Computing (2009) 85:85–104
DOI 10.1007/s00607-009-0034-y

Computational aspects of the expected differential probability of 4-round AES and AES-like ciphers

Joan Daemen · Mario Lamberger · Norbert Pramstaller · Vincent Rijmen · Frederik Vercauteren

Received: 31 October 2008 / Accepted: 27 January 2009 / Published online: 9 May 2009

Abstract In this paper we study the security of the Advanced Encryption Standard (AES) and AES-like block ciphers against differential cryptanalysis. Differential cryptanalysis is one of the most powerful methods for analyzing the security of block ciphers. Even though no formal proofs for the security of AES against differential cryptanalysis have been provided to date, some attempts to compute the maximum expected differential probability (MEDP) for two and four rounds of AES have been presented recently. In this paper, we will improve upon existing approaches in order to derive better bounds on the EDP for two and four rounds of AES based on a slightly simplified S-box. More precisely, we are able to provide the complete distribution of the EDP for two rounds of this AES variant with five active S-boxes and methods to improve the estimates for the EDP in the case of six active S-boxes.

Keywords Cryptography · Differential cryptanalysis · AES · Differential probability

Mathematics Subject Classification (2000) 94A60 · 11T71

J. Daemen
STMicroelectronics Belgium, Zaventem, Belgium

M. Lamberger (✉) · N. Pramstaller · V. Rijmen
IAIK, Graz University of Technology, Graz, Austria
e-mail: mario.lamberger@iaik.tugraz.at

V. Rijmen · F. Vercauteren
ESAT/COSIC, K.U. Leuven, Louvain, Belgium

1 Introduction

Although symmetric key primitives such as block ciphers are ubiquitously deployed throughout all cryptosystems, they do not come with a formal proof of security. In modern designs the resistance against differential [2] and linear [14] cryptanalysis is always a design objective. An important class of block ciphers are the so-called substitution-permutation networks (SPNs) of which the Advanced Encryption Standard (AES) [4] is the most prominent example.

In this paper, we want to study the distribution of the probability of differentials over two rounds of the AES. Our objective is to use this close study of the two round probabilities to obtain tighter bounds for the probability of differentials over four rounds of the AES. Bounds on the expected differential probability of 4-round differentials have earlier been investigated in [9,10,16,17].

Tight bounds on the probability of differentials over four rounds of AES are relevant in practice because there are several constructions which use four rounds of AES as a building block: the message authentication codes Pelican MAC [5], PC-MAC [15] and the stream cipher LEX [3].

Our detailed study of the full distribution of the expected differential probability of differentials over two rounds of a simplified variant of AES results in a better understanding of the interaction between the different components of the AES.

The paper is organized as follows. Sections 2 and 3 give an overview of the terminology used and of basic differential properties of the AES S-box. Section 4 introduces the main object of our study, the super box, and Sect. 5 gives an even more general structure. Sections 6–8 then systematically derive results on the distribution properties of the super box based on a simplified S-Box and in Sect. 9, we conclude.

2 Characteristics, differentials and probabilities

2.1 Notation

Let us denote by $\mathbb{F}_2$ the finite field with two elements. In all of the following we will make extensive use of the correspondence between the set $\{0, 1\}^n$, the vector space $\mathbb{F}_2^n$ and the finite field $\mathbb{F}_{2^n}$. Therefore, we will denote the bit-wise XOR between two elements simply by the addition "+".

The finite field $\mathbb{F}_{2^8}$ will play a crucial role in our discussion because it lies at the core of the AES substitution box [4]. We use the representation $\mathbb{F}_{2^8} \simeq \mathbb{F}_2[\theta]/(\theta^8 + \theta^4 + \theta^3 + \theta + 1)$. Whenever we use a finite field inversion in this paper, we implicitly mean inversion in $\mathbb{F}_{2^8}$ with respect to the above irreducible polynomial.

In our discussion we will encounter a superstructure that works on 32-bit values. Due to this superstructure, it will be convenient to consider these 32-bit values as elements of the vector space $\mathbf{V} = (\mathbb{F}_{2^8})^4$.

2.2 Terminology

A *differential* [12] of a function $f : \mathbb{F}_2^n \to \mathbb{F}_2^n$ is a pair $(a, b) \in \mathbb{F}_2^n \times \mathbb{F}_2^n$ such that $f(x + a) = f(x) + b$ for some $x \in \mathbb{F}_2^n$. We call a the *input difference* and b the *output*

difference. The *differential probability* $\mathrm{DP}_f(a,b)$ of a differential (a,b) with respect to $f(x)$ is defined as

$$\mathrm{DP}_f(a,b) = 2^{-n}\#\left\{x \in \mathbb{F}_2^n | f(x+a) = f(x)+b\right\}.$$

The *difference table* of a function f is the matrix containing all the differential probabilities

$$DT_f[a,b] = \mathrm{DP}_f(a,b).$$

Where it is clear from the context which function f is meant, we will often drop it from the notation.

If f is a function parameterized by a key k, we can also define the parameterized differential probability $\mathrm{DP}[k](a,b)$ in a straightforward way. Then, the *expected differential probability* (*EDP*) of a differential (a,b) is defined as the mean value of $\mathrm{DP}[k](a,b)$:

$$\mathrm{EDP}(a,b) = \mathbb{E}(\mathrm{DP}[k](a,b);k) = 2^{-|\mathcal{K}|}\sum_{k\in\mathcal{K}}\mathrm{DP}[k](a,b).$$

Here, k is assumed to be a uniformly distributed random variable taking values in $\mathcal{K}$. So for a function parameterized by a key, the difference table consists of the values

$$DT_f[a,b] = \mathrm{EDP}_f(a,b).$$

Let $B[k](x)$ denote a function composed of r steps $f^i[k^i](x)$ parameterized by r keys $k^1, k^2, \ldots, k^r \in \mathbb{F}_2^n$:

$$B[k](x) = \left(f^r[k^r] \circ \cdots \circ f^1[k^1]\right)(x).$$

A *characteristic* through $B[k](x)$ is a vector $Q = (b^0, b^1, \ldots, b^r)$ with $b^i \in \mathbb{F}_2^n$ for $i = 0, \ldots, r$. A characteristic $Q = (b^0, b^1, \ldots, b^r)$ is *in* a differential (a,b) if $b^0 = a$ and $b^r = b$. If we now consider the following set of equations

$$\begin{aligned} f^1[k^1](x+b^0) &= f^1[k^1](x) + b^1 \\ &\vdots \\ \left(f^R[k^r] \circ \cdots \circ f^1[k^1]\right)(x+b^0) &= \left(f^R[k^r] \circ \cdots \circ f^1[k^1]\right)(x) + b^r, \end{aligned} \tag{1}$$

then the parameterized differential probability $\mathrm{DP}_B[k](Q)$ of a characteristic Q with respect to $B[k](x)$ is defined as

$$\mathrm{DP}[k](Q) = 2^{-n}\#\left\{x \in \mathbb{F}_2^n \mid x \text{ satisfies (1)}\right\}.$$

It is well-known [12] that for Markov ciphers we have

$$\mathrm{DP}[k](a,b) = \sum_{Q \in (a,b)} \mathrm{DP}[k](Q) \tag{2}$$

$$\mathrm{EDP}(a,b) = \sum_{Q \in (a,b)} \mathrm{EDP}(Q). \tag{3}$$

3 The AES S-box

3.1 Definition

The AES S-box operates on $\mathbb{F}_{2^8}$ and can be described as

$$S(x) = L(x^{-1}) + q, \tag{4}$$

Here x^{-1} denotes the multiplicative inverse of x in $\mathbb{F}_{2^8}$, extended by 0 being mapped to 0. L is a linear transformation over $\mathbb{F}_2$ and q a constant. Note that L is not linear over $\mathbb{F}_{2^8}$ and can be expressed as a so-called *linearized polynomial* (cf. [13]). In this paper, we will study in detail the differential properties of an AES variant where the S-box is simplified to $S(x) = x^{-1}$ [with $S(0) = 0$]. We will call this the *naive* S-box S_n.

3.2 Properties

Since $S(x)$ and $S_\mathrm{n}(x)$ are invertible, we trivially obtain for both maps that $\mathrm{DP}(0,0) = 1$ and $\mathrm{DP}(a,0) = \mathrm{DP}(0,b) = 0$ for all $a, b \in \mathbb{F}_{2^8}^*$. For each nonzero input difference a the corresponding row of the difference tables of $S(x)$ and $S_\mathrm{n}(x)$ contains one time 2^{-6}, 126 times 2^{-7} and 129 times 0. The same holds true for each nonzero output difference and the corresponding columns of the difference tables.

From [6] (see also [1]) we have the following neat result how to characterize the distribution of $\mathrm{DP}_{S_\mathrm{n}}(a,b)$ when a and b are different from 0:

$$\mathrm{DP}_{S_\mathrm{n}}(a,b) = 2^{-8} \#\{x \in \mathbb{F}_{2^8} | x^{-1} + (x+a)^{-1} = b\}$$

If $x = a$ or $x = 0$ are not solutions to the equation $x^{-1} + (x+a)^{-1} = b$, then there is a bijection between the solutions and those of $y^2 + y + (ab)^{-1} = 0$. Such a quadratic equation in $\mathbb{F}_{2^8}$ has 2 solutions if $\mathrm{Tr}\left((ab)^{-1}\right) = 0$ and no solution otherwise. Here $\mathrm{Tr}(\cdot)$ denotes the trace map of $\mathbb{F}_{2^8}$ over $\mathbb{F}_2$.

If $ab = 1$, it is easy to see that the equation $x^{-1} + (a+x)^{-1} = a^{-1}$ has the four solutions $\{0, a, a\nu, a\nu^2\}$ with $\nu \in \mathbb{F}_{2^2} \setminus \mathbb{F}_2$ since such a ν satisfies $\nu^2 + \nu = 1$. Summarizing, we have:

$$\mathrm{DP}_{S_\mathrm{n}}(a,b) = \begin{cases} 2^{-6} & \text{iff } a = b^{-1} \\ 2^{-7} & \text{iff } a \neq b^{-1} \quad \text{and} \quad \mathrm{Tr}\left((ab)^{-1}\right) = 0 \\ 0 & \text{else} \end{cases} \tag{5}$$

As an easy consequence it follows that for all $c \in \mathbb{F}_{2^8}^*$, $\mathrm{DP}_{S_n}(a, b)$ is equal to $\mathrm{DP}_{S_n}(ac, bc^{-1})$. Because a differential (a, b) with $a = b^{-1}$ has two times the DP of the other differentials (having $DP > 0$), we also call it a *double differential*.

4 Super boxes

4.1 Definition

A *super box* is the minimal structure where the non-linear and the linear part of an SPN, that influence the differential probability, interact. A super box operates on $\mathbf{V} = \mathbb{F}_{2^8}^4$, mapping an array $a = [a_1, a_2, a_3, a_4]$ to an array $e = [e_1, e_2, e_3, e_4]$ and taking an array $k = [k_1, k_2, k_3, k_4]$ as key. The AES super box consists of the sequence of four transformations:

Substitution: $b_i = S(a_i)$
Mixing: $c = \mathbf{M}b$ with $\mathbf{M}$ the 4×4 matrix over $\mathbb{F}_{2^8}$ used in MixColumns
Key addition: $d = c + k$ with k the round key
Substitution: $e_i = S(d_i)$

If we consider two AES rounds, swap the steps ShiftRows and SubBytes in the first round, and remove the linear transformations before the first SubBytes transformation and after the second SubBytes transformation, then we obtain a map that can also be described as 4 parallel instances of the AES super box. It is well-known that this map has the same differential properties as two rounds of AES. Similarly, we define the naive super box by using in both Substitution steps S_n instead of S.

4.2 Properties

The matrix $\mathbf{M}$ in the Mixing step is given by:

$$\mathbf{M} = \begin{pmatrix} \theta & 1+\theta & 1 & 1 \\ 1 & \theta & 1+\theta & 1 \\ 1 & 1 & \theta & 1+\theta \\ 1+\theta & 1 & 1 & \theta \end{pmatrix} \tag{6}$$

where θ was the element used to define $\mathbb{F}_{2^8}$. $\mathbf{M}$ is sometimes called an *MDS matrix* because the associated linear code with generator matrix $G = [\mathrm{I}\ \mathbf{M}]$ is an MDS code (with dimension 4, length 8 and minimum distance 5).

The matrix $\mathbf{M}$ has optimal diffusion properties [4]. For example, let $\mathrm{w}(a)$ denote the number of non-zero values in $a = [a_1, a_2, a_3, a_4]$. The (differential) *branch number* $\mathcal{B}$ of a linear map $l(x)$ is defined as

$$\begin{aligned} \mathcal{B} &= \min_{a, b \neq a} \{\mathrm{w}(a + b) + \mathrm{w}(l(a) + l(b))\} \\ &= \min_{a \neq 0} \{\mathrm{w}(a) + \mathrm{w}(l(a))\}. \end{aligned}$$

It can be shown that the branch number of the Mixing map defined by $\mathbf{M}$ equals the minimum distance of the associated linear code. Since the code associated to $\mathbf{M}$ has maximal distance, the Mixing step has the highest possible branch number $\mathcal{B} = 5$ (cf. [4]).

4.3 The MEDP of the AES Super Box

For differentials (a, e) over a super box, we write

$$\mathrm{EDP}_{\text{super box}}(a, e) := \mathrm{EDP}_{32}(a, e).$$

The *maximum expected differential probability* (*MEDP*) of a super box is defined as

$$\mathrm{MEDP}_{32} = \max_{a, e \neq 0} \mathrm{EDP}_{32}(a, e).$$

Park et al. [17] give a bound on the MEDP of generalized super box structures. They consider super boxes with Substitution layers constructed from an arbitrary number of parallel S-boxes, which don't need to be all the same, and a Mixing layer using an arbitrary matrix. We state here their theorem for the less general case where each Substitution layer consists of 4 applications of the same S-box, and where the Mixing layer has branch number 5. We introduce the following notation:

$$\mathrm{E}_{5r}(x) = \sum_{y} (\mathrm{DP}(x, y))^5, \tag{7}$$

$$\mathrm{E}_{5c}(x) = \sum_{y} (\mathrm{DP}(y, x))^5. \tag{8}$$

Observe that the first quantity is the running sum of the fifth powers of all the entries in a *row* of the difference table, whereas the second quantity is obtained by summing the fifth powers of all the entries in a *column*. Now we can state the simplified form of [17, Theorem 1] as follows:

Theorem 1 *The* MEDP *of a differential over a super box is bounded by*

$$\mathrm{MEDP}_{32} \leq \max_{x} \{\max\{\mathrm{E}_{5r}(x), \mathrm{E}_{5c}(x)\}\}.$$

As stated already in Section 3.2, both for the AES S-box and the naive S-box, the rows and the columns of the difference tables are all equal up to a reordering of the values. Hence, the bound of Theorem 1 reduces to:

$$\mathrm{MEDP}_{32} \leq \sum_{y} (\mathrm{DP}(1, y))^5 = 2^{-30} + 126 \times 2^{-35} \approx 1.23 \times 2^{-28}.$$

For a naive super box, this bound is tight. It is reached for instance in the differential $([1, 0, 0, 0]; [\theta^{-1}, 1, 1, (1 + \theta)^{-1}])$. For an AES super box, this bound is not tight. Keliher and Sui [11] computed that $\mathrm{MEDP}_{32} \approx 1.656 \times 2^{-29}$.

Already earlier, Hong et al. [8] proved the bound

$$\mathrm{MEDP}_{32} \leq \left(\max_{x \neq 0,\ y} \mathrm{DP}(x, y) \right)^4. \tag{9}$$

This bound can be easily derived as a corollary of Theorem 1 since for all x we have

$$\sum_y \mathrm{DP}(x, y) = 1.$$

Hence, we get for all x:

$$\sum_y (\mathrm{DP}(x, y))^5 \leq \left(\max_y \mathrm{DP}(x, y) \right)^4 \sum_y \mathrm{DP}(x, y) = (\max_y \mathrm{DP}(x, y))^4.$$

Filling out $\max_y \mathrm{DP}(x, y) = 2^{-6}$ in (9) gives $\mathrm{MEDP}_{32} \leq 2^{-24}$. The results given in this section motivate the following conjecture:

Conjecture 1 For any number of rounds s, the MEDP of the AES reduced to s rounds is smaller than the MEDP of the AES variant using the naive S-box S_{n}, reduced to s rounds.

This is equivalent to stating that the real AES is more resistant against differential cryptanalysis than the variant using the naive S-box. Therefore, we focus on the naive S-box S_{n}.

5 Mega boxes

5.1 Definition

A *mega box* takes as input an array consisting of four elements of **V**. Each element is called a *column*. A mega box produces an output of the same dimensions and takes as key three arrays k^1, k^2, k^3, each consisting of 4 columns. The AES mega box consists of the sequence of four transformations:

Substitution: each of the 4 columns of the input is substituted by applying the AES super box parameterized by the key k^1
Mixing: the linear AES transformations ShiftRows, MixColumns and again ShiftRows are applied
Key addition: key k^2 is added by means of the binary XOR operation
Substitution: each column is substituted by applying the AES super box parameterized by the key k^3

If we consider four AES rounds, swap the steps ShiftRows and SubBytes in the first and the third round, remove the linear transformations before the first SubBytes transformation and after the fourth SubBytes transformation, then we obtain the same

map as the AES mega box. Again, this map has the same differential properties as four rounds of AES. (This is in fact the basis for the proof that characteristics over four rounds of AES have at least 25 active S-boxes [4].) Similarly, we define the naive mega box by using in both substitution steps the naive super box.

5.2 The MEDP of the AES mega box

The currently used bound for differentials over AES reduced to four rounds is the one that follows from applying the Hong et al. bound (9) to the mega box (cf. [11]):

$$\mathrm{MEDP}_{128} \leq (\mathrm{MEDP}_{32})^4 \approx 1.881 \times 2^{-114}$$

If the naive S-box is used, then this bound increases to 1.16×2^{-111}. The contribution of this paper is that we make progress towards a tighter bound for the case of the naive S-box by using the the same strategy that Park et al. used for two rounds, namely we compute E_{5r} and E_{5c} for the super box to derive tighter bounds on MEDP_{128}. In order to compute these values, we need to determine the entries of the difference table of the super box. The entries will be computed, or bounded, by extending the observations made in [6]. In fact, it is sufficient to determine the distribution of the entries over the rows and columns of the difference table. We will use this fact to reduce the computational complexity.

6 Characteristics and bundles

Applying (3), we get

$$\mathrm{EDP}_{32}(a, e) = \sum_{b,c,d} \mathrm{EDP}_{\mathrm{Sub.}}(a, b)\mathrm{EDP}_{\mathrm{Mix.}}(b, c)\mathrm{EDP}_{\mathrm{Key.}}(c, d)\mathrm{EDP}_{\mathrm{Sub.}}(d, e).$$

Since the Mixing step is linear,

$$\mathrm{EDP}_{\mathrm{Mix.}}(b, c) = 1 \Leftrightarrow c = \mathbf{M}b,$$

and zero otherwise. Similarly we have that $\mathrm{EDP}_{\mathrm{Key.}}(c, d)$ equals 1 if $c = d$, and equals zero otherwise. We obtain that

$$\mathrm{EDP}_{32}(a, e) = \sum_{b} \mathrm{EDP}_{32\,\mathrm{Sub.}}(a, b)\mathrm{EDP}_{32\,\mathrm{Sub.}}(\mathbf{M}b, e) \tag{10}$$

$$= \sum_{b} \prod_{i=1}^{4} \mathrm{DP}(a_i, b_i) \prod_{i=1}^{4} \mathrm{DP}((\mathbf{M}b)_i, e_i). \tag{11}$$

For a super box, the characteristics in a differential (a, e) are defined uniquely by a, e and the value of b. Further, if $a_i = 0$ then $\mathrm{DP}(a_i, 0) = 1$ and $\mathrm{DP}(a_i, b_i) = 0, \forall b_i \neq 0$.

Table 1 Number of bundles in any given differential (a, e)

w(a, e)	Number of bundles in (a, e)
5	1
6	251
7	64,015
8	16,323,805

We say that position i in a is *active* if $a_i \neq 0$. The *activity pattern* of a has a single bit for each position in a. The bit is 1 when the position is active, and 0 when it is not active. The activity pattern of the differential (a, e) is the pair of the activity patterns of a and e. The weight w(a, e) of a differential (a, e) is defined as w(a) + w(e).

In order to compute $\mathrm{EDP}_{32}(a, e)$ of a differential (a, e) we need to consider only the characteristics defined by values b such that the activity pattern of $(b, \mathbf{M}b)$ equals the activity pattern of (a, e). Observe that scalar multiplication doesn't change the activity pattern of a vector. Furthermore, MixColumns is linear over $\mathbb{F}_{2^8}$: $\gamma(\mathbf{M}b) = \mathbf{M}(\gamma b)$. Hence, the activity pattern of $(b, \mathbf{M}b)$ equals the activity pattern of $(\gamma b, \gamma\mathbf{M}b)$ for all $\gamma \in \mathbb{F}_{2^8}^*$. It follows that if $(a, b, \mathbf{M}b, e)$ is a characteristic through the AES super box, then $(a, \gamma b, \gamma\mathbf{M}b, e)$ is a characteristic through the AES super box for all $\gamma \in \mathbb{F}_{2^8}^*$.

We define a *bundle of characteristics* as follows.

Definition 1 *Let b be an arbitrary nonzero vector in* $\mathbf{V}$. *The corresponding* bundle of characteristics $B(b)$ *in some differential* (a, e)*, is the set of 255 characteristics defined as follows*:

$$B(b) = \{(a, \gamma b, \gamma\mathbf{M}b, e) | \gamma \in \mathbb{F}_{2^8}^*\}.$$

For any differential (a, e), the set of characteristics in (a, e) can be partitioned into a number of bundles of characteristics. A characteristic in a bundle $B(b)$ of the differential (a, e) is uniquely identified by the value of γ. The number of bundles in (a, e) depends on w(a, e), see Table 1.

7 Counting characteristics

7.1 S-box

We define $\mathrm{N}_8(a, b)$ as:

$$\mathrm{N}_8(a, b) = 1 \Leftrightarrow \mathrm{DP}(a, b) > 0$$
$$\mathrm{N}_8(a, b) = 0 \Leftrightarrow \mathrm{DP}(a, b) = 0.$$

Then we have for both S and S_n that

$$\mathrm{N}_8(a, b) \approx 2^7 \cdot \mathrm{DP}(a, b). \tag{12}$$

Approximation (12) is exact in 255 out of 256 cases. In the remaining cases we have $\mathrm{N}_8(a,b) = 2^6 \cdot \mathrm{DP}(a,b)$ due to the double differentials. Furthermore, because of (5) we have the following nice description for the quantity $\mathrm{N}_8(a,b)$ in the case of the naive S-box:

$$\mathrm{N}_8(a,b) = 1 - \mathrm{Tr}\,((ab)^{-1}).$$

7.2 Super box

Let us define the number of characteristics $Q \in (a,e)$ having $\mathrm{EDP}_{32}(Q) > 0$ by:

$$\mathrm{N}_{32}(a,e) = \#\{Q \in (a,e) | \mathrm{EDP}_{32}(Q) > 0\}.$$

Similarly as in the derivation of (11) we obtain that

$$\mathrm{N}_{32}(a,e) = \sum_b \prod_{i=1}^{4} \mathrm{N}_8(a_i,b_i) \prod_{i=1}^{4} \mathrm{N}_8((\mathbf{M}b)_i, e_i).$$

The EDP of a characteristic Q in (a,e) equals

$$\mathrm{EDP}_{32}(Q) = 2^{-7\mathrm{w}(a,e)} \times 2^t$$

where t is the number of double differentials in Q. Hence, we can write the EDP of the differential (a,b) as:

$$\mathrm{EDP}_{32}(a,e) = \sum_{Q \in (a,e)} \mathrm{EDP}_{32}(Q) = (\mathrm{N}_{32}(a,e) + \epsilon(a,e)) 2^{-7\mathrm{w}(a,e)}, \tag{13}$$

where $\epsilon(a,e)$ is a correction term that compensates for the error made in (12) due to the characteristics with one or more double differentials.

8 The difference table for the super box based on the naive S-box

8.1 General observations

For all non-zero c we have [4]:

$$\mathrm{DP}(a,b) = \mathrm{DP}(ca, c^{-1}b) \tag{14}$$

$$\mathrm{EDP}_{32}(a,e) = \mathrm{EDP}_{32}(ca, ce). \tag{15}$$

In fact, for any differential over AES reduced to r rounds and using the naive S-box, we have

$$\mathrm{EDP}_{128}(a,e) = \mathrm{EDP}_{128}(ca, c^{(-1)^r} e). \tag{16}$$

8.2 The case $\mathrm{w}(a) = 1$

8.2.1 Computing $\mathrm{N}_{32}(a, e)$

Due to the properties of $\mathbf{M}$ we know that $\mathrm{EDP}_{32}(a, e) > 0$ only if $\mathrm{w}(e) = 4$. Hence, a row in the difference table with $\mathrm{w}(a) = 1$ contains at most 255^4 entries different from zero. Further, because of the rotational symmetry of MixColumns and because of (15), all the rows with $\mathrm{w}(a) = 1$ have the same distribution of entries. Without loss of generality we can assume that $a_2 = a_3 = a_4 = 0$. Then,

$$\mathrm{N}_{32}(a, e) = \sum_{b_1} \mathrm{N}_8(a_1, b_1) \prod_{i=1}^{4} \mathrm{N}_8((\mathbf{M}b)_i, e_i).$$

The characteristics in (a, e) form one bundle. Hence, we are allowed to choose $b = [1, 0, 0, 0]$, $d = \mathbf{M}b = [\theta, 1, 1, 1 + \theta]$ and write:

$$\begin{aligned}\mathrm{N}_{32}(a, e) &= \sum_{\gamma \in \mathbb{F}_{2^8}^*} \mathrm{N}_8(a_1, \gamma) \prod_i \mathrm{N}_8(\gamma d_i, e_i) \\ &= \sum_{\gamma \in \mathbb{F}_{2^8}^*} (1 - \mathrm{Tr}\left((\gamma a_1 b_1)^{-1}\right)) \prod_i (1 - \mathrm{Tr}\left((\gamma d_i e_i)^{-1}\right)).\end{aligned}$$

This is equivalent to stating that $\mathrm{N}_{32}(a, e)$ equals the number of non-zero solutions for the following set of equations in γ^{-1}:

$$\begin{aligned}\mathrm{Tr}\left(\gamma^{-1}(a_1)^{-1}\right) &= 0, \\ \mathrm{Tr}\left(\gamma^{-1}(\theta e_1)^{-1}\right) &= 0, \\ \mathrm{Tr}\left(\gamma^{-1}(e_2)^{-1}\right) &= 0, \\ \mathrm{Tr}\left(\gamma^{-1}(e_3)^{-1}\right) &= 0, \\ \mathrm{Tr}\left(\gamma^{-1}((1 + \theta)e_4)^{-1}\right) &= 0.\end{aligned} \tag{17}$$

The non-zero solutions γ^{-1} to (17) can be described as the trace-orthogonal elements lying in the vector space spanned by the vectors in the (multi-)set

$$\mathrm{V} = \{a_1^{-1}, (\theta e_1)^{-1}, (e_2)^{-1}, (e_3)^{-1}, ((1 + \theta)e_4)^{-1}\}.$$

In the following, let $\langle \mathrm{V} \rangle$ denote the span of V as a $\mathbb{F}_2$ vector space. Therefore, the number of such non-zero γ^{-1} is exactly $2^{8-\dim(\langle \mathrm{V} \rangle)} - 1$.

When we compute the values of $\mathrm{N}_{32}(a, e)$ for a fixed a_1 and all 255^4 nonzero e, then the vector a_1^{-1} remains fixed and the vectors $(\theta e_1)^{-1}, (e_2)^{-1}, (e_3)^{-1},$

Table 2 The distribution of $N_{32}(a, e)$ for $w(a) = 1$

d	$Z(d)$	$N_{32}(a, e)$
1	1	127
2	10,160	63
3	5,760,720	31
4	412,723,584	15
5	3,809,756,160	7

$((1 + \theta)e_4)^{-1}$ take all possible nonzero values. Hence, we can compute the distribution of the $N_{32}(a, e)$ values as follows.

For a given d let $Z(d)$ be the number of $e = [e_1, e_2, e_3, e_4]$ resulting in a value $N_{32}(a, e) = 2^{8-d} - 1$. Let $E(m, n, d)$ be the number of $m \times n$ matrices ($m \leq n$) with rank d over $\mathbb{F}_2$. We have

$$E(m, n, d) = \prod_{i=0}^{d-1} (2^n - 2^i) \cdot \binom{m}{d}_2 = \prod_{i=0}^{d-1} \frac{(2^n - 2^i) \cdot (2^m - 2^i)}{2^d - 2^i},$$

where $\binom{m}{d}_2$ denotes the q-binomial coefficient with $q = 2$ (see for example [7,13]). Except for the fact that a_1^{-1} is fixed and that the e_i cannot be zero, this is almost what we want. We therefore have to employ an inclusion–exclusion principle to arrive at:

$$Z(d) = \frac{1}{255} \cdot \sum_{j=0}^{5-d} (-1)^j \binom{5}{j} E(5 - j, 8, d) \tag{18}$$

The values of $Z(d)$ for all possible dimensions d are given in Table 2.

8.2.2 Computing $\epsilon(a, e)$

The correction term $\epsilon(a, e)$ for a differential $([a_1, 0, 0, 0]; [e_1, e_2, e_3, e_4])$ can be computed by the following reasoning.

If we consider the 255 characteristics in one bundle, then we see that for each of the active S-boxes either the input difference or the output difference is fixed, while the other difference takes all nonzero values once. In terms of the difference table, we can say that all the values of a row, respectively column occur exactly once. Hence, for every active S-box there are 129 entries equal to zero, 126 entries equal to 2^{-7} and one entry equal to 2^6. The last one is the double differential.

From (5) and (17) we obtain that the double differentials occur in the characteristics corresponding to the values $\gamma \in V$. If $\dim(\langle V \rangle) = 1$, then $a_1 = \theta e_1 = e_2 = e_3 = (1 + \theta)e_4$, and therefore the double differentials of the 5 active S-boxes occur for the same value of γ. Hence, in this case $\epsilon(a, e)$ equals $2^5 - 1 = 31$. If $\dim(\langle V \rangle) = 2$, then the set V contains 2 or 3 different values, which we denote by A, B and $C = A + B$, since the third must be linearly dependent from A and B. Now the 5 double differentials

Table 3 The values of $\epsilon(a, e)$ of characteristics when $\dim(\langle \mathrm{V} \rangle) = 2$

Tr (B/A)	Tr (A/B)	Tr (A/C)	AAAAB (×5)	AAABB (× 10)	AAABC (× 10)	AABBC (× 15)
0	0	0	16	10	9	7
0	0	1	16	10	8	6
0	1	0	15	7	8	4
0	1	1	15	7	7	3
1	0	0	1	3	2	4
1	0	1	1	3	1	3
1	1	0	0	0	1	1
1	1	1	0	0	0	0

are distributed over 2 or 3 different characteristics. (In general, they are distributed over at least $\dim(\langle \mathrm{V} \rangle)$ different characteristics). Then, $\epsilon(a, e)$ of a characteristic depends on its number of double differentials, which equals the number of times its γ value occurs in the (multi)-set V. However, it is now also possible that a γ value leads to double differentials in some S-boxes and to DP $= 0$ in other S-boxes. In such a case, $\epsilon(a, e) = 0$.

Example 1 Let $\mathrm{V} = \{A, A, A, B, C\}$. Then the characteristic corresponding to $\gamma = A$ has three double differentials. The characteristic has $\mathrm{EDP}_{32} > 0$ if and only if $\mathrm{DP}(A, B^{-1}) > 0$ and $\mathrm{DP}(A, C^{-1}) > 0$, which corresponds to the conditions $\mathrm{Tr}\left(A^{-1}B\right) = 0$ and $\mathrm{Tr}\left(A^{-1}C\right) = 0$. The second condition is redundant, since:

$$\left(A^{-1}C\right) = \left(A^{-1}(A + B)\right) = \left(1\right) + \left(A^{-1}B\right) = \left(A^{-1}B\right).$$

If the trace condition is satisfied, then this characteristic results in $\epsilon(a, e) = 2^3 - 1$. Similarly, the characteristics corresponding to $\gamma = B$ and $\gamma = C$ may also result in $\epsilon(a, e) > 0$. Table 3 summarizes all the possibilities for the case $\dim(\langle \mathrm{V} \rangle) = 2$.

The previous computations have to be repeated for all possible values for $\epsilon(a, e)$ and for all possible dimensions of $\langle \mathrm{V} \rangle$. The computations for $\dim(\langle \mathrm{V} \rangle) > 2$ require tables similar to Table 3, which are too bulky to include in this paper. We have computed the full distribution both of the $\mathrm{N}_{32}(a, e)$ entries in Table 2 and the $\epsilon(a, e)$ values in Table 4. Together, they completely determine the distribution of EDP_{32} in the difference table for the naive super box for $\mathrm{w}(a) = 1$ based on Eq. (13) which is also illustrated in Fig. 1.

8.3 The case $\mathrm{w}(a) = 2$

We know then that $\mathrm{EDP}_{32}(a, e) > 0$ only if $\mathrm{w}(e) \geq 3$. Because of the rotational symmetry of MixColumns and because of (15), there are $2 \times 255 = 510$ different cases. Without loss of generality we can assume that $a_1 = 1$. We discuss here the 255

Table 4 Distribution of $\epsilon(a, e)$ for $\mathrm{w}(a) = 1$

d	# of e	$\epsilon(a, e)$	d	# of e	$\epsilon(a, e)$	d	# of e	$\epsilon(a, e)$
1	1	31	3	2,091,600	0	4	233,620,800	0
				1,568,880	1		126,056,160	1
2	2,280	0		1,568,880	1		30,118,080	2
	1,120	1		1,568,880	1		16,453,440	3
	320	2		622,560	2		5,597,760	4
	1,280	3		682,080	3		824,544	5
	960	4		437,760	4		52,800	6
	480	6		111,960	5			
	1,370	7		70,560	6	5	2,867,167,200	0
	640	8		104,040	7		840,636,960	1
	380	9		58,560	8		96,719,040	2
	700	10		12,720	9		5,075,520	3
	280	15					156,000	4
	350	16					1,440	5

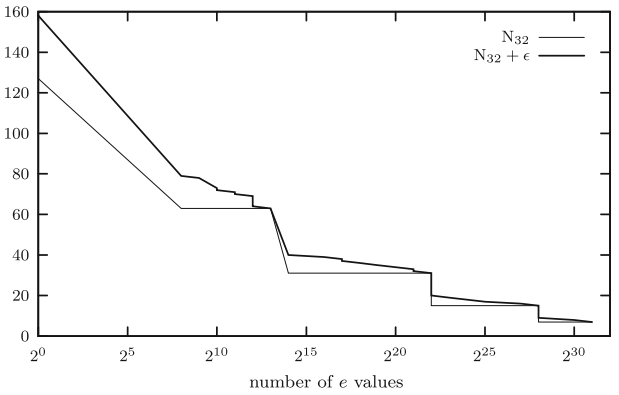

Fig. 1 Plot of the quantities N_{32} and $\mathrm{N}_{32} + \epsilon$

cases where $a_2 \neq 0$ and $a_3 = a_4 = 0$. The other 255 cases have $a_2 = a_4 = 0$ and $a_3 \neq 0$. We have:

$$\mathrm{N}_{32}(a, e) = \sum_{b_1, b_2} \mathrm{N}_8(a_1, b_1)\mathrm{N}_8(a_2, b_2) \prod_{i=1}^{4} \mathrm{N}_8((\mathbf{M}b)_i, e_i).$$

The difference table contains now two types of entries, depending on $\mathrm{w}(e)$:

$\mathrm{w}(e) = 3$: There are 4×255^3 values for e with $\mathrm{w}(e) = 3$. The resulting differentials have weight 5. The EDP_{32} values form a subset of the entries computed in Sect. 8.2.

$\mathrm{w}(e) = 4$: There are 255^4 values for e. The resulting differentials have weight 6. Computing these entries is the main topic of this section.

The characteristics of a differential with weight 6 can be divided into 251 different non-overlapping bundles [6]. Since $b_1 \neq 0$ in all characteristics with $\mathrm{EDP}_{32} > 0$, we introduce the new variable $t = b_2 \cdot b_1^{-1}$ and obtain

$$\begin{aligned}\mathrm{EDP}_{32}(a, e) = \sum_{t\in\mathcal{I}} \sum_{\gamma\in\mathbb{F}_{2^8}^*} &\mathrm{DP}(a_1, \gamma)\mathrm{DP}(a_2, t\gamma)\mathrm{DP}((\theta + (1+\theta)t)\gamma, e_1)\\ &\times\mathrm{DP}((1+\theta t)\gamma, e_2)\mathrm{DP}((1+t)\gamma, e_3)\mathrm{DP}(((1+\theta)+t)\gamma, e_4).\end{aligned}$$

Here t takes all values in $\mathcal{I} = \mathbb{F}_{2^8} \setminus \{0, 1, 1+\theta, \theta^{-1}, \theta\cdot(1+\theta)^{-1}\}$. Computing the EDP_{32} entries for all values of a_2 would require approximately 2^{48} table lookups and additions, which is not feasible on our current hardware. Therefore we will compute the N_{32} values as an estimate for the EDP_{32} values. In other words, we will ignore the correction terms $\epsilon(a, e)$. We have:

$$\begin{aligned}\mathrm{N}_{32}(a, e) = \sum_{t\in\mathcal{I}} \sum_{\gamma\in\mathbb{F}_{2^8}^*} &\mathrm{N}_8(a_1, \gamma)\mathrm{N}_8(a_2, t\gamma)\mathrm{N}_8((\theta + (1+\theta)t)\gamma, e_1)\\ &\times\mathrm{N}_8((1+\theta t)\gamma, e_2)\mathrm{N}_8((1+t)\gamma, e_3)\mathrm{N}_8(((1+\theta)+t)\gamma, e_4).\end{aligned}$$

Hence, $\mathrm{N}_{32}(a, e)$ is now the sum of 251 terms which can be computed as before:

$$\mathrm{N}_{32}(a, e) = \sum_{t\in\mathcal{I}} \left(2^{8-\dim(\langle \mathrm{V}(t)\rangle)} - 1\right),$$

where the sets $V(t)$ are given by

$$\begin{aligned}\mathrm{V}(t) = \{&1, (ta_2)^{-1}, ((\theta + (1+\theta)t)e_1)^{-1},\\ &((1+\theta t)e_2)^{-1}, ((1+t)e_3)^{-1}, (((1+\theta)+t)e_4)^{-1}\}.\end{aligned}$$

We will now compute bounds for the values

$$M := \sum_e \mathrm{N}_{32}(a, e)^5 = \sum_e \left(\sum_{t\in\mathcal{I}} \left(2^{8-\dim(V(t))} - 1\right)\right)^5.$$

for all values of a_2 and use them as estimates for $\mathrm{E}_{5r}(a) = \sum_e \mathrm{EDP}_{32}(a, e)^5$.

8.4 Bounding M

The basic idea in order to bound M is to split the set $\mathrm{V}(t)$ to facilitate the computation. A straightforward approach would be

$$\mathrm{V}(t) = S_1(t) \cup S_2(t),$$

where

$$\begin{aligned} S_1(t) &= \{1, (ta_2)^{-1}\}, \\ S_2(t) &= \{((\theta + (1+\theta)t)e_1)^{-1}, ((1+\theta t)e_2)^{-1}, \\ &\quad ((1+t)e_3)^{-1}, (((1+\theta)+t)e_4)^{-1}\}. \end{aligned}$$

From this we have

$$\dim(\langle S_2(t)\rangle) \leq \dim(\langle V(t)\rangle) \leq \dim(\langle S_1(t)\rangle) + \dim(\langle S_2(t)\rangle).$$

Since $\dim(\langle S_1(t)\rangle) = 2$ except when $t = a_2^{-1}$ we derive the following upper and lower bounds:

$$M \leq U := \sum_e \left(\sum_{t \in \mathcal{I}} 2^{8-\dim(\langle S_2(t)\rangle)} - 1 \right)^5 \tag{19}$$

$$M \geq L := \sum_e \left(\sum_{t \in \mathcal{I}} 2^{8-\dim(\langle S_2(t)\rangle)-2} - 1 \right)^5. \tag{20}$$

From (20) it is also easy to see that $L \leq 2^{-10}U$. Computing U or L still has a complexity of 2^{40} dimension computations. We can use a nice observation for the computation of the values U and L.

Proposition 1 *Let U be defined as in* (19)*. Then, we have*

$$U = 255 \cdot \sum_{\substack{e \\ e_1=1}} \left(\sum_{t \in \mathcal{I}} 2^{8-\dim(\langle S_2(t)\rangle)} - 1 \right)^5. \tag{21}$$

An analogous statement holds for L.

Proof The sum in (19) runs over all values $e = [e_1, \ldots, e_4]$. The contribution of each such tuple to U depends on $\dim(\langle S_2(t)\rangle)$. Now it is easy to see, that the dimension $\dim(\langle S_2(t)\rangle)$ does not change if we replace e by the tuple $[\lambda e_1, \ldots, \lambda e_4]$. This is because the multiplication with a non-zero element amounts to the multiplication of the vectors of $S_2(t)$ by an invertible matrix, and therefore does not change the dimension of the space spanned by these vectors.

We can identify all tuples $[\lambda e_1, \ldots, \lambda e_4]$ for $\lambda \neq 0$. This is exactly the definition of projective space, so we can choose a representative of these tuples by taking $e_1 = 1$ and multiply the contribution by 255.

Proposition 2 *Using the above approach, we can compute the bounds U and L to be*

$$L \approx 2^{80.16} \leq M \leq U \approx 2^{91.7}. \tag{22}$$

In order to arrive at a closer approximation we also computed the following alternative bounds U' and L' by splitting $V(t)$ into $S_1(t) \cup S_2(t)$ with

$$\begin{aligned} S_1(t) &= \{(ta_2)^{-1}\}, \\ S_2(t) &= \{1, ((\theta + (1+\theta)t)e_1)^{-1}, ((1+\theta t)e_2)^{-1}, \\ &\quad ((1+t)e_3)^{-1}, (((1+\theta)+t)e_4)^{-1}\}. \end{aligned}$$

Therefore, we have $\dim(\langle S_2(t)\rangle) \leq \dim(\langle V(t)\rangle) \leq \dim(\langle S_2(t)\rangle) + 1$ and thus

$$\begin{aligned} M &\geq \sum_e \left(\sum_{t \in \mathcal{I}} 2^{8-\dim(\langle S_2(t)\rangle)-1} - 1 \right)^5 \\ M &\leq \sum_e \left(\sum_{t \in \mathcal{I}} 2^{8-\dim(\langle S_2(t)\rangle)} - 1 \right)^5. \end{aligned} \tag{23}$$

Note, that these bounds have the advantage of being independent of a_2. Nevertheless, we cannot use the trick of Proposition 1 anymore, since $S_2(t)$ always contains 1 which makes the computation more difficult by a factor 2^8. We arrive at the following

Proposition 3 *Based on the inequalities in* (23)*, we can compute the following bounds*:

$$2^{80.68} \leq M \leq 2^{86.67}. \tag{24}$$

8.5 The case $\mathrm{w}(a) = 3$

We know then that $\mathrm{EDP}_{32}(a, e) > 0$ only if $\mathrm{w}(e) \geq 2$. Because of the rotational symmetry of MixColumns and because of (15), there are 255^2 different cases. Without loss of generality we can assume that $a_1 = 1$ and $a_4 = 0$. If $\mathrm{w}(e) = 4$, then

$$\mathrm{N}_{32}(a, e) = \sum_{t,s} \left(2^{8-\dim(V(t,s)} - 1 \right),$$

where the sum runs over 64015 values of (s, t) (see Table 1) and we consider the vector space spanned by the elements of

$$\begin{aligned} \mathrm{V}(t, s) = \{&1, (ta_2)^{-1}, (sa_3)^{-1}, ((\theta + (1+\theta)t + s)e_1)^{-1}, \\ &((1+\theta t + (1+\theta)s)e_2)^{-1}, ((1+t+\theta s)e_3)^{-1}, \\ &(((1+\theta)+t+s)e_4)^{-1}\}. \end{aligned}$$

It seems currently computationally too expensive to compute the distribution of $\mathrm{N}_{32}(a, e)$ for all possible values of a, e or $\mathrm{E}_{5r}(a)$ for all possible values of a.

8.6 The case $w(a) = 4$

Because of the rotational symmetry of MixColumns and because of (15), there are 255^3 different cases. Without loss of generality we can assume that $a_1 = 1$. If $w(e) = 4$, then

$$N_{32}(a, e) = \sum_{u,t,s} \left(2^{8-\dim(V(u,t,s)} - 1\right),$$

where the sum runs over 16323805 values of (s, t, u) (see Table 1) and we consider the vector space spanned by the elements of

$$\begin{aligned} V(u, t, s) = \{&1, (ta_2)^{-1}, (sa_3)^{-1}, (ua_4)^{-1}, \\ &((\theta + (1+\theta)t + s + u)e_1)^{-1}, ((1 + \theta t + (1+\theta)s + u)e_2)^{-1}, \\ &((1 + t + \theta s + (1+\theta)u)e_3)^{-1}, (((1+\theta) + t + s + \theta u)e_4)^{-1}\}. \end{aligned}$$

It seems currently computationally too expensive to compute the distribution of $N_{32}(a, e)$ for all possible values of a, e or $E_{5r}(a)$ for all possible values of a.

9 Conclusions and outlook

In this paper, we determined the distribution of the $EDP_{32}(a, e)$ values for all differentials with $w(a, b) = 5$ over the super box using the naive S-box. Observe that for the matrix $\mathbf{M}$ used in MixColumns, we only used the fact it is an MDS matrix. Hence, these results apply for any MDS matrix, in particular also for $\mathbf{M}^{-1}$. From these partial results, we can already draw some conclusions on the EDP_{128} of a fraction of the differentials over the mega box. Consider a differential

$$\begin{aligned} (x, y) = (&x_0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0; \\ &y_0, 0, 0, 0, y_4, 0, 0, 0, y_8, 0, 0, 0, y_{12}, 0, 0, 0) \end{aligned}$$

over a mega box. All characteristics in this differential have exactly 5 active super boxes, and each active super box has exactly 5 active S-boxes. Summing up the fifth powers of the EDP_{32} entries given in Fig. 1 results in the following bound:

$$EDP_{128}(x, y) \leq 2^{49.35} \times 2^{-175} = 2^{-125.65} \approx 5 \times 2^{-128}.$$

This bound is quite close to the theoretically optimal bound of 2×2^{-128}. It holds for all mega box differentials where five super boxes are active and each active super box has exactly 5 active S-boxes. Furthermore, Theorem 1 can be generalized to hold for a mega box, which implies that that the above bound also holds for all mega box differentials with more than five active super boxes where each active super box has exactly five active S-boxes.

Subsequently, we looked at super box differentials with weight 6. First, we split up $\mathrm{E}_{5r}(a)$ as follows:

$$\mathrm{E}_{5r}(a) = \sum_{\substack{e \\ \mathrm{w}(e)=3}} (\mathrm{EDP}_{32}(a,e))^5 + \sum_{\substack{e \\ \mathrm{w}(e)=4}} (\mathrm{EDP}_{32}(a,e))^5.$$

The first term is upper bounded by 4 times the $\mathrm{E}_{5r}(a')$ where a' is any difference with $\mathrm{w}(a') = 1$. We estimated the second term by bounding $M = \sum(\mathrm{N}_{32})^5$ for all differences a with activity pattern (1, 1, 0, 0). If the bound holds also for other differences a with weight 2 and if the correction terms $\epsilon(a, e)$ can be ignored, then this would result in the following bound

$$\mathrm{EDP}_{128}(x, y) \leq 4 \times 2^{-125.65} + 2^{86.67} \times 2^{-210} \approx 45 \times 2^{-128},$$

which would hold for all mega box differentials where exactly 5 super boxes are active and each active super box has at most 6 active S-boxes. It remains a topic of further research to compute the correction terms, to verify this last bound, and to compute bounds for the cases of 7 and 8 active S-boxes. This work would provide further arguments supporting Conjecture 1.

Acknowledgments The work in this paper has been supported in part by the IAP Programme P6/26 BCRYPT of the Belgian State (Belgian Science Policy).

References

1. Beth T, Ding C (1993) On Almost Perfect Nonlinear Permutations. In: EUROCRYPT. Lecture Notes in Computer Science, vol 765. Springer, Heidelberg, pp 65–76
2. Biham E, Shamir A (1990) Differential Cryptanalysis of DES-like Cryptosystems. In: Menezes A, Vanstone SA (eds) CRYPTO. Lecture Notes in Computer Science, vol 537. Springer, Heidelberg, pp 2–21
3. Biryukov A (2007) The design of a stream Cipher LEX. Selected areas in cryptography. Lecture Notes in Computer Science, vol 4356. Springer, Heidelberg, pp 67–75
4. Daemen J, Rijmen V (2002) The design of Rijndael: AES—the advanced encryption standard. Springer, Heidelberg
5. Daemen J, Rijmen V (2005) The Pelican MAC Function. Cryptology ePrint Archive, Report 2005/088. http://eprint.iacr.org/
6. Daemen J, Rijmen V (2006) Understanding two-round differentials in AES. In: De Prisco R, Yung M (eds) SCN. Lecture Notes in Computer Science, vol 4116. Springer, Heidelberg, pp 78–94
7. Fisher SD (1966) Classroom notes: matrices over a finite field. Am Math Mon 73(6):639–641
8. Hong S, Lee S, Lim J, Sung J, Cheon DH, Cho I (2000) Provable Security against Differential and Linear Cryptanalysis for the SPN Structure. In: Schneier B (ed) FSE. Lecture Notes in Computer Science, vol 1978. Springer, Heidelberg, pp 273–283
9. Keliher L, Meijer H, Tavares SE (2001) New method for upper bounding the maximum average linear hull probability for SPNs. In: Pfitzmann B (ed) EUROCRYPT. Lecture Notes in Computer Science, vol 2045. Springer, Heidelberg, pp 420–436
10. Keliher L (2004) Refined analysis of bounds related to linear and differential cryptanalysis for the AES. In: Dobbertin H, Rijmen V, Sowa A (eds) AES4 Conference. Lecture Notes in Computer Science, vol 3373. Springer, Heidelberg, pp 42–57
11. Keliher L, Sui J (2007) Exact maximum expected differential and linear probability for 2-round advanced encryption standard (AES). IET Inf Secur 1(2):53–57

12. Lai X, Massey JL, Murphy S (1991) Markov ciphers and differential cryptanalysis. In: Advances in Cryptology—EUROCRYPT '91 (Brighton, 1991). Lecture Notes in Computer Science, vol 547. Springer, Berlin, pp 17–38
13. Lidl R, Niederreiter H (1997) Finite fields, Encyclopedia of mathematics and its applications, 2nd edn. Cambridge University Press, Cambridge
14. Matsui M (1993) Linear Cryptoanalysis Method for DES Cipher. EUROCRYPT. In: Helleseth T (ed) Lecture Notes in Computer Science, vol 765. Springer, Heidelberg, pp 386–397
15. Minematsu K, Tsunoo Y (2006) Provably secure MACs from differentially-uniform permutations and AES-based implementations. In: Robshaw M (ed) FSE. Lecture Notes in Computer Science, vol 4047. Springer, Heidelberg, pp 226–241
16. Park S, Sung SH, Chee S, Yoon E-J, Lim J (2002) On the security of Rijndael-like structures against differential and linear cryptanalysis. In: Zheng Y (ed) ASIACRYPT. Lecture Notes in Computer Science, vol 2501. Springer, Heidelberg, pp 176–191
17. Park S, Sung SH, Lee S, Lim J (2003) Improving the upper bound on the maximum differential and the maximum linear hull probability for SPN structures and AES. In: Johansson T (ed) FSE. Lecture Notes in Computer Science, vol 2887. Springer, Heidelberg, pp 247–260