The Wide Trail Design Strategy

Joan Daemen[1] and Vincent Rijmen[2]

[1] ProtonWorld, Zweefvliegtuigstraat 10, B-1130 Brussel, Belgium
Joan.Daemen@protonworld.com
[2] CRYPTOMAThIC, Lei 8A, B-3001 Leuven, Belgium
Vincent.Rijmen@cryptomathic.com

Abstract. We explain the theoretical background of the wide trail design strategy, which was used to design Rijndael, the Advanced Encryption Standard (AES). In order to facilitate the discussion, we introduce our own notation to describe differential and linear cryptanalysis. We present a block cipher structure and prove bounds on the resistance against differential and linear cryptanalysis.

1 Introduction

The development of differential [2] and linear cryptanalysis [7] has led to several design theories for block ciphers. The most important requirement for a new cipher is that it resists state-of-the-art cryptanalytic attacks. Preferably, this can be demonstrated in a rigorous, mathematical way. The second requirement is a good performance and an acceptable 'cost', in terms of CPU requirements, memory requirements, ...

The Wide trail strategy is an approach to design the round transformations of block ciphers that combine efficiency and resistance against differential and linear cryptanalysis. The strategy has been used in the design of Rijndael, the block cipher which has been selected to become the Advanced Encryption Standard (AES). In this article we describe the application of the strategy to the design of a certain type of block ciphers only, but the strategy can easily be extended to more general block cipher structures. Moreover, the wide trail strategy can also be applied to the design of synchronous stream ciphers and hash functions.

In order to explain the wide trail strategy, we introduce our own notation for differential and linear cryptanalysis. We are convinced that a good notation helps to understand the reasonings, and our notation is suited very well to understand the wide trail strategy. We introduce a general block cipher model and explain how linear correlations and difference propagation probabilities are built up in block ciphers designed according to this model. Subsequently, we explain the basic principles of the wide trail strategy and introduce our new diffusion measure, the *branch number*. We explain its relevance in providing bounds for the probability of differential trails and the correlation of linear trails over two rounds. We then introduce a cipher structure that combines efficiency with high resistance against linear and differential cryptanalysis. The resistance against linear and differential cryptanalysis is based on a theorem that lower bounds the diffusion after four rounds of the cipher structure. In this paper, we emphasize the theoretical foundations of the wide trail design strategy. More explanation about the practical constructions can be found in [4].

In the following, the symbols $+$ and $\sum$ are used to denote bit-wise addition (XOR). The results can be generalized to other definitions for addition.

2 A General Block Cipher Model

We introduce a model for block ciphers that can be analyzed easily for their resistance against linear and differential cryptanalysis.

2.1 Key-alternating block ciphers

A *block cipher* transforms *plaintext blocks* of a fixed length n_{b} to *ciphertext blocks* of the same length under the influence of a key k. An *iterative block cipher* is defined as the application of a number of key-dependent Boolean permutations. The Boolean transformations are called the *round transformations*. Every application of a round transformation is called a *round*. We denote the number of rounds by r. We have:

$$\beta[k] = \rho^{(r)}[k^{(r)}] \circ \cdots \circ \rho^{(2)}[k^{(2)}] \circ \rho^{(1)}[k^{(1)}] \ . \tag{1}$$

In this expression, $\rho^{(i)}$ is called the i-th *round* of the block cipher and $k^{(i)}$ is called the i-th round key. For instance, the DES has 16 rounds. Every round uses the same round transformation, so we say there is only one round transformation. The round keys are computed from the cipher key. Usually, this is specified with an algorithm, called the *key schedule*.

A *key-alternating block cipher* is an iterative block cipher with the following properties:

- Alternation: the cipher is defined as the alternated application of key-independent round transformations and the application of a round key. The first round key is applied before the first round and the last round key is applied after the last round.
- Binary Key Addition: the round keys are applied by means of a simple XOR: to each bit of the intermediate state a round key bit is XORed.

We have:

$$\beta[k] = \sigma[k^{(r)}] \circ \rho^{(r)} \circ \sigma[k^{(r-1)}] \circ \cdots \circ \ \sigma[k^{(1)}] \circ \rho^{(1)} \circ \sigma[k^{(0)}] \ . \tag{2}$$

As, hopefully, will become clear soon, key-alternating block ciphers lend themselves very well to analysis with respect to the resistance against cryptanalysis.

2.2 The $\gamma\lambda$ round structure

In the wide trail strategy, the round transformations are composed of two invertible steps:

- γ: a local non-linear transformation. By local, we mean that any output bit depends on only a limited number of input bits and that neighboring output bits depend on neighboring input bits.
- λ: a linear mixing transformation providing high diffusion. What is meant by high diffusion will be explained in the following sections.

Hence we have a round transformation ρ:

$$\rho = \lambda \circ \gamma. \tag{3}$$

and refer to this as a $\gamma\lambda$ round transformation.

A typical construction for γ is the so-called *bricklayer mapping* consisting of a number of invertible S-boxes. In this construction, the bits of input vector a are partitioned into n_t m-bit *bundles* $a_i \in Z_2^m$ with $i \in \mathcal{I}$ by the so-called *bundle partition*. $\mathcal{I}$ is called the *index space*. Clearly, the inverse of γ consists of applying the inverse substitution boxes to the bundles. The block size of the cipher is given by $n_{\mathrm{b}} = mn_t$. In the case of the AES, the bundle size m is 8, hence bundles are bytes. This is by no means a necessity. For instance, Serpent [1] and Noekeon [5] also can be described in this framework, but have a bundle size of 4 bits. 3-WAY [3] uses 3-bit bundles.

For the purpose of this analysis, the S-boxes of γ need not to be specified. Since the use of different S-boxes for different bundles does not result in a plausible improvement of the resistance against known attacks, we propose to use the same S-box for all bundles. This allows to reduce the code size in software, and the required chip area in hardware implementations.

The transformation λ combines the bundles linearly: each bundle at the output is a linear function of bundles at the input. λ can be specified at the bit level by a simple $n_{\mathrm{b}} \times n_{\mathrm{b}}$ binary matrix M. We have

$$\lambda : b = \lambda(a) \Leftrightarrow b = Ma \tag{4}$$

λ can also be specified at the bundle level. For example, the bundles can be considered as elements in GF(2^m) with respect to some basis. In its most general form, we have:

$$\lambda : b = \lambda(a) \Leftrightarrow b_i = \sum_j \sum_{0 \le \ell < m} C_{i,j,\ell} {a_j}^{2^\ell} \tag{5}$$

In most instances a more simple linear function is chosen that is a special case of (5):

$$\lambda : b = \lambda(a) \Leftrightarrow b_i = \sum_j C_{i,j} a_j \tag{6}$$

Figure 1 gives a schematic representation of a $\gamma\lambda$ round transformation, followed by a key addition.

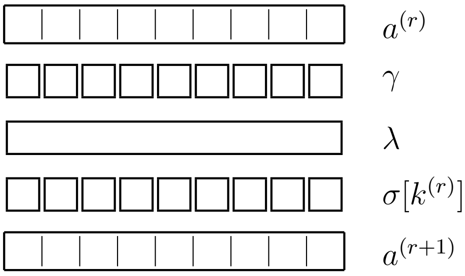

Fig. 1. Schematic representation of a $\gamma\lambda$ round, followed by a key addition.

3 Propagation in Key-alternating Block Ciphers

In the following subsections we describe the anatomy of correlations and difference propagations in key-alternating block ciphers. This is used to determine the number of rounds required to provide resistance against linear and differential cryptanalysis. We assume that the round transformations do not exhibit correlations with amplitude 1 or difference propagation with probability 1. The limitation to the key-alternating structure allows us to reason more easily about linear and differential trails as the effect of the key addition on the propagation is quite simple.

3.1 Differential cryptanalysis

We assume that the reader has a basic understanding of the principles of differential cryptanalysis as explained in [2]. We give a very short overview and introduce our notation.

Consider a couple of n-bit vectors a and a^* with bitwise difference $a+a^* = a'$. Let $b = h(a), b^* = h(a^*)$ and $b' = b + b^*$. The difference a' propagates to the difference b' through h. In general, b' is not fully determined by a' but depends on the value of a (or a^*).

Definition 1. *A* difference propagation probability $\mathrm{P}^h(a', b')$ *is defined as*

$$\mathrm{P}^h(a', b') = 2^{-n} \sum_a \delta(b' + h(a + a') + h(a)) \ . \tag{7}$$

Here $\delta(a)$ denotes the Kronecker delta function, which outputs zero, except when the input equals zero: $\delta(0) = 1$. If a pair is chosen uniformly from the set of all pairs (a, a^*) with $a + a^* = a'$, $\mathrm{P}^h(a', b')$ is the probability that $h(a) + h(a^*) = b'$.

Let β be a Boolean mapping operating on n-bit vectors that is composed of r mappings: $\beta = \rho^{(r)} \circ \rho^{(r-1)} \circ \ldots \circ \rho^{(2)} \circ \rho^{(1)}$. A *differential trail* A over a composed mapping consist of a sequence of $r + 1$ difference patterns:

$$Q = (q^{(0)}, q^{(1)}, q^{(2)}, \ldots, q^{(r-1)}, q^{(r)}) \ . \tag{8}$$

A differential trail has a *probability.* The probability of a differential trail is the number of values a_0 for which the difference patterns follow the differential trail divided by the number of possible values for a_0. This differential trail is composed of r differential steps $(q^{(i-1)}, q^{(i)})$, that have a propagation probability:

$$\mathrm{P}^{\rho^{(i)}}(q^{(i-1)}, q^{(i)}) \ . \tag{9}$$

Differential cryptanalysis exploits difference propagations $(q^{(0)}, q^{(r)})$ with large probabilities. The probability of difference propagation (a', b') is the sum of the probabilities of all r-round differential trails with initial difference a' and terminal difference b', i.e.,

$$\mathrm{P}(a', b') = \sum_{q^{(0)}=a', q^{(r)}=b'} \mathrm{P}(Q) \ . \tag{10}$$

3.2 Achieving low difference propagation probabilities

For a successful classical differential cryptanalysis attack, the cryptanalyst needs to know an input difference pattern that propagates to an output difference pattern over all but a few (2 or 3) rounds of the cipher, with a probability that is significantly larger than 2^{1-n_b}. To avoid this, we choose the number of rounds so that there are no differential trails with a probability above 2^{1-n_b}.

This strategy does not guarantee that there are no such difference propagations with a high probability. Equation (10) shows that in principle, many trails with each a low probability may add up to a difference propagation with high probability. As a matter of fact, for any Boolean mapping, a difference pattern at the input must propagate to some difference pattern at the output, and the sum of the difference propagation probabilities over all possible output differences is 1. Hence, there must be difference propagations with probability equal to or larger than 2^{1-n_b}. This also applies to the Boolean mapping formed by a cipher for a given value of the cipher key. Hence, the presence of difference propagations with a high probability over any number of rounds of the cipher is a mathematical fact which can't be avoided by design.

Let us analyze a difference propagation with probability y for a given key value. A difference propagation probability y means that there are exactly $y2^{n_\mathrm{b}-1}$ pairs with the given input difference pattern and the given output difference pattern. Each of these pairs follows a particular differential trail.

Assuming that the pairs are distributed over the trails according to a Poisson distribution, the expected number of pairs that, for a given key value, follow a differential trail with propagation probability 2^{-z}, is $2^{n_\mathrm{b}-1-z}$. Consider a differential trail with a propagation probability 2^{-z} smaller than 2^{1-n_b} that is followed by at least one pair. The probability that this trail is followed by more than one pair, is approximately $2^{n_\mathrm{b}-1-z}$. It follows that if there are no differential trails with a propagation probability above 2^{1-n_b}, the $y2^{n_\mathrm{b}-1}$ pairs that have the correct input difference pattern and output difference pattern, follow almost $y2^{n_\mathrm{b}-1}$ different differential trails.

If there are no differential trails with a low weight, difference propagations with a large probability are the result of multiple differential trails that happen to be followed by a pair in the given circumstances, i.e. for the given key value. For another key value, each of these individual differential trails may be followed by a pair, or not. This makes predicting the input difference patterns and output difference patterns that have large difference propagation probabilities practically infeasible. This is true if the key is known, and even more so if it is unknown.

We conclude that restricting the probability of difference propagations is a sound design strategy. However, it doesn't result in a proof of security.

3.3 Linear cryptanalysis

We assume again that the reader is familiar with the basic principles of linear cryptanalysis [7]. However, instead of using the notions *probability of a linear approximation*, and *deviation*, we prefer to use our own formalism, based on *correlation*.

Definition 2. *The correlation* $\mathrm{C}(f,g)$ *between two Boolean functions* $f(a)$ *and* $g(a)$ *is defined as*

$$\mathrm{C}(f,g) = 2 \cdot \mathrm{Prob}(f(a) = g(a)) - 1 \ . \tag{11}$$

It follows that $\mathrm{C}(f,g) = \mathrm{C}(g,f)$. A *parity* of a Boolean vector is a Boolean function that consists of the XOR of a number of bits. A parity is determined by the positions of the bits of the Boolean vector that are included in the XOR. The *selection pattern* w of a parity is a Boolean vector value that has a 1 in the components that are included in the parity and a 0 in all other components. Analogous to the inner product of vectors in linear algebra, we express the parity of vector a corresponding with selection pattern w as $w^{\mathrm{t}}a$. In this expression the t suffix denotes transposition of the vector w.

Note that for a vector a with n bits, there are 2^n different parities. The set of parities of a Boolean vector is in fact the set of all linear Boolean functions of that vector.

A *linear trail* U over a composed mapping consist of a sequence of $r+1$ *selection patterns*

$$U = (u^{(0)}, u^{(1)}, u^{(2)}, \ldots, u^{(r-1)}, u^{(r)}) \ . \tag{12}$$

This linear trail is composed of r linear steps $(u^{(i-1)}, u^{(i)})$, that have a correlation:

$$\mathrm{C}({u^{(i)}}^{\mathrm{t}} \rho^{(i)}(a), {u^{(i-1)}}^{\mathrm{t}} a)$$

The *correlation contribution* $\mathrm{C_p}$ of a linear trail is the product of the correlation of all its steps:

$$\mathrm{C_p}(U) = \prod_i C^{\rho^{(i)}}_{u^{(i)} u^{(i-1)}} \ . \tag{13}$$

3.4 Achieving low correlation amplitudes

For a successful linear cryptanalysis attack, the cryptanalyst needs to know an input parity and an output parity after all but a few rounds of the cipher that have a correlation with an amplitude that is significantly larger than $2^{-n_{\mathrm{b}}/2}$. To avoid this, we choose the number of rounds so that there are no linear trails with a correlation contribution above $n_{\mathrm{k}}^{-1} 2^{-n_{\mathrm{b}}/2}$.

This does not guarantee that there are no high correlations over r rounds. From Parseval's equality, it follows that for any output parity, the sum of the squared amplitudes of the correlations with all input parities is 1. In the assumption that the output parity is equally correlated to all $2^{n_{\mathrm{b}}}$ possible input parities, the correlation to each of these input parities has amplitude $2^{-n_{\mathrm{b}}/2}$. In practice it is very unlikely that such a uniform distribution will be attained and correlations will exist that are orders of magnitude higher than $2^{-n_{\mathrm{b}}/2}$. This also applies to the Boolean mapping formed by a cipher for a given value of the cipher key. Hence, the presence of high correlations over (all but a few rounds of) the cipher is a mathematical fact that can't be avoided by design.

However, in the absence of local clustering of linear trails, high correlations can only occur as the result of 'constructive interference' of many linear trails that share the same initial and final selection patterns. Specifically, any such correlation with an amplitude above $2^{-n_{\mathrm{b}}/2}$ must be the result of at least n_{k} different linear trails. The condition that a linear trail in this set contributes constructively to the resulting correlation imposes a linear relation on the round key bits. From the point that more than n_{k} linear trails are combined, it is very unlikely that all such conditions can be satisfied by choosing the appropriate cipher key value.

The strong key-dependence of this interference makes it very unlikely that if a specific output parity has a high correlation with a specific input parity for a given key, that this will also be the case for another value of the key. In other words, although it follows from Parseval's Theorem that high correlations over the cipher will exist whatever the number of rounds, the strong round key dependence of interference makes locating the input and output selection patterns for which high correlations occur practically infeasible. This is true if the key is known, and even more so if it is unknown.

Again we conclude that restricting the amplitude of the correlation between input parities and output parities is a sound design strategy. However, it doesn't result in a proof of security.

3.5 Weight of a trail

γ is a bricklayer mapping consisting of S-boxes. It is easy to see that the correlation over γ is the product of the correlations over the different S-box positions for the given input and output selection patterns. We define the *weight* of a correlation as the negative logarithm of its amplitude. The correlation weight for an input selection pattern and output selection pattern is the sum of the correlation weights of the different S-Box positions. If the output selection pattern is non-zero for a particular S-box position or bundle, we call this S-box or bundle *active*.

Similarly, the weight of the difference propagation over γ is defined as the negative logarithm of its probability. The weight of the difference propagation over γ is given by the sum of the weights of the difference propagations of the S-box positions for the given input difference pattern and output difference pattern. If the input difference pattern is non-zero for a particular S-box position or bundle, we call this S-box or bundle *active*.

The correlation contribution of a linear trail is the product of the correlation of all its steps. The weight of such a trail is defined as the sum of the weights of its steps. As the weight of a step is the sum of the weights of its active S-box positions, the weight of a linear trail is the sum of that of its active S-boxes. An upper limit to the correlation is a lower limit to the weight per S-box. Hence, the weight of a linear trail is equal to or larger than the number of active bundles in all its selection patterns times the minimum (correlation) weight per S-box. We call the number of active bundles in a pattern or a trail its *bundle weight*.

A differential trail is defined by a series of difference patterns. The weight of such a trail is the sum of the weights of the difference patterns of the trail. Completely analogous to linear trails, the weight of a differential trail is equal to or larger than the number of active S-boxes times the minimum (differential) weight per S-box.

3.6 Wide trails

The reasoning above suggests two possible mechanisms to eliminate low-weight trails:

1. Choose S-boxes with high minimum differential and correlation weight.
2. Design the round transformation such a way that there are no relevant trails with low bundle weight.

The maximum correlation amplitude of an m-bit invertible S-box is above $2^{m/2}$ yielding an upper bound for the minimum (correlation) weight of $n/2$. The maximum difference propagation probability is at least 2^{m-2}, yielding an upper bound for the minimum (differential) weight of $m - 2$. This seems to suggest that one should take large S-boxes. This is not the approach we follow in the wide trail design strategy.

> Instead of spending most of the resources on large S-boxes, the wide trail strategy aims at designing the round transformation(s) such that there are no trails with a low bundle weight. In ciphers designed by the wide trail strategy, a relatively large amount of resources is spent in the linear step to provide high multiple-round diffusion.

4 Diffusion

Diffusion is the term introduced by C. Shannon to denote the quantitative spreading of information [9]. Diffusion is a rather vague concept the exact meaning of which strongly depends on the context in which it is used. We will explain now what we mean by diffusion in the context of the wide trail strategy.

Inevitably, the mapping γ provides some interaction between the different bits within the bundles that may be referred to as diffusion. However, it does not provide any inter-bundle interaction: difference propagation and correlation over γ stays confined within the bundles. In the context of the wide trail strategy, it is not this kind of diffusion we are interested in. We use the term diffusion to indicate properties of a mapping that increase the minimum bundle weight of linear and differential trails. In this sense, all diffusion is realized by λ. γ does not provide any diffusion at all.

For single-round trails, obviously the bundle weight of a single round trail, differential or linear, is equal to the number of active bundles at its input. It follows that the minimum bundle weight of a single-round trail is 1, independent of λ. The situation becomes interesting as soon as we consider two-round trails.

4.1 Branch Numbers and Two-Round Trails

In two-round trails, the bundle weight is the sum of the number of active bundles in the (selection or difference) patterns at the beginning of the first and the input of the second round. We will see that the bundle weight of two-round trails can be expressed elegantly by using *branch numbers.*

Consider a partition α that divides the different bit positions of a state into n_α sets called α-sets. An example of this is the partition that divides the bits in a number of bundles. The weight of a state value with respect to a partition α is equal to the number of α-sets that have at least one non-zero bit. This is denoted by $w_\alpha(a)$. If this is applied to a difference pattern a', $w_\alpha(a')$ is the number of active α-sets in a'. Applied to a selection pattern v, $w_\alpha(v)$ is the number of active α-sets in v. If α is the partition that forms the bundles, $w_\alpha(a)$ is the number of active bundles in the pattern a and is denoted by $w_b(a)$.

We make a distinction between the differential and the linear branch number of a transformation.

Definition 3. *The* differential branch number *of a transformation ϕ with respect to a partition α is defined by*

$$\mathcal{B}_d(\phi, \alpha) = \min_{a, b \neq a} \{w_\alpha(a \oplus b) + w_\alpha(\phi(a) \oplus \phi(b))\} \tag{14}$$

For a linear transformation $\lambda(a) \oplus \lambda(b) = \lambda(a \oplus b)$, and (14) reduces to:

$$\mathcal{B}_d(\lambda, \alpha) = \min_{a' \neq 0} \{w_\alpha(a') + w_\alpha(\lambda(a'))\} \ . \tag{15}$$

An upper bound for the differential branch number of a Boolean transformation ϕ with respect to a partition α is given by n_α, since the output difference corresponding to an input difference with a single non-zero bundle can have at most weight n_α. Therefore, the differential branch number of ϕ with respect to α is upper bounded by

$$\mathcal{B}_d(\phi, \alpha) \leq n_\alpha + 1. \tag{16}$$

Analogous to the differential branch number, we can define the linear branch number.

Definition 4. *The* linear branch number *of a transformation ϕ with respect to a is given by*

$$\mathcal{B}_l(\phi, \alpha) = \min_{\alpha, \beta, c(\alpha^{\mathrm{t}} x, \beta^{\mathrm{t}} \phi(x)) \neq 0} \{w_\alpha(\alpha) + w_\alpha(\beta)\} \tag{17}$$

Many of the following discussions are valid both for differential and linear branch numbers and both $\mathcal{B}_d$ and $\mathcal{B}_l$ are denoted simply by $\mathcal{B}$. Moreover, in many cases the partition is clear from the context and $\mathcal{B}(\phi, \alpha)$ is expressed as $\mathcal{B}(\phi)$.

4.2 Some properties

In general, the linear and differential branch number of a transformation with respect to a partition are not equal. From the symmetry of Definition 3 and 4 it follows that the branch number of a transformation and that of its inverse are the same. Moreover, we have the following properties:

- a (differential or selection) pattern a is not affected by a key addition and hence its weight $w_\alpha(a)$ is not affected. This property holds independently of the partition α.
- a bricklayer permutation compatible with α cannot turn an active α-subset into a non-active one or vice versa. Hence, it does not affect the weight $w_\alpha(a)$.

Assume we have a transformation ϕ composed of a transformation ϕ_1 and a bricklayer transformation ϕ_α operating on α-subsets, i.e., $\phi = \phi_\alpha \circ \phi_1$. As ϕ_α does not affect the number of active α-subsets in a propagation pattern, the branch number of ϕ and ϕ_1 are the same. More general, if propagation of patterns is analyzed at the level of α-subsets, bricklayer transformations compatible with α may be ignored.

If we apply this to the bundle weight of a $\gamma\lambda$ round transformation ρ, it follows immediately that the (linear or differential) bundle branch number of ρ is that of its linear part λ.

4.3 A two-round propagation theorem

The following theorem relates the value of $\mathcal{B}(\lambda)$ to a bound on the number of active bundles in a trail. The proof is valid both for linear and differential trails: in the case of linear trails $\mathcal{B}$ stands for $\mathcal{B}_l$ and in the case of differential trails $\mathcal{B}$ stands for $\mathcal{B}_d$.

Theorem 1 (Two-Round Propagation Theorem).
For a key-alternating block cipher with a $\gamma\lambda$ round structure the number of active bundles of any two-round trail is lower bounded by the (bundle) branch number of λ.

Proof. Figure 2 depicts two rounds. Since the transformations γ and $\sigma[k]$ operate on each bundle individually, they do not affect the propagation of patterns. Hence it follows that $w_b(a^{(1)}) + w_b(a^{(2)})$ is only bounded by the properties of the linear transformation λ of the first round. Definition 3 and 4 imply that the sum of the active bundles before and after λ of the first round is lower bounded by $\mathcal{B}(\lambda)$. □

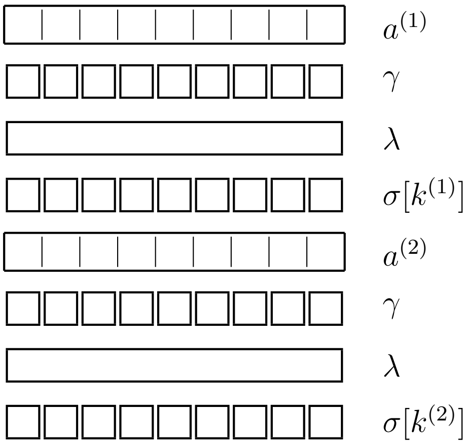

Fig. 2. Transformations relevant in the proof of Theorem 1.

5 An Efficient Key-Alternating Structure

In trails of more than two rounds, the desired diffusion properties of ρ are less trivial. It is clear than any $2n$-round trail can be decomposed in n 2-round trails and hence that its bundle weight is lower bounded by n times the branch number of ρ. The 'greedy' approach to eliminate low-weight trails is to consider Theorem 1 only and to design a round transformation with a maximum branch number. However, transformations that provide high branch numbers have a tendency to have a high implementation cost. More efficient designs can be achieved in the following way. We build a key-alternating block cipher that consists of an alternation of two different round transformations defined by:

$$\rho^a = \theta \circ \gamma \tag{18}$$

$$\rho^b = \Theta \circ \gamma \tag{19}$$

The transformation γ is defined as before and operates on n_b m-bit bundles.

5.1 The diffusion transformation θ

With respect to θ, the bundles of the state are grouped into a number of *columns* by a partition Ξ of the index space $\mathcal{I}$. We denote a column by ξ and the number of columns by n_Ξ. The column containing an index i is denoted by $\xi(i)$ and the number of indices in a column ξ by n_ξ. The size of the columns relates to the block length by

$$m \sum_{\xi \in \Xi} n_\xi = m n_t .$$

θ is a bricklayer mapping with component mappings that each operate on a column. Within each column, bundles are linearly combined. We have

$$\theta : b = \theta(a) \Leftrightarrow b_i = \sum_{j \in \xi(i)} C_{i,j} a_j \qquad (20)$$

The bricklayer transformation θ only needs to realize diffusion within the columns and has hence an implementation cost that is much lower.

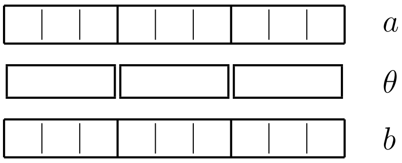

Fig. 3. The diffusion transformation θ

Similar to active bundles, we can speak of active columns. The number of active columns of a propagation pattern a is denoted by $w_s(a)$. The round transformation $\rho^a = \theta \circ \gamma$ is a bricklayer transformation operating independently on a number of columns. Taking this bricklayer structure into account we can extend the result of Section 4.1 slightly. The branch number of θ is given by the minimum branch number of its component transformations. Applying (16) to the component mappings defined by the matrices C_ξ results in the following upper bound:

$$\mathcal{B}(\theta) \leq \min_\xi n_\xi + 1. \qquad (21)$$

Hence, the smallest column imposes the upper limit for the branch number. The Two-Round Propagation Theorem (Theorem 1) implies the following Lemma.

Lemma 1. *The bundle weight of any two-round trail in which the first round has a $\gamma\theta$ round transformation is lower bounded by $N\mathcal{B}(\theta)$, where N is the number of active columns at the input of the second round.*

Proof. Theorem 1 can be applied to each of the component mappings of the bricklayer mapping ρ^a separately. For each active column there are at least $\mathcal{B}(\theta)$ active bundles in the two-round trail. □

5.2 The linear transformation Θ

Θ mixes bundles across columns.

$$\Theta : b = \Theta(a) \Leftrightarrow b_i = \sum_j C_{i,j} a_j \tag{22}$$

The goal of Θ is to provide inter-column diffusion. The design criterion for Θ is to have a high branch number with respect to Ξ. This is denoted by $\mathcal{B}(\Theta, \Xi)$ and called its *column branch number.*

5.3 A lower bound on the bundle weight of 4-round trails

The combination of the bundle branch number of θ and the column branch number of Θ allows us to prove a lower bound on the bundle weight of any trail over 4 rounds starting with ρ^a.

Theorem 2 (Four-round Propagation Theorem for $\theta\Theta$ construction). *For a key-alternating block cipher with round transformations as defined in (18) and (19), the bundle weight of any trail over*

$$\rho^b \circ \rho^a \circ \rho^b \circ \rho^a$$

is lower bounded by $\mathcal{B}(\theta) \times \mathcal{B}(\Theta, \Xi)$.

Proof. Figure 4 depicts four rounds. As the key additions play no role in the propagation of patterns, they have been left out. It is easy to see that the linear transformation of the fourth round plays no role. The sum of the number of active columns in $a^{(2)}$ and $a^{(3)}$ is lower bounded by $\mathcal{B}(\Theta, \Xi)$. According to Lemma 1, for each active column in $a^{(2)}$ there are at least $\mathcal{B}(\theta)$ active bundles in the corresponding columns of $a^{(1)}$ and $a^{(2)}$. Similarly, for each active column in $a^{(3)}$ there are at least $\mathcal{B}(\theta)$ active bundles in the corresponding columns of $a^{(3)}$ and $a^{(4)}$. Hence the total number of active bundles is lower bounded by $\mathcal{B}(\theta) \times \mathcal{B}(\Theta, \Xi)$. □

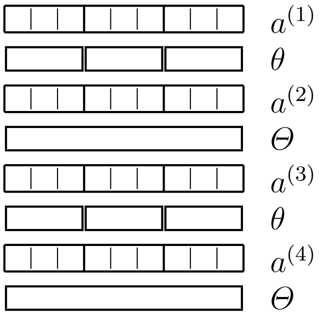

Fig. 4. Relevant transformations for the proof of Theorem 2.

5.4 An efficient construction for Θ

As opposed to θ, Θ does not operate on different columns independently and hence may have a much higher implementation cost. In this we present a construction of Θ in terms of θ and bundle transpositions denoted by π. We define

$$\Theta = \pi \circ \theta \circ \pi \ . \tag{23}$$

In the following we will define π and prove that if π is well chosen the column branch number of Θ can be made equal to the bundle branch number of θ.

The bundle transposition π The bundle transposition π is defined as

$$\pi : b = \pi(a) \Leftrightarrow b_i = a_{p(i)} \ , \tag{24}$$

with $p(i)$ a permutation of the index space $\mathcal{I}$. The inverse of π is defined by $p^{-1}(i)$. Observe that a bundle transposition π does not affect the bundle weight of a propagation pattern and hence that the branch number of a transformation is not affected if it is composed with π.

Contrary to θ, π provides *inter-column diffusion.* Intuitively, good diffusion for π would mean that it distributes the different bundles of a column to as many different columns as possible. We say π is *diffusion-optimal* if the different bundles in each column are distributed over all different columns. More formally, we have:

Definition 5. *π is* diffusion-optimal *if and only if*

$$\forall i, j \in \mathcal{I}, i \neq j : (\xi(i) = \xi(j)) \Rightarrow (\xi(p(i)) \neq \xi(p(j))). \tag{25}$$

It is easy to see that this implies the same condition for π^{-1}. A diffusion-optimal bundle transposition π implies

$$w_s(\pi(a)) \geq \max_{\xi}(w_b(a_\xi)) \ .$$

Therefore a diffusion-optimal transformation can only exist if $n_\Xi \geq \max_i(n_{\xi_i})$. In words, π can only be diffusion-optimal if there are at least as many columns as there are bundles in the largest column. If π is diffusion-optimal, we can prove that the column branch number of the mapping Θ is lower bounded by the branch number of θ.

Lemma 2. *If π is a diffusion-optimal transposition of bundles, the column branch number of $\pi \circ \phi \circ \pi$ is lower bounded by the bundle branch number of ϕ*

Proof. We refer to Figure 5 for the notations used in this proof. We have to demonstrate that

$$w_s(a) + w_s(d) \geq \mathcal{B}(\phi) \ .$$

For any active column in b, the number of active bundles in that column and the corresponding column of c is at least $\mathcal{B}(\phi)$. π moves all active bundles in an active column of c to different columns in d and π^{-1} moves all active bundles in an active column of b to different columns in a. It follows that the sum of the number of active columns in a and in d is lower bounded by the bundle branch number of ϕ. □

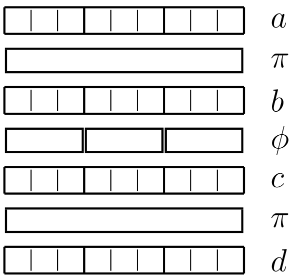

Fig. 5. Transformations relevant in the proof of Lemma 2.

6 Using Identical Round Transformations

The efficient structure described in Section 5 uses two different round transformations. It is possible to define a block cipher structure with only one round transformation, that achieves the same bound. This is the round structure used in the AES and related ciphers. The advantage of having a single round transformation is a reduction in software code in software implementations and chip area in dedicated hardware implementations. For this purpose, λ is composed of two types of the mappings:

- θ: the linear bricklayer mapping that provides high local diffusion, as defined in Section 5.1, and
- π: the transposition mapping that provides high *dispersion*, as defined in Section 5.4.

Hence we have for the round transformation:

$$\rho^c = \theta \circ \pi \circ \gamma \tag{26}$$

Figure 6 gives a schematic representation of the different transformations of a round. These component transformation are defined in such a way that they impose strict lower bounds on the number of active S-boxes in four-round trails. For two-round trails it can be seen that the number of active bundles is lower bounded by $\mathcal{B}(\rho^c) = \mathcal{B}(\lambda) = \mathcal{B}(\theta)$. For four rounds, we can prove the following important theorem:

Theorem 3 (Four-Round Propagation Theorem).
For a key-iterated block cipher with a $\gamma\pi\theta$ round transformation and diffusion-optimal π, the number of active S-boxes in a four-round trail is lower bounded by $(\mathcal{B}(\theta))^2$.

Proof. Firstly, we show that the transformation formed by 4 applications of the round transformation ρ^c as defined in (26) is equivalent to four rounds of the construction with ρ^a and ρ^b as defined in (18) and (19). For simplicity, we leave

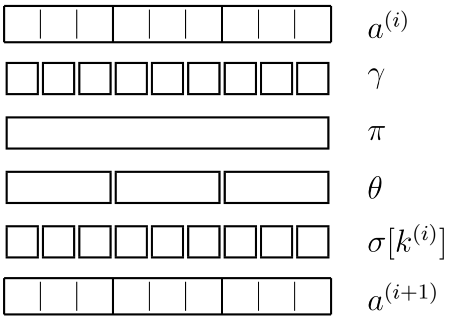

Fig. 6. Schematic representation of the different steps of a $\gamma\pi\theta$ round transformation, followed by a key addition.

out the applications of the key additions, but the proof works in the same way if the key additions are present. Let $\mathcal{A}$ be defined as:

$$\begin{aligned}\mathcal{A} &= \rho^c \circ \rho^c \circ \rho^c \circ \rho^c \\ &= (\theta \circ \pi \circ \gamma) \circ (\theta \circ \pi \circ \gamma) \circ (\theta \circ \pi \circ \gamma) \circ (\theta \circ \pi \circ \gamma)\ \ .\end{aligned}$$

γ is a bricklayer mapping, operating on every bundle separately and operating independently of the bundle's position. Therefore γ commutes with π, which only moves the bundles to different positions. We get:

$$\begin{aligned}\mathcal{A} &= (\theta \circ \gamma) \circ (\pi \circ \theta \circ \pi \circ \gamma) \circ (\theta \circ \gamma) \circ (\pi \circ \theta \circ \pi \circ \gamma) \\ &= \rho^a \circ \rho^b \circ \rho^a \circ \rho^b\ \ ,\end{aligned}$$

with Θ of ρ^b defined exactly as in (23). Now we can apply Lemma 2 and Theorem 2 to finish the proof. □

7 Conclusion and Open Problems

We have shown how the application of the wide trail design strategy leads to the definition of a round transformation as the one used in Rijndael. The proposed round transformation allows us to give provable bounds on the correlation of linear trails and the weight of differential trails while at the same time allowing efficient implementations.

An interesting open problem is the effect of *trail clustering.* Theorems 1, 2 and 3 give lower bounds on the weight of trails. As mentioned in Section 3, the probability of input-output difference propagations as well as the correlation between input parities and output parities are a sum over many trails. If the trails follow indeed a Poisson distribution, then the results can be applied straightforwardly. However, it has already been observed that in some cases, the trails *don't* follow a Poisson distribution. Instead, they tend to cluster and as a result the probability of a difference propagation can be significantly higher [6]. A similar effect for correlations has been studied in [8]. It remains an open problem whether trail clustering occurs and impacts the security for the cipher structure described here.

References

1. R. Anderson, E. Biham, and L. R. Knudsen. Serpent. In *Proceedings of the first AES candidate conference*, Ventura, August 1998.
2. E. Biham and A. Shamir. Differential cryptanalysis of DES-like cryptosystems. *Journal of Cryptology*, 4(1):3–72, 1991.
3. J. Daemen, R. Govaerts, and J. Vandewalle. A new approach to block cipher design. In Vaudenay [10], pages 18–32.
4. J. Daemen, L. R. Knudsen, and V. Rijmen. Linear frameworks for block ciphers. *Designs, Codes and Cryptography*, 22(1):65–87, January 2001.
5. J. Daemen, M. Peeters, G. V. Assche, and V. Rijmen. Noekeon. In *First open NESSIE Workshop*, Leuven, November 2000.
6. L. R. Knudsen. Truncated and higher order differentials. In B. Preneel, editor, *Fast Software Encryption '94*, volume 1008 of *Lecture Notes in Computer Science*, pages 196–211. Springer-Verlag, 1995.
7. M. Matsui. Linear cryptanalysis method for DES cipher. In T. Helleseth, editor, *Advances in Cryptology, Proceedings of Eurocrypt '93*, volume 765 of *Lecture Notes in Computer Science*, pages 386–397. Springer-Verlag, 1994.
8. K. Nyberg. Linear approximation of block ciphers. In A. D. Santis, editor, *Advances in Cryptology, Proceedings of Eurocrypt '94*, volume 950 of *Lecture Notes in Computer Science*, pages 439–444. Springer-Verlag, 1995.
9. C. E. Shannon. Communication theory of secrecy systems. *Bell Syst. Tech. Journal*, 28:656–715, 1949.
10. S. Vaudenay, editor. *Fast Software Encryption '98*, volume 1372 of *Lecture Notes in Computer Science*. Springer-Verlag, 1998.