Polytopic Cryptanalysis

Tyge Tiessen

DTU Compute, Technical University of Denmark, Kgs. Lyngby, Denmark
tyti@dtu.dk

Abstract. Standard differential cryptanalysis uses statistical dependencies between the difference of two plaintexts and the difference of the respective two ciphertexts to attack a cipher. Here we introduce polytopic cryptanalysis which considers interdependencies between larger sets of texts as they traverse through the cipher. We prove that the methodology of standard differential cryptanalysis can unambiguously be extended and transferred to the polytopic case including impossible differentials. We show that impossible polytopic transitions have generic advantages over impossible differentials. To demonstrate the practical relevance of the generalization, we present new low-data attacks on round-reduced DES and AES using impossible polytopic transitions that are able to compete with existing attacks, partially outperforming these.

1 Introduction

Without doubt is differential cryptanalysis one of the most important tools that the cryptanalyst has at hand when trying to evaluate the security of a block cipher. Since its conception by Biham and Shamir [2] in their effort to break the Data Encryption Standard [26], it has been successfully applied to many block ciphers such that any modern block cipher is expected to have strong security arguments against this attack.

The methodology of differential cryptanalysis has been extended several times with a number of attack vectors, most importantly truncated differentials [19], impossible differentials [1,20], and higher-order differentials [19,22]. Further attacks include the boomerang attack [29], which bears some resemblance of second-order differential attacks, and differential-linear attacks [24].

Nonetheless many open problems remain in the field of differential cryptanalysis. Although the concept of higher-order differentials is almost 20 years old, it has not seen many good use cases. One reason has been the difficulty of determining the probability of higher-order differentials accurately without evaluating Boolean functions with prohibitively many terms. Thus the common use case remains probability 1 higher-order differentials where we know that a derivative of a certain order has to evaluate to zero because of a limit in the degree of the function.

Another open problem is the exact determination of the success probability of boomerang attacks and their extensions. It has correctly been observed that the correlation between differentials must be taken into account to accurately determine the success probability [25]. The true probability can otherwise deviate arbitrarily from the estimated one.

Starting with Chabaud and Vaudenay [13], considerable effort has gone into shedding light on the relation and interdependencies of various cryptographic attacks (see for example [5,6,30]). With this paper, we offer a generalized view on the various types of differential attacks that might help to understand both the interrelation between the attacks as well as the probabilities of the attacks better.

Our contribution

In this paper we introduce polytopic cryptanalysis. It can be viewed as a generalization of standard differential cryptanalysis which it embeds as a special case. We prove that the definitions and methodology of differential cryptanalysis can unambiguously be extended to polytopic cryptanalysis, including the concept of impossible differentials. Polytopic cryptanalysis is general enough to even encompass attacks such as higher-order differentials and might thus be valuable as a reference framework.

For impossible polytopic transitions, we show that they exhibit properties that allow them to be very effective in scenarios where ordinary impossible differentials fail. This is mostly due to a generic limit in the diffusion of any block cipher that guarantees that only a negligible number of all polytopic transitions is possible for a sufficiently high choice of dimension. This also makes impossible polytopic transitions ideal for low-data attacks where standard impossible differentials usually have a high data complexity.

Finally we prove that polytopic cryptanalysis is not only theoretically intriguing but indeed relevant for practical cryptanalysis by demonstrating competitive impossible polytopic attacks on round-reduced DES and AES that partly outperform existing low-data attacks and offer different trade-offs between time and data complexity.

In the appendix, we further prove that higher-order differentials can be expressed as truncated polytopic transitions and are hence a special case of these. Thus higher-order differentials can be expressed in terms of a collection of polytopic trails just as differentials can be expressed as a collection of differential trails. A consequence of this is that it is principally possible to determine lower bounds for the probability of a higher-order differential by summing over the probabilities of a subset of the polytopic trails which it contains.

Related work

To our knowledge, the concept of polytopic transitions is new and has not been used in cryptanalysis before. Nonetheless there is other work that shares some similarities with polytopic cryptanalysis.

Higher-order differentials [22] can in some sense also be seen as a higher-dimensional version of a differential. However, most concepts of ordinary differentials do not seem to extend to higher-order differentials, such as characteristics or iterated differentials.

The idea of using several differentials simultaneously in an attack is not new (see for example [4]). However as opposed to assuming independence of the differentials, which does not hold in general (see [25]), we explicitly take their correlation into account and use it in our framework.

Another type of cryptanalysis that uses a larger set of texts instead of a single pair is integral cryptanalysis (see for example [3,15]), in which structural properties of the cipher are used to elegantly determine a higher-order derivative to be zero without relying on bounds in the degree. These attacks can be considered a particular form of higher-order differentials.

Finally decorrelation theory [28] also considers relations between multiple plaintext-ciphertext pairs but takes a different direction by considering security proofs based on a lack of correlation between the texts.

Organization of the paper

In Section 2, notation and concepts necessary for polytopic cryptanalysis are introduced. It is demonstrated how the concepts of differential cryptanalysis naturally extend to polytopic cryptanalysis. We also take a closer look at the probability of polytopic transitions and applicability of simple polytopic cryptanalysis.

In Section 3, we introduce impossible polytopic transitions. We show that impossible polytopic transitions offer some inherent advantages over impossible differentials and are particularly interesting for low-data attacks. We show that, given an efficient method to determine the possibility of a polytopic transition, generic impossible polytopic attack always exist.

In Section 4, we demonstrate the practicability of impossible polytopic transition attacks. We present some attacks on DES and AES that are able to compete with existing attacks with low-data complexity, partially outperforming these.

Furthermore, in Appendix B truncated polytopic transitions are introduced. We then give a proof that higher-order differentials are a special case of these. The cryptanalytic ramifications of the fact that higher-order differentials consist of polytopic trails are then discussed.

Notation

We use $\mathbb{F}_2^n$ to denote the n-dimensional binary vector space. To identify numbers in hexadecimal notation we use a typewriter font as in `3af179`. Random variables are denoted with bold capital letters ($\mathbf{X}$). We will denote d-difference (introduced later) by bold Greek letters ($\boldsymbol{\alpha}$) and standard differences by Roman (i.e., non-bold) Greek letters (α).

2 Polytopes and polytopic transitions

Classical differential cryptanalysis utilizes the statistical interdependency of two texts as they traverse through the cipher. When we are not interested in the absolute position of the two texts in the state space, the difference between the two texts completely determines their relative positioning.

But there is no inherent reason that forces us to be restricted to only using a pair of texts. Let us instead consider an ordered set of texts as they traverse through the cipher.

Definition 1 (s-polytope). *An s-polytope in $\mathbb{F}_2^n$ is an s-tuple of values in $\mathbb{F}_2^n$.*

Similar to differential cryptanalysis, we are not so much interested in the absolute position of these texts but the relations between the texts. If we choose one of the texts as the point of reference, the relations between all texts are already uniquely determined by only considering their differences with respect to the reference text. If we thus have $d+1$ texts, we can describe their relative positioning by a tuple of d differences (see also Fig. 1).

Definition 2 (d-difference). *A d-difference over $\mathbb{F}_2^n$ is a d-tuple of values in $\mathbb{F}_2^n$ describing the relative position of the texts of a $(d+1)$-polytope from one point of reference.*

When we reduce a $(d+1)$-polytope to a corresponding d-difference, we loose the information of the absolute position of the polytope. A d-difference thus corresponds to an equivalence class of $(d+1)$-polytopes where polytopes are equivalent if and only if they can be transformed into each other by simple shifting in state space. We will mostly be dealing with these equivalence classes.

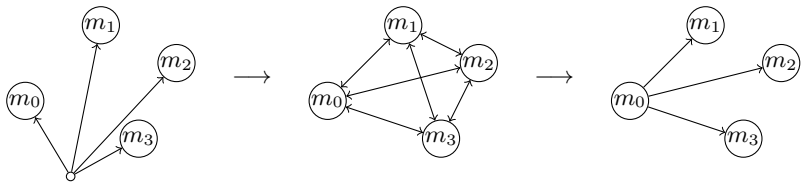

Fig. 1. Depiction of three views of a polytope with four vertices.

In principal there are many d-differences that correspond to one $(d+1)$-polytope depending on the choice of reference text and the order of the differences. As a convention we will construct a d-difference from a $(d+1)$-polytope as follows:

Convention. For a $(d+1)$-polytope $(m_0, m_1, \ldots, m_d)$, the corresponding d-difference is created as $(m_0 \oplus m_1, m_0 \oplus m_2, \ldots, m_0 \oplus m_d)$.

This means, we use the first text of the polytope as the reference text and write the differences in the same order as the remaining texts of the polytope. We will call the reference text the *anchor* of the d-difference. Hence if we are given a d-difference and the value of the anchor, we can reconstruct the corresponding $(d+1)$-polytope uniquely.

Example. Let (m_0, m_1, m_2, m_3) be a 4-polytope in $\mathbb{F}_2^n$. Then $(m_0 \oplus m_1, m_0 \oplus m_2, m_0 \oplus m_3)$ is the corresponding 3-difference with m_0 as the anchor.

In the following, we will now show that we can build a theory of polytopic cryptanalysis in which the same methodology as in standard differential cryptanalysis applies. Standard differential cryptanalysis is contained in this framework as a special case.

A short note regarding possible definitions of difference: in this paper we restrict ourselves to XOR-differences as the most common choice. Most, if not all, statements in this paper naturally extend to other definitions of difference, e.g., in modular arithmetic.

The equivalent of a differential in polytopic cryptanalysis is the polytopic transition. We use d-differences for the definition.

Definition 3 (Polytopic transition with fixed anchor). *Let $f : \mathbb{F}_2^n \to \mathbb{F}_2^q$. Let $\boldsymbol{\alpha}$ be a d-difference $(\alpha_1, \alpha_2, \ldots, \alpha_d)$ over $\mathbb{F}_2^n$ and let $\boldsymbol{\beta}$ be the d-difference $(\beta_1, \beta_2, \ldots, \beta_d)$ over $\mathbb{F}_2^q$. By the $(d+1)$-polytopic transition $\boldsymbol{\alpha} \xrightarrow[x]{f} \boldsymbol{\beta}$ we denote that f maps the polytope corresponding to $\boldsymbol{\alpha}$ with anchor x to a polytope corresponding to $\boldsymbol{\beta}$. More precisely, we have $\boldsymbol{\alpha} \xrightarrow[x]{f} \boldsymbol{\beta}$ if and only if*

$$\begin{aligned} f(x \oplus \alpha_1) \oplus f(x) &= \beta_1 \\ \text{and } f(x \oplus \alpha_2) \oplus f(x) &= \beta_2 \\ &\ldots \\ \text{and } f(x \oplus \alpha_d) \oplus f(x) &= \beta_d. \end{aligned}$$

Building up on this definition, we can now define the probability of a polytopic transition under a random anchor.

Definition 4 (Polytopic transition). *Let f, $\boldsymbol{\alpha}$, and $\boldsymbol{\beta}$ again be as in Definition 3. The probability of the $(d+1)$-polytopic transition $\boldsymbol{\alpha} \xrightarrow{f} \boldsymbol{\beta}$ is then defined as:*

$$\Pr\left(\boldsymbol{\alpha} \xrightarrow{f} \boldsymbol{\beta}\right) := \Pr_{\mathbf{X}}\left(\boldsymbol{\alpha} \xrightarrow[\mathbf{X}]{f} \boldsymbol{\beta}\right) \tag{1}$$

where $\mathbf{X}$ is a random variable, uniformly distributed on $\mathbb{F}_2^n$. We will at times also write $\boldsymbol{\alpha} \longrightarrow \boldsymbol{\beta}$ if the function is clear from the context or not important.

Note that this definition coincides with the definition of the differential probability when differences between only two texts (2-polytopes) are considered.

Let $f : \mathbb{F}_2^n \to \mathbb{F}_2^n$ now be a function that is the repeated composition of round functions $f_i : \mathbb{F}_2^n \to \mathbb{F}_2^n$:

$$f := f_r \circ \cdots \circ f_2 \circ f_1. \tag{2}$$

Similarly to differential cryptanalysis, we can now define trails of polytopes:

Definition 5 (Polytopic trail). *Let f be as in Eq. (2). A polytopic trail on f is an $(r+1)$-tuple of d-differences $(\boldsymbol{\alpha}_0, \boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_r)$ written as*

$$\boldsymbol{\alpha}_0 \xrightarrow{f_1} \boldsymbol{\alpha}_1 \xrightarrow{f_2} \cdots \xrightarrow{f_r} \boldsymbol{\alpha}_r. \tag{3}$$

The probability of such a polytopic trail is defined as

$$\Pr_{\mathbf{X}}\left(\boldsymbol{\alpha}_0 \xrightarrow[\mathbf{X}]{f_1} \boldsymbol{\alpha}_1 \text{ and } \boldsymbol{\alpha}_1 \xrightarrow[f_1(\mathbf{X})]{f_2} \boldsymbol{\alpha}_2 \text{ and } \cdots \text{ and } \boldsymbol{\alpha}_{r-1} \xrightarrow[f_{r-1}\circ\cdots\circ f_1(\mathbf{X})]{f_r} \boldsymbol{\alpha}_r\right) \tag{4}$$

where $\mathbf{X}$ is a random variable, distributed uniformly on $\mathbb{F}_2^n$.

Similarly to differentials, it is possible to partition a polytopic transition over a composed function into all polytopic trails that feature the same input and output differences as the polytopic transition.

Proposition 1. *The probability of a polytopic transition $\boldsymbol{\alpha}_0 \xrightarrow{f} \boldsymbol{\alpha}_r$ over a function $f : \mathbb{F}_2^n \to \mathbb{F}_2^n, f = f_r \circ \cdots \circ f_2 \circ f_1$ is the sum of the probabilities of all polytopic trails $(\boldsymbol{\alpha}_0, \boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_r)$ which it contains:*

$$\Pr\left(\boldsymbol{\alpha}_0 \xrightarrow{f} \boldsymbol{\alpha}_r\right) = \sum_{\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_{r-1}} \Pr\left(\boldsymbol{\alpha}_0 \xrightarrow{f_1} \boldsymbol{\alpha}_1 \xrightarrow{f_2} \cdots \xrightarrow{f_{r-1}} \boldsymbol{\alpha}_{r-1} \xrightarrow{f_r} \boldsymbol{\alpha}_r\right) \tag{5}$$

where $\boldsymbol{\alpha}_0, \ldots, \boldsymbol{\alpha}_r$ are d-differences and as such lie in $\mathbb{F}_2^{dn}$.

Proof. If we fix the initial value of the anchor, we also fix the trail that the polytope has to take. The set of polytopic trails gives us thus a partition of the possible anchor values and in particular a partition of the anchors for which the output polytope is of type $\boldsymbol{\alpha}_r$. Using the above definitions we thus get:

$$\begin{aligned}
\Pr\left(\boldsymbol{\alpha}_0 \xrightarrow{f} \boldsymbol{\alpha}_r\right) &= \Pr_{\mathbf{X}}\left(\boldsymbol{\alpha}_0 \xrightarrow[\mathbf{X}]{f} \boldsymbol{\alpha}_r\right) \\
&= 2^{-n} \cdot \left|\left\{x \in \mathbb{F}_2^n \;\middle|\; \boldsymbol{\alpha}_0 \xrightarrow[x]{f} \boldsymbol{\alpha}_r\right\}\right| \\
&= 2^{-n} \cdot \sum_{\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_{r-1}} \left|\left\{x \in \mathbb{F}_2^n \;\middle|\; \boldsymbol{\alpha}_0 \xrightarrow[x]{f_1} \boldsymbol{\alpha}_1,\ \boldsymbol{\alpha}_1 \xrightarrow[f_1(x)]{f_2} \boldsymbol{\alpha}_2, \ldots \right.\right. \\
&\qquad\qquad \left.\left. \ldots,\ \boldsymbol{\alpha}_{r-1} \xrightarrow[f_{r-1}\circ\cdots\circ f_1(x)]{f_r} \boldsymbol{\alpha}_r\right\}\right| \\
&= \sum_{\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_{r-1}} \Pr_{\mathbf{X}}\left(\boldsymbol{\alpha}_0 \xrightarrow[\mathbf{X}]{f_1} \boldsymbol{\alpha}_1 \text{ and } \boldsymbol{\alpha}_1 \xrightarrow[f_1(\mathbf{X})]{f_2} \boldsymbol{\alpha}_2 \text{ and } \ldots \right. \\
&\qquad\qquad \left. \ldots \text{ and } \boldsymbol{\alpha}_{r-1} \xrightarrow[f_{r-1}\circ\cdots\circ f_1(\mathbf{X})]{f_r} \boldsymbol{\alpha}_r\right) \\
&= \sum_{\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_{r-1}} \Pr\left(\boldsymbol{\alpha}_0 \xrightarrow{f_1} \boldsymbol{\alpha}_1 \xrightarrow{f_2} \cdots \xrightarrow{f_{r-1}} \boldsymbol{\alpha}_{r-1} \xrightarrow{f_r} \boldsymbol{\alpha}_r\right)
\end{aligned}$$

which proves the proposition. □

To be able to calculate the probability of a differential trail, it is common in differential cryptanalysis to make an assumption on the independence of the round transitions. This is usually justified by showing that the cipher is a Markov cipher and by assuming the stochastic equivalence hypothesis (see [23]). As we will mostly be working with impossible trails where these assumptions are not needed, we will assume for now that this independence holds and refer the interested reader to Appendix A where the Markov model is adapted to polytopic cryptanalysis.

Under the assumption that the single round transitions are independent, we can work with polytopic transitions just as with differentials:

1. The probability of a polytopic transition is the sum of the probabilities of all polytopic trails with the same input and output d-difference.
2. The probability of a polytopic trail is the product of the probabilities of the 1-round polytopic transitions that constitute the trail.

We are thus principally able to calculate the probability of a polytopic transition over many rounds by knowing how to calculate the polytopic transition over single rounds.

Now to calculate the probability of a 1-round polytopic transition, we can use the following observations:

3. A linear function maps a d-difference with probability 1 to the d-difference that is the result of applying the linear function to each single difference in the d-difference.
4. Addition of a constant to the anchor leaves the d-difference unchanged.
5. The probability of a polytopic transition over an S-box layer is the product of the polytopic transitions for each S-box.

We are thus able to determine probabilities of polytopic transitions and polytopic trails just as we are used to from standard differential cryptanalysis.

A note on correlation, diffusion and the difference distribution table

When estimating the probability of a polytopic transition a first guess might be that it is just the product of the individual 1-dimensional differentials. For a 3-polytopic transition we might for example expect:

$$\Pr\Big((\alpha_0, \alpha_1) \longrightarrow (\beta_0, \beta_1)\Big) \stackrel{?}{=} \Pr\Big(\alpha_0 \longrightarrow \beta_0\Big) \cdot \Pr\Big(\alpha_1 \longrightarrow \beta_1\Big).$$

That this is generally *not* the case is a consequence of the following lemma.

Lemma 1. *Let $f : \mathbb{F}_2^n \to \mathbb{F}_2^n$. For a given input d-difference $\boldsymbol{\alpha}$ the number of output d-differences to which $\boldsymbol{\alpha}$ is mapped with non-zero probability is upper bounded by 2^n.*

Proof. This is just a result of the fact that the number of anchors for the transition is limited to 2^n:

$$\left|\left\{\boldsymbol{\beta} \in \mathbb{F}_2^{dn} \;\middle|\; \Pr\left(\boldsymbol{\alpha} \xrightarrow{f} \boldsymbol{\beta}\right) > 0\right\}\right| = \left|\left\{\boldsymbol{\beta} \in \mathbb{F}_2^{dn} \;\middle|\; \exists x \in \mathbb{F}_2^n : \boldsymbol{\alpha} \xrightarrow[x]{f} \boldsymbol{\beta}\right\}\right| \leq 2^n$$

□

One implication of this limitation of possible output d-differences is a correlation between differentials: the closer the distribution of differences of a function is to a uniform distribution, the stronger is the correlation of differentials over that function.

Example. Let us take the AES 8-bit S-box (denoted by S here) which is differentially 4-uniform. Consider the three differentials, $\mathtt{7} \xrightarrow{S} \mathtt{166}$, $\mathtt{25} \xrightarrow{S} \mathtt{183}$, and $\mathtt{25} \xrightarrow{S} \mathtt{1}$ which have probabilities 2^{-6}, 2^{-6}, and 2^{-7} respectively. The probabilities of the polytopic transitions of the combined differentials deviate strongly from the product of the single probabilities:

$$\Pr\left((\mathtt{7},\mathtt{25}) \xrightarrow{S} (\mathtt{166},\mathtt{183})\right) = 2^{-6} > \Pr\left(\mathtt{7} \xrightarrow{S} \mathtt{166}\right) \cdot \Pr\left(\mathtt{25} \xrightarrow{S} \mathtt{183}\right) = 2^{-12}$$
$$\Pr\left((\mathtt{7},\mathtt{25}) \xrightarrow{S} (\mathtt{166},\mathtt{1})\right) = 0 < \Pr\left(\mathtt{7} \xrightarrow{S} \mathtt{166}\right) \cdot \Pr\left(\mathtt{25} \xrightarrow{S} \mathtt{1}\right) = 2^{-13}.$$

Another consequence of Lemma 1 is that it sets an inherent limit to the maximal diffusion possible over one round. A one d-difference can at most be mapped to 2^n possible d-differences over one round, the number of possible d-differences reachable can only increase by a factor of 2^n over each round. Thus when starting from one d-difference, after one round at most 2^n d-differences are possible, after two rounds at most 2^{2n} differences are possible, after three rounds at most 2^{3n} are possible and generally after round r at most 2^{rn} d-differences are possible.

In standard differential cryptanalysis, the number of possible output differences for a given input difference is limited by the state size of the function. This is no longer true for d-differences: if the state space is $\mathbb{F}_2^n$, the space of d-differences is $\mathbb{F}_2^{dn}$. The number of possible d-differences thus increases exponentially with the dimension d. This has a consequence for the size of the difference distribution table (DDT). For an 8-bit S-box, a classical DDT has a size of 2^{16} entries, i.e., 64 kilobytes. But already the DDT for 3-differences has a size of 2^{48}, i.e., 256 terabytes. Fortunately though, a third consequence of Lemma 1 is that the DDT table is sparse for $d > 1$. As a matter of fact, we can calculate any row of the DDT with a time complexity of 2^n by trying out all possible values for the anchor.

Relation to decorrelation theory. Decorrelation theory [28] is a framework that can be used to design ciphers which are provably secure against a range of attacks including differential and linear cryptanalysis. A cipher is called perfectly decorrelated of order d when the image of any d-tuple of distinct plaintexts is

uniformly distributed on all d-tuples of ciphertexts with distinct values under a uniformly distributed random key. It can for example be proved that a cipher which is perfectly decorrelated of order 2 is secure against standard differential and linear cryptanalysis.

When we consider $(d+1)$-polytopes in polytopic cryptanalysis, we can naturally circumvent security proofs for order-d perfectly decorrelated ciphers. The boomerang attack [29] for example – invented to break an order-2 perfectly decorrelated cipher – can be described as a 4-polytopic attack.

Limitations of simple polytopic cryptanalysis

Can simple polytopic cryptanalysis, i.e., using a single polytopic transition, outperform standard differential cryptanalysis? Unfortunately this is generally not the case as is shown in the following.

Definition 6. *Let $\boldsymbol{\alpha} \longrightarrow \boldsymbol{\beta}$ be a $(d+1)$-polytopic transition with d-differences $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. Let $\boldsymbol{\alpha}' \longrightarrow \boldsymbol{\beta}'$ be a d'-difference with $d' \leq d$. We then write $(\boldsymbol{\alpha}', \boldsymbol{\beta}') \sqsubseteq (\boldsymbol{\alpha}, \boldsymbol{\beta})$ if and only if for each $i \in [1, d']$ there exists $j \in [1, d]$ such that the ith differences in $\boldsymbol{\alpha}'$ and $\boldsymbol{\beta}'$ correspond to the jth differences in $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$.*

Using this notation, we have the following lemma:

Lemma 2. *Let $\boldsymbol{\alpha} \longrightarrow \boldsymbol{\beta}$ be a $(d+1)$-polytopic transition and let $\boldsymbol{\alpha}' \longrightarrow \boldsymbol{\beta}'$ be a $(d'+1)$-polytopic transition with $d' \leq d$ and $(\boldsymbol{\alpha}', \boldsymbol{\beta}') \sqsubseteq (\boldsymbol{\alpha}, \boldsymbol{\beta})$. Then*

$$\Pr(\boldsymbol{\alpha} \longrightarrow \boldsymbol{\beta}) \leq \Pr(\boldsymbol{\alpha}' \longrightarrow \boldsymbol{\beta}'). \tag{6}$$

Proof. This follows directly from the fact that $\boldsymbol{\alpha} \xrightarrow[x]{f} \boldsymbol{\beta}$ implies $\boldsymbol{\alpha}' \xrightarrow[x]{f} \boldsymbol{\beta}'$. □

In words, the probability of a polytopic transition is always at most as high as the probability of any lower dimensional polytopic transition that it contains. In particular, it can never have a higher probability than any standard differential that it contains.

It can in some instances still be profitable to use a single polytopic transition instead of a standard differential that it contains. This is the case when the probability of the polytopic transition is the same as (or close to) the probability of the best standard differential it contains. Due to the fact that the space of d-differences is much larger than that of standard differentials (2^{dn} vs. 2^n), one set of texts that follows the polytopic transition is usually enough to distinguish the biased distribution from a uniform distribution as opposed to standard differentials where at least two are needed. Nonetheless the cryptanalytic advantages of polytopic cryptanalysis lie elsewhere as we will see in the next sections.

3 Impossible polytopic cryptanalysis

Impossible differential cryptanalysis makes use of differentials with probability zero to distinguish a cipher from an ideal cipher. In this section, we extend the definition to encompass polytopic transitions.

Impossible polytopic cryptanalysis offers distinct advantages over standard impossible differential cryptanalysis that are a result of the exponential increase in the size of the space of d-differences with increasing dimension d. This not only allows impossible $(d+1)$-polytopic attacks using just a single set of $d+1$ chosen plaintexts, it also allows generic distinguishing attacks on $(d-1)$-round block ciphers whenever it is computationally easy to determine whether a $(d+1)$-polytopic transition is possible or not. We elaborate this in more detail later in this section.

Definition 7. *An impossible $(d+1)$-polytopic transition is a $(d+1)$-polytopic transition that occurs with probability zero.*

In impossible differential attacks, we use knowledge of an impossible differential over $r-1$ rounds to filter out wrong round key guesses for the last round: any round key that decrypts a text pair such that their difference adheres to the impossible differential has to be wrong. The large disadvantage of this attack is that it always requires a large number of text pairs to sufficiently reduce the number of possible keys. This is due to the fact that the filtering probability corresponds to the fraction of the impossible differentials among all differentials. Unfortunately for the attacker, most ciphers are designed to provide good diffusion, such that this ratio is usually low after a few rounds.

This is exactly where the advantage of impossible polytopic transitions lies. Due to the exponential increase in the size of the space of d-differences (from $\mathbb{F}_2^n$ to $\mathbb{F}_2^{dn}$) and the limitation of the diffusion to maximally a factor of 2^n (see Lemma 1), the ratio of possible $(d+1)$-polytopic transitions to impossible $(d+1)$-polytopic transitions will be low for many more rounds than possible for standard differentials. In fact, by increasing the dimension d of the polytopic transition, it can be assured that the ratio of possible to impossible polytopic transitions is close to zero for an almost arbitrary number of rounds.

An impossible $(d+1)$-polytopic attack could then proceed as follows. Let n be the block size of the cipher and let l be the number of bits in the last round key.

1. Choose a d and a d-difference such that the ratio of possible to impossible $(d+1)$-polytopic transitions is lower than 2^{-l-1}.
2. Get the r-round encryption of $d+1$ plaintexts chosen such that they adhere to the input d-difference.
3. For each guess of the round key k_r decrypt the last round. If the obtained d-difference after the $(r-1)$th round is possible, keep the key as a candidate. Otherwise discard it.

Clearly this should leave us on average with one round key candidate which is bound to be the correct one. In practice, an attack would likely be more complex, e.g., with only partially guessed round keys and tradeoffs in the filtering probability and the data/time complexities.

While the data complexity is limited to $d+1$ chosen plaintexts (and thus very low), the time complexity is harder to determine and depends on the difficulty

of determining whether an obtained $(r-1)$-round $(d+1)$-polytopic transition is possible or not. The straightforward approach is to precompute a list of possible d-differences after round $r-1$. Both the exponentially increasing memory requirements and the time of the precomputation limit this approach though. In spite of this, attacks using this approach are competitive with existing low-data attacks as we show in Section 4.

One possibility to reduce the memory complexity is to use a meet-in-the-middle approach where one searches for a collision in the possible d-differences reachable from the input d-difference and the calculated d-difference after round $(r-1)$ at a round somewhere in the middle of the cipher. This however requires to repeat the computation for every newly calculated d-difference and thus limits its use in the scenario where we calculated a new d-difference after round $(r-1)$ for each key guess (not in a distinguishing attack though).

Clearly any method that could efficiently determine the impossibility of most impossible polytopic transitions would prove extremely useful in an attack. Intuitively it might seem that it is generally a hard problem to determine the possibility of a polytopic transition. As a matter of fact though, there already exists a cryptographic technique that provides a very efficient distinguisher for certain types of polytopic transitions, namely higher-order differentials which are shown in Appendix B to correspond to truncated polytopic transitions. This raises the hope that better distinguishing techniques could still be discovered.

There is one important further effect of the increase in the size of the difference space: it allows us to restrict ourselves to impossible d-differences on only a part of the state. It is even possible to restrict the d-difference to a d-difference in one bit and still use it for efficient filtering.[1] In Section 4 we will use these techniques in impossible polytopic attacks to demonstrate the validity of the attacks and provide a usage scenario.

Wrong keys and random permutations

Note that while impossible polytopic attacks – just like impossible differential attacks – do not require the stochastic equivalence hypothesis, practical attacks require another hypothesis: the wrong-key randomization hypothesis. This hypothesis states that when decrypting one or several rounds with a wrong key guess creates a function that behaves like a random function. For our setting, we formulate it is as following:

Wrong-key randomization hypothesis. When decrypting one or multiple rounds of a block cipher with a wrong key guess, the resulting polytopic transition probability will be close to the transition probability over a random permutation for almost all key guesses.

Let us therefore take a look at the polytopic transition probabilities over random functions and random permutation. To simplify the treatment, we make the following definition:

[1] In standard differential cryptanalysis, this would require a probability 1 truncated differential.

Definition 8 (Degenerate d-difference). *Let $\boldsymbol{\alpha}$ be a d-difference over $\mathbb{F}_2^n$: $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_d)$. We call $\boldsymbol{\alpha}$* degenerate *if there exists an i with $1 \leq i \leq d$ with $\alpha_i = 0$ or if there exists a pair i, j with $1 \leq i < j \leq d$ and $\alpha_i = \alpha_j$. Otherwise we call $\boldsymbol{\alpha}$* non-degenerate.

Clearly if and only if a d-difference $\boldsymbol{\alpha}$ is degenerate, there exist two texts in the underlying $(d+1)$-polytope that are identical. To understand the transition probability of a degenerate d-difference it is thus sufficient to evaluate the transition probability of a non-degenerate d'-difference ($d' < d$) that contains the same set of texts. For the following two propositions, we will thus restrict ourselves to non-degenerate d-differences.

Proposition 2 (Distribution over random function). *Let $\boldsymbol{\alpha}$ be a non-degenerate d-difference over $\mathbb{F}_2^n$. Let $\mathbf{F}$ be a uniformly distributed random function from $\mathbb{F}_2^n$ to $\mathbb{F}_2^m$. The image of $\boldsymbol{\alpha}$ is then uniformly distributed over all d-difference over $\mathbb{F}_2^m$. In particular $\Pr\left(\boldsymbol{\alpha} \xrightarrow{F} \boldsymbol{\beta}\right) = 2^{-md}$ for any d-difference $\boldsymbol{\beta} \in (\mathbb{F}_2^m)^d$.*

Proof. Let $(m_0, m_1, \ldots, m_d)$ be a $(d+1)$-polytope that adheres to $\boldsymbol{\alpha}$. Then the polytope $(\mathbf{F}(m_0), \mathbf{F}(m_1), \ldots, \mathbf{F}(m_d))$ is clearly uniformly randomly distributed on $(\mathbb{F}_2^m)^{d+1}$ and accordingly $\boldsymbol{\beta}$ with $\boldsymbol{\alpha} \xrightarrow{\mathbf{F}} \boldsymbol{\beta}$ is distributed uniformly randomly on $(\mathbb{F}_2^m)^d$. □

For the image of a d-difference over a random permutation, we have a similar result:

Proposition 3 (Distribution over random permutation). *Let $\boldsymbol{\alpha}$ be a non-degenerate d-difference over $\mathbb{F}_2^n$. Let $\mathbf{F}$ be a uniformly distributed random permutation on $\mathbb{F}_2^n$. The image of $\boldsymbol{\alpha}$ is then uniformly distributed over all non-degenerate d-difference over $\mathbb{F}_2^n$.*

Proof. Let $(m_0, m_1, \ldots, m_d)$ be a $(d+1)$-polytope that adheres to $\boldsymbol{\alpha}$. As $\boldsymbol{\alpha}$ is non-degenerate, all m_i are distinct. Thus the polytope $(\mathbf{F}(m_0), \mathbf{F}(m_1), \ldots, \mathbf{F}(m_d))$ is clearly uniformly randomly distributed on all $(d+1)$-polytopes in $(\mathbb{F}_2^m)^{d+1}$ with distinct values. Accordingly $\boldsymbol{\beta}$ with $\boldsymbol{\alpha} \xrightarrow{F} \boldsymbol{\beta}$ is distributed uniformly randomly on all non-degenerate d-differences over $\mathbb{F}_2^n$. □

As long as $d \ll 2^n$, we can thus well approximate the probability $\Pr\left(\boldsymbol{\alpha} \xrightarrow{\mathbf{F}} \boldsymbol{\beta}\right)$ by 2^{-dn} when $\boldsymbol{\beta}$ is non-degenerate.

In the following, these proposition will be useful when we try to estimate the probability that a partial decryption with a wrong key guess will still give us a possible intermediate d-difference. We will then always assume that the wrong-key randomization hypothesis holds and that the probability of getting a particular d-difference on m bits is the same as if we had used a random permutation, i.e., it is 2^{-dm} (as our d is always small).

4 Impossible polytopic attacks on DES and AES

Without much doubt are the Data Encryption Standard (DES) [26] and the Advanced Encryption Standard (AES) [16] the most studied and best cryptanalyzed block ciphers. Any cryptanalytic improvement on these ciphers should thus be a good indicator of the novelty and quality of a new cryptanalytic attack. We believe that these ciphers thus pose ideal candidates to demonstrate that the generalization of differential cryptanalysis to polytopic cryptanalysis is not a mere intellectual exercise but useful for practical cryptanalysis.

In the following, we demonstrate several impossible polytopic attacks on reduced-round versions of DES and AES that make only use of a very small set of chosen plaintexts. The natural reference frame for these attacks are hence low-data attacks. In Table 1 and in Table 2 we compare our attacks to other low-data attacks on round-reduced versions of DES and AES respectively. We should mention here that [12] only states attacks on 7 and 8 rounds of DES. It is not clear whether the techniques therein could also be used to improve complexities of meet-in-the-middle attacks for 5- and 6-round versions of that cipher.

We stress here that in contrast to at least some of the other low-data attacks, our attacks make no assumption on the key schedule and work equally well with independent round keys. In fact, all of our attacks are straight-forward applications of the ideas developed in this paper. There is likely still room for improvement of these attacks using details of the ciphers and more finely controlled trade-offs.

In all of the following attacks, we determine the possibility or impossibility of a polytopic transition by deterministically generating a list of all d-differences that are reachable from the starting d-difference, i.e., we generate and keep a list of all possible d-differences. The determination of these lists is straightforward using the rules described in Section 2. The sizes of these lists are the limiting factors of the attacks both for the time and the memory complexities.

4.1 Attacks on the DES

For a good reference for the DES, we refer to [21]. A summary of the results for DES can be found in Table 1.

A 5-round attack. Let us start with an impossible 4-polytopic attack on 5-round DES. We split our input 3-difference into two parts, one for the left 32 state bits and one for the right 32 state bits. Let us denote the left 3-difference as (α, β, γ). For the right half we choose the 3-difference $(0, 0, 0)$. This allows us to pass the first round for free (as can be seen in Fig. 2).

The number of possible 3-differences after the second round depends now on our choice of α, β, and γ. To keep this number low, clearly it is good to choose the differences to activate as few S-boxes as possible. We experimentally tried out different natural choices and chose the values

$$(\alpha, \beta, \gamma) = (\texttt{02000000}, \texttt{04000000}, \texttt{06000000}).$$

Table 1. Comparison table of low-data attacks on round-reduced DES. Data complexity is measured in number of required known plaintexts (KP) or chosen plaintexts (CP). Time complexity is measured in round-reduced DES encryption equivalents. Memory complexity is measured in plaintexts (8 bytes). For the other attacks no memory requirements were explicitly specified in the publications. They should be low though. The attacks of this paper are in bold.

Rounds	Attack Type	Time	Data	Memory	Source
5	Differential	$> 2^{11.7}$	64 CP	-	As in [18]
	Linear	$> 2^{13.8}$	72 CP	-	As in [18]
	MitM	$2^{45.5}$	1 KP	-	From [14]
	MitM	$2^{37.9}$	28 KP	-	From [18]
	MitM	2^{30}	8 CP	-	From [18]
	Imp. polytopic	$\mathbf{2^{13.2}}$	**4 CP**	$\mathbf{2^9}$	**This paper**
6	Differential	$2^{13.7}$	256 CP	-	As in [18]
	Linear	$2^{13.9}$	>104 KP	-	As in [18]
	MitM	$2^{51.8}$	1 KP	-	From [18]
	Imp. polytopic	$\mathbf{2^{32.2}}$	**4 CP**	$\mathbf{2^{10}}$	**This paper**
	Imp. polytopic	$\mathbf{2^{18.4}}$	**48 CP**	$\mathbf{2^9}$	**This paper**
7	MitM Sieve	2^{53}	1 KP	-	From [12]
	Imp. polytopic	$\mathbf{2^{43}}$	**16 CP**	$\mathbf{2^{43}}$	**This paper**
	Imp. polytopic	$\mathbf{2^{37.8}}$	**48 CP**	$\mathbf{2^{10}}$	**This paper**
8	MitM Sieve	2^{53}	16 KP	-	From [12]

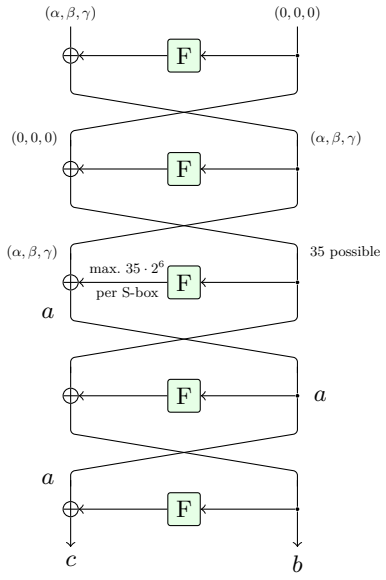

Fig. 2. Outline of the 5-round attack on DES.

All of these three differences only activate S-box 2 in round 2. With this choice we get 35 possible 3-differences after round 2. Note that the left 3-difference is still (α, β, γ) after round 2 while the 35 variations only appear in the right half.

As discussed earlier, the maximal number of output d-differences for a fixed input d-difference is inherently limited by the size of the domain of the function. A consequence of this is that for any of the 35 3-differences after round 2 the possible number of output 3-differences of any S-box in round 3 is limited to 2^6 as shown in Fig. 2. But by guessing the 6 bits of round key 5 that go into the corresponding S-box in round 5, we can determine the 3-difference in the same four output bits of round 3 now coming from the ciphertexts. For the right guess of the 6 key bits, the determined 3-difference will be possible. For a wrong key guess though, we expect the 3-difference to take a random value in the set of all 3-differences on 4 bits.

But the size of the space of 3-differences in these four output bits is now $2^{4 \cdot 3} = 2^{12}$. Thus when fixing one of the 35 possible 3-differences after round 2, we expect on average to get one suggestion for the 6 key bits in that S-box. Repeating this for every S-box, we get on average one suggestion for the last round key for each of the 35 possible 3-differences after round 2, leaving us with an average of 35 key candidates for the last round key.

What are the complexities of the attack? Clearly we only need 4 chosen plaintexts. For the time complexity we get the following: For each of the 35 possible 3-differences after round 2, we have to determine the 2^6 possible output 3-differences and for each of these, we have to see in the list of possible 3-differences obtained from the key guesses whether there is a guess of the 6 key bits that gives us exactly that 3-difference. Thus we have a total of $35 \cdot 8 \cdot 2^6 = 2^{14.2}$ steps each of which should be easier than one round of DES encryption. This leaves us with a time complexity of $\approx 2^{12}$ 5-round DES encryptions equivalents. But to completely determine the DES key we need 8 additional bits that are not present in the last round key. As we expect on average maximally 35 round keys, we are left with trying out the $35 \cdot 2^8 = 2^{13.2}$ full key candidates, setting the time complexity of this attack to that value.

The only memory requirement in this attack is storing the list of possible 3-differences for each key guess in each S-box. This should roughly be no more than 2^{12} bytes.

A 6-round attack. The 6-round attack proceeds exactly as the 5-round attack, with the only difference being that instead of determining the possible 3-difference output of each S-box in round 3, we do the same in round 4 and thus have to repeat the attack for every possible 3-difference after round 3.

Experimental testing revealed that it is beneficial for this attack to choose a different choice of α, β, and γ, namely

$$(\alpha, \beta, \gamma) = (\mathtt{20000000}, \mathtt{40000000}, \mathtt{60000000}),$$

which now activates S-box 1 instead of S-box 2 as it gives us the lowest number of 3-differences after round 3. For this choice, we get a number of 48 possible

3-differences after round 2 and $2^{24.12}$ possible 3-differences after round 3. Now substituting 35 with this number in the previous attack, gives us the time complexity for this 6-round attack.

A note regarding the memory requirement of this attack: As we loop over the $2^{24.12}$ possible 3-differences after round 3, we are not required to store all of them at any time. By doing the attack while creating these possible 3-differences we can keep the memory complexity nearly as low as before, namely to roughly 2^{13} bytes.

A 7-round attack. Unfortunately extending from 6 to 7 rounds as done when going from 5 to 6 rounds is not possible, due to the prohibitively large number of possible 3-differences after round 4. Instead we use a different angle.

It is well known that when attacking r-round DES, guessing the appropriate 36 round key bits of the last round key and the appropriate 6 bits of the round key in round $r-1$ allows us to determine the output state bits of an S-box of our choice after round $r-3$. We will thus restrict ourselves to looking for impossible d-differences in only one S-box. We choose S-box 1 here.

In order to have a sufficiently high success rate, we need to increase the dimension of our polytopic transitions to increase the size of the d-difference space of the four output bits of the S-box of our choice. For this attack we choose $d = 15$ giving us a 15-difference space size of 2^{60} in four bits.

For our choice of input 15-difference, we again leave all differences in the right side to 0, while choosing for the 15-difference on the left side:

$$\begin{aligned}(&\texttt{00000002}, \texttt{00000004}, \texttt{00000006}, \texttt{02000000}, \texttt{02000002}, \texttt{02000004},\\ &\texttt{02000006}, \texttt{04000000}, \texttt{04000002}, \texttt{04000004}, \texttt{04000006}, \texttt{06000000},\\ &\texttt{06000002}, \texttt{06000004}, \texttt{06000006})\end{aligned}$$

which only activates S-boxes 2 and 8. For this choice of input 15-difference we get 1470 possible 15-differences after round 2 and $2^{36.43}$ possible 15-differences after round 3.

For each of these $2^{36.43}$ possible 15-differences after round 3, we calculate the 2^6 possible output 15-differences of S-box 1 after round 4. Now having precomputed a list of possible 15-differences in the output bits of S-box 1 after round 4 for each of the 2^{42} guessed key bits of round 7 and 6, we can easily test whether we get a collision. What is the probability of this? The 15-difference space size in the four bits is 2^{60} and, we get maximally 2^{42} possible 15-differences from the key guesses. This leaves us with a chance of 2^{-18} that we find a 15-difference in that list. Thus for each of the $2^{36.43}$ possible 15-differences after round 3, we expect on average at most 2^{-12} suggestions for the guessed 42 key bits, a total of $2^{24.43}$ suggestions.

What are the complexities for this attack? Clearly again, the data complexity is 16 chosen plaintexts. For the time complexity, for each of the $2^{42.42}$ possible 4-bit 15-differences obtained after round 4, we have to see whether it is contained in the list of 2^{42} 3-differences which we obtained from the key guesses. To do

this efficiently, we first have to sort the list which should take $2^{42} \cdot 42 = 2^{47.4}$ elementary steps. Assuming that a 7-round DES encryption takes at least 42 elementary steps, we can upperbound this complexity with 2^{42} DES encryption equivalents. As finding an entry in a list of 2^{42} entries also takes approximately 42 elementary steps, this leaves us with a total time complexity of at most 2^{43} 7-round DES encryption equivalents. As each suggestion gives us 42 DES key bits and as the list of suggestions has a size of $2^{24.23}$, we can find the correct full key with $2^{38.23}$ 7-round DES trial encryptions which is lower than then the previously mentioned time complexity and can thus be disregarded.

The data complexity is determined by the size of the list of 4-bit 15-differences generated from the key guesses. This gives us a memory requirement of $2^{42}(15 \cdot 4 + 42)$ bits $\approx 2^{46}$ bytes.

Extension of the attacks using more plaintexts. The attacks for 5 and 6 rounds can be extended by one round at the cost of a higher data complexity. The extension can be made at the beginning of the cipher in the following way.

Let us suppose we start with a 3-difference $(\delta_1, \delta_2, \delta_3)$ in the left half and the 3-difference (α, β, γ) in the right half. If we knew the output 3-difference of the round function in the first round, we could choose $(\delta_1, \delta_2, \delta_3)$ accordingly to make sure that we end up at the starting position of the original attack. Thus by guessing this value and repeating the attack for each guessed value of this 3-difference we can make sure we still retrieve the key.

Fortunately the values of (α, β, γ) are already chosen to give a minimal number of possible 3-difference in the round function. Thus the time complexity only increases by this value, i.e., 35 and 48. The data complexity increases even less. As it turns out, 12 different values for the left half of the input text are enough to generate all of the 35 resp. 48 3-differences. Thus the data complexity only increases to 48 chosen plaintexts.

We should mention that the same technique can be used to extend the 7-round attack to an 8-round attack. But this leaves us with the same time complexity as the 8-round attack in [12], albeit at a much higher data cost.

Experimental results. To verify the correctness of the above attacks and their complexities, we implemented the 5-round and 6-round attacks that use 4 chosen plaintexts. We ran the attacks on a single core of an Intel Core i5-4300U processor. We ran the 5-round attack 100000 times which took about 140 seconds. The average number of suggested round keys was 47 which is slightly higher than the expected number of 35. The suggested number of round keys was below 35 though in 84 percent of the cases and below 100 in 95 percent of the cases. In fact, the raised average is created by a few outliers in the distribution: taking the average on all but the 0.02 percent worst cases, we get 33.1 round key suggestions per case. While this shows that the estimated probability is generally good, it also demonstrates that the wrong-key randomization hypothesis has to be handled with care.

Running the six-round attack 10 times, an attack ran an average time of 10 min and produced an average of $2^{22.3}$ candidates for the last round key. As expected, the correct round key was always in the list of candidate round keys for both the 5-round and 6-round attacks.

4.2 Attacks on the AES

For a good reference for the AES, we refer to [16]. A summary of the results for AES can be found in Table 2.

Table 2. Comparison table of low-data attacks on round-reduced AES. Data complexity is measured in number of required chosen plaintexts (CP). Time complexity is measured in round-reduced AES encryption equivalents. Memory complexity is measured in plaintexts (16 bytes). The column 'keyschedule' denotes whether the attacks use the AES key schedule. All attacks that rely on the keyschedule are attacks on AES-128. The attacks of this paper are in bold.

Rounds	Attack Type	Time	Data	Memory	Keyschedule	Source
4	Guess & Det.	2^{120}	1 KP	2^{120}	Yes	As in [10]
	Diff. MitM	2^{104}	3 CP	1	Yes	As in [8,9]
	Guess & Det.	2^{80}	2 CP	2^{80}	Yes	As in [10]
	Guess & Det.	2^{32}	4 CP	2^{24}	Yes	As in [10]
	Imp. polytopic	$\mathbf{2^{38}}$	**8 CP**	$\mathbf{2^{15}}$	**No**	**This paper**
5	MitM	2^{64}	8 CP	2^{56}	Yes	As in [17], Sec. 7.5.1
	Imp. polytopic	$\mathbf{2^{70}}$	**15 CP**	$\mathbf{2^{41}}$	**No**	**This paper**

A 4-round attack. We first present here an impossible 8-polytopic attack on 4-round AES. For the input 7-difference, we choose a 7-difference that activates only the first byte, i.e., that is all-zero in all other bytes. Such a 7-difference can be mapped after round 1 to at most 2^8 different 7-differences. If we restrict ourselves to the 7-differences in the first column after round 2, we can then at most have 2^{16} different 7-differences in this column. In particular, we can have at most have 2^{16} different 7-differences in the first byte. For a depiction of this, see Fig. 3.

If we now request the encryptions of 8 plaintexts that adhere to our chosen start 7-difference, we can now determine the corresponding 7-difference after round 2 in the first byte by guessing 40 round key bits of round keys 3 and 4. If this 7-difference does not belong to the set of 2^{16} possible ones, we can discard the key guess as wrong.

How many guesses of the 40 key bits, do we expect to survive the filtering? There are 2^{56} possible 7-difference on a byte and only 2^{16} possible ones coming from our chosen input 7-difference. This leaves a chance of 2^{-40} for a wrong key guess to produce a correct 7-difference. We thus expect on average 2 suggestions

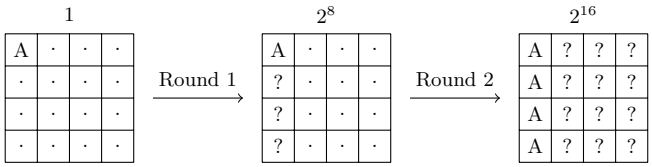

Fig. 3. Diffusion of the starting 7-difference for the 4-round attack on AES. The letter A shows a byte position in which a possible 7-difference is non-zero and known. A dot indicates a byte position where the 7-difference is known to be zero. A question mark indicates a byte position where arbitrary values for the 7-differences are allowed. In total there are 2^{16} different 7-differences possible in the first column after the second round.

for the 40 key bits, among them the right one. To determine the remaining round key bits, we can use the same texts, only restricting ourselves to different columns.

The data complexity of the attack is limited to 8 chosen plaintexts. The time complexity is dominated by determining the 7-difference in the first byte after round 2 for each guess of the 40 key bits and checking whether it is among the 2^{16} possible ones. This can be done in less than 16 table lookups on average for each key guess. Thus the time complexity corresponds to $2^{40} \cdot 2^{-2} = 2^{38}$ 4-round AES encryption equivalents, assuming one 4-round encryption corresponds to $4 \cdot 16$ table lookups. The memory complexity is limited to a table of the 2^{16} allowed 7-difference in one byte, corresponding to 2^{19} bytes or 2^{15} plaintext equivalents.

A 5-round attack. In this attack, we are working with 15-polytopes and trace the possible 14-differences one round further than in the 4-round attack. Again we choose our starting 14-difference such that it only activates the first byte. After two rounds we then have maximally 2^{40} different 14-differences on the whole state. If we restrict ourselves to only the first column of the state after round 3, we then get an additional 2^{32} possible 14-differences in this column for each of the 2^{40} possible 14-differences after round 2, resulting in a total of 2^{72} possible 14-differences in the first column after round 3. This is depicted in Fig. 4. In particular again, we can have at most have 2^{72} different 14-differences in the first byte.

Let us suppose now we have the encrypted values of a 15-polytope that adheres to our starting 14-difference. We can then again find the respective 14-difference in the first byte after the third round by guessing 40 key bits in round keys 4 and 5. There are in total 2^{112} different 14-differences in one byte. The chance of a wrong key guess to produce one of the possible 2^{72} 14-differences is thus 2^{-40}. We thus expect on average 2 suggestions for the 40 key bits, among them the right one. To determine the remaining round key bits, we can again use the same texts but restricting ourselves to a different column.

To lower the memory complexity of this attack it is advantageous to not store the 2^{72} possible 14-differences but store for each of the 2^{40} key guesses

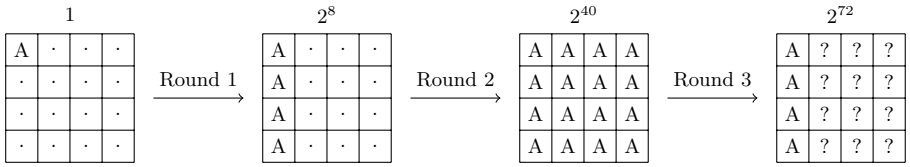

Fig. 4. Diffusion of the starting 14-difference for the 5-round attack on AES. The letter A shows a byte position in which a possible 14-difference is non-zero and known. A dot indicates a byte position where the 14-difference is known to be zero. A question mark indicates a byte position where arbitrary values for the 14-differences are allowed. In total there are 2^{72} different 14-differences possible in the first column after the third round.

the obtained 14-difference. This gives a memory complexity of $2^{40} \cdot (14 + 5)$ bytes corresponding to 2^{41} plaintext equivalents. The time complexity is then dominated by constructing the 2^{72} possible 14-differences and testing whether they correspond to one of the key guesses. This should not take more than the equivalent of $2^{72} \cdot 16$ table lookups resulting in a time complexity of 2^{70} 5-round AES encryption equivalents. The data complexity is restricted to the 15 chosen plaintexts needed to construct one 15-polytope corresponding to the starting 14-difference.

5 Conclusion

In this paper, we developed and studied polytopic cryptanalysis. We were able to show that the methodology and notation of standard cryptanalysis can be unambiguously extended to polytopic cryptanalysis, including the concept of impossible differentials. Standard differential cryptanalysis remains as a special case of polytopic cryptanalysis.

For impossible polytopic transitions, we demonstrated that both the increase in the size of the space of d-differences and the inherent limit in the diffusion of d-differences in a cipher allow them to be very effective in settings where ordinary impossible differentials fail. This is the case when the number of rounds is so high that impossible differentials do no longer exist or when the allowed data complexity is too low.

Finally we showed the practical relevance of this framework by demonstrating novel low-data attacks on DES and AES that are able to compete with existing attacks.

Acknowledgements

The author thanks Christian Rechberger, Stefan Kölbl, and Martin M. Lauridsen for fruitful discussions. The author also thanks Dmitry Khovratovich and the anonymous reviewers for comments that helped to considerably improve the quality of the paper.

References

1. Biham, E., Biryukov, A., Shamir, A.: Cryptanalysis of Skipjack reduced to 31 rounds using impossible differentials. J. Cryptology 18(4), 291–311 (2005)
2. Biham, E., Shamir, A.: Differential cryptanalysis of DES-like cryptosystems. J. Cryptology 4(1), 3–72 (1991)
3. Biryukov, A., Shamir, A.: Structural cryptanalysis of SASAS. J. Cryptology 23(4), 505–518 (2010)
4. Blondeau, C., Gérard, B.: Multiple differential cryptanalysis: Theory and practice. In: Joux, A. (ed.) Fast Software Encryption, FSE 2011. Lecture Notes in Computer Science, vol. 6733, pp. 35–54. Springer (2011)
5. Blondeau, C., Leander, G., Nyberg, K.: Differential-linear cryptanalysis revisited. In: Cid, C., Rechberger, C. (eds.) Fast Software Encryption, FSE 2014. Lecture Notes in Computer Science, vol. 8540, pp. 411–430. Springer (2015)
6. Blondeau, C., Nyberg, K.: Links between truncated differential and multidimensional linear properties of block ciphers and underlying attack complexities. In: Nguyen, P.Q., Oswald, E. (eds.) Advances in Cryptology - EUROCRYPT 2014. Lecture Notes in Computer Science, vol. 8441, pp. 165–182. Springer (2014)
7. Bogdanov, A., Knudsen, L.R., Leander, G., Paar, C., Poschmann, A., Robshaw, M.J.B., Seurin, Y., Vikkelsoe, C.: PRESENT: An ultra-lightweight block cipher. In: Paillier, P., Verbauwhede, I. (eds.) Cryptographic Hardware and Embedded Systems - CHES 2007. Lecture Notes in Computer Science, vol. 4727, pp. 450–466. Springer (2007)
8. Bouillaguet, C., Derbez, P., Dunkelman, O., Fouque, P., Keller, N., Rijmen, V.: Low-data complexity attacks on AES. IEEE Transactions on Information Theory 58(11), 7002–7017 (2012)
9. Bouillaguet, C., Derbez, P., Dunkelman, O., Keller, N., Rijmen, V., Fouque, P.: Low data complexity attacks on AES. Cryptology ePrint Archive, Report 2010/633 (2010), `http://eprint.iacr.org/`
10. Bouillaguet, C., Derbez, P., Fouque, P.: Automatic search of attacks on round-reduced AES and applications. In: Rogaway, P. (ed.) Advances in Cryptology - CRYPTO 2011. Lecture Notes in Computer Science, vol. 6841, pp. 169–187. Springer (2011)
11. Cannière, C.D., Dunkelman, O., Knezevic, M.: KATAN and KTANTAN - A family of small and efficient hardware-oriented block ciphers. In: Clavier, C., Gaj, K. (eds.) Cryptographic Hardware and Embedded Systems - CHES 2009,. Lecture Notes in Computer Science, vol. 5747, pp. 272–288. Springer (2009)
12. Canteaut, A., Naya-Plasencia, M., Vayssière, B.: Sieve-in-the-middle: Improved MITM attacks. In: Canetti, R., Garay, J.A. (eds.) Advances in Cryptology - CRYPTO 2013. Lecture Notes in Computer Science, vol. 8042, pp. 222–240. Springer (2013)
13. Chabaud, F., Vaudenay, S.: Links between differential and linear cryptanalysis. In: Santis, A.D. (ed.) Advances in Cryptology - EUROCRYPT '94. Lecture Notes in Computer Science, vol. 950, pp. 356–365. Springer (1995)
14. Chaum, D., Evertse, J.: Crytanalysis of DES with a reduced number of rounds: Sequences of linear factors in block ciphers. In: Williams, H.C. (ed.) Advances in Cryptology - CRYPTO '85. Lecture Notes in Computer Science, vol. 218, pp. 192–211. Springer (1986)
15. Daemen, J., Knudsen, L.R., Rijmen, V.: The block cipher square. In: Biham, E. (ed.) Fast Software Encryption, FSE '97. Lecture Notes in Computer Science, vol. 1267, pp. 149–165. Springer (1997)

16. Daemen, J., Rijmen, V.: The Design of Rijndael: AES - The Advanced Encryption Standard. Information Security and Cryptography, Springer (2002)
17. Derbez, P.: Meet-in-the-Middle Attacks on AES. Ph.D. thesis, Ecole Normale Supérieure de Paris - ENS Paris (Dec 2013), https://tel.archives-ouvertes.fr/tel-00918146
18. Dunkelman, O., Sekar, G., Preneel, B.: Improved meet-in-the-middle attacks on reduced-round DES. In: Srinathan, K., Rangan, C.P., Yung, M. (eds.) Progress in Cryptology - INDOCRYPT 2007. Lecture Notes in Computer Science, vol. 4859, pp. 86–100. Springer (2007)
19. Knudsen, L.R.: Truncated and higher order differentials. In: Preneel, B. (ed.) Fast Software Encryption, FSE '94. Lecture Notes in Computer Science, vol. 1008, pp. 196–211. Springer (1995)
20. Knudsen, L.R.: DEAL - a 128-bit block cipher. Technical Report 151, Department of Informatics, University of Bergen, Norway (February 1998), submitted as an AES candidate by Richard Outerbridge
21. Knudsen, L.R., Robshaw, M.: The Block Cipher Companion. Information Security and Cryptography, Springer (2011)
22. Lai, X.: Higher order derivatives and differential cryptanalysis. In: Blahut, R.E., Daniel J. Costello, J., Maurer, U., Mittelholzer, T. (eds.) Communications and Cryptography, Two Sides of One Tapestry, pp. 227–233. Kluwer Academic Publishers (1994)
23. Lai, X., Massey, J.L.: Markov ciphers and differential cryptanalysis. In: Davies, D.W. (ed.) Advances in Cryptology - EUROCRYPT '91. Lecture Notes in Computer Science, vol. 547, pp. 17–38. Springer (1991)
24. Langford, S.K., Hellman, M.E.: Differential-linear cryptanalysis. In: Desmedt, Y. (ed.) Advances in Cryptology - CRYPTO '94. Lecture Notes in Computer Science, vol. 839, pp. 17–25. Springer (1994)
25. Murphy, S.: The return of the cryptographic boomerang. IEEE Transactions on Information Theory 57(4), 2517–2521 (2011)
26. National Institute of Standards and Technology: Data Encryption Standard. Federal Information Processing Standard (FIPS), Publication 46, U.S. Department of Commerce, Washington D.C. (January 1977)
27. Shirai, T., Shibutani, K., Akishita, T., Moriai, S., Iwata, T.: The 128-bit blockcipher CLEFIA (extended abstract). In: Biryukov, A. (ed.) Fast Software Encryption, FSE 2007. Lecture Notes in Computer Science, vol. 4593, pp. 181–195. Springer (2007)
28. Vaudenay, S.: Decorrelation: A theory for block cipher security. J. Cryptology 16(4), 249–286 (2003)
29. Wagner, D.: The boomerang attack. In: Knudsen, L.R. (ed.) Fast Software Encryption, FSE '99. Lecture Notes in Computer Science, vol. 1636, pp. 156–170. Springer (1999)
30. Wagner, D.: Towards a unifying view of block cipher cryptanalysis. In: Roy, B.K., Meier, W. (eds.) Fast Software Encryption, FSE 2004. Lecture Notes in Computer Science, vol. 3017, pp. 16–33. Springer (2004)

A Markov model in polytopic cryptanalysis

To develop the Markov model, we first need to introduce keys in the function over which the transitions take place. We will thus restrict our discussion to product

ciphers i.e., block ciphers that are constructed through repeated composition of round functions. In contrast to Eq. (2), each round function f^i is now keyed with its own round key k_i which itself is derived from the key k of the cipher via a key schedule.[2] We can then write the block cipher f_k as:

$$f_k := f^r_{k_r} \circ \cdots \circ f^2_{k_2} \circ f^1_{k_1}. \tag{7}$$

The first assumption that we now need to make, is that the round keys are independent. The second assumption is that the product cipher is a Markov cipher. Here we adopt the notion of a Markov cipher from [23] to polytopic cryptanalysis:

Definition 9. *A product cipher is a $(d+1)$-polytopic Markov cipher if and only if for all round functions f^i, for any $(d+1)$-polytopic transition $\boldsymbol{\alpha} \longrightarrow \boldsymbol{\beta}$ for that round function and any fixed inputs $x, y \in \mathbb{F}_2^n$, we have*

$$\Pr_{\mathbf{K}}\left(\boldsymbol{\alpha} \xrightarrow[x]{f^i_{\mathbf{K}}} \boldsymbol{\beta}\right) = \Pr_{\mathbf{K}}\left(\boldsymbol{\alpha} \xrightarrow[y]{f^i_{\mathbf{K}}} \boldsymbol{\beta}\right) \tag{8}$$

where $\mathbf{K}$ is a random variable distributed uniformly over the spaces of round keys.

In words, a cipher is a $(d+1)$-polytopic Markov cipher if and only if the probabilities of 1-round $(d+1)$-polytopic transitions do not depend on the specific anchor as long as the round key is distributed uniformly at random. For $d = 1$, the definition coincides with the classical definition.

Just as with the standard definition of Markov ciphers, most block ciphers are $(d+1)$-polytopic Markov ciphers for any d as the round keys are usually added to any part of the state that enters the non-linear part of the round function (for a counterexample, see [11]). Examples of $(d+1)$-polytopic Markov ciphers are SPN ciphers such as AES [16] or PRESENT [7], and Feistel ciphers such as DES [26] or CLEFIA [27]. We are not aware of any cipher that is Markov in the classical definition but not $(d+1)$-polytopic Markov.

In the following, we extend the central theorem from [23] (Theorem 2) to the case of $(d+1)$-polytopes.

Theorem 1. *Let $f_k = f^r_{k_r} \circ \cdots \circ f^1_{k_1}$ be a $(d+1)$-polytopic Markov cipher with independent round keys, chosen uniformly at random and let $\boldsymbol{\delta}_0, \boldsymbol{\delta}_1, \ldots, \boldsymbol{\delta}_r$ be a series of d-differences such that $\boldsymbol{\delta}_0$ is the input d-difference of round 1 and $\boldsymbol{\delta}_i$ is the output d-difference of round i of some fixed input $(d+1)$-polytope. The series $\boldsymbol{\delta}_0, \boldsymbol{\delta}_1, \ldots, \boldsymbol{\delta}_r$ then forms a Markov chain.*

The following proof follows the lines of the original proof from [23].

Proof. We limit ourselves here to showing that

$$\Pr_{\mathbf{K}_1,\mathbf{K}_2}\left(\boldsymbol{\delta}_1 \xrightarrow[f^1_{\mathbf{K}_1}(x)]{f^2_{\mathbf{K}_2}} \boldsymbol{\delta}_2 \;\middle|\; \boldsymbol{\delta}_0 \xrightarrow[x]{f^1_{\mathbf{K}_1}} \boldsymbol{\delta}_1\right) = \Pr_{\mathbf{K}_2}\left(\boldsymbol{\delta}_1 \xrightarrow[z]{f^2_{\mathbf{K}_2}} \boldsymbol{\delta}_2\right) \tag{9}$$

[2] For a clearer notation, we moved the index from subscript to superscript.

where x and z are any elements from $\mathbb{F}_2^n$ and $\mathbf{K}_1$ and $\mathbf{K}_2$ are distributed uniformly at random over their respective round key spaces and the conditioned event has positive probability. The theorem then follows easily by induction and application of the same arguments to the other rounds.

For any $x, z \in \mathbb{F}_2^n$, we now have

$$\begin{aligned}
&\Pr_{\mathbf{K}_1,\mathbf{K}_2}\left(\boldsymbol{\delta}_1 \xrightarrow[f^1_{\mathbf{K}_1}(x)]{f^2_{\mathbf{K}_2}} \boldsymbol{\delta}_2 \text{ and } \boldsymbol{\delta}_0 \xrightarrow[x]{f^1_{\mathbf{K}_1}} \boldsymbol{\delta}_1\right)\\
&= \sum_{y\in\mathbb{F}_2^n} \Pr_{\mathbf{K}_1,\mathbf{K}_2}\left(\boldsymbol{\delta}_1 \xrightarrow[y]{f^2_{\mathbf{K}_2}} \boldsymbol{\delta}_2 \text{ and } \boldsymbol{\delta}_0 \xrightarrow[x]{f^1_{\mathbf{K}_1}} \boldsymbol{\delta}_1 \text{ and } f^1_{\mathbf{K}_1}(x) = y\right)\\
&= \sum_{y\in\mathbb{F}_2^n} \Pr_{\mathbf{K}_2}\left(\boldsymbol{\delta}_1 \xrightarrow[y]{f^2_{\mathbf{K}_2}} \boldsymbol{\delta}_2\right) \cdot \Pr_{\mathbf{K}_1}\left(\boldsymbol{\delta}_0 \xrightarrow[x]{f^1_{\mathbf{K}_1}} \boldsymbol{\delta}_1 \text{ and } f^1_{\mathbf{K}_1}(x) = y\right)\\
&= \Pr_{\mathbf{K}_2}\left(\boldsymbol{\delta}_1 \xrightarrow[z]{f^2_{\mathbf{K}_2}} \boldsymbol{\delta}_2\right) \cdot \sum_{y\in\mathbb{F}_2^n} \Pr_{\mathbf{K}_1}\left(\boldsymbol{\delta}_0 \xrightarrow[x]{f^1_{\mathbf{K}_1}} \boldsymbol{\delta}_1 \text{ and } f^1_{\mathbf{K}_1}(x) = y\right)\\
&= \Pr_{\mathbf{K}_2}\left(\boldsymbol{\delta}_1 \xrightarrow[z]{f^2_{\mathbf{K}_2}} \boldsymbol{\delta}_2\right) \cdot \Pr_{\mathbf{K}_1}\left(\boldsymbol{\delta}_0 \xrightarrow[x]{f^1_{\mathbf{K}_1}} \boldsymbol{\delta}_1\right)
\end{aligned}$$

where the second equality comes from the independence of keys K_1 and K_2 and the third equality comes from the Markov property of the cipher. From this, Eq. (9) follows directly. □

The important consequence of the fact that the sequence of d-differences forms a Markov chain is that, just as in standard differential cryptanalysis, the average probability of a particular polytopic trail with respect to independent random round keys is the product of the single polytopic 1-round transitions of which it consists. We then have the following result:

Corollary 1. *Let f_k, $f^i_{k_i}, 1 \le i \le r$ be as before. Let $\boldsymbol{\alpha}_0 \xrightarrow{f_1} \boldsymbol{\alpha}_1 \xrightarrow{f_2} \cdots \xrightarrow{f_r} \boldsymbol{\alpha}_r$ be an r-round $(d+1)$-polytopic trail. Then*

$$\Pr\left(\boldsymbol{\alpha}_0 \xrightarrow{f^1_{\mathbf{K}_1}} \boldsymbol{\alpha}_1 \xrightarrow{f^2_{\mathbf{K}_2}} \cdots \xrightarrow{f^r_{\mathbf{K}_r}} \boldsymbol{\alpha}_r\right) = \prod_{i=1}^{r} \Pr\left(\boldsymbol{\alpha}_{i-1} \xrightarrow{f^i_{\mathbf{K}_i}} \boldsymbol{\alpha}_i\right) \tag{10}$$

where $x \in \mathbb{F}_2^n$ and the $\mathbf{K}_i$ are uniformly randomly distributed on their respective spaces.

Proof. This is a direct consequence of the fact that d-differences form a Markov chain. □

In most attacks though, we are attacking one fixed key and can not average the attack over all keys. Thus the following assumption is necessary:

Hypothesis of stochastic equivalence. *Let f be as above. The hypothesis of stochastic equivalence then refers to the assumption that the probability of*

any polytopic trail $\boldsymbol{\alpha}_0 \xrightarrow{f_1} \boldsymbol{\alpha}_1 \xrightarrow{f_2} \cdots \xrightarrow{f_r} \boldsymbol{\alpha}_r$ *is roughly the same for the large majority of keys:*

$$\Pr\left(\boldsymbol{\alpha}_0 \xrightarrow{f^1_{\mathbf{K}_1}} \boldsymbol{\alpha}_1 \xrightarrow{f^2_{\mathbf{K}_2}} \cdots \xrightarrow{f^r_{\mathbf{K}_r}} \boldsymbol{\alpha}_r\right) \approx \Pr\left(\boldsymbol{\alpha}_0 \xrightarrow{f^1_{k_1}} \boldsymbol{\alpha}_1 \xrightarrow{f^2_{k_2}} \cdots \xrightarrow{f^r_{k_r}} \boldsymbol{\alpha}_r\right) \quad (11)$$

for almost all tuples of round keys $(k_1, k_2, \ldots, k_r)$.

B Truncated polytopic transitions and higher-order differentials

In this section, we extend the definition of truncated differentials to polytopic transitions and prove that higher-order differentials are a special case of these. We then gauge the cryptographic ramifications of this.

In accordance with usual definitions for standard truncated differentials (see for example [6], we define:

Definition 10. *A truncated* d*-difference is an affine subspace of the space of* d*-differences. A truncated* $(d+1)$*-polytopic transition is a pair* (A, B) *of truncated* d*-differences, mostly denoted as* $A \xrightarrow{f} B$*. The probability of a truncated* $(d+1)$*-polytopic transition* (A, B) *is defined as the probability that an input* d*-difference chosen uniformly randomly from* A *maps to a* d*-difference in* B*:*

$$\Pr\left(A \xrightarrow{f} B\right) := |A|^{-1} \sum_{\substack{\boldsymbol{\alpha} \in A \\ \boldsymbol{\beta} \in B}} \Pr\left(\boldsymbol{\alpha} \xrightarrow{f} \boldsymbol{\beta}\right) \quad (12)$$

As the truncated input d-difference is usually just a single d-difference, the probability of a truncated differential is then just the probability that this input d-difference maps to any of the output d-differences in the output truncated d-difference. With a slight abuse of notation, we will denote the truncated polytopic transition then also as $\boldsymbol{\alpha} \xrightarrow{f} B$ where $\boldsymbol{\alpha}$ is the single d-difference of the input truncated d-difference.

A particular case of a truncated d-difference is the case where the individual differences of the d-differences always add up to the same value. This is in fact just the kind of d-differences one is interested in when working with higher-order derivatives. We refer here to Lai's original paper on higher-order derivatives [22] and Knudsen's paper on higher-order differentials [19] for reference and notation.

Theorem 2. *A differential of order* t *is a special case of a truncated* 2^t*-polytopic transition. In particular, its probability is the sum of the probabilities of all* 2^t*-polytopic trails that adhere to the truncated* 2^t*-polytopic transition.*

Proof. Let $f : \mathbb{F}_2^n \to \mathbb{F}_2^n$. Let $(\alpha_1, \ldots, \alpha_t)$ be the set of linearly independent differences that are used as the base for our derivative. Let $L(\alpha_1, \ldots, \alpha_t)$ denote the linear space spanned by these differences. Let furthermore β be the output

difference we are interested in. The probability of the t-th order differential $\Delta_{\alpha_1,\ldots,\alpha_t} f(\mathbf{X}) = \beta$ is then defined as the probability that

$$\sum_{\gamma \in L(\alpha_1,\ldots,\alpha_t)} f(\mathbf{X} \oplus \gamma) = \beta \tag{13}$$

holds with $\mathbf{X}$ being a random variable, uniformly distributed on $\mathbb{F}_2^n$.

Let B now be the truncated $(2^t - 1)$-difference defined as

$$B := \left\{ (\delta_1, \ldots, \delta_{2^t-1}) \;\middle|\; \sum_{i=1}^{2^t-1} \delta_i = \beta \right\}.$$

Let $\gamma_1, \gamma_2, \ldots, \gamma_{2^t-1}$ be an arbitrary ordering of the non-zero elements of the linear space $L(\alpha_1, \ldots, \alpha_t)$ and let $\boldsymbol{\alpha} = (\gamma_1, \ldots, \gamma_{2^t-1})$ be the $(2^t - 1)$-difference consisting of these. We will then show that the probability of the t-th order differential $(\alpha_1, \ldots, \alpha_t, \beta)$ is equal to the the probability of the truncated 2^t-polytopic transition $\boldsymbol{\alpha} \xrightarrow{f} B$. With $\mathbf{X}$ being a random variable, uniformly distributed on $\mathbb{F}_2^n$, we have

$$\begin{aligned}
& \Pr\left(\boldsymbol{\alpha} \xrightarrow{f} B\right) \\
= \; & \Pr_{\mathbf{X}}\left(\sum_{i=1}^{2^t-1} \Big(f(\mathbf{X} \oplus \gamma_i) \oplus f(\mathbf{X})\Big) = \beta\right) \\
= \; & \Pr_{\mathbf{X}}\left(\sum_{i=1}^{2^t-1} \Big(f(\mathbf{X} \oplus \gamma_i)\Big) \oplus f(\mathbf{X}) = \beta\right) \\
= \; & \Pr_{\mathbf{X}}\left(\sum_{\gamma \in L(\alpha_1,\ldots,\alpha_t)} \Big(f(\mathbf{X} \oplus \gamma)\Big) = \beta\right)
\end{aligned}$$

which proves the theorem. □

Example. Let α_1 and α_2 be two differences with respect to which we want to take the second order derivative and let β be the output value we are interested in. The probability that $\Delta_{\alpha_1,\alpha_2} f(\mathbf{X}) = \beta$ for uniformly randomly chosen $\mathbf{X}$ is then nothing else than the probability that the 3-difference $(\alpha_1, \alpha_2, \alpha_1 \oplus \alpha_2)$ is mapped to a 3-difference $(\beta_1, \beta_2, \beta_3)$ with $\beta_1 \oplus \beta_2 \oplus \beta_3 = \beta$.

This theoretical connection between truncated and higher-order differentials has an interesting consequence: a higher-order differentials can be regarded as the union of polytopic trails. This principally allows us to determine lower bounds for the probability of higher-order differentials by summing over the probabilities of a subset of all polytopic trails that it contains – just as we are used to from standard differentials.

As shown in Lemma 2, the probability of a $(d+1)$-polytopic trail is always at most as high as the probability of the worst standard differential trail that it contains. A situation in which the probability of a higher-order differential at the same time is dominated by a single polytopic trail and has a higher probability than any ordinary differential can thus never occur. To find a higher-order differential with a higher probability than any ordinary differential for a given cipher, we are thus always forced to sum over many polytopic trails. Whether this number can remain manageable for a large number of rounds will require further research and is beyond the scope of this paper.