Plateau Characteristics

Joan Daemen[1] and Vincent Rijmen[2]

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

Abstract. Plateau characteristics are a special type of characteristics whose probability depends on the key and can have only 2 values. For a (usually small) subset of the keys it has a non-zero probability and for all other keys its probability is zero. In this paper we prove that for a large group of ciphers, including the AES, all two-round characteristics are plateau characteristics. For the AES and other ciphers with a similar structure we show that the vast majority of characteristics over 4 or more rounds are plateau characteristics. In the case of the AES, for most keys there are two-round characteristics with fixed-key probability equal to $32/2^{32}$ while the Maximum Expected Differential Probability (MEDP) of two-round differentials is at most $13.25/2^{32}$.

Keywords: Differential cryptanalysis, AES

1 Introduction

In this paper we study the probability of *characteristics* [6] over (reduced-round versions of) block ciphers, where the difference is the bitwise XOR, and apply our results to the AES and a simplified variant.

It has been reported before that the fixed-key probability of characteristics depends on the value of the key [5, 7, 3]. We define *plateau characteristics*, where the dependency on the value of the key is very structured. The fixed-key probability of these characteristics is either zero, or 2^h, with h a value that depends only on the characteristic and not on the key. We show that for a large class of ciphers, all two-round characteristics and a fraction of the more-round characteristics are plateau characteristics. This fraction is very close to 100% for the AES and for other ciphers with a diffusion mapping based on Maximum Distance Separable (MDS) codes and with S-boxes that have '2' or '4' as maximal entries in their XOR-tables (4-uniform S-boxes).

Our results don't affect the MEDP of characteristics, but show that the distribution of the key-dependent probability is not narrow and hence the widely made assumption that it can be approximated by the EDP, is not justified.

Applying our results to the AES, we see that for almost all values of the key there are two-round characteristics with a fixed-key probability equal to $32/2^{32}$, while the MEDP of two-round characteristics is at most $4/2^{32}$ [12] and the MEDP of two-round *differentials* [17] is at most $13.25/2^{32}$ [14, 23, 24].

After introducing some basic definitions in Section 2 and presenting our motivation for this work in Section 3, Section 4 introduces plateau characteristics over two rounds. Characteristics over more than two rounds are studied in Section 5. We present some further observations in Section 6. We analyze the key-dependent probability of all two-round characteristics over the AES in Section 7. We discuss more-round characteristics over the AES in Section 8 and the impact on differentials in Section 9. We conclude in Section 10.

2 Definitions

We denote a *differential* [17] over a map by (a, b) and assume that it is clear from the context which map we mean. We call a the input difference and b the output difference. The *probability* of a differential is denoted by $\mathrm{DP}(a, b)$. For a keyed map, we can define a differential probability $\mathrm{DP}[k](a, b)$ for each value k of the key. We define the *expected differential probability* (EDP) of a differential as the average of the differential probability $\mathrm{DP}(a, b)$ over all keys.

A *composed function* consists of a sequence of maps, called steps. A characteristic through a composed function is a sequence of differences a, b, ... The sequence consists of an input difference a, followed by the output differences of all the steps of the composed function. A characteristic over a keyed composed map has a differential probability $\mathrm{DP}[k](Q)$ for each value k of the key. The EDP of a characteristic is the average of its $\mathrm{DP}[k]$ over all keys. In this paper, we only consider characteristics with $\mathrm{EDP} > 0$.

Definition 1 ([12]). *The* weight *of a differential or a characteristic is minus the binary logarithm of their EDP.*

$$\text{weight}(a,b) = -\log_2 EDP(a,b), \qquad \text{weight}(Q) = -\log_2 EDP(Q) \ .$$

3 Motivation

An important example of composed maps is formed by block ciphers. Modern block ciphers are designed to resist differential cryptanalysis. Often they come with provable bounds on EDP values of either characteristics [12] or differentials [22, 20]. For older ciphers like the DES, the EDP of a differential can be estimated by the EDP of one characteristic, the *dominating characteristic* [6]. For many modern block ciphers, there are no dominating characteristics, and this estimate can no longer be used. It is also tempting to invoke the

Hypothesis 1 (Hypothesis of Stochastic Equivalence [17]) *For all differentials* (a, b)*, it holds that for most values of the key* k*,* $DP[k](a,b) \approx EDP(a,b)$*.*

The main use of the Hypothesis of Stochastic Equivalence is that it allows one to construct proofs of security. The proofs give bounds on the expected data complexity of differential attacks where the attacker uses exactly one differential. The hypothesis has been observed to hold in toy ciphers [16], but not in ciphers that are used in practice, for instance DES [7, 15] and IDEA [9]. In any case, the hypothesis can't be applied to the probability of characteristics. While $\mathrm{DP}[k]$ of a characteristic is always a multiple of $2^{1-\mathrm{n_b}}$, with $\mathrm{n_b}$ the input size of the map, the EDP of a characteristic can take much smaller values.

When we examine the resistance of the AES and related ciphers against differential cryptanalysis, the Hypothesis of Stochastic Equivalence is of little use. For instance, for characteristics over 4 or more rounds, the EDP values are already below $2^{-150} \ll 2^{1-\mathrm{n_b}} = 2^{-127}$. In this paper we describe completely the distribution of DP of all characteristics over two rounds of the AES. We also give results for four and more rounds. Our analysis is not based on any hypothesis. The following example illustrates in a simple way the effects we want to examine for the AES.

Example 1 Consider the keyed map $E[k]$, defined as

$$E[k](x) = \rho^{-1}(k + \rho(x)), \tag{1}$$

where ρ is an arbitrary invertible transformation [13]. Let $MEDP(\rho)$ denote the maximum $\mathrm{EDP}(a, b)$ of a differential over ρ, where only nonzero values for a are considered. Then $MEDP(\rho)$ is an upper bound for the EDP of a differential over E. Since $E[0]$ is the identity transformation, for all differences a the differential (a, a) over the map E has fixed-key probability $\mathrm{DP}[0](a, a) = 1$. This property holds whatever value $MEDP(\rho)$ takes.

This example is contrived. In practice we don't expect $\mathrm{DP}[k]$ to deviate this strongly from EDP. However, we observe effects that go in this direction. We found that for several ciphers, including the AES, $\mathrm{DP}[k]$ has a distribution with a surprisingly rich structure.

4 Two-round plateau characteristics

In this section we show that for a large class of ciphers, two-round characteristics have a $\mathrm{DP}[k]$ that can take only two values.

4.1 Planar differentials and maps

As customary, we consider ordered pairs of inputs [6], but we denote them using curled braces '{}' in order to avoid confusion with differentials. Let $F_{(a,b)}$ denote the set containing the inputs x for which the pair $\{x, x + a\}$ is a right pair for the differential (a, b) over an unkeyed map. Let $G_{(a,b)}$ denote the set containing the corresponding outputs. Similarly, let $F_Q[k]$ denote the set containing the inputs x for which the pair $\{x, x + a\}$ is a right pair for the characteristic Q over a keyed map. Let $G_Q[k]$ denote the set containing the corresponding outputs. We introduce the concept of planar differentials:

Definition 2. *A differential* (a, b) *is* planar *if* $F_{(a,b)}$ *and* $G_{(a,b)}$ *form affine subspaces [1]:*

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

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

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

If $F_{(a,b)}$ contains an element x, then it also contains $x \oplus a$. Hence if $F_{(a,b)}$ is not empty, then $a \in U_{(a,b)}$. The number of elements in $F_{(a,b)}$ is $2^{\dim(U_{(a,b)})}$, so $\dim(U_{(a,b)}) = \mathrm{n_b} - \mathrm{weight}(a,b)$. Similarly, we have $b \in V_{(a,b)}$ and $\dim(V_{(a,b)}) = \mathrm{n_b} - \mathrm{weight}(a,b)$. We can now prove the following lemmas.

Lemma 1. *A differential (a,b) which has exactly two right pairs, is planar.*

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

Lemma 2. *A differential (a,b) which has exactly four right pairs, is planar.*

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

Lemma 3. *Any differential with $DP = 1$ is a planar differential.*

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

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

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

Any map for which all non-trivial differentials have $\mathrm{DP}(a,b) \leq 2^{2-\mathrm{n_b}}$ is planar. Such maps are called differentially 4-uniform [21]. Now we give two lemmas on planar differentials over composed maps. The first lemma applies for instance to a substitution step in a block cipher, consisting of the parallel application of some S-boxes.

Lemma 4. *Let $y = \alpha(x)$ be a map consisting of a set of parallel maps $y_i = \alpha_i(x_i)$ with $x = (x_0, x_1, \ldots, x_t)$ and $y = (y_0, y_1, \ldots, y_t)$. A differential (a,b) for which the differentials (a_i, b_i) are planar, is planar.*

We have

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

with $\times$ denoting the direct product [1]. The following lemma applies to a sequence of maps.

Lemma 5. *If (a, b) is a planar differential of α, then for any pair of affine maps L_1 and L_2 with L_1 invertible, the differential $(L_1(a), L_2(b))$ is planar over $L_2 \circ \alpha \circ L_1^{-1}$.*

Examples of ciphers in which single-round differentials are planar are the AES, but also 3-Way [8], SHARK [25], Square [10], Camellia [4], Serpent [2] and Noekeon [11]. Some other popular maps that are planar, are the majority function $f(x, y, z) = xy \oplus xz \oplus yz$ and the 'if' function $g(x, y, z) = xy \oplus (\neg x)z$ or any map that is quadratic in GF(2).

4.2 Plateau characteristics

Similar to the concept of plateaued functions [28], for which the Walsh spectrum takes only two values (in absolute value), we introduce here *plateau characteristics* as characteristics for which the DP$[k]$ takes only two values (where one value is always zero). The *height* of a plateau characteristic determines how high the non-zero DP$[k]$ value of the plateau characteristic is.

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

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

The height of a plateau characteristic can be bounded as follows. Firstly, height$(Q) \leq \mathrm{n_b}$. Secondly, height(Q) is maximal when all but one key have DP equal to zero. Denoting the number of keys by $2^{\mathrm{n_k}}$, we obtain that in this case EDP equals $2^{-\mathrm{n_k}}$ times the non-zero DP value. Taking the logarithm, we obtain $-\mathrm{weight}(Q) = -\mathrm{n_k} + \mathrm{height}(Q) - \mathrm{n_b}$. Hence, we have in all cases $\mathrm{height}(Q) \leq \mathrm{n_k} + \mathrm{n_b} - \mathrm{weight}(Q)$. We can now prove the following result on an $\mathrm{n_b}$-bit map consisting of two steps and an addition with an $\mathrm{n_b}$-bit key in between (hence $\mathrm{n_k} = \mathrm{n_b}$).

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

Proof. The proof is based on geometrical arguments [1]. For right pairs the values at the output of the first step are in $G_{(a,b)}$. The values at the input of the second step are in $F_{(b,c)}$, or equivalently, the values at the output of the first step are in $k \oplus F_{(b,c)}$. It follows that the values at the output of the first step are in:

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

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

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

with $V_{(a,b)}$ and $U_{(b,c)}$ vector spaces. We start by deriving the condition that H is non-empty. First we translate the affine subspaces in the equation by the vector v:

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

$v \oplus H$ is a translated version of H and has the same number of elements. Now $v \oplus H$ is non-empty iff there is a vector $x \in V_{(a,b)}$ and a vector $y \in U_{(b,c)}$ such that $x = k \oplus u \oplus v \oplus y$ or formally, iff

$$\exists x \in V_{(a,b)}, y \in U_{(b,c)} : k \oplus u \oplus v = x \oplus y \ .$$

This is equivalent to saying that $(k \oplus u \oplus v) \in (V_{(a,b)} \oplus U_{(b,c)})$. If we denote $u \oplus v \oplus (V_{(a,b)} \oplus U_{(b,c)})$ by K_Q, this corresponds to saying that H is non-empty iff $k \in K_Q$. Consider now the case that H is non-empty and let w be an element of H. Clearly, w is an element of both $G_{(a,b)}$ and $k \oplus F_{(b,c)}$. It follows that $w \oplus k \oplus F_{(b,c)} = U_{(b,c)}$ and hence $G_{(a,b)} = w \oplus V_{(a,b)}$ and $k \oplus F_{(b,c)} = w \oplus U_{(b,c)}$. We have

$$H = (w \oplus V_{(a,b)}) \cap (w \oplus U_{(b,c)}) \ .$$

Translation by w yields:

$$w \oplus H = V_{(a,b)} \cap U_{(b,c)} \ .$$

Let $W_Q = V_{(a,b)} \cap U_{(b,c)}$. The number of pairs in H is $2^{\dim(W_Q)}$ if $k \in K_Q$ and zero otherwise. The number of elements in K_Q is determined by the dimension of $V_{(a,b)} \oplus U_{(b,c)}$. We use the subspace dimension theorem: $\dim(U) + \dim(V) = \dim(U \oplus V) + \dim(U \cap V)$. This gives:

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

If we now denote $\mathrm{height}(Q) = \dim(W_Q)$, we have $\mathrm{DP}(Q) = 2^{\mathrm{height}(Q) - \mathrm{n_b}}$ for $2^{2\mathrm{n_b} - \mathrm{weight}(Q) - \mathrm{height}(Q)}$ keys on the total number of $2^{\mathrm{n_b}}$ keys, and zero for all other keys. □

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

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

4.3 Plateau characteristics in super boxes

Several ciphers that use S-boxes and linear transformations can also be described using the structure of a *super box*.

Definition 5. *A* super box *maps an array a of $\mathrm{n_t}$ elements a_i to an array e of $\mathrm{n_t}$ elements e_i. Each of the elements has size $\mathrm{n_s}$. A super box takes a key k of size $\mathrm{n_t} \times \mathrm{n_s} = \mathrm{n_b}$. It consists of the sequence of four transformations (or* steps*):*

- $b_i = S[a_i]$*: $\mathrm{n_t}$ parallel applications of a $\mathrm{n_s}$-bit S-box*
- $c = \mathrm{M}(b)$*: a linear map with branch number $\mathrm{n_t} + 1$ [10]*
- $d = c \oplus k$*: key addition*
- $e_i = S[d_i]$*: $\mathrm{n_t}$ parallel applications of a $\mathrm{n_s}$-bit S-box*

The S-boxes in the two S-box steps may also be all different. A characteristic over the super box can specify that one or more of the S-boxes have input difference 0. Such S-boxes have always output difference 0, with probability 1. When computing the probability of a characteristic, only the S-boxes with non-zero input difference need to be taken into account. They are called *active S-boxes*.

We can prove the following upper bound on height(Q) for plateau characteristics in super boxes.

Theorem 2. *Let Q be a plateau characteristic over a super box. Let the sets γ_j denote all possible selections of* $\mathrm{n_t}$ *S-boxes from the super box. Then*

$$\mathrm{height}(Q) \leq \mathrm{n_b} - \max_j \left(\sum_{i \in \gamma_j} \mathrm{weight}(x_i, y_i) \right) ,$$

where (x_i, y_i) denotes a differential over an S-box.

Proof. Since $\mathrm{height}(Q) = \dim(V_{(a,b)} \cap U_{(d,e)})$, we have $\mathrm{height}(Q) \leq \dim(V_{(a,b)})$. Taking for γ_1 the selection of the $\mathrm{n_t}$ S-boxes of the first step, we have from Definition 2:

$$\dim(V_{(a,b)}) = \mathrm{n_b} - \mathrm{weight}(a,b) = \mathrm{n_b} - \sum_{i=1}^{\mathrm{n_t}} \mathrm{weight}(a_i, b_i) \ . \tag{2}$$

Secondly, observe that the vectors $(b,d) = (b, \mathrm{M}(b))$ are code vectors of a linear code over $\mathrm{GF}(2^{\mathrm{n_s}})$ with length $2\mathrm{n_t}$ and dimension $\mathrm{n_t}$. Since the branch number of M equals $\mathrm{n_t} + 1$, the minimal distance between code vectors is $\mathrm{n_t} + 1$, hence the code is MDS. Any $\mathrm{n_t}$ symbols of the codewords may be taken as message symbols [19]. Hence, we can always pick $\mathrm{n_t}$ out of the symbols, consider them as the message symbols (input) and compute the check symbols (output) from them. This leads to the definition of alternative vector spaces V' which bound height(Q) as before. The new definition of input and output leads to a new definition of input difference a', output difference e' and intermediate differences b', d'. This in turn leads to a definition of a new vector space $V' = V_{a',b'}$, which bounds height(Q) in the same was as in (2). □

Hence, given a characteristic over a super box, one chooses the n_t S-box differentials with the highest weight and adds them. The bound for the height is n_b minus this weight.

Theorem 3. *Let SB be a super box with $n_t = 4$. Then we have the following bounds on the height of characteristics where all active S-boxes have weight $n_s - 1$.*

$$\begin{aligned} \textit{5 active boxes:}\ & \text{height}(Q) \leq 3 \\ \textit{6 active boxes:}\ & \text{height}(Q) \leq 2 \\ \textit{7 or 8 active boxes:}\ & \text{height}(Q) = 1 \end{aligned}$$

The theorem can be proven by going through all possibilities and counting. There is also a link to the existence of codes: a characteristic with i active boxes and height h exists only if there is a binary linear code with length i, distance $i - 3$ and dimension h, which contains the vector $(1, 1, 1, \ldots, 1)$.

5 Plateau characteristics over more than 2 rounds

In this section we derive conditions for characteristics over more than 2 rounds to be plateau characteristics. For ciphers with an AES-like block structure (the AES, Square, SHARK, ...), the results of this section cover the majority of the characteristics. For ciphers without the block structure (3-Way, Serpent, Noekeon, ...), only a small fraction of the characteristics is covered.

We can extend the *planar* property of differentials to characteristics:

Definition 6. *A characteristic Q is* input-planar *(respectively output-planar) if for all values of the key it holds that $F_Q[k]$ (respectively $G_Q[k]$) is either $\emptyset$ or an affine subspace.*

Lemma 6. *Plateau characteristics with height 1 or 2 are both input-planar and output-planar.*

Proof. Let Q be a plateau characteristic. Then $F_Q[k] = G_Q[k] = \emptyset$ or $\#F_Q[k] = \#G_Q[k] = 2^{\text{height}(Q)}$. If $\text{height}(Q)$ is 1 or 2, then the proofs of Lemma 1 and Lemma 2 can be extended to the case of characteristics. □

$$
\begin{array}{ccc}
 & Q_1 & \\
F_{Q_1} & \longrightarrow & G_{Q_1}
\end{array}
\qquad\qquad
\begin{array}{rcl}
 & (q_{n-1}, q_n) & \\
F_{(q_{n-1},q_n)} & \longrightarrow & G_{(q_{n-1},q_n)} \\
H = G_{Q_1} \cap F_{(q_{n-1},q_n)} & \longrightarrow & G_Q \subseteq G_{(q_{n-1},q_n)} \\
\text{proj.} \downarrow & & \downarrow \text{ proj.} \\
H_i & \longrightarrow & O_i
\end{array}
$$

Fig. 1. Notation used in the proof of Theorem 4.

For a cipher with S-boxes that are differentially 4-uniform or differentially 2-uniform we have the following result.

Theorem 4 (Planar Characteristic Extension Theorem). *Let $Q = (Q_1, q_n)$ be a characteristic composed of an output-planar characteristic $Q_1 = (q_0, \ldots, q_{n-1})$, followed by a (one-step) differential (q_{n-1}, q_n). If all S-boxes in (q_{n-1}, q_n) are active, then Q is output-planar.*

Proof. In this proof, we drop $[k]$ from the notation, which we illustrate in Figure 1. If we look at the output of Q_1, the elements of the right pairs of Q_1, form the affine subspace V_{Q_1}. Since the differential (q_{n-1}, q_n) is planar, the elements of the right pairs of Q form an affine subspace $H = V_{Q_1} \cap U_{(q_{n-1},q_n)}$. We denote by H_i, $0 \leq i < \mathrm{n_t}$, the projection of H onto the coordinate i: H_i contains the inputs of one S-box in the last step, for the right pairs of Q. Since H_i is a projection of an affine subspace, it is also an affine subspace. We denote by O_i the corresponding outputs of the S-box. The set O_i is nothing else than the projection of the set G_Q on the coordinate i.

Since the S-boxes are differentially 4-uniform (or 2-uniform), H_i and O_i contain 1, 2, or 4 elements and O_i is also an affine subspace. Using Lemma 4, we conclude that G_Q is also an affine subspace. □

Note that this theorem also holds for a characteristic composed of a planar differential in which all S-boxes are active followed by an input-planar characteristic. Planar plateau characteristics can be composed to plateau characteristics. Theorem 1 can easily be extended.

Theorem 5 (Planar Characteristic Composition Theorem).
Let $Q = (q_0, q_1, \ldots, q_{i-1}, q_i, \ldots, q_n)$ be a characteristic over a map consisting of n steps with a key addition as i-th step. If the characteristics $Q_1 = (q_1, q_2, \ldots, q_{i-1})$ and $Q_2 = (q_i, \ldots, q_n)$

are plateau characteristics with Q_1 output-planar and Q_2 input-planar, then Q is a plateau characteristic with $\text{height}(Q) = \dim(V_{Q_1} \cap U_{Q_2})$.

The proof is similar to the proof of Theorem 1. From this theorem follows this corollary:

Corollary 1. *Let $Q = (Q_1, q_n)$ be a characteristic composed of a characteristic $Q_1 = (q_0, \ldots, q_{n-1})$, followed by a (one-round) planar differential (q_{n-1}, q_n). If Q_1 is a plateau characteristic with height 1 or 2, then Q is a plateau characteristic.*

This is a special case of Theorem 1: according to Lemma 6, Q_1 is output-planar and the differential can be seen as a (one-round) input-planar plateau characteristic.

This extension of a plateau characteristic by a single round can be performed iteratively: an r-round plateau characteristic with height 1 or 2 can be extended by an arbitrary number of rounds, as long as the appended differentials are planar.

Plateau characteristics with height larger than 2, are in general neither input-planar nor output-planar. For instance, assume that we have a plateau characteristic Q consisting of an output-planar characteristic Q_1 followed by an input-planar characteristic Q_2 with $Q = (Q_1, Q_2)$. Then it follows from the proof of Theorem 1 that the elements of the right pairs of Q form an affine subspace H at the junction of Q_1 and Q_2. The set H is transformed to the set $G_Q[k]$ at the output. $G_Q[k]$ is a subset of $G_{Q_2}[k]$. $G_{Q_2}[k]$ is an affine subspace, but $G_Q[k]$ in general is not an affine subspace.

6 Further observations

6.1 Effect of the key schedule

When we consider more than two rounds, the round keys are typically not independent. They are related by means of the key schedule. The key schedule doesn't change whether a characteristic is a plateau characteristic. The only visible effect of the key schedule is on the size of the set K_Q which contains the keys for which $\text{DP}[k](Q) > 0$ (see the proof of Theorem 1). The key schedule determines which values are possible for the expanded key. If relatively many of the expanded keys are in K_Q, then the average of $\text{DP}[k](Q)$ will be larger than $\text{EDP}(Q)$.

6.2 Impact on the DP of differentials

The dependence of the DP of characteristics on the key value means that the DP of differentials also depends on the key. Assume we have a differential for which all characteristics are plateau characteristics. If we denote the characteristics that contribute to a differential (a, b) by Q_i we have:

$$\mathrm{DP}[k](a,b) = 2^{-\mathrm{n_b}} \sum_{i|k \in K_{Q_i}} 2^{\mathrm{height}(Q_i)} \ . \tag{3}$$

Hence this value varies per key k depending on the number of affine subspaces K_{Q_i} it is in.

7 Two-round characteristics in the AES

We now apply the results of the previous sections to the AES. We also compute the heights of all two-round characteristics for the AES and for a simplified variant. Since the sequence of two rounds of Square is equivalent to the sequence of two rounds of the AES, the distribution of the heights of characteristics is the same in both cases.

7.1 The AES Super box

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

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

Here x^{-1} denotes the multiplicative inverse of x in $\mathrm{GF}(2^8)$, extended with 0 being mapped to 0. L is a linear transformation over $GF(2)$ and q a constant. Note that L is not linear over $\mathrm{GF}(2^8)$ and can be expressed as a so-called *linearized polynomial* [18].

The *AES super box* is a super box where the elements are bytes and $\mathrm{n_t} = 4$ and M is the multiplication with the MixColumns matrix, which we denote by $\mathrm{M_c}$. If we consider two AES rounds, swap the steps ShiftRows and SubBytes in the first round, and remove all linear operations before the first application of SubBytes and after the second application of SubBytes, then we obtain a map that can also be described as 4 parallel instances of the AES super box.

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

7.2 Characteristics in the AES super box

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

Applying Theorem 2 to the AES results in the following bounds. It holds always that $\mathrm{height}(Q) \leq 8$. If all active S-boxes have weight 7, then $\mathrm{height}(Q) \leq 4$. Only if at most 3 S-boxes have weight 7, $\mathrm{height}(Q)$ can be larger than 4. Theorem 3 further decreases the bounds when all S-boxes have weight 7.

We have determined the weight and height of all characteristics over the AES super box. An overview of the results is given in Table 1. Because of the large number of characteristics, the entries in the table were not computed by checking the height of each characteristic individually. We used the following observations to speed up the computations. Let $Q = (a, b, d, e)$ be a characteristic over the super box.

Lemma 7. *For all non-zero a, b: $a \in U_{(a,b)}$ and $b \in V_{(a,b)}$.*

Lemma 8. *For differentials with weight 7 over a single S-box $U_{(a,b)} = \{0, a\}$ and $V_{(a,b)} = \{0, b\}$.*

Hence $U_{(a,b)}$ is independent of the output difference b and $V_{(a,b)}$ of the input difference a.

Lemma 9. *Let $Q' = (a', b, d, e')$ be a characteristic in which all S-box differentials have weight 6. Then for all $Q = (a, b, d, e)$,*

$$W_Q \subseteq W_{Q'} \ .$$

Proof. Remember that $W_Q = \mathrm{M_c}(V_{(a,b)}) \cap U_{(d,e)}$. From Lemma 7 and Lemma 8, we have that $V_{(a,b)} \subseteq V_{(a',b)}$ and $U_{(d,e)} \subseteq U_{(d',e)}$. □

Consequently, it is only needed to check the height of each characteristic with (a, e) chosen such that all active S-boxes have weight 6, and then to evaluate the effect of increasing the

weight of the active S-boxes by one. For this last step, only one out of the 126 possible differences a_i, respectively e_i, needs to be tried for each active S-box.

7.3 Observations

We see in Table 1 that the characteristic weight ranges from 30 to 56 and the height from 1 to 5. It follows from the data in the table that the ratio

$$\mathrm{DP}[k](Q)/\mathrm{EDP}(Q) = 2^{\mathrm{height}(Q)-32+\mathrm{weight}(Q)} \tag{5}$$

ranges from 1 to 2^{25}. We call characteristics for which the ratio is 1 *flat characteristics* because for these the equality $\mathrm{DP}[k](Q) = \mathrm{EDP}(Q)$ holds for all keys. Table 1 shows that there are in total $2^{20.9}$ flat characteristics: those with weight 30 and height 2, and those with weight 31 and height 1.

The characteristics for which the ratio is 2^{25} are the characteristics with weight 56 and height 1. Only for a fraction $2^{-32.9}$ of the characteristics this ratio is smaller than 2^{25}. Since the sequence of two AES rounds can be described as the parallel application of 4 super boxes, it follows that for most characteristics over two rounds of the AES, there are keys with $\mathrm{DP}[k](Q) = 2^{100}\mathrm{EDP}(Q)$. The characteristics that are the most interesting for standard differential attacks, are the characteristics with the lowest weight. They are in the top rows of the table. We see that exactly these characteristics have the highest heights, hence the most variation between DP values for different keys. Characteristics with height 5 have a DP equal to $32/2^{32}$, which is almost three times higher than the maximal MEDP of a differential ($13.25/2^{32}$ [14, 23, 24]).

There are 72 characteristics of height 5 and weight 30. These characteristics have nonzero $\mathrm{DP}[k]$ for a fraction $2^{32-30-5} = 2^{-3}$ of all keys. For a given key this results in an expected value of 9 such characteristics with $\mathrm{DP}[k] = 2^5/2^{32}$. Similarly, there are 2^{11} characteristics of height 5 and weight 31 resulting in an expected value of 2^7 such characteristics with $\mathrm{DP}[k] = 2^5/2^{32}$. This totals to an expected number of 137 characteristics with $\mathrm{DP}[k] = 2^5/2^{32}$ per key for the AES super box. For 2 AES rounds, this is 548.

The table shows also that it is easy to find characteristics Q_1, Q_2 with $\mathrm{EDP}(Q_1) < \mathrm{EDP}(Q_2)$ and $\mathrm{height}(Q_1) > \mathrm{height}(Q_2)$.

7.4 An AES variant

If we remove the linear transformation L and the constant q from the S-box (4), we obtain a cipher with a simpler algebraic structure than the AES. We call this variant the *naive AES*. We computed the heights of all characteristics over the naive AES super box. The results are summarized in Table 2. The comparison with the results in Table 1 shows us something about the effect of adding L. (It can be shown that the choice of q has no impact here.) For instance, we see that for the naive AES super box, there are no characteristics with height 5. The characteristics where all active S-boxes have weight 6, have always an even-numbered height, and those with exactly one S-box with weight 7, have always on odd-numbered height.

Although the results on characteristics do not necessarily translate to results on differentials, it appears that L has an impact on the differential properties of two and more rounds of the AES. To the best of our knowledge, this is the first demonstration of the impact of L.

8 Characteristics over four or more rounds in the AES

Four-round AES can be described with a super box-like structure, where again $\mathrm{n_t} = 4$ but now the elements are 32-bit words [26, 27]. The AES super boxes we defined before serve now as (key-dependent) S-boxes. The super box-like structure is constructed as follows. Firstly, rewrite the first two rounds and the last two rounds as the parallel applications of 4 AES super boxes each. It can shown that the remaining linear transformations 'in the middle' form together a map with branch number 5. The remaining 'outer' linear transformations can be ignored.

A characteristic over such a super box-like structure consists of 5 to 8 smaller characteristics, each over an AES super box. If the characteristics over the AES super boxes of the first step are output-planar and the characteristics over the AES super boxes of the second step are input-planar, then according to Theorem 5 the four-round characteristic is a plateau characteristic. These conditions are fulfilled if the characteristics over the AES super boxes

Table 1. Number of characteristics (binary logarithm) per number of active S-boxes, weight and height for the AES super box.

No. active S-boxes	characteristic weight	height 1	2	3	4	5
5	30	—	12.6	12.6	10.6	6.2
	31	20.9	22.1	21.2	18.1	11.0
	32	29.8	30.0	28.2	23.4	—
	33	37.1	36.9	33.7	26.4	—
	34	43.2	42.9	36.2	—	—
	35	48.0	47.5	—	—	—
6	36	20.7	15.6	8.3	3.8	—
	37	30.3	24.2	16.3	11.6	—
	38	38.6	31.5	23.1	17.5	—
	39	46.1	38.1	28.9	—	—
	40	52.6	44.0	33.4	—	—
	41	58.3	49.3	—	—	—
	42	62.7	53.4	—	—	—
7	42	27.0	15.7	5.3	—	—
	43	36.8	24.3	13.1	—	—
	44	45.3	31.7	19.5	—	—
	45	53.1	38.0	24.9	—	—
	46	60.0	43.5	—	—	—
	47	66.3	48.0	—	—	—
	48	71.7	50.9	—	—	—
	49	75.9	—	—	—	—
8	48	32.0	14.7	1.0	—	—
	49	41.9	23.7	9.0	—	—
	50	50.7	31.4	15.0	—	—
	51	58.7	38.3	—	—	—
	52	66.0	44.5	—	—	—
	53	72.7	49.9	—	—	—
	54	78.7	54.1	—	—	—
	55	83.7	—	—	—	—
	56	87.9	—	—	—	—
total		87.9	55.0	36.6	26.6	11.0

Table 2. Number of characteristics (binary logarithm) per number of active S-boxes, weight and height for the naive super box.

No. active S-boxes	characteristic weight	height 1	2	3	4
5	30	—	12.6	—	13.0
	31	21.9	—	22.3	—
	32	29.9	29.6	28.8	—
	33	37.1	37.0	33.5	—
	34	43.2	42.9	—	—
	35	48.0	47.5	—	—
6	36	—	20.8	—	12.8
	37	30.3	—	22.4	—
	38	38.7	30.1	29.0	—
	39	46.1	37.7	34.2	—
	40	52.6	44.0	—	—
	41	58.3	49.3	—	—
	42	62.7	53.4	—	—
7	42	—	27.0	—	11.0
	43	36.8	—	20.8	—
	44	45.3	28.9	27.3	—
	45	53.1	36.4	31.9	—
	46	60.0	42.4	—	—
	47	66.3	46.9	—	—
	48	71.7	—	—	—
	49	75.9	—	—	—
8	48	—	32.0	—	13.0
	49	41.9	—	23.0	—
	50	50.7	31.4	29.6	—
	51	58.7	39.0	34.7	—
	52	66.0	45.3	—	—
	53	72.7	50.6	—	—
	54	78.7	54.4	—	—
	55	83.7	—	—	—
	56	87.9	—	—	—
total		87.9	55.1	36.0	14.6

– have height 1 or 2 (by Lemma 6), or
– have 4 active S-boxes at the output, respectively at the input (by Theorem 4).

This implies that most of the four-round characteristics over the AES are plateau characteristics. We have not determined the distribution of the heights, but Theorem 2 and a straightforward generalization of Theorem 3 apply. Since the overwhelming majority of the characteristics over the AES super box have height 1, we expect that also the vast majority of the characteristics over 4 rounds of the AES will have height 1. Corollary 1 would then imply that the vast majority of characteristics over more than 4 AES rounds are plateau characteristics with height 1.

9 DP of differentials in the AES

The exact distribution of $\mathrm{DP}[k](a, b)$ depends on the relative positions of the affine subspaces K_{Q_i} and the height of the characteristics. Flat characteristics add a constant term and do not contribute to the variability. The larger the height of a characteristic, the more it contributes to the variability. In the AES super box there is at most one flat characteristic per differential with 5 active S-boxes and none for differentials with more than 5 active S-boxes.

The AES has no flat plateau characteristics over 4 rounds or more, and the vast majority of characteristics has height equal to 1. In the assumption that the affine subspaces K_{Q_i} are independent, the distribution of the $\mathrm{DP}[k]$ of any four-round differential is the convolution of a huge number of distributions with a high peak in 0 and a very small peak in 2^{-127}. We conjecture that this gives rise to a Poisson distribution.

10 Conclusions and further work

We believe that an analysis of resistance against differential cryptanalysis needs to take into account more than the average behavior of a key-dependent map.

In this paper, we showed that the $\mathrm{DP}[k]$ of certain characteristics is distributed in a very structured way. For characteristics over the AES super box, we showed that $\mathrm{DP}[k] \in 0, 2^{h-32}$, with h an integer value between 1 and 5. We think that the results are somewhat surprising

and deserve to be investigated in further detail. We have illustrated our analysis only on the AES, but several other ciphers using differentially 4-uniform S-boxes will show a similar behavior.

It would be interesting to find out what the maximum $\mathrm{DP}[k]$ is for characteristics over more than 2 rounds. If the impact on the $\mathrm{DP}[k]$ of differentials over more than 2 rounds can be investigated, then this could lead to new insights about the security margin of the AES and other ciphers.

References

1. American Mathematical Society, *Algebra*, ISBN 0821816462, 1999.
2. R.A. Anderson, E. Biham, L.R. Knudsen, "Serpent," *Proc. of the 1st AES candidate conference,* CD-1: Documentation, August 20–22, 1998, Ventura.
3. K. Aoki, "Maximum non-averaged differential probability," *Selected Areas in Cryptography SAC '98, LNCS 1556*, Springer-Verlag, 1998, pp. 118–130.
4. K. Aoki, T. Ichikawa, M. Kanda, M. Matsui, S. Moriai, J. Nakajima, T. Tokita, "Camellia: a 128-bit block cipher suitable for multiple platforms — Design and analysis," *Selected Areas in Cryptography 2000, LNCS 2012,* D. Stinson, S. Tavares, Eds., Springer-Verlag, 2000, pp. 39–56.
5. I. Ben-Aroya, E. Biham, "Differential cryptanalysis of Lucifer," *Advances in Cryptology, Proc. Crypto'93, LNCS 773*, D. Stinson, Ed., Springer-Verlag, 1994, pp. 187–199.
6. E. Biham and A. Shamir, "Differential cryptanalysis of DES-like cryptosystems," *Journal of Cryptology*, Vol. 4, No. 1, 1991, pp. 3–72.
7. A. Canteaut, "Differential cryptanalysis of Feistel ciphers and differentially δ-uniform mappings," *Workshop record of Selected Areas in Cryptography SAC '97*, 1997, pp. 172–184.
8. J. Daemen, R. Govaerts and J. Vandewalle, "A new approach to block cipher design," *Proc. of Fast Software Encryption 1993, LNCS 809,* R. Anderson, Ed. Springer-Verlag, 1994, pp. 18–32.
9. J. Daemen, R. Govaerts and J. Vandewalle, "Weak keys of IDEA," *Advances in Cryptology, Proc. Crypto'93, LNCS 773*, D. Stinson, Ed., Springer-Verlag, 1994, pp. 224–231.
10. J. Daemen, L.R. Knudsen and V. Rijmen, "The block cipher Square," *Fast Software Encryption '97, LNCS 1267*, E. Biham, Ed., Springer-Verlag, 1997, pp. 149–165.
11. J. Daemen, M. Peeters, G. Van Assche and V. Rijmen, "Nessie Proposal: the block cipher Noekeon," Submitted to Nessie.
12. J. Daemen, V. Rijmen, *The design of Rijndael — AES, The Advanced Encryption Standard,* Springer-Verlag, 2002.

13. O. Dunkelman, N. Keller, "A new criterion for nonlinearity of block ciphers," *CT-RSA 2006, LNCS 3860*, D. Pointcheval, Ed., Springer-Verlag, 2006, pp. 295–312.
14. L. Keliher and J. Sui, "Exact maximum expected differential and linear probability for 2-round Advanced Encryption Standard (AES)," Cryptology ePrint archive, Report 2005/321, 2005, http://eprint.iacr.org.
15. L.R. Knudsen, "Iterative characteristics of DES and s^2-DES," *Advances in Cryptology, Proc. CRYPTO'92, LNCS 746*, E.F. Brickell, Ed., Springer-Verlag, 1993, pp. 497–511.
16. L.R. Knudsen and J.E. Mathiassen, "On the role of key schedules in attacks on iterated ciphers," *ESORICS 2004, LNCS 3193,* Springer-Verlag, 2004, pp. 322–334.
17. X. Lai, J.L. Massey and S. Murphy, "Markov ciphers and differential cryptanalysis," *Advances in Cryptology, Proc. Eurocrypt'91, LNCS 547*, D.W. Davies, Ed., Springer-Verlag, 1991, pp. 17–38.
18. R. Lidl, H. Niederreiter, *Introduction to finite fields and their applications,* Cambridge University Press, 1986 (Reprinted 1988).
19. F.J. MacWilliams and N.J.A. Sloane, *The theory of error-correcting codes*, North-Holland Publishing Company, 1978.
20. M. Matsui, "New block encryption algorithm Misty," *Fast Software Encryption '97, LNCS 1267,* E. Biham, Ed., Springer-Verlag, 1997, pp. 64–74.
21. K. Nyberg, "Differentially uniform mappings for cryptography," *Advances in Cryptology, Proc. Eurocrypt'93, LNCS 765*, T. Helleseth, Ed., Springer-Verlag, 1993, pp. 55-64.
22. K. Nyberg and L.R. Knudsen, "Provable security against a differential attack," *Journal of Cryptology*, Vol. 8, No. 1, 1995, pp. 27–38.
23. S. Park, S.H. Sung, S. Chee, E. Yoon and J. Lim, "On the security of Rijndael-like structures against differential and linear cryptanalysis," *Asiacrypt 2002, LNCS 2501*, Y. Zheng, Ed., Springer-Verlag, 2002, pp. 176–191.
24. S. Park, S.H. Sung, S. Lee and J. Lim, "Improving the upper bound on the maximum differential and the maximum linear hull probability for SPN structures and AES," *Fast Software Encryption '03, LNCS 2887,* T. Johansson, Ed., Springer-Verlag, 2003, pp. 247–260.
25. V. Rijmen, J. Daemen, B. Preneel, A. Bosselaers, E. De Win, "The cipher SHARK," *Fast Software Encryption '96, LNCS 1039*, D. Gollmann, Ed., Springer-Verlag, 1996, pp. 99–111.
26. V. Rijmen, "Cryptanalysis and design of iterated block ciphers," *Doctoral Dissertation*, October 1997, K.U.Leuven.
27. Toshiba corporation, "Specification of Hierocrypt-L1," available from the NESSIE homepage, URL: `http://cryptonessie.org`.
28. Y. Zheng, X.M. Zhang, "Plateaued functions," *Advances in Cryptology, ICICS '99, LNCS 1726*, Springer-Verlag, 1999, pp. 284–300.