Proving the Security of AES Substitution-Permutation Network

Thomas Baignères* and Serge Vaudenay

EPFL
http://lasecwww.epfl.ch

Abstract. In this paper we study the substitution-permutation network (SPN) on which AES is based. We introduce AES*, a SPN identical to AES except that fixed S-boxes are replaced by random and independent permutations. We prove that this construction resists linear and differential cryptanalysis with 4 inner rounds only, despite the huge cumulative effect of multipath characteristics that is induced by the symmetries of AES. We show that the DP and LP terms both tend towards $1/(2^{128}-1)$ very fast when the number of round increases. This proves a conjecture by Keliher, Meijer, and Tavares. We further show that AES* is immune to any iterated attack of order 1 after 10 rounds only, which substantially improves a previous result by Moriai and Vaudenay.

Keywords: Differential Cryptanalysis, Linear Cryptanalysis, Differentials, Linear Hulls, Provable Security, AES.

1 Preamble

When we refer to "cryptanalysis", we usually think about its destructive side which consists in breaking cryptographic algorithms. Cryptanalysis however means "cryptographic analysis", which includes a constructive side that consists in proving the security of a system or the soundness of a construction. However, this last side has not received as much attention for block ciphers. Indeed, security proofs often rely on arguments derived from previous cryptanalytic attacks.

We can use linear and differential cryptanalysis [2, 3, 25, 24] (respectively denoted LC and DC) to illustrate this statement. If C denotes a block cipher, DC and LC have a complexity which is inversely proportional to the *differential probability*[1] (DP) [30] and to the *linear probability*[2] (LP) [4] terms respectively. When using an r-round Markov cipher [21], one can prove that the DP (resp. LP) is expressed as the sum of the product of the DP's (resp. LP's) in all possible inner chains of differences [37] (resp. masks). We thus usually refer to multipath characteristics or differentials [21] (resp. linear hull [31]). Typically, attacks make a heuristic approximation of the DP (resp. LP) by considering only one (single path) characteristic. If the LP or DP of such a characteristic is significant enough, then

* Supported by the Swiss National Science Foundation, 200021-107982/1.

[1] Given an input/output difference of (a, b), $\mathrm{DP}(a, b) = \Pr[\mathsf{C}(X) \oplus \mathsf{C}(X \oplus a) = b]$.

[2] Given input/output masks (a, b), $\mathrm{LP}(a, b) = (2\Pr[a \cdot X = b \cdot \mathsf{C}(X)] - 1)^2$.

B. Preneel and S. Tavares (Eds.): SAC 2005, LNCS 3897, pp. 65–81, 2006.

an attack can definitely be performed. In that situation, the cumulative effect of the differentials (resp. linear hull) can only make the attack work better than expected. Similar approximations are also made in *security proofs* of block ciphers. This could be acceptable only if one could make sure that, among the differentials (resp. linear hull), one single path characteristic is overwhelming (so that the rest can be neglected). Although this is actually the case for DES, this does not appear to be true for AES [6, 7]. Indeed, the argument saying that all DP and LP terms are at most 2^{-300} on 8 rounds [6, pp. 30–31] obviously cannot be true. Since for any a, the sum over all the 2^{128} values $\mathrm{DP}(a, b)$ (resp. $\mathrm{LP}(a, b)$) is equal to 1, at least one value of the DP is larger than 2^{-128}. Obviously, symmetries in AES are likely to lead to a considerable cumulative effect when considering many equivalent characteristics. Therefore, proving that there is no single path characteristic with a significant DP (resp. LP) is not sufficient to prove the resistance of a block cipher against DC (resp. LC).

In practice, however, differentials and linear hull are rarely taken into consideration in security proofs, as evaluating the true DP or LP is computationally not practical for a typical block cipher. One natural solution is to try to upper bound these terms. This approach was chosen by Keliher, Meijer, and Tavares [19, 18] who showed that the LP of AES is upper bounded by 2^{-75} for 7 or more rounds. Park et al. showed [33, 34] that the DP and LP for four rounds are respectively bounded by 1.144×2^{-111} and 1.075×2^{-106}. Finally, in a recent work [17], Keliher shows that the bound on the LP is 1.778×2^{-107} for 8 or more rounds.

Another solution is to adopt a Luby-Rackoff-like approach. In their seminal work [22], Luby and Rackoff showed how to construct a pseudo-random permutation from any pseudo-random function. They provided an example based on the Feistel scheme [8] (because it is the one on which DES is based). Since then, the security of Feistel ciphers with *random* and *independent* round functions received a considerable amount of attention (see [29, 35, 38, 27, 9], to name only a few). Although Substitution-Permutation Network (SPN) schemes security has already been widely studied (see for example [5, 13, 14, 19]), only a few papers adopted a Luby-Rackoff-like approach to study the one on which AES is based (see for example Moriai-Vaudenay [28] and Keliher-Meijer-Tavares [20]).

In this paper, we analyze the security of the SPN on which AES is based, where fixed S-boxes are replaced by *random* and *independent* permutations. This scheme, that we call AES^*, is introduced in Section 2, together with some of its properties with respect to the LP and DP terms. This includes a discussion about keyed operations, i.e., over the subkey addition and over the substitution boxes layer. In Section 3, we give an expression of the expected LP over AES^*, depending on the input/output masks and the number of rounds (see Theorem 6). Using this result, we prove a conjecture made by Keliher, Meijer, and Tavares in [20], namely that all DP's and LP's converge towards $1/(2^{128} - 1)$ as the number of rounds increases (see Theorem 8). This means that AES^* behaves exactly like the perfect cipher (as far as LC is concerned) when the number of rounds is high enough. The rest of Section 3 shows how to reduce the computational complexity of the expression given in Theorem 6 by exploiting some of the symmetries of AES^* (see Theorem 12). We

conclude the section by exhibiting results of practical experiments. We give the expected LP over AES* for several number of rounds and several S-box sizes, and deduce that AES* is protected against LC after four inner rounds only[3]. Section 4 shows how these results extend to differential cryptanalysis. In Section 5 we generalize the results on LC by considering *any* iterated attack of order 1 [39]. Recall that these kind of attacks are very similar to LC, except that the bit of information extracted from each plaintext/ciphertext pair is not necessarily computed by a linear masking of text bits, but can be derived using any type of *projection* (in the sense of [41,1]). Experimental results show that after 10 rounds, AES* is immune to iterated attacks of order 1. This substantially improves a previous result of Moriai and Vaudenay, who showed that 384 were sufficient [28]. Finally, we show in Section 6 by derandomization techniques that all security results on AES* remain valid when the random S-boxes are replaced by S-boxes with perfect pairwise decorrelation.

2 Preliminaries

2.1 Description of AES

AES [6,7] is a block cipher with a fixed block length of 128 bits, supporting 128, 192, and 256 bit keys. It is iterated, meaning that it is made of a succession of r identical rounds. It is byte oriented, meaning that each round takes as an input a 128 bit long state that is considered as a one-dimensional vector of 16 bytes. Each round also takes a round key as an input, issued from a key schedule that takes as an input the main 128 bit key. We do not detail the key schedule here, since we will assume that all round keys are randomly selected and independent. Each round is a succession of four operations (we use the notations of [7]): `SubBytes`, that applies a fixed S-box to each of the 16 input state bytes, `ShiftRows`, which shifts the rows of the state (considered as a 4×4 array) over four different offsets, `MixColumns`, that applies a linear transformation to each state columns, and `AddRoundKey`, which is a bitwise XOR between the subkey and the state. AES is made of 10 rounds (for a 128 bit key), where the last one does not include the `MixColumns` operation. The first round is preceded by a additional `AddRoundKey` operation.

2.2 Introducing AES*

In the subsequent, we will be considering a family of block ciphers based on AES. This family, that we call AES*, almost follows the same SPN as AES, except for the last round, which excludes *both* linear operations (that is, `MixColumns` and `ShiftRows`). Although this modification does not have any influence on the security results, it simplifies the notations. Moreover, it does not involve a fixed S-box. Following a Luby-Rackoff-like approach, each S-box will be considered as an independent permutation chosen uniformly at random. Consequently, we denote by `SubBytes`* the confusion step in AES*. In that sense, AES is a particular

[3] Note that this result does take hulls effect into consideration.

instance of AES* where all the S-boxes have been chosen to be the one defined in the specifications.

Clearly, a truly random S-box following the XOR of a random byte is equivalent to a truly random S-box. Hence, we can completely ignore the addition of round keys in AES*.

2.3 States, Activity Patterns, and Notations

We denote by $\mathrm{GF}(q)$ the finite field with q elements and by $\mathcal{S}$ the set of AES* states, so that $\mathcal{S} = \mathrm{GF}(q)^{16}$. In the case of AES, $q = 2^8$. AES* states (or equivalently, masks on AES* states) will generally be denoted by bold small letters such as $\boldsymbol{a}$,$\boldsymbol{b}$, etc. An arbitrary state $\boldsymbol{a}$ is a vector which can be viewed as a four by four array with elements in $\mathrm{GF}(q)$ denoted $(a_{i,j})_{1\leq i,j\leq 4}$. The four by four array of $\{0,1\}^{16}$ with 0's at the positions where the entry of $\boldsymbol{a}$ is 0, and with 1's where the entry of $\boldsymbol{a}$ non-zero, is called the *activity pattern* [6] or *support* corresponding to the state $\boldsymbol{a}$. Supports will either be denoted by Greek letters or by $\mathrm{SUPP}(\cdot)$. For example, the support corresponding to a state $\boldsymbol{a}$ will either be denoted $\boldsymbol{\alpha}$ or $\mathrm{SUPP}(\boldsymbol{a})$. The (i,j) entry in the array $\boldsymbol{\alpha}$ will be denoted $\alpha_{i,j}$. The Hamming weight of a support $\boldsymbol{\alpha}$, denoted $|\boldsymbol{\alpha}|$, is the number of 1's in this support (i.e., $|\boldsymbol{\alpha}| = \sum_{i,j} \alpha_{i,j}$), so that $0 \leq |\boldsymbol{\alpha}| \leq 16$. When $|\boldsymbol{\alpha}| = 0$ it means that $\boldsymbol{a}$ is zero, whereas when $|\boldsymbol{\alpha}| = 16$, it means that all entries of $\boldsymbol{a}$ are non-zero. In the latter case, we say that $\boldsymbol{a}$ is a state of *full-support*. The set of states limited to some specific support $\boldsymbol{\alpha}$ will be denoted $\mathcal{S}_{|\boldsymbol{\alpha}}$, and thus $\#\mathcal{S}_{|\boldsymbol{\alpha}} = \sigma^{|\boldsymbol{\alpha}|}$, with $\sigma = q - 1$. The set of states of full-support will be denoted $\mathcal{S}_{\mathrm{full}}$ so that $\#\mathcal{S}_{\mathrm{full}} = \sigma^{16}$.

2.4 The Scalar Product and the LP Coefficient

The scalar product of a state (plaintext) $\boldsymbol{x}$ and a state (mask) $\boldsymbol{a}$ is usually defined as the exclusive-or between several bits of $\boldsymbol{x}$, chosen according to a pattern specified by a mask $\boldsymbol{a}$, which depends on the way the elements of $\mathrm{GF}(q)$ are represented. We prefer here an equivalent definition in terms of the trace function[4] Tr, defined from $\mathrm{GF}(q)$ onto $\mathrm{GF}(2)$ by $\mathrm{Tr}(x) = x + x^2 + x^4 + x^8 + \cdots + x^{128}$. If $\boldsymbol{a}$ and $\boldsymbol{b}$ are two arbitrary states of AES, we define the scalar product of $\boldsymbol{a}$ and $\boldsymbol{b}$ as $\boldsymbol{a} \bullet \boldsymbol{b} = \sum_{i,j} \mathrm{Tr}(a_{i,j} b_{i,j})$. We use the following well known linear algebra property.

Lemma 1. *Let* M *denote an arbitrary* 16 *by* 16 *matrix of elements in* $\mathrm{GF}(q)$*, representing a linear transformation on* AES *states*[5]*. Let* $\boldsymbol{x}$ *be an input state to this linear transformation* M *and let* $\boldsymbol{b}$ *be a non-zero output mask. Then* $\boldsymbol{b} \bullet (\mathsf{M} \times \boldsymbol{x}) = (\mathsf{M}^T \times \boldsymbol{b}) \bullet \boldsymbol{x}$*.*

The efficiency of a linear cryptanalysis can be measured by means of the *linear probability* [4]. With our definition of the scalar product, this quantity is defined

[4] One advantage of this variant is that it does not depend on the way we represent $\mathrm{GF}(q)$. Namely, even if we represent the cells of AES states by the Zech logarithm, we can still define the scalar product in the same way.

[5] AES states are indeed considered as column vectors.

in the following way (here we use the notation introduced in [26]): if $\boldsymbol{a}$ and $\boldsymbol{b}$ are two states and C is some fixed permutation on $\mathcal{S}$, then $\mathrm{LP}^{\boldsymbol{C}}(\boldsymbol{a},\boldsymbol{b}) = (2\Pr_{\boldsymbol{X}\in\mathcal{S}}[\boldsymbol{a}\bullet\boldsymbol{X} = \boldsymbol{b}\bullet\mathsf{C}(\boldsymbol{X})] - 1)^2$, where the probability holds over the uniform distribution of $\boldsymbol{X}$.

2.5 Expected LP over Keyed Operations in AES*

A round key, or simply a key, is an AES state. The only keyed operation in AES is `AddRoundKey`. As stated in Section 2.2, we can ignore this operation in AES*. We can thus consider the choice of the random S-boxes as the only keyed operation in AES*. The following lemma evaluates the average LP over all possible random S-boxes.

Lemma 2. *Let $a, b \in \mathrm{GF}(q) \setminus \{0\}$ be two non-zero input/output masks on the uniformly distributed random S-box S^* and let $\sigma = q - 1$. The average* LP *value over all possible random S-boxes is independent of a and b, and is $\mathrm{E}_{\mathsf{S}^*}[\mathrm{LP}^{\mathsf{S}^*}(a,b)] = \sigma^{-1}$.*

Proof. See Lemma 14 in [39] for a direct proof. One can also use the explicit distribution of the LP of S^* [20], deduced from results available in [32]. □

Note that for any S-box S we have $\mathrm{LP}^S(a,0) = \mathrm{LP}^S(0,b) = 0$ (for non-zero a and b) and $\mathrm{LP}^S(0,0) = 1$. From this, we derive the expected LP over `SubBytes`*.

Lemma 3. *Let $\boldsymbol{a}$ and $\boldsymbol{b}$ be two non-zero masks in $\mathrm{GF}(q)^{16}$, and let $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ be their respective supports. Let $\sigma = q - 1$. We have $\mathrm{E}[\mathrm{LP}^{\mathtt{SubBytes}^*}(\boldsymbol{a},\boldsymbol{b})] = \sigma^{-|\boldsymbol{\alpha}|}$ if $\boldsymbol{\alpha} = \boldsymbol{\beta}$ and 0 otherwise, where the mean is taken over all possible uniformly distributed and independent random S-boxes.*

3 Expected LP on AES*

3.1 From Sums over Masks to Sums over Supports

The complexity of computing the expected LP of AES is prohibitive for the reason that, once input/output masks are given, one has to sum over all possible intermediate masks in order to take into account every possible characteristic. We will see that AES* provides just enough randomization for this sum to be made over intermediate supports. Consequently, we will need to count the number of possible states succession corresponding to some given succession of supports.

Definition 4. *Let* LT *denote the linear transformation of* AES*, i.e., the operation corresponding to* `MixColumns` ∘ `ShiftRows`*. Let $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ be two supports and*

$$\mathrm{N}[\boldsymbol{\alpha},\boldsymbol{\beta}] = \#\{(\boldsymbol{a},\boldsymbol{b}) \in \mathcal{S}_{|\boldsymbol{\alpha}} \times \mathcal{S}_{|\boldsymbol{\beta}} \ : \ \mathsf{LT}^T \times \boldsymbol{b} = \boldsymbol{a}\},$$

where states $\boldsymbol{a}$ and $\boldsymbol{b}$ are considered as column vectors here. $\mathrm{N}[\boldsymbol{\alpha},\boldsymbol{\beta}]$ is the number of ways of connecting a support $\boldsymbol{\alpha}$ to a support $\boldsymbol{\beta}$ through LT*.*

From now on, we will consider an r-round version of AES^*, with $r > 1$. Round $i \in \{1, \ldots, r\}$ will be denoted by $\mathtt{Round}_i^*$, where the last round $\mathtt{Round}_r^*$ excludes the linear transformation LT. With these notations, $\mathsf{AES}^* = \mathtt{Round}_r^* \circ \cdots \circ \mathtt{Round}_1^*$. The input/output masks on $\mathtt{Round}_i^*$ will usually be denoted $\boldsymbol{c}_{i-1}$ and $\boldsymbol{c}_i$ respectively, while their corresponding supports will be denoted $\boldsymbol{\gamma}_{i-1}$ and $\boldsymbol{\gamma}_i$. Consequently, $\boldsymbol{c}_0$ and $\boldsymbol{c}_r$ will respectively denote the input and the output masks on a r-rounds version of AES^*. Using Lemma 3, we can derive the expected LP over one round and extend it to the full AES^*.

Lemma 5. *Let $\boldsymbol{c}_{i-1}$ and $\boldsymbol{c}_i$ be two non-zero masks in $\mathrm{GF}(q)^{16}$ of support $\boldsymbol{\gamma}_{i-1}$ and $\boldsymbol{\gamma}_i$ respectively. Let $\sigma = q - 1$. For $1 \leq i < r$, the expected linear probability over $\mathtt{Round}_i^*$ is given by $\mathrm{E}[\mathrm{LP}^{\mathtt{Round}_i^*}(\boldsymbol{c}_{i-1}, \boldsymbol{c}_i)] = \sigma^{-|\boldsymbol{\gamma}_{i-1}|}$ if $\boldsymbol{\gamma}_{i-1} = \mathrm{SUPP}(\mathsf{LT}^T \times \boldsymbol{c}_i)$ and 0 otherwise. Similarly, the expected* LP *over the last round is given by $\mathrm{E}[\mathrm{LP}^{\mathtt{Round}_r^*}(\boldsymbol{c}_{r-1}, \boldsymbol{c}_r)] = \sigma^{-|\boldsymbol{\gamma}_{r-1}|}$ if $\boldsymbol{\gamma}_{r-1} = \boldsymbol{\gamma}_r$ and 0 otherwise.*

Proof. We first consider the case where $1 \leq i < r$. Using Lemma 1, we have $\mathrm{E}[\mathrm{LP}^{\mathtt{Round}_i^*}(\boldsymbol{c}_{i-1}, \boldsymbol{c}_i)] = \mathrm{E}[\mathrm{LP}^{\mathtt{SubBytes}^*}(\boldsymbol{c}_{i-1}, \mathsf{LT}^T \times \boldsymbol{c}_i)]$. Lemma 3 allows to conclude. The proof for the $i = r$ case is similar, except that we don't make use of Lemma 1 as the last round excludes LT. □

Theorem 6. *Let $\boldsymbol{c}_0$ and $\boldsymbol{c}_r$ be two masks in $\mathrm{GF}(q)^{16}$ of support $\boldsymbol{\gamma}_0$ and $\boldsymbol{\gamma}_r$ respectively. Let $\sigma = q - 1$. The expected linear probability over $r > 1$ rounds of AES^*, when $\boldsymbol{c}_0$ is the input mask and $\boldsymbol{c}_r$ the output mask is*

$$\mathrm{E}[\mathrm{LP}^{\mathsf{AES}^*}(\boldsymbol{c}_0, \boldsymbol{c}_r)] = \sigma^{-|\boldsymbol{\gamma}_r|} \times \left(\mathcal{M}^{r-1}\right)_{\boldsymbol{\gamma}_0, \boldsymbol{\gamma}_r} ,$$

where $\mathcal{M}$ is a $2^{16} \times 2^{16}$ square matrix, indexed by pairs of masks $(\boldsymbol{\gamma}_{i-1}, \boldsymbol{\gamma}_i)$, such that $\mathcal{M}_{\boldsymbol{\gamma}_{i-1}, \boldsymbol{\gamma}_i} = \sigma^{-|\boldsymbol{\gamma}_{i-1}|}\mathrm{N}[\boldsymbol{\gamma}_{i-1}, \boldsymbol{\gamma}_i]$.

Proof. Following Nyberg [31], $\mathrm{E}[\mathrm{LP}^{\mathsf{AES}^*}(\boldsymbol{c}_0, \boldsymbol{c}_r)] = \sum \prod_{i=1}^{r} \mathrm{E}[\mathrm{LP}^{\mathtt{Round}_i^*}(\boldsymbol{c}_{i-1}, \boldsymbol{c}_i)]$, where the sum is taken over all intermediate masks $c_1, \ldots, c_{r-1}$. Using the results (and the notations) of Lemma 5 this gives

$$\mathrm{E}[\mathrm{LP}^{\mathsf{AES}^*}(\boldsymbol{c}_0, \boldsymbol{c}_r)] = \sum_{\substack{\boldsymbol{c}_1, \ldots, \boldsymbol{c}_{r-1} \\ \boldsymbol{\delta}_1, \ldots, \boldsymbol{\delta}_{r-1}}} \sigma^{-|\boldsymbol{\gamma}_{r-1}|} \mathbf{1}_{\boldsymbol{\gamma}_{r-1} = \boldsymbol{\gamma}_r} \prod_{i=1}^{r-1} \sigma^{-|\boldsymbol{\gamma}_{i-1}|} \mathbf{1}_{\substack{\boldsymbol{\gamma}_{i-1} = \mathrm{SUPP}(\mathsf{LT}^T \times \boldsymbol{c}_i) \\ \boldsymbol{\delta}_i = \boldsymbol{\gamma}_i}} ,$$

where the sum is also taken over all possible intermediate supports. Taking $\boldsymbol{\delta}_0 = \boldsymbol{\gamma}_0$ and $\boldsymbol{\delta}_r = \boldsymbol{\gamma}_r$ and including the sum over the $\boldsymbol{c}_i$'s in the product, we obtain $\mathrm{E}[\mathrm{LP}^{\mathsf{AES}^*}(\boldsymbol{c}_0, \boldsymbol{c}_r)] = \sigma^{-|\boldsymbol{\delta}_r|} \sum_{\boldsymbol{\delta}_1, \ldots, \boldsymbol{\delta}_{r-2}} \prod_{i=1}^{r-1} \sigma^{-|\boldsymbol{\delta}_{i-1}|}\mathrm{N}[\boldsymbol{\delta}_{i-1}, \boldsymbol{\delta}_i]$. The definition of the product of square matrices concludes the proof. □

Using supports drops the matrix size from 2^{128} down to 2^{16}. As one matrix multiplication roughly takes $(2^{16})^3$ field operations[6] and, using a square and multiply technique, $\log r$ such multiplications are needed, the overall number of operations needed to compute $\mathcal{M}^{r-1}$ is roughly equal to 2^{50} (for 8 rounds) by using 2×2^{32}

[6] Using Strassen's algorithm, the complexity drops to $(2^{16})^{\log 7}$ field operations [40].

multiple precision rational number registers. This is still pretty hard to implement using ordinary hardware. Nevertheless, from one computation of $\mathcal{M}^{r-1}$ we could deduce all expected linear probability over all possible input/output masks almost for free. In section 3.3, we show how to exploit symmetries of table $\mathrm{N}[\cdot,\cdot]$ in order to further reduce the matrix size.

3.2 Towards the True Random Cipher

For any non-zero mask $\boldsymbol{c}$, $\mathrm{LP}^{\mathsf{AES}^*}(\boldsymbol{c},0) = \mathrm{LP}^{\mathsf{AES}^*}(0,\boldsymbol{c}) = 0$ and $\mathrm{LP}^{\mathsf{AES}^*}(0,0) = 1$. Thus, the $2^{16}\times 2^{16}$ square matrix $\mathcal{M}$ of Theorem 6 has the following shape

$$\mathcal{M} = \left(\begin{array}{c|c} 1 & \mathbf{0} \\ \hline \mathbf{0} & \mathcal{M}' \end{array}\right) \tag{1}$$

where $\mathcal{M}'$ is a $(2^{16}-1)\times(2^{16}-1)$ square matrix, indexed by non-zero supports. We can now notice from Theorem 6 that $\mathrm{E}[\mathrm{LP}^{\mathsf{AES}^*}(\boldsymbol{c}_0,\boldsymbol{c}_2)] = \sigma^{-|\boldsymbol{\gamma}_2|}\mathcal{M}'_{\boldsymbol{\gamma}_0,\boldsymbol{\gamma}_2}$ for any non-zero supports $\boldsymbol{c}_0$ and $\boldsymbol{c}_2$. Recall that $\sum_{\boldsymbol{c}_2}\mathrm{E}[\mathrm{LP}^{\mathsf{AES}^*}(\boldsymbol{c}_0,\boldsymbol{c}_2)] = 1$. Hence

$$1 = \sum_{\boldsymbol{c}_2}\sigma^{-|\boldsymbol{\gamma}_2|}\mathcal{M}'_{\boldsymbol{\gamma}_0,\boldsymbol{\gamma}_2} = \sum_{\boldsymbol{\gamma}_2}\sigma^{|\boldsymbol{\gamma}_2|}\sigma^{-|\boldsymbol{\gamma}_2|}\mathcal{M}'_{\boldsymbol{\gamma}_0,\boldsymbol{\gamma}_2} = \sum_{\boldsymbol{\gamma}_2}\mathcal{M}'_{\boldsymbol{\gamma}_0,\boldsymbol{\gamma}_2}\,.$$

We also note that $\mathcal{M}'_{\boldsymbol{\gamma}_0,\boldsymbol{\gamma}_2} \geq 0$ for any $\boldsymbol{\gamma}_0$ and $\boldsymbol{\gamma}_2$.

Lemma 7. *The matrix $\mathcal{M}'$ defined by* (1) *is the transition matrix of a Markov chain, whose set of states is the set of non-zero supports and whose transition probability from a non-zero support $\boldsymbol{\gamma}$ to a non-zero support $\boldsymbol{\gamma}'$ is given by $\mathcal{M}_{\boldsymbol{\gamma},\boldsymbol{\gamma}'}$.*

The transition graph of the Markov chain is the directed graph whose vertices are the σ non-zero supports and such that there is an edge from $\boldsymbol{\gamma}$ to $\boldsymbol{\gamma}'$ when $\mathcal{M}_{\boldsymbol{\gamma},\boldsymbol{\gamma}'} > 0$. From the study of supports propagation [6] (which is based on the MDS criterion), it clearly appears that from any graph state, there is a path towards the graph state corresponding to the full support $\boldsymbol{\gamma}_{\mathrm{full}}$ (for example, two steps are required to go from a support of Hamming weight 1 to $\boldsymbol{\gamma}_{\mathrm{full}}$). Moreover, from the graph state corresponding to $\boldsymbol{\gamma}_{\mathrm{full}}$ one can reach any graph state. Hence, from each graph state there is a sequence of arrows leading to *any* other graph state. This means that the corresponding Markov chain is *irreducible* [12]. Since there is an arrow from $\boldsymbol{\gamma}_{\mathrm{full}}$ to itself, one can find a sequence of arrows leading from any graph state to any graph state, of *any* (yet long enough) length. This means the Markov chain is *aperiodic*. We can deduce that there exists exactly one stationary distribution (see for example chapter 5 in [12]), i.e., a $1\times(2^{16}-1)$ row vector $\boldsymbol{\pi} = (\pi_{\boldsymbol{\gamma}})_{\boldsymbol{\gamma}\neq\mathbf{0}}$ indexed by non-zero supports such that $\pi_{\boldsymbol{\gamma}} \geq 0$ for all non-zero $\boldsymbol{\gamma}$ with $\sum_{\boldsymbol{\gamma}\neq 0}\pi_{\boldsymbol{\gamma}} = 1$, and such that $\pi\mathcal{M}' = \pi$ (which is to say that $\pi_{\boldsymbol{\gamma}'} = \sum_{\boldsymbol{\gamma}\neq\mathbf{0}}\pi_{\boldsymbol{\gamma}}\mathcal{M}'_{\boldsymbol{\gamma},\boldsymbol{\gamma}'}$ for all non zero $\boldsymbol{\gamma}'$). It is easy to show that the row vector $\boldsymbol{\pi}$ indexed by non-zero supports such that $\pi_{\boldsymbol{\gamma}} = \sigma^{|\boldsymbol{\gamma}|}(q^{16}-1)^{-1}$ is a stationary distribution of the Markov chain described by the transition matrix $\mathcal{M}'$. Indeed,

$$\sum_{\boldsymbol{\gamma}\neq\mathbf{0}}\pi_{\boldsymbol{\gamma}} = \frac{1}{q^{16}-1}\sum_{\boldsymbol{\gamma}\neq\mathbf{0}}\left(\sum_{s=1}^{16}\mathbf{1}_{s=|\boldsymbol{\gamma}|}\right)\sigma^{|\boldsymbol{\gamma}|} = \frac{1}{q^{16}-1}\sum_{s=1}^{16}\binom{16}{s}\sigma^s = 1,$$

and therefore $\boldsymbol{\pi}$ is a probability distribution. Moreover, for any non-zero $\boldsymbol{\gamma}'$, $(\boldsymbol{\pi}\mathcal{M}')_{\boldsymbol{\gamma}'} = (q^{16}-1)^{-1}\sum_{\boldsymbol{\gamma}\neq\mathbf{0}} \mathrm{N}[\boldsymbol{\gamma},\boldsymbol{\gamma}'] = (q^{16}-1)^{-1}\sigma^{|\boldsymbol{\gamma}'|} = \pi_{\boldsymbol{\gamma}'}$, as the sum is simply the number of non-zero states that can be connected to some non-zero support $\boldsymbol{\gamma}'$ through LT, which is exactly the number of states of support equal to γ', as each state of support $\boldsymbol{\gamma}'$ has one and only one preimage through LT.

It is known [11] that $(\mathcal{M}'^r)_{\boldsymbol{\gamma},\boldsymbol{\gamma}'} \to \pi_{\boldsymbol{\gamma}'}$ when $r \to \infty$. As $\mathrm{E}[\mathrm{LP}^{\mathsf{AES}^*}(\boldsymbol{c_0},\boldsymbol{c_r})] = \sigma^{-|\boldsymbol{\gamma}_r|}(\mathcal{M}'^{r-1})_{\boldsymbol{\gamma}_0,\boldsymbol{\gamma}_r}$ for non-zero masks $\boldsymbol{c_0}$ and $\boldsymbol{c_r}$, we have proven the following theorem (which corresponds to the conjecture in [20]).

Theorem 8. *Let $\boldsymbol{c}_0$ and $\boldsymbol{c}_r$ be two non-zero masks in* $\mathrm{GF}(q)^{16}$*. Then*

$$\lim_{r\to\infty} \mathrm{E}[\mathrm{LP}^{\mathsf{AES}^*}(\boldsymbol{c}_0,\boldsymbol{c}_r)] = \frac{1}{q^{16}-1}\,. \tag{2}$$

We conclude this discussion by wondering *how fast* does the expected LP of AES^* tends towards $(q^{16}-1)^{-1}$. As $\mathcal{M}'$ is the transition matrix of a finite irreducible and aperiodic chain, the Perron-Frobenius Theorem [11] states that $\lambda_1 = 1$ is an eigenvalue of $\mathcal{M}'$, while the remaining eigenvalues $\lambda_2,\ldots,\lambda_m$ satisfy $|\lambda_j| < 1$. Assuming that $\lambda_1 > |\lambda_2| \geq \cdots \geq |\lambda_m|$, the rate of the convergence depends on $|\lambda_2|$. If we let λ be any real value such that $1 > \lambda > |\lambda_2|$, we deduce that for any non-zero masks $\boldsymbol{c}_0$ and $\boldsymbol{c}_r$, $\mathrm{E}[\mathrm{LP}^{\mathsf{AES}^*}(\boldsymbol{c}_0,\boldsymbol{c}_r)] = \frac{1}{q^{16}-1} + O(\lambda^r)$ when $r \to \infty$.

Note that the same results can be obtained on AES itself with independent round keys using almost the same proof. The only change is that one needs to prove that for any non-zero masks $\boldsymbol{a}$ and $\boldsymbol{b}$, there is a number of rounds r such that $\mathrm{LP}^{\mathtt{Round}_r\circ\cdots\circ\mathtt{Round}_1}(\boldsymbol{a},\boldsymbol{b}) \neq 0$. Equivalently, we can prove it with DP's by using results by Wernsdorf et al. [15, 42].

3.3 Combinatorial Tables on Supports

We will see that, thanks to the properties of LT, $\mathrm{N}[\boldsymbol{\gamma}_{i-1},\boldsymbol{\gamma}_i]$ only depends on the weights of the diagonals of $\boldsymbol{\gamma}_{i-1}$ and of the columns of $\boldsymbol{\gamma}_i$. We introduce notations to deal with Hamming weights of columns and diagonals. If $\boldsymbol{\gamma}_i$ is the ith support in a characteristic, we denote by $\mathbf{c}_i = (\mathsf{c}_{i,1},\mathsf{c}_{i,2},\mathsf{c}_{i,3},\mathsf{c}_{i,4})$ the vector of the four weights of $\boldsymbol{\gamma}_i$'s columns. Similarly, we denote by $\mathbf{d}_i = (\mathsf{d}_{i,1},\mathsf{d}_{i,2},\mathsf{d}_{i,3},\mathsf{d}_{i,4})$ the four weights of $\boldsymbol{\gamma}_i$'s diagonals. What we mean by columns and diagonals should be clear from Figure 1. Finally, we denote by $\mathbf{w}_j^i = (\mathbf{d}_i,\mathbf{c}_j)$ the *weight pattern* of a pair of supports $(\boldsymbol{\gamma}_i,\boldsymbol{\gamma}_j)$. Note that $|\mathbf{w}_j^i| = |\boldsymbol{\gamma}_i| + |\boldsymbol{\gamma}_j|$ and that this weight pattern only includes the weights of the diagonals of $\boldsymbol{\gamma}_i$ and of the columns of $\boldsymbol{\gamma}_j$. Consequently, if $\boldsymbol{\gamma}_{i-1}$ and $\boldsymbol{\gamma}_i$ are two successive masks in a characteristic, $\mathbf{w}_i^{i-1}$

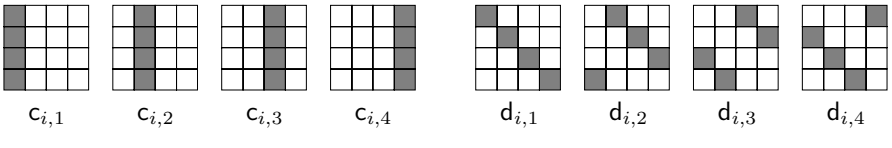

Fig. 1. The four column's and diagonal's weights of a state $\boldsymbol{\gamma}_i$

contains enough information to compute $\mathrm{N}[\boldsymbol{\gamma}_{i-1}, \boldsymbol{\gamma}_i]$ (as we will see in Corollary 10). We now recall a known fact about the weight distribution of MDS codes.

Theorem 9 (Theorem 7.4.1 in [16]). *Let $\mathcal{C}$ be an $[n, k, d]$ MDS code over $\mathrm{GF}(q)$. For $i = 0, \ldots, n$, the number A_i of codewords of weight i is given by $A_0 = 1$, $A_i = 0$ for $1 \leq i < d$ and $A_i = \binom{n}{i} \sum_{j=0}^{i-d} (-1)^j \binom{i}{j} (q^{i+1-d-j} - 1)$ for $d \leq i \leq n$, where $d = n - k + 1$.*

The `MixColumns` operation is a linear multipermutation [36], as the set of all codewords $(\boldsymbol{a}, \texttt{MixColumns}(\boldsymbol{a}))$ is a $[8, 4, 5]$ MDS code.

Corollary 10. *Let $\boldsymbol{\gamma}_{i-1}$ and $\boldsymbol{\gamma}_i$ be two successive supports of a characteristic and let $\mathbf{w}_i^{i-1} = (\mathbf{d}_{i-1}, \mathbf{c}_i)$ be their weight pattern. We have*

$$\mathrm{N}[\boldsymbol{\gamma}_{i-1}, \boldsymbol{\gamma}_i] = \prod_{s=1}^{4} \frac{A_{\mathsf{d}_{i-1,s}+\mathsf{c}_{i,s}}}{\binom{8}{\mathsf{d}_{i-1,s}+\mathsf{c}_{i,s}}}.$$

Thus, $\mathbf{w}_i^{i-1}$ is sufficient to compute $\mathrm{N}[\boldsymbol{\gamma}_{i-1}, \boldsymbol{\gamma}_i]$ so that we will denote this value by $\mathrm{N}[\mathbf{w}_i^{i-1}]$. By symmetry, it is clear that an arbitrary permutation applied on both the diagonal's and column's weights of $\mathbf{w}_i^{i-1}$ will not change the value of $\mathrm{N}[\mathbf{w}_i^{i-1}]$, i.e., if two weight patterns $\mathbf{w} = (\mathbf{d}, \mathbf{c})$ and $\mathbf{w}' = (\mathbf{d}', \mathbf{c}')$ are such that

$$(\mathsf{d}_1, \mathsf{d}_2, \mathsf{d}_3, \mathsf{d}_4, \mathsf{c}_1, \mathsf{c}_2, \mathsf{c}_3, \mathsf{c}_4) = (\mathsf{d}'_{\pi(1)}, \mathsf{d}'_{\pi(2)}, \mathsf{d}'_{\pi(3)}, \mathsf{d}'_{\pi(4)}, \mathsf{c}'_{\pi(1)}, \mathsf{c}'_{\pi(2)}, \mathsf{c}'_{\pi(3)}, \mathsf{c}'_{\pi(4)})$$

for some permutation π of $[1, 4]$, then $\mathrm{N}[\mathbf{w}] = \mathrm{N}[\mathbf{w}']$. It is natural to consider such weight patterns as equivalent and to choose a unique representative for each equivalence class. We arbitrarily choose to take the greatest element in the sense of the lexicographic order as the representative and denote it $\overline{\mathbf{w}}$. The number of elements in the equivalence class of $\overline{\mathbf{w}}$ will be denoted $\mathrm{C}[\overline{\mathbf{w}}]$. By the end of this section, we will be summing over weight patterns of supports surrounding the linear transformation LT (Theorem 12) instead of supports between individual rounds (Theorem 6). It will be natural to link both concepts. Given two successive weight patterns $\mathbf{w}_i^{i-1} = (\mathbf{d}_{i-1}, \mathbf{c}_i)$ and $\mathbf{w}_{i+1}^{i} = (\mathbf{d}_i, \mathbf{c}_{i+1})$, we denote by $\mathrm{P}[\mathbf{w}_i^{i-1}, \mathbf{w}_{i+1}^{i}]$ the number of supports $\boldsymbol{\gamma}$ (between rounds i and $i+1$) compatible with these weight patterns, i.e., the number of supports $\boldsymbol{\gamma}$ of weight pattern $(\mathbf{d}, \mathbf{c})$ such that $\mathbf{d} = \mathbf{d}_i$ and $\mathbf{c} = \mathbf{c}_i$ (see Figure 2). In other words, table $\mathrm{P}[\cdot, \cdot]$ gives the number of possible supports with given Hamming weights of the columns and of the diagonals. We note that by shifting columns, this is equivalent to counting 4×4 binary matrices with given weights for every row and column. Consequently, $\mathrm{P}[\mathbf{w}_i^{i-1}, \mathbf{w}_{i+1}^{i}]$ remains unchanged by permuting the weight of the diagonals given by $\mathbf{c}_i$ and/or the weight of the columns given by $\mathbf{d}_i$.

Lemma 11. *Let $(\boldsymbol{\gamma}_{i-1}, \boldsymbol{\gamma}_i, \boldsymbol{\gamma}_{i+1})$ be a characteristic of supports on two rounds, let $\mathbf{w}_i^{i-1} = (\mathbf{d}_{i-1}, \mathbf{c}_i)$ and $\mathbf{w}_{i+1}^{i} = (\mathbf{d}_i, \mathbf{c}_{i+1})$ be the weight pattern of $(\boldsymbol{\gamma}_{i-1}, \boldsymbol{\gamma}_i)$ and $(\boldsymbol{\gamma}_i, \boldsymbol{\gamma}_{i+1})$ respectively, and let $\overline{\mathbf{w}}_i^{i-1}$ and $\overline{\mathbf{w}}_{i+1}^{i}$ be their representatives. Then $\mathrm{N}[\mathbf{w}_i^{i-1}] = \mathrm{N}[\overline{\mathbf{w}}_i^{i-1}]$, $\mathrm{P}[\mathbf{w}_i^{i-1}, \mathbf{w}_{i+1}^{i}] = \mathrm{P}[\overline{\mathbf{w}}_i^{i-1}, \overline{\mathbf{w}}_{i+1}^{i}]$, and $|\mathbf{w}_i^{i-1}| = |\overline{\mathbf{w}}_i^{i-1}|$.*

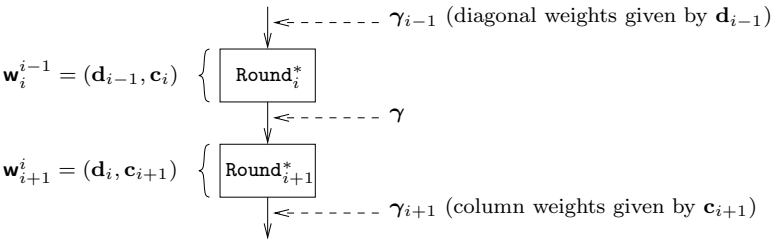

Fig. 2. Given $\mathbf{w}_i^{i-1}$ and $\mathbf{w}_{i+1}^{i}$, there are $\mathrm{P}[\mathbf{w}_i^{i-1}, \mathbf{w}_{i+1}^{i}]$ compatible supports $\boldsymbol{\gamma}$'s

3.4 From Sums over Supports to Sums over Weight Pattern Representatives

Theorem 12. *Let $\boldsymbol{c}_0$ and $\boldsymbol{c}_r$ be two masks in $\mathrm{GF}(q)^{16}$ of support $\boldsymbol{\gamma}_0$ and $\boldsymbol{\gamma}_r$ respectively. Let $\mathbf{d}_0$ denote the weight vector of $\boldsymbol{\gamma}_0$'s diagonals and let $\mathbf{c}_r$ denote the weight vector of $\boldsymbol{\gamma}_r$'s columns. Let $\sigma = q - 1$. Let $\mathcal{L}$ be the square matrix indexed by weight patterns representatives, defined by*

$$\mathcal{L}_{\overline{\mathbf{u}},\overline{\mathbf{v}}} = \mathrm{R}[\overline{\mathbf{u}}]\mathrm{P}[\overline{\mathbf{u}},\overline{\mathbf{v}}]\mathrm{R}[\overline{\mathbf{v}}] \quad \textit{where} \quad \mathrm{R}[\mathbf{u}] = \sqrt{\sigma^{\frac{1}{2}|\mathbf{u}|}\mathrm{C}[\mathbf{u}]\mathrm{N}[\mathbf{u}]}\ .$$

Finally, let $\mathcal{U}(\mathbf{d}_0)$ and $\mathcal{V}(\mathbf{c}_r)$ be the column vectors indexed by weight patterns representatives, defined by

$$\mathcal{U}(\mathbf{d}_0)_{\overline{\mathbf{v}}} = \sigma^{-\frac{1}{2}|\mathbf{d}_0|}\mathrm{R}[\overline{\mathbf{v}}]\mathrm{C}[\overline{\mathbf{v}}]^{-1} \sum_{\mathbf{u}=(\mathbf{d},\mathbf{c})} \mathbf{1}_{\overline{\mathbf{u}}=\overline{\mathbf{v}}}\mathbf{1}_{\mathbf{d}=\mathbf{d}_0} \quad \textit{and}$$

$$\mathcal{V}(\mathbf{c}_r)_{\overline{\mathbf{v}}} = \sigma^{-\frac{1}{2}|\mathbf{c}_r|}\mathrm{R}[\overline{\mathbf{v}}]\mathrm{C}[\overline{\mathbf{v}}]^{-1} \sum_{\mathbf{u}=(\mathbf{d},\mathbf{c})} \mathbf{1}_{\overline{\mathbf{u}}=\overline{\mathbf{v}}}\mathbf{1}_{\mathbf{c}=\mathbf{c}_r}\ .$$

Then the expected linear probability over $r > 1$ rounds of AES^* is

$$\mathrm{E}[\mathrm{LP}^{\mathsf{AES}^*}(\boldsymbol{c}_0, \boldsymbol{c}_r)] = \mathcal{U}(\mathbf{d}_0)^T \times \mathcal{L}^{r-2} \times \mathcal{V}(\mathbf{c}_r)\ .$$

Proof. For simplicity, $\mathrm{E}[\mathrm{LP}^{\mathsf{AES}^*}(\boldsymbol{c}_0, \boldsymbol{c}_r)]$ will simply be denoted $\mathrm{ELP}(\boldsymbol{c}_0, \boldsymbol{c}_r)$ and we will consider the case where $r > 2$. In Theorem 6, we had

$$\mathrm{ELP}(\boldsymbol{c}_0, \boldsymbol{c}_r) = \sigma^{-|\boldsymbol{\gamma}_r|} \sum_{\boldsymbol{\gamma}_1,\ldots,\boldsymbol{\gamma}_{r-2}} \prod_{i=1}^{r-1} \sigma^{-|\boldsymbol{\gamma}_{i-1}|}\mathrm{N}[\boldsymbol{\gamma}_{i-1}, \boldsymbol{\gamma}_i].$$

We notice that $2\sum_{i=1}^{r} |\boldsymbol{\gamma}_{i-1}| = |\mathbf{w}_0^r| + \sum_{i=1}^{r-1} |\mathbf{w}_i^{i-1}|$, where we used the fact that, as we do not need to take into account characteristics that give a zero linear probability, $\boldsymbol{\gamma}_{r-1} = \boldsymbol{\gamma}_r$ (see Lemma 5). From this and from Corollary 10, we deduce that $\mathrm{ELP}(\boldsymbol{c}_0, \boldsymbol{c}_r) = \sigma^{-\frac{1}{2}|\mathbf{w}_0^r|} \sum_{\boldsymbol{\gamma}_1,\ldots,\boldsymbol{\gamma}_{r-2}} \prod_{i=1}^{r-1} \mathrm{D}[\mathbf{w}_i^{i-1}]$, where $\mathrm{D}[\mathbf{w}] = \sigma^{\frac{1}{2}|\mathbf{w}|}\mathrm{N}[\mathbf{w}]$. As we want to consider weight patterns instead of supports, we introduce a new sum and permute both sums to obtain

$$\mathrm{ELP}(\boldsymbol{c}_0, \boldsymbol{c}_r) = \sigma^{-\frac{1}{2}|\mathbf{w}_0^r|} \sum_{\mathbf{u}_1^0,\ldots,\mathbf{u}_{r-1}^{r-2}} \left(\sum_{\boldsymbol{\gamma}_1,\ldots,\boldsymbol{\gamma}_{r-2}} \prod_{j=1}^{r-1} \mathbf{1}_{\mathbf{w}_j^{j-1}=\mathbf{u}_j^{j-1}} \right) \prod_{i=1}^{r-1} \mathrm{D}[\mathbf{u}_i^{i-1}]\ .$$

Denoting $\mathbf{u}_j^{j-1} = (\mathbf{d}'_{j-1}, \mathbf{c}'_j)$, it is easy to show that

$$\sum_{\gamma_1,\ldots,\gamma_{r-2}} \prod_{j=1}^{r-1} \mathbf{1}_{\mathbf{u}_j^{j-1}=\mathbf{w}_j^{j-1}} = \mathbf{1}_{\mathbf{d}'_0=\mathbf{d}_0} \mathbf{1}_{\mathbf{c}'_{r-1}=\mathbf{c}_{r-1}} \prod_{j=1}^{r-2} \sum_{\gamma_j} \mathbf{1}_{(\mathbf{d}_j,\mathbf{c}_j)=(\mathbf{d}'_j,\mathbf{c}'_j)} \cdot$$

As, by definition, $\mathrm{P}[\mathbf{u}_j^{j-1}, \mathbf{u}_{j+1}^{j}] = \sum_{\boldsymbol{\gamma}_j} \mathbf{1}_{(\mathbf{d}_j,\mathbf{c}_j)=(\mathbf{d}'_j,\mathbf{c}'_j)}$, this gives

$$\mathrm{ELP}(\boldsymbol{c}_0, \boldsymbol{c}_r) = \sigma^{-\frac{1}{2}|\mathbf{w}_0^r|} \sum_{\substack{\mathbf{u}_1^0,\ldots,\mathbf{u}_{r-1}^{r-2} \\ \mathbf{d}'_0=\mathbf{d}_0}} \mathbf{1}_{\mathbf{c}'_{r-1}=\mathbf{c}_{r-1}} \mathrm{D}[\mathbf{u}_{r-1}^{r-2}] \prod_{i=1}^{r-2} \mathrm{D}[\mathbf{u}_i^{i-1}] \mathrm{P}[\mathbf{u}_i^{i-1}, \mathbf{u}_{i+1}^{i}] \,.$$

We denote $\mathcal{L}_{\overline{\mathbf{u}},\overline{\mathbf{v}}} = \mathrm{C}[\overline{\mathbf{u}}]^{\frac{1}{2}} \mathrm{D}[\overline{\mathbf{u}}]^{\frac{1}{2}} \mathrm{P}[\overline{\mathbf{u}}, \overline{\mathbf{v}}] \mathrm{C}[\overline{\mathbf{v}}]^{\frac{1}{2}} \mathrm{D}[\overline{\mathbf{v}}]^{\frac{1}{2}}$ and $\mathrm{F}[\overline{\mathbf{u}}] = \mathrm{D}[\overline{\mathbf{u}}]^{\frac{1}{2}} \mathrm{C}[\overline{\mathbf{u}}]^{-\frac{1}{2}}$. Using Lemma 11, the last expression becomes

$$\mathrm{ELP}(\boldsymbol{c}_0, \boldsymbol{c}_r) = \sigma^{-\frac{1}{2}|\mathbf{w}_0^r|} \sum_{\mathbf{u}_1^0, \mathbf{u}_{r-1}^{r-2}} \mathbf{1}_{\mathbf{d}'_0=\mathbf{d}_0} \mathbf{1}_{\mathbf{c}'_{r-1}=\mathbf{c}_{r-1}} \mathrm{F}[\overline{\mathbf{u}}_1^0] \mathrm{F}[\overline{\mathbf{u}}_{r-1}^{r-2}] (\mathcal{L}^{r-2})_{\overline{\mathbf{u}}_1^0, \overline{\mathbf{u}}_{r-1}^{r-2}} \,.$$

Introducing $(\mathcal{U}(\mathbf{d}_0))_{\overline{\mathbf{u}}_1^0}$ and $(\mathcal{V}(\mathbf{c}_{r-1}))_{\overline{\mathbf{u}}_{r-1}^{r-2}}$ in the previous expression leads (as $\boldsymbol{c}_{r-1} = \boldsymbol{c}_r$) to the announced result. □

In order to evaluate the complexity of the matrix multiplication of Theorem 12, we need to evaluate the size of the matrices, i.e., the number of equivalence classes. There are $20475 \approx 2^{14.33}$ such classes. Yet, it is not necessary to consider those equivalence classes for which $\mathrm{N}[\cdot]$ is 0. It can be checked that the number of remaining equivalence classes is $1001 \approx 2^{10}$. The computation of $\mathcal{L}^{r-1}$ therefore roughly takes $2^{30} \cdot \log r$ operations, which is feasible on standard computers.

3.5 Experimental Linear Hull for Various S-Box Sizes

Theorems 6 and 12 remain valid with several sizes of S-boxes. We implemented the computation of Theorem 12 with various sizes, our experimental results were obtained using GMP [10]. They are shown in Table 1. It appears that 4 rounds

Table 1. $\max_{\boldsymbol{a},\boldsymbol{b}} \mathrm{E}[\mathrm{LP}^{\mathsf{AES}^*}(\boldsymbol{a}, \boldsymbol{b})]$ for various number of rounds r and S-box sizes

r	2	3	4	5	6	7	8	9
3 bits	$2^{-13.2294}$	$2^{-19.6515}$	$2^{-44.9177}$	$2^{-44.9177}$	$2^{-47.3861}$	$2^{-47.9966}$	$2^{-47.9999}$	$2^{-48.0}$
4 bits	$2^{-17.6276}$	$2^{-27.3482}$	$2^{-62.5102}$	$2^{-62.5102}$	$2^{-63.9852}$	$2^{-63.9999}$	$2^{-63.9999}$	$2^{-64.0}$
5 bits	$2^{-21.8168}$	$2^{-34.6793}$	$2^{-79.2671}$	$2^{-79.2671}$	$2^{-79.9999}$	$2^{-79.9999}$	$2^{-79.9999}$	$2^{-80.0}$
6 bits	$2^{-25.9091}$	$2^{-41.8409}$	$2^{-95.6364}$	$2^{-95.6364}$	$2^{-95.9999}$	$2^{-95.9999}$	$2^{-96.0}$	$2^{-96.0}$
7 bits	$2^{-29.9547}$	$2^{-48.9207}$	$2^{-111.8189}$	$2^{-111.8189}$	$2^{-111.9999}$	$2^{-111.9999}$	$2^{-112.0}$	$2^{-112.0}$
8 bits	$2^{-33.9774}$	$2^{-55.9605}$	$2^{-127.9096}$	$2^{-127.9096}$	$2^{-127.9999}$	$2^{-127.9999}$	$2^{-128.0}$	$2^{-128.0}$

are enough to provide security against LC. We do not provide any result for the case where the S-box acts on 2 bit elements as it is impossible to find a 4×4 matrix with elements in $\mathrm{GF}(2^2)$ such that `MixColumns` stays a multipermutation. A second independent implementation of the computation was implemented in Maple [23] in order to obtain perfect results instead of floating point numbers. It was used for the masks presenting the maximum expected LP in Table 1.

4 Expected DP on AES*

Just as the efficiency of LC can be measured by means of LP's, the efficiency of DC can be measured by means of DP's [30]. If C is some fixed permutation on $\mathcal{S}$ and if $\boldsymbol{a}$ and $\boldsymbol{b}$ are two masks, the differential probability is given by $\mathrm{DP}^{\mathsf{C}}(\boldsymbol{a}, \boldsymbol{b}) = \Pr_{\boldsymbol{X} \in \mathcal{S}}[\mathsf{C}(\boldsymbol{X} \oplus \boldsymbol{a}) = \mathsf{C}(\boldsymbol{X}) \oplus \boldsymbol{b}]$, where the probability holds over the uniform distribution of $\boldsymbol{X}$. Here, $\boldsymbol{a}$ (resp. $\boldsymbol{b}$) represents the input (resp. output) difference between the pair of plaintexts (resp. ciphertexts). The computations that we performed on the expected LP of AES^* can be applied, with almost no modification, in order to compute the expected DP. The major modification concerns Lemma 1. We provide here its version for the DP.

Lemma 13. *Let* M *denote an arbitrary* 16 *by* 16 *matrix of elements in* $\mathrm{GF}(q)$*, representing a linear transformation on* AES *states (considered as column vectors). If the difference between two inputs of this transformation is equal to* $\boldsymbol{a}$*, then the output difference is equal to* $\mathsf{M} \times \boldsymbol{a}$*.*

We now follow the steps that lead to the final result on the LP coefficient and see whether they apply to the DP coefficient. Lemma 2 applies to the DP coefficient, and therefore, it is also the case for Lemma 3 (where we use the independence of the the 16 inputs on the S-boxes in order to obtain a product of DP, instead of using Matsui's Piling-up Lemma). Because the relation between an input difference on the linear transformation of AES and its output difference is not the same as in the case where we considered input/output masks, it looks as if we must replace LT^T by LT^{-1} in Definition 4. But according to Theorem 9, the actual *values* of $\mathrm{N}[\cdot]$ do not depend on which multipermutation is used, it just needs to be one. In other words, replacing LT^T by LT^{-1} in the definition of $\mathrm{N}[\cdot]$ does not change its entries. The computations on the LP coefficient thus still apply for the DP coefficient. Theorems 6, 8, and 12 apply to the DP, the numerical results given in Table 1 being exactly the same.

5 Extension to Iterated Attacks of Order 1

In the Luby-Rackoff model [22], an adversary $\mathcal{A}$ has an unlimited computational power, but has limited access to an oracle $\mathcal{O}$. The oracle implements either an instance of a given cipher C (such as AES^*) or of the perfect cipher C^*, the objective of the adversary being to guess which is the case (see Figure 3). Eventually, the adversary will output 1 (resp. 0) if his guess is that the oracle implements C (resp. C^*). Denoting by $\Pr[\mathcal{A}^{\mathcal{O}} \to 1]$ the probability

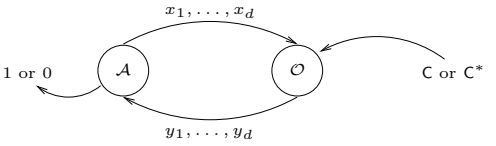

Fig. 3. An adversary $\mathcal{A}$ limited to d questions to an oracle $\mathcal{O}$

that the adversary outputs 1 depending on the oracle $\mathcal{O}$, his ability to distinguish C from C^* is measured by means of the advantage $\mathrm{Adv}_{\mathcal{A}} = |\Pr[\mathcal{A}^{\mathsf{C}} \rightarrow 1] - \Pr[\mathcal{A}^{\mathsf{C}^*} \rightarrow 1]|$. The most powerful adversary will select his d queries depending on the previous answers of the oracle. Such an adversary is called a *d-limited adaptative distinguisher* [39]. The advantage of the best distinguisher of this type is such that $\mathrm{Adv}_{\mathcal{A}} = \frac{1}{2} \parallel [\mathsf{C}]^d - [\mathsf{C}^*]^d \parallel_a$, where $[\mathsf{C}]^d$ is the d-wise distribution matrix[7] of the random permutation C over $\mathcal{S}$, and where $\parallel M \parallel_a = \max_{x_1} \sum_{y_1} \cdots \max_{x_d} \sum_{y_d} |M_{(x_1,\ldots,x_d),(y_1,\ldots,y_d)}|$ for any $\#\mathcal{S}^d \times \#\mathcal{S}^d$ matrix M (Theorem 11 in [39]). Proving the resistance of C against such a $2d$-limited distinguisher is sufficient to prove its resistance against any iterated attacks of order d (Theorem 18 in [39]). Using Theorem 14, we bound the advantage of the best 2-limited adaptative distinguisher and deduce the number rounds necessary to resist any iterated attacks of order 1.

Theorem 14. *Let C be a random permutation over $\{0,1\}^n$. If ϵ is the non-negative value such that $\epsilon = \max_{\boldsymbol{a} \neq 0, \boldsymbol{b}} \mathrm{E}[\mathrm{DP}^{\mathsf{C}}(\boldsymbol{a}, \boldsymbol{b})] - \frac{1}{2^n - 1}$, we have $\parallel [\mathsf{C}']^2 - [\mathsf{C}^*]^2 \parallel_a \leq 2^n \varepsilon$ where $\mathsf{C}'(x) = \mathsf{C}(x \oplus K_1) \oplus K_2$ with independent and uniformly distributed K_1 and K_2.*

Proof. Let $x_1, x_2, y_1, y_2 \in \{0,1\}^n$. Starting from the definition of $[\mathsf{C}']^2$, we have

$$[\mathsf{C}']^2_{(x_1,x_2),(y_1,y_2)} = \sum_{\mathsf{c}} \Pr_{K_1,K_2} \begin{bmatrix} \mathsf{c}(x_1 \oplus K_1) = y_1 \oplus K_2 \\ \mathsf{c}(x_2 \oplus K_1) = y_2 \oplus K_2 \end{bmatrix} \Pr[\mathsf{C} = \mathsf{c}],$$

as C is independent from (K_1, K_2). Furthermore, we have

$$\Pr_{K_1,K_2} \begin{bmatrix} \mathsf{c}(x_1 \oplus K_1) = y_1 \oplus K_2 \\ \mathsf{c}(x_2 \oplus K_1) = y_2 \oplus K_2 \end{bmatrix} = \sum_{u,v} \mathbf{1}_{\substack{x_1 \oplus x_2 = u \\ y_1 \oplus y_2 = v}} \Pr_{K_1,K_2} \begin{bmatrix} \mathsf{c}(K_1) = K_2 \\ \mathsf{c}(u \oplus K_1) = v \oplus K_2 \end{bmatrix}$$

where the probability in the sum is equal to

$$2^{-2n} \sum_{k_1,k_2} \mathbf{1}_{\substack{\mathsf{c}(k_1 \oplus u) = k_2 \oplus v \\ \mathsf{c}(k_1) = k_2}} = 2^{-n} \Pr_{K_1}[\mathsf{c}(K_1) \oplus \mathsf{c}(K_1 \oplus u) = v] = 2^{-n} \mathrm{DP}^{\mathsf{c}}(u, v) \; .$$

Therefore, $[\mathsf{C}']^2_{(x_1,x_2),(y_1,y_2)} = 2^{-n} \mathrm{E}_{\mathsf{C}}[\mathrm{DP}^{\mathsf{C}}(x_1 \oplus x_2, y_1 \oplus y_2)]$. As the sum of the $\mathrm{DP}^{\mathsf{C}^*}$ on the input mask is 1 (as C^* is a permutation), $\mathrm{E}_{\mathsf{C}^*}[\mathrm{DP}^{\mathsf{C}^*}(x_1 \oplus x_2, y_1 \oplus$

[7] Recall that the d-wise distribution matrix of a random function F is such that $[F]^d_{(x_1,\ldots,x_d),(y_1,\ldots,y_d)}$ is the probability that $F(x_i) = y_i$ for all $i = 1, \ldots, d$.

Table 2. Values of ϵ depending of the number of rounds r

r	2	3	4	5	6	7	8	9	10
ϵ	$2^{-33.98}$	$2^{-55.96}$	$2^{-131.95}$	$2^{-131.95}$	$2^{-152.17}$	$2^{-174.74}$	$2^{-200.39}$	$2^{-223.93}$	$2^{-270.82}$

$y_2)] = \frac{1}{2^n-1}$ when $x_1 \neq x_2$ (when $x_1 = x_2$, the DP value is always 0, except when $y_1 \oplus y_2$ is also 0, in which case DP is 1). From the last two equations we deduce $[\mathsf{C}']^2_{(x_1,x_2),(y_1,y_2)} - [\mathsf{C}^*]^2_{(x_1,x_2),(y_1,y_2)} = 2^{-n}\left(\mathrm{E}_{\mathsf{C}}[\mathrm{DP}^{\mathsf{C}}(x_1 \oplus x_2, y_1 \oplus y_2)] - \frac{1}{2^n-1}\right)$, and thus, by definition of the $||\cdot||_a$ norm, $||[\mathsf{C}']^2 - [\mathsf{C}^*]^2||_a$ is upper bounded by $2^{-n}\sum_{y_1,y_2}\max_{x_1\neq x_2}|\mathrm{E}_{\mathsf{C}}[\mathrm{DP}^{\mathsf{C}}(x_1 \oplus x_2, y_1 \oplus y_2)] - (2^n-1)^{-1}| = 2^n\varepsilon$. □

Such an ϵ always exists, as the maximum DP (or LP) value is always larger or equal to $1/(2^n-1)$. Experimental results on ϵ (obtained both with our GMP and Maple implementations) are given in Table 2 for several number of rounds. We conclude that provable security is achieved for 10 rounds of AES^* (which substantially improves [28], where it is shown that 384 rounds are enough).

6 Derandomizing the S-Boxes

We note that all results presented so far hold if replace the uniformly distributed random S-box S^* by *any* random S-box S, provided that it satisfies $\mathrm{E}_{\mathsf{S}}[\mathrm{LP}^{\mathsf{S}}(a,b)] = \sigma^{-1}$ (which is proved for S^* in Lemma 2). According to Lemma 14 in [39],

$$\mathrm{E}_{\mathsf{S}}[\mathrm{LP}^{\mathsf{S}}(a,b)] = q^{-2}\sum_{\substack{x_1,x_2\\y_1,y_2}}(-1)^{(x_1\oplus x_2)\bullet a+(y_1\oplus y_2)\bullet b}\Pr[\mathsf{S}(x_1)=y_1,\mathsf{S}(x_2)=y_2].$$

Hence, $\mathrm{E}_{\mathsf{S}}[\mathrm{LP}^{\mathsf{S}}(a,b)]$ only depends on the pairwise distribution. If S has a perfect pairwise decorrelation, we deduce $\mathrm{E}_{\mathsf{S}}[\mathrm{LP}^{\mathsf{S}}(a,b)] = \sigma^{-1}$. In order to construct such a variant of AES, one can just insert a `MulRoundKey` operation before each `AddRoundKey` of AES, with independent subkeys, where `MulRoundKey` is the component-wise product in $\mathrm{GF}(q)$ of an input state and a subkey, i.e., considering the three states $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{k}$ as a one-dimensional vectors of 16 bytes,

$$\boldsymbol{b} = \texttt{MulRoundKey}(\boldsymbol{a}, \boldsymbol{k}) \quad \Leftrightarrow \quad b_i = a_i \times k_i \quad \text{for } i = 1, \ldots, 16.$$

Note that all the component of a subkey $\boldsymbol{k}$ used in a `MulRoundKey` operation have to be non-zero to preserve bijectivity.

7 Discussion and Conclusion

We studied the SPN on which AES is based using a Luby-Rackoff-like approach. Following [20] and [28], we considered that the only "round function" that can reasonably be replaced by a random one is the S-box. We chose to replace the S-boxes by random and *independent* permutations. In this model, we computed the

exact (i.e., using neither heuristic approximations nor bounds) hull and differential average probabilities. Clearly, a better model (i.e., intuitively closer to the real AES) would be to choose *one* permutation at random and use it throughout the whole cipher, although it is not clear to us that one can easily prove similar security results in that case. Obviously, we cannot draw direct consequences on the security of AES. At least we get some increased confidence in its high-level structure and claim that AES with independent keys has no useful linear hull nor differentials, unless the S-box structure selection is really unfortunate. We also pushed the analysis further by studying iterated attacks of order 1. We showed that ten inner rounds are sufficient to ensure the security of AES* against any attack of this kind. Finally, we proved the (non-surprising) convergence of AES* towards the perfect cipher (as far as LC and DC are concerned) as the number of rounds increases, which was only conjectured so far.

Acknowledgments. We would like to thank the anonymous referees, Pascal Junod, and Matthieu Finiasz for helpful comments, as well as Ralph Wernsdorf for quite useful references.

References

[1] T. Baignères, P. Junod, and S. Vaudenay. How far can we go beyond linear cryptanalysis? In P.J. Lee, editor, *Advances in Cryptology* - ASIACRYPT *'04*, volume 3329 of *LNCS*, pages 432–450. Springer-Verlag, 2004.

[2] E. Biham and A. Shamir. Differential cryptanalysis of DES-like cryptosystems. *Journal of Cryptology*, 4:3–72, 1991.

[3] E. Biham and A. Shamir. Differential cryptanalysis of the full 16-round DES. In E.F. Brickell, editor, *Advances in Cryptology* - CRYPTO *'92*, volume 740 of *LNCS*, pages 487–496. Springer-Verlag, 1993.

[4] F. Chabaud and S. Vaudenay. Links between differential and linear cryptanalysis. In A. De Santis, editor, *Advances in Cryptology* - EUROCRYPT *'94*, volume 950 of *LNCS*, pages 356–365. Springer-Verlag, 1995.

[5] Z.G. Chen and S.E. Tavares. Towards provable security of substitution-permutation encryption networks. In S.E. Tavares and H. Meijer, editors, *Selected Areas in Cryptography,* SAC *'98*, volume 1556 of *LNCS*, pages 43–56. Springer-Verlag, 1999.

[6] J. Daemen and V. Rijmen. AES proposal: Rijndael. NIST AES Proposal, 1998.

[7] J. Daemen and V. Rijmen. *The Design of Rijndael.* Information Security and Cryptography. Springer-Verlag, 2002.

[8] H. Feistel. Cryptography and computer privacy. *Scientific American*, 228:15–23, 1973.

[9] H. Gilbert and M. Minier. New results on the pseudorandomness of some block-cipher constructions. In M. Matsui, editor, *Fast Software Encryption* - FSE *'01*, volume 2355 of *LNCS*, pages 248–266. Springer-Verlag, 2002.

[10] GMP. GNU Multiple Precision arithmetic library. `http://www.swox.com/gmp`.

[11] G. Grimmett and D. Stirzaker. *Probability and Random Processes.* Oxford University Press, 3d edition, 2001.

[12] O. Häggström. *Finite Markov Chains and Algorithmic Applications.* London Mathematical Society Student Texts. Cambridge University Press, 2002.

[13] H.M. Heys and S.E. Tavares. Substitution-permutation networks resistant to differential and linear cryptanalysis. *Journal of Cryptology*, 9(1):1–19, 1996.
[14] S. Hong, S. Lee, J. Lim, J. Sung, D. Cheon, and I. Cho. Provable security against differential and linear cryptanalysis for the SPN structure. In B. Schneier, editor, *Fast Software Encryption* - FSE*'00*, volume 1978 of *LNCS*, pages 273–283. Springer-Verlag, 2001.
[15] G. Hornauer, W. Stephan, and R. Wernsdorf. Markov ciphers and alternating groups. In T. Helleseth, editor, *Advances in Cryptology* - EUROCRYPT*'93*, volume 765 of *LNCS*, pages 453–460. Springer-Verlag, 1994.
[16] W.C. Huffman and V.S. Pless. *Fundamentals of Error-Correcting Codes*. Cambridge University Press, 2003.
[17] L. Keliher. Refined analysis of bounds related to linear and differential cryptanalysis for the AES. In H. Dobbertin, V. Rijmen, and A. Sowa, editors, *Fourth Conference on the Advanced Encryption Standard* - AES*4*, volume 3373 of *LNCS*, pages 42–57. Springer-Verlag, 2005.
[18] L. Keliher, H. Meijer, and S.E. Tavares. Improving the upper bound on the maximum average linear hull probability for Rijndael. In S. Vaudenay and A.M. Youssef, editors, *Selected Areas in Cryptography,* SAC*'01*, volume 2259 of *LNCS*, pages 112–128. Springer-Verlag, 2001.
[19] L. Keliher, H. Meijer, and S.E. Tavares. New method for upper bounding the maximum average linear hull probability for SPNs. In B. Pfitzmann, editor, *Advances in Cryptology* - EUROCRYPT*'01*, volume 2045 of *LNCS*, pages 420–436. Springer-Verlag, 2001.
[20] L. Keliher, H. Meijer, and S.E. Tavares. Toward the true random cipher: On expected linear probability values for SPNs with randomly selected S-boxes. In V. Bhargava, H.V. Poor, V. Tarokh, and S. Yoon, editors, *Communication, Information and Network Security*, pages 123–146. Kluwer Academic Publishers, 2003.
[21] X. Lai, J. Massey, and S. Murphy. Markov ciphers and differential cryptanalysis. In D.W. Davies, editor, *Advances in Cryptology* - EUROCRYPT*'91*, volume 547 of *LNCS*, pages 17–38. Springer-Verlag, 1991.
[22] M. Luby and C. Rackoff. How to construct pseudorandom permutations from pseudorandom functions. *SIAM Journal on Computing*, 17(2):373–386, 1988.
[23] Maplesoft. Maple 9. `http://www.maplesoft.com/`.
[24] M. Matsui. The first experimental cryptanalysis of the Data Encryption Standard. In Y.G. Desmedt, editor, *Advances in Cryptology* - CRYPTO*'94*, volume 839 of *LNCS*, pages 1–11. Springer-Verlag, 1994.
[25] M. Matsui. Linear cryptanalysis method for DES cipher. In T. Helleseth, editor, *Advances in Cryptology* - EUROCRYPT*'93*, volume 765 of *LNCS*, pages 386–397. Springer-Verlag, 1994.
[26] M. Matsui. New structure of block ciphers with provable security against differential and linear cryptanalysis. In D. Gollmann, editor, *Fast Software Encryption* - FSE*'96*, volume 1039 of *LNCS*, pages 205–218. Springer-Verlag, 1996.
[27] U. Maurer and K. Pietrzak. The security of many-round Luby-Rackoff pseudorandom permutations. In E. Biham, editor, *Advances in Cryptology* - EUROCRYPT*'03*, volume 2656 of *LNCS*, pages 544–561. Springer-Verlag, 2003.
[28] S. Moriai and S. Vaudenay. On the pseudorandomness of top-level schemes of block ciphers. In T. Okamoto, editor, *Advances in Cryptology* - ASIACRYPT*'00*, volume 1976 of *LNCS*, pages 289–302. Springer-Verlag, 2000.
[29] M. Naor and O. Reingold. On the construction of pseudorandom permutations: Luby-Rackoff revisited. *Journal of Cryptology*, 12(1):29–66, 1999.

[30] K. Nyberg. Perfect nonlinear S-boxes. In D.W. Davies, editor, *Advances in Cryptology* - EUROCRYPT *'91*, volume 547 of *LNCS*, pages 378–386. Springer-Verlag, 1991.

[31] K. Nyberg. Linear approximation of block ciphers. In A. De Santis, editor, *Advances in Cryptology* - EUROCRYPT *'94*, volume 950 of *LNCS*, pages 439–444. Springer-Verlag, 1995.

[32] L. O'Connor. Properties of linear approximation tables. In B. Preneel, editor, *Fast Software Encryption* - FSE *'94*, volume 1008 of *LNCS*, pages 131–136. Springer-Verlag, 1995.

[33] 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. In Y. Zheng, editor, *Advances in Cryptology* - ASIACRYPT *'02*, volume 2501 of *LNCS*, pages 176–191. Springer-Verlag, 2002.

[34] 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. In T. Johansson, editor, *Fast Software Encryption* - FSE *'03*, volume 2887 of *LNCS*, pages 247–260. Springer-Verlag, 2003.

[35] J. Patarin. Security of random Feistel schemes with 5 or more rounds. In M. Franklin, editor, *Advances in Cryptology* - CRYPTO *'04*, volume 3152 of *LNCS*, pages 106–122. Springer-Verlag, 2004.

[36] S. Vaudenay. On the need for multipermutations: Cryptanalysis of MD4 and SAFER. In B. Preneel, editor, *Fast Software Encryption* - FSE *'94*, volume 1008 of *LNCS*, pages 286–297. Springer-Verlag, 1995.

[37] S. Vaudenay. On the security of CS-cipher. In L. Knudsen, editor, *Fast Software Encryption* - FSE *'99*, volume 1636 of *LNCS*, pages 260–274. Springer-Verlag, 1999.

[38] S. Vaudenay. On the Lai-Massey scheme. In L. Kwok Yan, O. Eiji, and X. Chaoping, editors, *Advances in Cryptology* - ASIACRYPT *'99*, volume 1716 of *LNCS*, pages 8–19. Springer-Verlag, 2000.

[39] S. Vaudenay. Decorrelation: a theory for block cipher security. *Journal of Cryptology*, 16(4):249–286, 2003.

[40] J. von zur Gathen and J. Gerhard. *Modern Computer Algebra*. Cambridge University Press, 2nd edition, 2003. First published 1999.

[41] D. Wagner. Towards a unifying view of block cipher cryptanalysis. In B. Roy and W. Meier, editors, *Fast Software Encryption* - FSE *'04*, volume 3017 of *LNCS*, pages 16–33. Springer-Verlag, 2004.

[42] R. Wernsdorf. The round functions of Rijndael generate the alternating group. In J. Daemen and V. Rijmen, editors, *Fast Software Encryption* - FSE *'02*, volume 2365 of *LNCS*, pages 143–148. Springer-Verlag, 2002.