 3b_0$. So every inner trail with this activity pattern is fully determined by the value of b_0. It follows that there can be at most 255 trails in the differential.

This can be generalized to any activity pattern with 5 active S-boxes. If $|b, d|, |b', d'|$ are two inner trails of a differential with 5 active S-boxes, then there exists a γ such that $b_i = \gamma b'_i$ and $d_i = \gamma d'_i$. We generalize this to the concept of inner trail bundles:

Definition 5. *The* inner trail bundle*, or* bundle*, $B(u)$ associated with the vector u, is the set of inner trails $|u, v|$ defined as follows:*

$$B(u) = \{|\gamma u, \gamma(\mathrm{M_c} u)| | \gamma \in GF(256) \backslash \{0\}\},$$

where the multiplications by γ are scalar multiplications.

All inner trails in a bundle have the same activity pattern. We call the bundles of a differential (a, b) the bundles with the same activity pattern as (a, b). The number of bundles of a differential (a, b) depends only on the number of active S-boxes. A differential with 5 active S-boxes has a single bundle, a differential with 6 active S-boxes has 251 bundles, see Appendix A.

Bundles are similar to, but different from, the 5-lists defined in [10, 11]. Both bundles and 5-lists group sets of 255 inner trails. Bundles with 5 active S-boxes correspond with the 5-lists of type 1. In bundles with more than 5 active S-boxes the ratios between the inner differences are fixed, while in 5-lists of type 2, a number of inner differences are fixed. Their goal is also different: the concept of 5-lists helps in efficiently finding bounds, while bundles help to gain insight in the distribution of trails in differentials.

Table 2. Mean (left) and variance (right) of the number of trails for a differential given α and β.

α, β	1	2	3	4
1	63.25	31.50	15.69	7.81
2	31.38	15.63	7.78	3.88
3	15.44	7.69	3.83	1.91
4	7.47	3.72	1.85	0.92

α, β	1	2	3	4
1	16.00	15.89	10.86	6.38
2	11.91	9.85	6.11	3.40
3	6.83	5.33	3.19	1.73
4	3.54	2.70	1.59	0.85

5.2 Number of trails in a bundle

In this section we derive formulas to estimate the number of trails in a bundle for the AES super box. Let (a, e) be a differential over the AES super box and $B(u)$ a bundle of (a, e). Define α as the dimension of the vector space generated by

$$\{y_0^{-1}, y_1^{-1}, y_2^{-1}, y_3^{-1}\}.$$

with $y = (\mathrm{M_c} u) * e$, where $*$ denotes the componentwise product. Also here, we adopt the convention that $0^{-1} = 0$. Furthermore, define β as the number of different non-zero elements in the following set of couples:

$$\{(a_0, u_0), (a_1, u_1), (a_2, u_2), (a_3, u_3)\}$$

Let (a, e) be a randomly selected differential and $B(u)$ a randomly selected bundle of (a, e) then the number of trails in the bundle $B(u)$ of the differential (a, e) can be described as a stochastic variable with the expected value and variance given by:

$$\langle \mathrm{N_t} \rangle = \left(\frac{127}{255}\right)^{\beta} (2^{8-\alpha} - 1) \ , \tag{15}$$

$$\mathrm{var}(\mathrm{N_t}) = (2^{8-\alpha} - 1) \left(\frac{127}{255}\right)^{\beta} \left[1 - \left(\frac{127}{255}\right)^{\beta} + (2^{8-\alpha} - 2) \left(\left(\frac{63}{127}\right)^{\beta} - \left(\frac{127}{255}\right)^{\beta} \right) \right] . \tag{16}$$

In the following section, we derive these formulas for the special case of one active S-box in the first round. A more general derivation is given in Appendix B.

5.3 One active S-box in the first round

Consider a differential (a, e) with activity pattern $(1000; 1111)$. This differential has only a single bundle, represented by $u = (1, 0, 0, 0)$. We have $(u; \mathrm{M_c}(u)) = (1, 0, 0, 0; 2, 1, 1, 3)$. It follows from Section 3 that an inner trail $|b, d|$ with activity pattern $(1000; 1111)$ is compatible with (a, e) if $(L^{-1}(b_0))^{-1}$ is a solution of $\mathrm{Tr}(a_0^{-1} x) = 0$ and each of the d_i^{-1} is a solution of $\mathrm{Tr}(e_i^{-1} x) = 0$. Using the relations between the inner differences in a bundle, we can reduce the latter four conditions to conditions in γ^{-1}:

$$\mathrm{Tr}((2e_0)^{-1}\gamma^{-1}) = \mathrm{Tr}(e_1^{-1}\gamma^{-1}) = \mathrm{Tr}(e_2^{-1}\gamma^{-1}) = \mathrm{Tr}((3e_3)^{-1}\gamma^{-1}) = 0 \ . \tag{17}$$

This can be seen as the intersection of 4 vector spaces over GF(2) which forms a subspace of dimension $8 - \alpha$, where α is the dimension of the vector space generated by the elements of the set $C = \{(2e_0)^{-1}, e_1^{-1}, e_2^{-1}, (3e_3)^{-1}\}$. The number of trails satisfying these four conditions is then $2^{8-\alpha} - 1$.

If $e = (z/2, z, z, z/3)$ for any nonzero value z, then $C = \{z^{-1}\}$ resulting in $\alpha = 1$ and hence there are 127 trails satisfying these four conditions. These conditions interact in an easy to analyze manner and we call them *sharp*.

The remaining condition is harder to analyse. Therefore we call it *blurred*. The effect of a blurred condition can be modeled with statistics. We model its presence as a sampling process. The space sampled are the 255 inner trails. Due to the blurred condition, 127 out of the 255 inner trails may give rise to a trail. These are called the good ones, the 128 others the bad ones. The joint sharp conditions take a sample with size $2^{8-\alpha} - 1$. This gives rise to a hypergeometric distribution $H(\mathrm{N_t}; n, m, N)$ [16] with the following parameters:

- Number of ways for a good selection $n = 127$.
- Number of ways for a bad selection $m = 255 - 127 = 128$.
- Sample size N: $2^{8-\alpha} - 1$.

Denoting the event that one inner trail gives rise to a trail (the outcome of a single sampling) by x_i, we obtain $\langle x_i \rangle = n/(m + n)$. Since $\mathrm{N_t} = \sum_i x_i$,

$$\langle \mathrm{N_t} \rangle = \frac{n}{m+n} N = \frac{127}{255}(2^{8-\alpha} - 1).$$

This gives formula (15). For the variance, we obtain:

$$\mathrm{var}(\mathrm{N_t}) = \frac{mnN(m + n - N)}{(m+n)^2(m+n-1)} = \frac{128 \times 127(2^{8-\alpha} - 1)(256 - 2^{8-\alpha})}{255^2 254} ,$$

which corresponds to (16). The exact distributions of the number of trails per differential for all four values of α are given in Appendix D.

5.4 Any bundle

Every differential (a, e) imposes on the bundle $B(u)$ a number of sharp conditions, determined by e and u, and a number of blurred conditions, determined by a and u. The sharp conditions result in the condition that γ^{-1} is in a vector space of dimension $8 - \alpha$ ranging from 4 to 7.

We denote the number of blurred conditions by β. For the vast majority of differentials, β equals the number of active S-boxes in a. β is smaller only when two a_i values are the same and the corresponding u_i in the bundle are also equal. Hence a reduction of β occurs much less often than a reduction of α. Both α and β range from 1 to 4 limited by $\alpha + \beta \leq 5$.

We have conducted a large number of experiments that confirm the mean and variance predicted by (15) and (16) for any combination of α and β.

5.5 Adding bundles

For a differential (a, e) with more than 5 active S-boxes, we can partition the inner trails with the same activity pattern in bundles. The mean number of trails in a differential is the sum of the mean number of trails in these bundles. For the variance of the number of trails, the sum of the variances in the bundles gives a good idea.

Given a differential (a, e), we can compute for each of its bundles the value of (α,β) and count for each of the combinations (α, β) how many times they occur, denoted by $N_{(a,e)}(\alpha, \beta)$. From this *profile*, we can easily compute the mean number of trails.

This analysis and the distributions of $N_{(a,e)}(\alpha, \beta)$ become more involved as the number of active S-boxes grows. On the other hand, the mean number of trails and the weight of the trails grows and the variance on the EDP shrinks. In the following subsection we describe the relatively simple case of a differential with 6 active S-boxes.

Table 3. Distribution of α-profiles for differentials with activity pattern (1110; 1110)

α-profile			number of	mean	standard deviation	
$\alpha = 3$	$\alpha = 2$	$\alpha = 1$	couples		theory	exp.
250	1	0	21	965.2	28.42	25.65
249	2	0	1501	969.1	28.47	25.14
248	3	0	31170	973.1	28.53	25.15
247	4	0	2175	977.0	28.58	25.16
246	5	0	29907	981.0	28.63	25.23
250	0	1	3	973.1	28.42	23.28
249	1	1	248	977.0	28.47	25.01

5.6 Example: differentials with activity pattern (1110; 1110)

There are in total 251 bundles with activity pattern (1110; 1110). The profile of α over the bundles, is completely determined by the couple $(e_1/e_0, e_2/e_0)$. Over all 255^2 values of this couple, Table 3 gives for each of α-profiles the number of couples. For the profile of β the values of a_0, a_1 and a_2 matter. If they are three different values, β is always equal to 3. For this case, Table 3 gives the theoretical mean and standard deviation of the number of trails (assuming independence between the bundles). If two of the values a_0, a_1 and a_2 are equal, β will be 2 for at most one bundle and 3 for all other. If they are all three equal, β will be 2 for at most three bundles, or maybe 1 for a single bundle and 3 for all the others. In principle, the α-profile and β-profile combine to a two-dimensional profile. In the worst case, the small values of β occur in bundles with a small value of α. All in all, there are only few bundles where β is smaller than the number of active S-boxes in a, hence the impact on the mean and the variance is small. We have experimentally checked the distributions by computing the number of trails for a large set of differentials with the given activity pattern and α-profiles. The measured mean values coincide with the theoretically predicted values. The measured standard deviations, also listed in Table 3 are systematically smaller than the theoretical ones, implying that the number of trails in the bundles of a differential are not independent.

For the other differentials with 6 active S-boxes, the average and standard deviation of the number of trails have similar values as those of the treated case. For differentials with 7 active S-boxes the average is above 120000 and the standard deviation is below 400. For differentials with 8 active S-boxes the average is above 15000000 and the standard deviation is below 4000. These large numbers of trails per differential are compensated by the small value of the EDP of the trails. Trails with 6 active S-boxes have an EDP between 2^{-42} and 2^{-36}, trails with 7 active S-boxes between 2^{-49} and 2^{-42} and trails with 8 active S-boxes between 2^{-56} and 2^{-48}. Among the trails that contribute to a given n-box differential, the vast majority has the minimum EDP value. We conclude that for differentials with more than 5 active S-boxes, due to the large mean and small standard deviation of the number of trails and the small EDP value of the individual trails, the EDP has very narrow distributions.

6 Distribution of EDP

In this section we will give a number of applications of our results on the number of trails per differential: the EDP value of 5-box differentials, the maximum EDP value for 2-round differentials over AES and a new truncated differential attack.

6.1 In 5-box differentials

The distributions for the number of trails can be converted to distributions of the EDP by taking into account the weights of the trail. In this section we will treat the distribution of EDP in 5-box differentials. Differentials with more than 5 active S-boxes have an EDP that is very close to the value 2^{-32}. The bundle of a 5-box differential (a, e) contains 255 inner trails $|b, d|$. The combination of a differential over the AES super box and an inner trail determines the differentials over the S-boxes.

Distributed over the 255 inner trails, there are exactly 5 S-box differentials with weight 6, which we will refer to as *double S-box differentials*. All other S-box differentials have weight 7, and we refer to them as *single S-box differentials*. As only a subset of these inner trails gives rise to trails contributing to the differential, in general not all 5 double S-box differentials occur in trails. A trail with no double S-box differentials has weight 35 or contributes 2^{-35} to the EDP of the differential. A trail with one double S-box differential has weight 34 and hence contributes 2^{-34} to the differential. In other words, it contributes twice as much. In general, a trail with i double S-box differentials contributes 2^i times 2^{-35} to the EDP of the differential.

6.2 Differentials with the maximum EDP value

The maximum EDP value obtained in [11] occurs for exactly 12 differentials over the AES super box. Due to the rotational symmetry of the AES super box, they come in 3 sets, where the differentials in a set are just rotated versions of each other. It is no surprise that they are differentials with 5 active S-boxes, where the deviations from the average value 2^{-32} are largest. Moreover, they have $\alpha = 1$ and $\beta = 1$ for which the expected number of trails is the highest over all differentials with 5 active S-boxes, as is clear from Figure 1 in Appendix D.

The bundles indicated by $(u; \mathrm{M_c}(u))$ are $(1, 0, 0, 0; 2, 1, 1, 3)$, $(1, 1, 0, 0; 1, 3, 0, 2)$ and $(1, 1, 1, 0; 0, 0, 2, 3)$. The corresponding differentials are: $(x, 0, 0, 0; y/2, y, y, y/3)$, $(x, x, 0, 0; y, y/3, 0, y/2)$ and $(x, x, x, 0; 0, 0, y/2, y/3)$, with $x = 75_\mathtt{x}$ and $y = 41_\mathtt{x}$.[6] For these differentials, the number of trails is 75: 74 trails with weight 35 and one with weight 30, resulting in EDP value $2^{-30} + 74 \times 2^{-35} = 13.25 \times 2^{-32}$. Clearly all five double S-box differentials are in the same trail. Note that there are differentials with 5 active S-boxes that have 82 trails (see Appendix D but these have a lower EDP value due to the fact that the double S-box differentials are not in the same trail. In Appendix E we discuss the effect of L on the maximum EDP value.

6.3 Truncated differential distinguishers

Since the EDP is only an average value, it can be expected that there are differentials with $\mathrm{DP} > 13.25 \times 2^{-32}$ for some values of the key. For each value of the key, the differentials with maximal DP are not necessarily those with maximal EDP. More particularly, given the input difference $(A, 0, 0, 0)$, for most keys there will be an output difference that has a higher DP than output difference $(B/2, B, B, B/3)$. This suggests that a differential attack based on the assumption that the differential (a, b) with the maximum EDP has the highest number of pairs among all differentials with input difference a, will not work.

We now describe a simple key-recovering attack exploiting a distinguisher that works for all keys and that is based on the properties discovered in our study. Consider a number

[6] Remember that in our definition of the AES super box, we omit the affine transformation in the S-boxes of the second round.

of AES rounds, preceded by an initial key addition and followed by a final key addition. The basic attack is on two rounds, that can be reduced to the AES super box. Denote the final key by k and the output after the final key addition by f: $f = e + k$.

Compute the output of n input pairs with only a difference in the first S-box: a_0. We now partition the n pairs in 256 sets according to the value of $g = f_1 + 2f_0$. We have $g = e_1 + 2e_0 + k_1 + 2k_0$.

In the output differences with $e_1 = 2e_0$, the first two S-boxes impose the same condition on γ, reducing the number of candidate inner trails from 255 to 127. In the other output differences, the first two S-boxes impose different conditions on γ, reducing the number of candidate inner trails from 255 to 63. The differentials with $e_1 = 2e_0$ turn out to have on the average an EDP that is twice as large as differentials with $e_1 \neq 2e_0$. The truncated differential consists of the sum of all differentials starting in a and with $e_1 = 2e_0$. As said, for a given key the DP values of differentials may diverge from their EDP values. For the DP$[k]$ of our truncated differential this is however not the case: its DP$[k]$ is the sum of the DP$[k]$ values of a huge number of differentials, cancelling out the variations over the keys. It follows that the expected number of pairs for the value of g corresponding with $e_0 = 2e_1$ is approximately two times as high as for the other values. This occurs for $g = k_1 + 2k_0$. If n is large enough, the peak in the correct value of g is the largest with high probability and reveals the value of $k_1 + 2k_0$. If the number of pairs is 2^{14}, the error probability is 0.01%. The partitioning in 256 sets can be performed for any pair of two output S-boxes, giving 6 ways in total. This gives the complete final key k.

This attack can be readily extended to 3 rounds with the same complexity. Extension to 4 rounds can be done by additionally guessing 4 bytes of the initial key and applying 2^{25} plaintexts from which for each key guess 2^{14} pairs with appropriate input difference can be extracted. Extension to 5 rounds requires guessing 4 more key bytes in the final key, resulting in a complexity of 2^{25} chosen plaintexts and workload of $2^{64}2^{15} = 2^{76}$. The distinguisher is presented only as an illustration. It does not yet result in an improved cryptanalysis of AES.

Acknowledgements

We thank anonymous referees for their helpful comments.

References

1. R.A. Anderson, E. Biham, L.R. Knudsen, “Serpent”, *Proc. of the 1st AES candidate conference,* CD-1: Documentation, August 20–22, 1998, Ventura.
2. K. Aoki, T. Ichikawa, M. Kanda, M. Matsui, S. Moriai, J. Nakajima, T. Tokita, “Camellia: a 128-bit block cipher suitable for multiple platforms — Design and analysis,” *Selected Areas in Cryptography 2000, LNCS 2012,* D. Stinson, S. Tavares, Eds., Springer-Verlag, 2000, pp. 39–56.
3. E. Biham and A. Shamir, “Differential Cryptanalysis of DES-like Cryptosystems,” *Journal of Cryptology*, Vol. 4, No. 1, 1991, pp. 3–72.
4. J. Daemen, R. Govaerts and J. Vandewalle, “A New Approach to Block Cipher Design,” *Proc. of Fast Software Encryption 1993, LNCS 809,* R. Anderson, Ed. Springer-Verlag, 1994, pp. 18–32.
5. J. Daemen, L.R. Knudsen and V. Rijmen, “The block cipher Square,” *Fast Software Encryption ’97, LNCS 1267*, E. Biham, Ed., Springer-Verlag, 1997, pp. 149–165.
6. J. Daemen, M. Peeters, G. Van Assche and V. Rijmen, “Nessie Proposal: the block cipher Noekeon,” Submitted to Nessie.

7. J. Daemen, V. Rijmen, *The design of Rijndael — AES, The Advanced Encryption Standard,* Springer-Verlag, 2002.
8. W.V.D. Hodge, D. Pedoe, *Methods of Algebraic Geometry: Volume 1*, Cambridge University Press, 1994.
9. IEEE P1363, Standard specifications for public key cryptography – Annex A: Number-theoretic background (Draft version 13), November 12, 1999.
10. L. Keliher, H. Meijer, and S. Tavares, "Improving the upper bound on the maximum average linear hull probability for Rijndael," Advances in Cryptology, Selected Areas in Cryptography '01, LNCS 2259, S. Vaudenay, A.M. Youssef, Eds., Springer-Verlag, 2001, pp. 112–128.
11. L. Keliher and J. Sui, "Exact maximum expected differential and linear probability for 2-round advanced encryption standard (AES)," Cryptology ePrint archive, Report 2005/321, 2005, http://eprint.iacr.org.
12. L.R. Knudsen and J.E. Mathiassen, "On the role of key schedules in attacks on iterated ciphers," *ESORICS 2004, LNCS 3193,* Springer-Verlag, 2004, pp. 322–334.
13. X. Lai, J.L. Massey and S. Murphy, "Markov Ciphers and Differential Cryptanalysis," *Advances in Cryptology, Proc. Eurocrypt'91, LNCS 547*, D.W. Davies, Ed., Springer-Verlag, 1991, pp. 17–38.
14. R. Lidl, H. Niederreiter, *Introduction to finite fields and their applications,* Cambridge University Press, 1986 (Reprinted 1988).
15. F.J. MacWilliams and N.J.A. Sloane, *The Theory of Error-Correcting Codes*, North-Holland Publishing Company, 1978.
16. Mathworld, `http://mathworld.wolfram.com/`.
17. K. Nyberg, "Differentially uniform mappings for cryptography," *Advances in Cryptology, Proc. Eurocrypt'93, LNCS 765*, T. Helleseth, Ed., Springer-Verlag, 1994, pp. 55–64.
18. S. Park, S.H. Sung, S. Chee, E-J. Yoon and J. Lim, "On the security of Rijndael-like structures against differential and linear cryptanalysis," Advances in Cryptology, Proceedings of Asiacrypt '02, LNCS 2501, Y. Zheng, Ed., Springer-Verlag, 2002, pp. 176–191.
19. S. Park, S.H. Sung, S. Lee and J. Lim, "Improving the upper bound on the maximum differential and the maximum linear hull probability for SPN structures and AES," Fast Software Encryption '03, LNCS 2887, T. Johansson, Ed., Springer-Verlag, 2003, pp. 247–260.
20. V. Rijmen, J. Daemen, B. Preneel, A. Bosselaers, E. De Win, "The cipher SHARK," *Fast Software Encryption '96, LNCS 1039*, D. Gollmann, Ed., Springer-Verlag, 1996, pp. 99–111.

A Number of bundles

The total number of inner trails is $2^{32}-1$. Each bundle has 255 inner trails, so the total number of bundles is

$$\frac{2^{32}-1}{2^8-1} = 2^{24}+2^{16}+2^8+1$$

The number of bundles with a given activity pattern is completely determined by the number of active S-boxes in the activity pattern. If we denote the number of bundles for an activity pattern with x active S-boxes by $\mathrm{BN}(x)$, we have:

- 5-box: $\mathrm{BN}(5) = 1$
- 6-box: $\mathrm{BN}(6) = 255 - 4\mathrm{BN}(5) = 251$
- 7-box: $\mathrm{BN}(7) = 255^2 - 4\mathrm{BN}(6) - 6\mathrm{BN}(5) = 64015$
- 8-box: $\mathrm{BN}(8) = 255^3 - 4\mathrm{BN}(7) - 6\mathrm{BN}(6) - 4\mathrm{BN}(5) = 16323805$

The number of trails with i active S-boxes is $\binom{8}{i}255BN_i127^i$. The total number of trails is 2.8×10^{26}.

B Derivation of (15) and (16)

We generalize the sampling model introduced in Section 5.3, approximating the different blurred conditions as being independent. The space sampled is now the set of β-component vectors where each of the components can take any nonzero value in $\mathrm{GF}(2^8)$. There are 255^β such vectors. A good selection is one in which the first component satisfies the first condition, the second component satisfies the second condition and so on. There are 127^β such vectors. Denoting by x_{it} the event that inner trail i satisfies condition t, we obtain:

$$\langle \mathrm{N_t} \rangle = \sum_{i=1}^{N} \langle x_i \rangle = \sum_{i=1}^{N} \langle x_{i1} \rangle \langle x_{i2} \rangle \cdots \langle x_{i\beta} \rangle = N \left(\frac{n}{n+m} \right)^\beta$$

The variance satisfies

$$\mathrm{var}(\mathrm{N_t}) = \sum_{i=1}^{N} \mathrm{var}(x_i) + \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} \mathrm{Cov}(x_i, x_j).$$

Since x_i takes only the values 0, 1, x_i is a Bernoulli variable, and

$$\mathrm{var}(x_i) = \langle x_i \rangle (1 - \langle x_i \rangle) \tag{18}$$

$$\mathrm{Cov}(x_i, x_j) = \langle x_i x_j \rangle - \langle x_i \rangle \langle x_j \rangle \tag{19}$$

$$\langle x_i \rangle = \left(\frac{n}{n+m} \right)^\beta . \tag{20}$$

Since two trails of the same bundle differ in the value of each of their components, we have:

$$\langle x_i x_j \rangle = \left(\frac{n(n-1)}{(n+m)(n+m-1)} \right)^\beta . \tag{21}$$

Putting everything together results in (16).

C Plateau trail distributions

Because of the large number of trails, the entries in Table 4 can clearly not be computed by checking the height of each trail individually. Let $Q = (a, b, d, e)$ be a trail over the super box. Firstly, remember that $\mathrm{height}(Q) = \dim(W_Q)$ with $W_Q = V_{(a,b)} \cap \mathrm{M_{ci}}(U_{(d,e)})$. Definition 2 implies that for single S-box differentials, $V_{(a_i,b_i)} = \{b_i\}$. Hence, for these S-boxes, the value of a_i doesn't matter. Similar for the difference e_i of single S-box differentials in the second round. Secondly, let $Q' = (a', b, d, e')$ be the trail with the same inner trail $|b, d|$ as Q and in which all S-box differentials are double S-box differentials. Since for these differentials $\{b_i\} \subset V_{(a'_i,b_i)}$, we have that $V_{(a,b)} \subseteq V_{(a',b)}$. Similarly $U_{(d,e)} \subseteq U_{(d',e)}$ and hence

$$W_Q \subseteq W'_Q . \tag{22}$$

Consequently, it is only needed to check the height of each inner trail with (a, e) chosen such that all active S-boxes have weight 6, and then to evaluate the effect of increasing the weight of the active S-boxes by one. For this last step, only one out of the 126 possible differences a_i, respectively e_i, needs to be tried for each active S-box.

Table 4 shows that there are in total 2^{20} flat trails: those with weight 30 and height 2, and those with weight 31 and height 1. The trails for which the ratio is 2^{25} are the

Table 4. Number of trails (binary logarithm) per number of active S-boxes, weight and height.

box	trail			height		
weight	weight	1	2	3	4	5
5	30	—-	12.6	12.6	10.8	6.7
	31	20.0	22.0	21.6	19.3	14.9
	32	29.0	30.0	29.2	26.5	21.7
	33	36.5	36.9	35.8	32.6	27.4
	34	42.8	42.9	41.3	37.8	31.2
	35	47.6	47.5	45.4	41.5	—-
6	36	20.7	16.1	9.5	4.2	1.0
	37	30.3	25.2	18.3	13.1	10.0
	38	38.6	33.0	25.8	20.8	17.5
	39	46.0	39.9	32.3	27.8	23.9
	40	52.6	46.1	38.0	34.0	28.9
	41	58.3	51.3	42.8	39.2	—-
	42	62.7	55.2	46.6	42.9	—-
7	42	26.9	20.8	13.2	3.0	—-
	43	36.7	30.4	22.6	12.2	—-
	44	45.3	38.7	30.7	20.1	—-
	45	53.0	46.2	37.8	27.0	—-
	46	60.0	52.8	44.1	32.9	—-
	47	66.3	58.7	49.5	37.5	—-
	48	71.7	63.5	53.4	—-	—-
	49	75.9	66.8	—-	—-	—-
8	48	31.9	25.9	18.5	10.0	0.0
	49	41.9	35.7	28.1	19.3	9.0
	50	50.7	44.3	36.4	27.3	16.5
	51	58.7	52.0	43.8	34.2	22.9
	52	66.0	59.0	50.4	40.2	27.9
	53	72.7	65.2	56.0	44.9	—-
	54	78.7	70.6	60.4	—-	—-
	55	83.7	74.8	—-	—-	—-
	56	87.9	—-	—-	—-	—-
total		87.9	74.9	60.5	45.4	31.7

trails with weight 56 and height 1, or weight 55 and height 2, or weight 54 and height 3, or weight 53 and height 4, or weight 52 and height 5. Only for 1/18 of the trails this ratio is smaller than 2^{25}. The table shows also that it is easy to find trails Q_1, Q_2 with $\mathrm{EDP}(Q_1) < \mathrm{EDP}(Q_2)$ and $\mathrm{height}(Q_1) > \mathrm{height}(Q_2)$.

D Distributions of the number of trails per differential for $\beta = 1$

We have experimentally verified the distributions of the number of trails per differential for all 16 combinations of α and β. For the combination of (α, β) equal to $(1, 1)$, $(2, 1)$, $(3, 1)$, $(4, 1)$ and $(1, 2)$ we were able to do this exhaustively, covering all possible cases. As a side result we found for these values of (α, β) the minimum and maximum values for the number of trails per differential, listed in Table 5.

For the other values of (α, β), the number of combinations becomes too large to compute exhaustively. Still, our sampling experiments confirm the shape predicted by formulas

Table 5. Minimum and maximum number of trails in differentials with 5 active S-boxes given (α, β)

(α, β)	minimum	maximum
$(1, 1)$	48	82
$(2, 1)$	14	48
$(3, 1)$	3	29
$(4, 1)$	0	15
$(1, 2)$	10	56

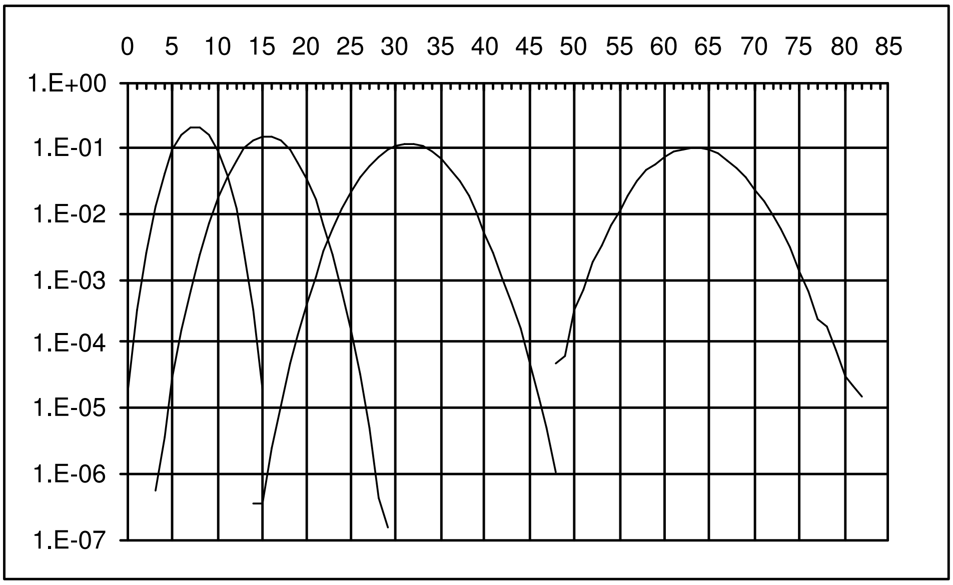

Fig. 1. distributions of the number of trails per differential for $\beta = 1$ and for α ranging from 4 (leftmost) to 1 (rightmost).

(15) and (16). As α and β grow, the mean and variance of the distributions shrink. Clearly, the majority of differentials with 5 active S-boxes and $\alpha = 1$ and $\beta = 1$ have more trails than any differential with 5 active S-boxes where $\alpha + \beta$ has a higher value.

Figure 1 depicts the four distributions for $\beta = 1$ on a logarithmic scale. The distributions appear as slightly skewed parabolas, which is the typical shape of hypergeometric distributions.

E The effect of L on the distribution of trails over differentials

In this section we study a variant of the AES super box with L removed. Consider the number of trails in a bundle $B(u)$ of a given differential (a, e). The absence of L makes that the conditions imposed by the first-round S-box differentials are sharp instead of blurred, see Section 5.3. The number of trails can be characterized by a single parameter α ranging from 1 to the number of active S-boxes in the differential:

$$\alpha = \dim\{x_0^{-1}, x_1^{-1}, x_2^{-1}, x_3^{-1}, y_0^{-1}, y_1^{-1}, y_2^{-1}, y_3^{-1}\}. \qquad (23)$$

with $x = u * a$ and $y = (\mathrm{M_c} u) * e$. The number of trails is $2^{8-\alpha} - 1$. This means that for differentials with 5 active S-boxes, the number of trails is completely determined by α. for differentials with more active S-boxes, we can compute from α-profiles as in Section 5.6. The α-profile of a differential completely determines its number of trails.

The maximum EDP occurs for trails with 5 active S-boxes and $\alpha = 1$. Since the condition of all active S-boxes coincide, the number of trails is 127 and the double S-box differentials are in the same trail. This gives an EDP value of $2^{-30} + 126 \times 2^{-35} = 19.75 \times 2^{32}$. Hence the introduction of L reduces the maximum EDP value by 33% to 13.25.

As said, the differentials with this EPD value are the differentials with 5 active S-boxes and $\alpha = 1$. For a given activity pattern with 5 active S-boxes, there are 255 such differentials. For example, the ones for activity pattern $(1000; 1111)$ are those of the form $(z, 0, 0, 0; z/2, z, z, z/3)$. As there are in total 65 5-box activity patterns, the total number of differentials with the maximum EDP is 16575. The introduction of L reduces the number of differentials that have the maximum EDP value to 12.