J. Math. Crypt. **1** (2007), 221–242

DOI 10.1515 / JMC.2007.011

Probability distributions of correlation and differentials in block ciphers

Joan Daemen and Vincent Rijmen

Communicated by Ed Dawson

Abstract. We study the probability distributions of difference propagation probabilities and input-output correlations for functions and block ciphers of given dimensions, for several of them for the first time. We show that these parameters have distributions that are well-studied in the field of probability such as the normal, Poisson and extreme value distributions. The results of this paper can be used to estimate how much effort will be required to generate functions satisfying certain criteria. The distributions we derive for block ciphers illustrate the significant difference between fixed-key parameters and averaged parameters.

Keywords. Block ciphers, probability distributions, differential cryptanalysis, linear cryptanalysis.

AMS classification. 94A60, 68P25.

1 Introduction

Differential and linear cryptanalysis are the two most powerful general purpose cryptographic attacks known to date. In this section, we briefly review both attacks and motivate our work.

1.1 Differential and linear cryptanalysis

In their basic form, both attacks retrieve key information from the last *round* by analyzing (large) amounts of plaintext/ciphertext pairs. The key information acquired is then used to find even more key bits until the full key is found.

Differential cryptanalysis is a chosen-plaintext attack where plaintexts are applied in pairs that have a fixed difference [1]. The attack exploits the non-uniformity in the distribution of differences in the outputs of a map α, when pairs of inputs with a fixed difference are applied. The non-uniformity exploited can be a differential with a high probability, or, for more advanced versions of the attack, a differential with probability zero, or a combination of differentials with a high probability.

In a first type of differential attack, α equals the block cipher. The information on the ciphertext (output) pairs and plaintext (input) pairs is used to derive information on the key (input). If the distribution of output differences has a large peak with value P, the amount of plaintext/ciphertext pairs for the attack to be successful is proportional to P^{-1}. P is called the *differential probability (DP)*. In a second type of differential

Second author: Part of this work was supported by Cryptomathic.

attack, α is only a part of the block cipher. The map α is selected in such a way that its inputs and outputs can be computed from the plaintext and ciphertext and a 'small' number of key bits (typically 10 to 20 key bits). As in the first type of attack, the required amount of plaintext/ciphertext pairs is proportional to DP^{-1}. In general, DP depends on the key. Hence, the probability of success given a certain amount of plaintext/ciphertext pairs is key-dependent. Often one approximates the DP of a differential by the average of the DP over all keys, called the expected DP (EDP).

Linear cryptanalysis is a known-plaintext attack [8]. It exploits the correlation between linear combinations of input bits and linear combinations of output bits of a non-linear map α. In a first type of linear attack, α equals the block cipher. The information on the ciphertext (output) and the plaintext (input) is used to derive information on the key (input). If the correlation between input and output equals C, the required amount of known plaintexts is proportional to C^{-2}. In a second type of linear attack, α is only a part of the block cipher. The map α is selected in such a way that its inputs and outputs can be computed from the plaintext and ciphertext and a 'small' number of key bits (typically 10 to 20 key bits). If there is a linear combination of input bits and output bits of α that correlate to zero with a correlation value C while all other linear combinations have a negligible correlation, then it is possible to recover the key bits involved in the computation. In this attack the required amount of known plaintext for the attack to be successful is proportional to C^{-2}. The quantity C^2 is generally denoted by the term *linear probability* (*LP*). In general, LP depends on the key and hence the probability of success given a certain amount of known plaintext is key-dependent. Often one approximates the LP of an approximation by the average of the LP over all keys, called the expected LP (EDP).

1.2 Motivation

We study the distributions of DP of differentials and LP of linear approximations over all possible vector Boolean functions and permutations maps with given dimensions. We also study the distributions of the EDP of differentials and ELP of hulls over block ciphers. Finally we characterize the distributions of the maximum of these properties over all differentials or approximations of a given mapping.

Cipher designers sometimes try to reduce the risk for undesired structural properties by including randomly generated substitution tables, or S-boxes, see e.g. [2, 3].

These S-boxes need to satisfy criteria on the DP and LP. The distributions we derive here allow to estimate the expected number of S-boxes that need to be generated —and hence the work factor— before one will satisfy the chosen criteria.

The results of this paper show what we can expect to achieve for the distribution of EDP and DP over all the differentials. We show that the distribution of the EDP stays very close to 2^{-n} for all differentials, while the DP values can deviate significantly. Similar conclusions hold for the ELP and LP values.

Our results are complementary to earlier work by O'Connor et al. about the DP and LP distributions of fixed maps [6, 10, 11]. We additionally consider key-dependent maps and give the distributions of DP and LP values averaged over all keys.

1.3 Overview

The remainder of this paper is organized as follows. After introducing vector Boolean functions, permutations, block ciphers and probability distributions in Section 2, we study the distributions of DP and EDP in Section 3 and the distribution of correlation, LP and ELP in Section 4. In Section 5 we study the distribution of the maximum cardinality and EDP over many differentials, the maximum LP over many approximations and the maximum ELP over many hulls. This is followed by a note on the applicability of our results on difference operations other than the bitwise XOR in Section 6. Section 7 contains the conclusions, Appendix A lists the probability distributions we encounter in this paper.

2 Preliminaries

A *Boolean vector* is a vector with bits as coordinates. The *bitwise binary addition* of two Boolean vectors a and b of the same dimension is a Boolean vector whose coordinates consist of the binary addition (addition modulo 2) of the corresponding bits of a and b. We denote this operation by $+$.

A *Boolean function* $b = f(a)$ is a function that maps a Boolean vector to a bit.

$$f : \mathrm{GF}(2)^n \to \mathrm{GF}(2) : a \mapsto b = f(a) \ . \tag{2.1}$$

The *imbalance* $\mathrm{Imb}(f)$ of a Boolean function f is the number of inputs that it maps to 0 minus the number of inputs that it maps to 1 divided by two. The imbalance can have any integer value and ranges from -2^{n-1} to 2^{n-1}. We have:

$$\mathrm{Imb}(f) = \frac{1}{2}\left(\#\{a|f(a) = 0\} - \#\{a|f(a) = 1\}\right) \ . \tag{2.2}$$

A Boolean function with imbalance 0 is called *balanced*.

A *vector Boolean function* $b = \alpha(a)$ is a function that maps a Boolean vector to another Boolean vector:

$$\alpha : \mathrm{GF}(2)^n \to \mathrm{GF}(2)^m : a \mapsto b = \alpha(a) \ . \tag{2.3}$$

This vector Boolean function has n input bits and m output bits. A vector Boolean function can be specified by its *definition table*: an array containing the output value for each of the 2^n possible input values.

Each bit of the output of a vector Boolean function is itself a Boolean function of the input vector. These are the *coordinate Boolean functions* of the vector Boolean function.

A *vector Boolean transformation* is a vector Boolean function with the same number of input bits as output bits. A *vector Boolean permutation* is an invertible vector Boolean transformation and maps all input values to different output values.

There are 2^{m2^n} n-bit to m-bit vector Boolean functions. In the following, we will study for certain properties the distribution over all vector Boolean functions. The probability space is formed by the 2^{m2^n} n-bit to m-bit vector Boolean functions, where each function has the same probability to occur. We will also study the distributions over the $2^n!$ different n-bit permutations.

A *block cipher* $\mathcal{B}$ with block length n and key length h is an array of 2^h vector Boolean permutations operating on n-bit Boolean vectors. Each key value k determines a vector Boolean permutation denoted by $\mathcal{B}[k]$. We also refer to $\mathcal{B}[k]$ as a *fixed-key (block) cipher*. The probability space of block ciphers with block length n and key length h is the set of $(2^n!)^{2^h}$ samples of size 2^h that can be drawn from the space of n-bit permutations.

All distributions in this paper are discrete. In most cases the variable can take only values that are integer multiples of some value ϵ. We denote the probability that a given discrete variable X has value x by $\Pr(X = x)$. The cumulative distribution of such a variable is given by

$$D(x) = \Pr(X < x) = \Pr(X = x - \epsilon) + D(x - \epsilon) .$$

In several cases we can approximate the distribution of a discrete variable by a continuous distribution. For a continuous distribution we similarly have a cumulative distribution $D(x) = \Pr(X < x)$. Due to the continuous nature, in general we have $\Pr(X = x) = 0$. Instead we have a density function $P_X(x) = \frac{dD(x)}{dx}$. If a the distribution of a discrete variable is approximated by a continuous distribution, we have:

$$\Pr(X = x) \approx \epsilon P_X(x)$$

and

$$\Pr(X < x) \approx D(x - \epsilon/2) .$$

This is known as the *continuity correction* [13]. Given a discrete variable X with $\epsilon = e$ and density $P_X(x)$, a variable $Y = aX$ with a some constant has $\epsilon = ae$. The density $P_Y(y) = P_X(y/a)/a$ as $\Pr(Y = y) = \Pr(X = y/a)$ and hence $eaP_Y(y) = eP_X(y/a)$. Appendix A lists some well-known distributions that we will use in this paper.

3 Differential probability (DP) values

In this section we study the distributions related to the DP of differentials in vector Boolean functions, permutations and block ciphers.

3.1 Terminology related to differentials

A *pair* is an unordered set of two Boolean vectors of the same dimension: $\{v, u\} = \{u, v\}$. The *difference of a pair* is a Boolean vector with value $u + v$, where $+$ denotes the bitwise difference or XOR. In the context of difference propagation in a vector Boolean function α, we speak of an *input difference* of a pair of vectors $\{v, u\}$ and of its output difference $\alpha(u) + \alpha(v)$. For a given n, there are $2^n - 1$ possible non-zero input differences. For each non-zero input difference, there are 2^{n-1} pairs with that input difference.

A *vector Boolean function differential*, or differential for short, consists of an input difference a and an output difference b and is denoted by (a,b). The *differential probability (DP)* of the differential (a,b) is given by the number of pairs that have input difference a and output difference b, divided by the total number of pairs with input difference a:

$$\mathrm{DP}(a,b) = \#\left\{\{v,u\} | v+u=a \text{ and } \alpha(v)+\alpha(u)=b\right\}/2^{n-1} . \tag{3.1}$$

Note that the differential probability of a vector Boolean function differential can take only a limited number of values: it is either zero or a multiple of 2^{1-n}. In the following, we will study the distribution of DP(a,b), where the probability space is formed by the 2^{m2^n} n-bit to m-bit vector Boolean functions. Hence DP(a,b) becomes a stochastic variable.

It is often more convenient to work with the *cardinality* of the differential, because this term avoids confusion between the stochastic variable DP(a,b) and its probability distribution. Furthermore, using the term cardinality, we emphasize the discrete character of this quantity.

Definition 3.1. The cardinality of a vector Boolean function differential $N(a,b)$ is the number of pairs with input difference a that have output difference b.

$$N(a,b) = \#\left\{\{v,u\} | v+u=a \text{ and } \alpha(v)+\alpha(u)=b\right\} . \tag{3.2}$$

Hence the cardinality equals the DP times 2^{n-1}. An *impossible differential* is a differential with DP (or cardinality) equal to 0.

A differential with an input difference equal to 0 also has output difference 0 and is called a *trivial* differential. The trivial differential has differential probability 1 and cardinality 2^{n-1}. For a permutation, all differentials $(a,0)$ with $a \neq 0$ are impossible differentials. The only possible differential of the form $(a,0)$ is the trivial differential. In the remainder of this document we will use the term differential to mean nontrivial differential.

We call the cardinality of a differential over a block cipher where the key is fixed to a specific value a *fixed-key cardinality* of a differential over that block cipher. Since a block cipher where the key is fixed, is simply a permutation, the fixed-key cardinality of a differential over a block cipher has the same distribution as the cardinality of a permutation (both are the same stochastic variable). We denote it by the symbol $N[k](a,b)$.

We define a second type of differential over a block cipher.

Definition 3.2. A *block cipher differential* (a,b) over the block cipher $\mathcal{B}$ is the combination of all vector Boolean function differentials (a,b) over the permutations defined by the fixed-key block ciphers. The *EDP* of a block cipher differential is defined as the sum over all keys of the fixed-key cardinalities of the differential (a,b), divided by the number of possible keys and the number of pairs with input difference a:

$$\mathrm{EDP}(a,b) = 2^{1-n-h} \sum_k N[k](a,b) .$$

The explicit distinction between vector Boolean function differentials and DP on the one hand, and block cipher differentials and EDP on the other hand, may seem artificial at first. However, the analysis in Sections 3.2 and 3.3 will show that the distributions of the DP and the EDP are quite different from one another. Therefore we think it is useful to explicitly make the distinction.

3.2 The cardinality of a differential

The cardinality of a vector Boolean function differential is determined as follows.

Lemma 3.3. *For a vector Boolean function differential (a,b) with fixed a and b, the distribution of the cardinality $N(a,b)$ over all n-bit to m-bit vector Boolean functions is binomial:*

$$\Pr\left(N(a,b)=i\right) = \left(2^{-m}\right)^{i}\left(1-2^{-m}\right)^{2^{n-1}-i}\binom{2^{n-1}}{i} .$$

Proof. Over all choices of the vector Boolean function α, each of the 2^n input values v is mapped to each of the 2^m output values w equally many times. Hence the difference $\alpha(v)+\alpha(v+a)$ takes each value equally many times. Given (a,b), taking a pair with a difference a is an experiment that is successful if the output difference is b. The number of experiments (pairs) is 2^{n-1} and the probability of success is 2^{-m}. The number of successes has the binomial distribution. □

The binomial distribution is often approximated by a Poisson distribution or a normal distribution. Both approximations improve as n grows.

Corollary 3.4. *If m is small, we have:*

$$\Pr(N(a,b)=i) \approx Z\left(\frac{i-2^{n-m-1}}{\sqrt{2^{n-m-1}(1-2^{-m})}}\right) .$$

with $Z\,()$ denoting a normal distribution (see Appendix A).

Theorem 3.5. *For $n \geq 5$ and $n-m$ small, we have (see Appendix A):*

$$\Pr\left(N(a,b)=i\right) \approx e^{-2^{n-m-1}}\frac{2^{(n-m-1)i}}{i!} = \text{Poisson}(i;2^{n-m-1}) .$$

For Boolean functions we have $m = 1$ and hence the cardinality $N(a,b)$ has a distribution close to a normal distribution with $\mu(N) = 2^{n-2}$ and $\sigma^2(N) = 2^{n-3}$.

Corollary 3.6. *Over the Boolean transformations, the cardinality has the following distribution:*

$$\Pr\left(N(a,b)=i\right) \approx \text{Poisson}(i;\frac{1}{2}) = \frac{e^{-\frac{1}{2}}}{i!2^i} .$$

The entries in the definition table of a permutation are not independent from one another. Taking this restriction into account would strongly complicate the analysis. Fortunately, the case of permutations was rigorously studied and described in [10, 6]. It turns out that in the computation of the distribution it is sufficient to replace the probability of success by $1/(2^n - 1)$ for nonzero output differences b and by 0 for $b = 0$. For large n this has a negligible effect on the cardinality of differentials (a, b) with $b \neq 0$ and hence Corollary 3.6 for transformations is also valid for permutations.

3.3 EDP of block cipher differentials

The distribution of the EDP of a block cipher differential looks quite different:

Theorem 3.7. *The continuous approximation of the distribution over all block ciphers of the EDP of a block cipher differential has a density which is very close to a normal density with mean $\mu(\text{EDP}) = 2^{-n}$ and standard deviation $\sigma(\text{EDP}) = 2^{-n}2^{(1-h)/2}$. The values with non-zero probability in the discrete distribution are the multiples of $\epsilon = 2^{1-n-h}$.*

Proof. The EDP of a block cipher differential is determined by the sum of 2^h independent variables. For all reasonable values of h, 2^h is large enough to invoke the central limit theorem. The individual variables have the distribution of Corollary 3.6, i.e. with mean 2^{-1} and variance 2^{-1} resulting in mean and variance both equal to 2^{1-h} for the sum, resulting in a standard deviation of $2^{(1-h)/2}$. Division by 2^{h+n-1} yields the mean and standard deviation for the EDP. □

We conclude that the distribution of the EDP differs significantly from zero only for values extremely close to the mean value 2^{-n}, in contrast to the distribution of the DP.

In the remainder of this paper, we will abbreviate statements similar to Theorem 3.7 by writing: The density function of the EDP of a block cipher differential has a normal shape with mean 2^{-n} and variance $2^{-2n}2^{1-h}$. The distribution has $\epsilon = 2^{1-n-h}$.

4 Correlation and LP values

In this section we study the distributions of correlation and LP of linear approximations over vector Boolean functions, permutations and block ciphers.

4.1 Terminology related to correlation

A *parity* of a Boolean vector is a binary Boolean function that consists of the binary addition of a number of its coordinates. A parity is determined by the indices of the bits of the Boolean vector that are included in the binary addition.

The *selection vector* u of a parity is a Boolean vector that has a 1 in the bit positions that are included in the parity and a 0 in all other positions. Analogously to the inner product of vectors in linear algebra, we express the parity of vector a corresponding with selection vector u as $u^{\mathrm{T}}a$. In this expression the T suffix denotes transposition of the vector u.

A vector Boolean function *(linear) approximation* α consists of an n-bit input selection vector v and an m-bit output selection vector u and is denoted by (v,u). An approximation with both the input selection vector and the output selection vector equal to 0 is called a *trivial* approximation. The imbalance $\text{Imb}(v,u)$ of an approximation (v,u) over a function α is the imbalance of the Boolean function given by:

$$v^{\mathrm{T}}a + u^{\mathrm{T}}\alpha(a) \ .$$

The *correlation* of an approximation is its imbalance divided by 2^{n-1}:

$$C(v,u) = \text{Imb}(v,u)/2^{n-1} = 2^{1-n} \times \text{Imb}(v,u) \ . \tag{4.1}$$

The correlation ranges from -1 to $+1$. A correlation with value -1 means that the parities defined by v and u are each others complement and value $+1$ means that they are equal. Several authors work with the *bias* [8]. The bias of an approximation is its correlation divided by two. We prefer to follow here the terminology of e.g. [9] and work with the correlation. By working with correlation, we avoid the Piling-Up Lemma and eliminate factors 2^i that appear in equations when using bias [8]. Furthermore, the quantity denoted here by correlation, corresponds exactly to the definition of correlation in other fields of mathematics, e.g. probability theory [7].

The *linear probability* (or rather *potential*) *(LP)* of an approximation $\text{LP}(v,u)$ is the square of its correlation and ranges from 0 to 1. We call an approximation with zero correlation an *impossible approximation*. It is well known, see e.g. [4], that for any vector Boolean function and for all u:

$$\sum_{v} \text{LP}(v,u) = 1 \ . \tag{4.2}$$

The approximation with output selection vector 0 and input selection vector 0 is the only possible *trivial approximation*. It has imbalance 2^{n-1} and correlation 1. Approximations $(v,0)$ with $v \neq 0$ are impossible approximations. For permutations, all approximations $(0,u)$ with $u \neq 0$ are also impossible approximations. The correlation of an approximation over an n-bit permutation is an integer multiple of 2^{n-2}. In the remainder of this document we will use the term approximation to mean nontrivial approximation.

In a block cipher we can consider the correlation (LP) of an approximation for a fixed key. We define a hull:

Definition 4.1. A *hull* (v,u) is the combination of the approximations (v,u) for all keys. The ELP of a hull (v,u) is the average of the LP values of the approximations (v,u) over all keys.

The average correlation of a hull gives no indication about the complexity of a linear attack. Therefore, we only talk about the ELP of a hull.

4.2 Correlation of a vector Boolean function approximation

We start with a result on the imbalance of an approximation (v,u).

Theorem 4.2. *The imbalance* $\mathrm{Imb}(v, u)$ *of an* n*-bit to* m*-bit vector Boolean function approximation has the following distribution:*

$$\Pr(\mathrm{Imb}(v, u) = z) = 2^{-2^n} \binom{2^n}{2^{n-1} + z} .$$

Proof. We start by computing the number of vector Boolean functions for which an approximation (v, u) has imbalance z. For a given Boolean function f, the number of n-bit to m-bit vector Boolean functions α that satisfy

$$v^{\mathrm{T}} a + u^{\mathrm{T}} \alpha(a) = f(a)$$

is independent of the choice of f and is equal to $2^{m2^{n-1}}$. So the number of vector Boolean functions that satisfy $\mathrm{Imb}(v, u) = z$ is equal to $2^{m2^{n-1}}$ times the number of Boolean functions $f(a)$ with imbalance z. Dividing by the total number of n-bit to m-bit vector Boolean functions results in the given distribution for the imbalance. □

Observe that the distribution of the imbalance is equal to a binomial distribution with parameters $p = 1/2$ and $\mathrm{N} = 2^n$, where the mean has been translated to the origin (see Appendix A).

Corollary 4.3. *For* $n \geq 5$*, the result of Theorem 4.2 can be approximated by:*

$$\Pr(\mathrm{Imb}(v, u) = z) \approx Z\left(\frac{z}{2^{(n-2)/2}}\right) ,$$

for z *an integer and 0 otherwise.*

This corollary follows from the normal approximation of the binomial distribution [5, 12]. For the correlation this yields $C(v, u) = 2^{-n+1}\mathrm{Imb}(v, u)$, hence $\mu(C) = 0$ and $\sigma(C) = 2^{-n/2}$.

Theorem 4.2 can also be used to compute the mean and the variance of the LP.

Corollary 4.4. *The LP of an* n*-bit to* m*-bit vector Boolean function approximation has mean* $\mu(\mathrm{LP}) = 2^{-n}$ *and, when* $n \geq 5$*, variance* $\sigma^2(\mathrm{LP}) \approx 2 \times 2^{-2n}$*.*

Proof. Since $\mathrm{LP}(v, u) = C^2(v, u) = 2^{-2n+2}\mathrm{Imb}^2(v, u)$, we have:

$$\begin{aligned} \mu(\mathrm{LP}) &= 2^{-2n+2}(\sigma^2(\mathrm{Imb}) - (\mu(\mathrm{Imb}))^2) \\ \sigma^2(\mathrm{LP}) &= 2^{-4n+4}\mu(\mathrm{Imb}^4) - (\mu(\mathrm{LP}))^2 . \end{aligned}$$

The fourth moment of a binomially distributed variable X is given by [12]:

$$\mu(X^4) = \mathrm{N}p(1-p)\left(3p^2(2-\mathrm{N}) + 3p(\mathrm{N}-2) + 1\right) .$$

Hence, we obtain:

$$\begin{aligned} \mu(\mathrm{LP}) &= 2^{-2n+2} \times 2^{n-2} = 2^{-n} \\ \sigma^2(\mathrm{LP}) &= 2^{-4n+4} \times \left(2^{n-2}\left(3/4(2-2^n) + 3/2(2^n-2) + 1\right)\right) - 2^{-2n} \\ &\approx 2^{-4n+4} \times 3 \times 2^{2n-4} - 2^{-2n} = 2 \times 2^{-2n} . \end{aligned}$$

□

4.3 Correlation of a permutation approximation

We first derive the distribution of the imbalance of an n-bit vector Boolean permutation approximation. This distribution was already given in [11] but the proof was missing due to page limit restrictions.

Lemma 4.5 ([11]). *The imbalance* $\mathrm{Imb}(v,u)$ *of an n-bit permutation approximation has the following distribution:*

$$\Pr(\mathrm{Imb}(v,u) = 2x) = \frac{\binom{2^{n-1}}{2^{n-2}+x}^2}{\binom{2^n}{2^{n-1}}} \; .$$

Proof. We start by computing the fraction of permutations for which an approximation (v,u) has imbalance $\mathrm{Imb}(v,u) = z$. Consider the Boolean function g defined by

$$g(a) = u^{\mathrm{T}}\alpha(a) \; . \tag{4.3}$$

Clearly, $g(a)$ is an output parity of α. Since α is a permutation, $g(a)$ is balanced. Its definition table contains 2^{n-1} zeroes and 2^{n-1} ones.

A vector Boolean permutation for which one output parity has been fixed to a function g can be constructed as follows. Complement the output parity with an $(n-1)$-bit permutation for the part of the definition table with $g(a) = 0$ and an $(n-1)$-bit permutation for the part of the definition table with $g(a) = 1$. It follows that the number of such vector Boolean permutations is independent from the particular function $g(a)$ and only depends on the dimension n. Hence the fraction of vector Boolean permutations that satisfy $\mathrm{Imb}(v,u) = z$ is equal to the number of balanced Boolean functions g that satisfy

$$\mathrm{Imb}(g(a) + v^{\mathrm{T}}a) = z \; , \tag{4.4}$$

divided by the total number of balanced Boolean functions.

We compute now the number of balanced Boolean functions that satisfy (4.4). Partition the definition table of $g(a)$ in two halves: D_0 for which $v^{\mathrm{T}}a = 0$ and D_1 for which $v^{\mathrm{T}}a = 1$. The total imbalance of $g(a) + v^{\mathrm{T}}a$ is given by the imbalance of $g(a)$ restricted to D_0 (called x) minus the imbalance of $g(a)$ restricted to D_1 (called y). As $g(a)$ is balanced, we have $x + y = 0$ and so $y = -x$. The imbalance of $g(a) + v^{\mathrm{T}}a$ is hence given by $2x$. It follows that in a vector Boolean permutation all approximations have an even imbalance.

The number of balanced Boolean functions $g(a)$ for a given value of x is:

$$\binom{2^{n-1}}{2^{n-2}+x}\binom{2^{n-1}}{2^{n-2}-x} = \binom{2^{n-1}}{2^{n-2}+x}^2 \; . \tag{4.5}$$

If we divide this by the total number of balanced Boolean functions, we obtain the probability distribution of the imbalance. □

In [11], it is proven that the number of approximations with correlation equal to 0, tends to zero when n grows. Additionally, some upper bounds are derived on the maximum correlation amplitude over all approximations of a permutation. The distributions for correlations and maximum LP values we derive in the remainder of this section and following sections, confirm these results.

Lemma 4.6. *The density function of the imbalance* $\mathrm{Imb}(v, u)$ *of an* n*-bit permutation approximation with* $n \geq 5$ *has a normal shape with mean* 0 *and variance* 2^{n-2}. *The discrete distribution has* $\epsilon = 2$.

Proof. We start with the expression of Lemma 4.5. If 2^{n-1} is large, we have:

$$\binom{2^{n-1}}{2^{n-2}+x} \approx 2^{2^{n-1}} Z\left(\frac{x}{2^{(n-3)/2}}\right), \tag{4.6}$$

and

$$\binom{2^n}{2^{n-1}} \approx 2^{2^n} \frac{2^{-(n-2)/2}}{\sqrt{2\pi}}. \tag{4.7}$$

Working this out yields:

$$\Pr(\mathrm{Imb}(v,u) = 2x) \approx Z\left(\frac{x}{2^{(n-4)/2}}\right). \tag{4.8}$$

Substituting x by $z/2$ gives the desired distribution. □

Theorem 4.7. *The density function of the correlation of an* n*-bit permutation approximation with* $n \geq 5$ *has a normal shape with mean 0 and variance* 2^{-n}.

This follows immediately from Lemma 4.6. The same holds for a transformation approximation. Both distributions differ in the fact that $\epsilon = 2^{1-n}$ for transformations while $\epsilon = 2^{2-n}$ for permutations.

4.4 ELP of hulls

The distribution of the ELP of a hull is the same as the distribution of the EDP of a block cipher differential.

Theorem 4.8. *The density function of the ELP of a hull has a normal shape with mean* 2^{-n} *and standard deviation* $2^{-n}2^{(1-h)/2}$.

Proof. The fixed-key LP of a hull has mean 2^{-n} and standard deviation approximately $2^{\frac{1}{2}-n}$ (Corollary 4.4). Application of the central limit theorem results in the given distribution. □

As in the case for EDP of block cipher differentials, the ELP of hulls stays very close to its mean value 2^{-n}, while this is not the case for the LP.

5 Maximum cardinality, EDP, ELP and LP

In [10, 6] bounds have been proven for the maximum DP over permutations. In this section, we derive the shape of the distribution of the maximum cardinality, the EDP, the ELP and the LP for Boolean vector functions and block ciphers.

5.1 Distribution of maxima

We now derive an expression for the cumulative distribution of the maximum of a very large set of variables with identical distributions, that decrease exponentially for large x. We model the cumulative distribution of these variables as:

$$D(x) = 1 - e^{-f(x)} , \tag{5.1}$$

with $f(x)$ a function that increases in a sub-exponential way. Let the number of variables be denoted by 2^y and let $D_{\text{max}}(x)$ denote the cumulative distribution of their maximum. We know from order statistics [5, 16] that the cumulative distribution of the maximum of a number of independent variables is the product of the cumulative distributions of these variables. Hence we have:

$$D_{\text{max}}(x) = D(x)^{2^y} = (1 - e^{-f(x)})^{2^y} \approx e^{-2^y e^{-f(x)}} = e^{-e^{\ln(2)y - f(x)}} . \tag{5.2}$$

For a continuous variable x, let p be the solution for x in $f(x) = \ln(2)y$ and let w be 1 divided by the derivative of $f(x)$ in p. We approximate the function $\ln(2)y - f(x)$ by a linear function around the point where its value is 0:

$$D_{\text{max}}(x) \approx e^{-e^{\frac{p-x}{w}}} . \tag{5.3}$$

This distribution has been well studied in probability theory and is known as the *extreme value* distribution, Fisher-Tippett distribution or log-Weibull distribution [5, 15]. The corresponding density $P_{\text{max}}(x)$ is depicted in Figure 1. Its peak is in p and its width is proportional to w. This distribution has $\mu(X) = p + w\gamma$ with $\gamma \approx 0.58$ and $\sigma(X) = \frac{\pi}{\sqrt{6}}w \approx 1.3w$. Clearly, the validity of (5.3) depends on the quality of the linear approximation of $f(x)$ around $(p, 0)$.

5.2 Maximum cardinality

We can now apply this to the maximum cardinality of a large set of vector Boolean function differentials because their cardinalities have identical distributions. However, they are not independent variables. An important relation between them is:

$$\sum_b N(a, b) = 2^{n-1}, \ \forall a . \tag{5.4}$$

Moreover, when α is a permutation, (5.4) also holds when summing over a and keeping b constant. Assuming independence allows us to apply the results of Section 5.1. The experimental data that we present in Figure 2 confirm that this assumption is justified.

Assumption 5.1. The joint distributions of a set of cardinalities $N(a_i, b_i)$ can be approximated by assuming that the cardinalities are statistically independent.

$$\begin{aligned} &\Pr(N(a_1, b_1) = N_1, N(a_2, b_2) = N_2, \dots) \\ &\qquad\qquad \approx \quad \Pr(N(a_1, b_1) = N_1) \times \Pr(N(a_2, b_2) = N_2) \times \dots \end{aligned}$$

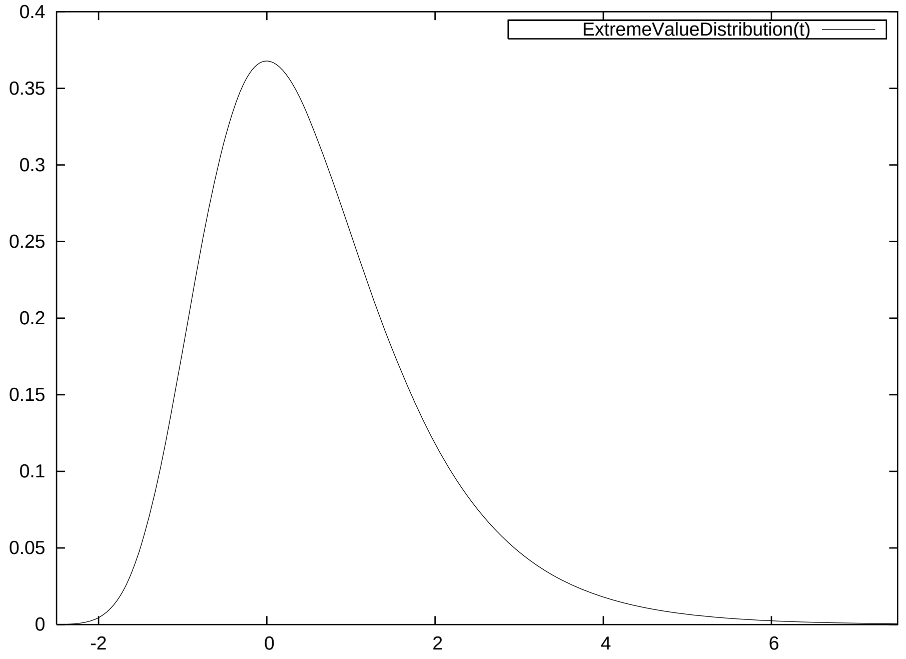

Figure 1. Extreme value distribution with $p = 0$ and $w = 1$.

The cumulative distribution of the cardinality of a differential is given by:

$$D(x) = 1 - e^{-f(x)} = \Pr(N(a,b) \leq x) \; . \tag{5.5}$$

For vector Boolean function differentials, the approximate distribution of the maximum cardinality with a given input difference is given by (5.2) with $y = m$, and for the maximum overall cardinality by (5.2) with $y = m + n$. The distributions are also valid for vector Boolean permutations, with $n = m$, approximating $2^n - 1$ by 2^n.

If we consider in a block cipher the maximum over all keys of the fixed-key cardinality of a differential (a,b), then the independence of the variables holds. Hence (5.2) applies, with $y = h$. Using Assumption 5.1, we obtain for the maximum cardinality over all keys and differentials (5.2) with $y = 2n + h$.

Theorem 3.5 implies that the cardinality of a permutation differential has a distribution that is close to a Poisson distribution with $\lambda = 1/2$.

We now derive the function $f(x)$ corresponding with a Poisson distribution. For the cumulative distribution of a variable with the Poisson distribution with given λ we have

$$D(x) = \sum_{i=0}^{x-1} \text{Poisson}(i;\lambda) = 1 - \sum_{i \geq x} \text{Poisson}(i;\lambda) \; . \tag{5.6}$$

For $x \gg \lambda$, this can be closely approximated by [5]:

$$D(x) \approx 1 - \big(1 + \frac{\lambda}{x}\big)\text{Poisson}(x;\lambda) \approx 1 - \text{Poisson}(x;\lambda) = 1 - e^{-\lambda}\frac{\lambda^x}{x!} \; . \tag{5.7}$$

If we use the Stirling approximation for the factorial [5, 17], we obtain the following expression for $f(x)$:

$$f(x) \approx \frac{1}{2}\ln(2\pi) + \lambda + x\ln(x) - (1 + \ln(\lambda))x + \frac{1}{2}\ln(x) \; . \tag{5.8}$$

Combined with Equation (5.2), this gives an expression for the distribution of the maximum cardinality over 2^y differentials. If we consider the maximum over all differentials (a,b) of a Boolean permutation with given a we have $y = n$. For the maximum over all differentials we have $y = 2n$ and for the maximum over all differentials and all keys of a block cipher, we have $y = 2n + h$. The Poisson distribution is discrete in nature, so this approximation is only valid for integer values of x. For the time being, we make abstraction of this and compute the parameters p and w as if x is a continuous variable.

$$\ln(2)y = \frac{1}{2}\ln(2\pi) + \lambda + p\ln(p) - (\ln(\lambda) + 1)p + \frac{1}{2}\ln(p) \; , \tag{5.9}$$

or equivalently:

$$p = \frac{\ln(2)y - \frac{1}{2}\ln(2\pi p) - \lambda}{\ln(\frac{p}{\lambda}) - 1} \; , \tag{5.10}$$

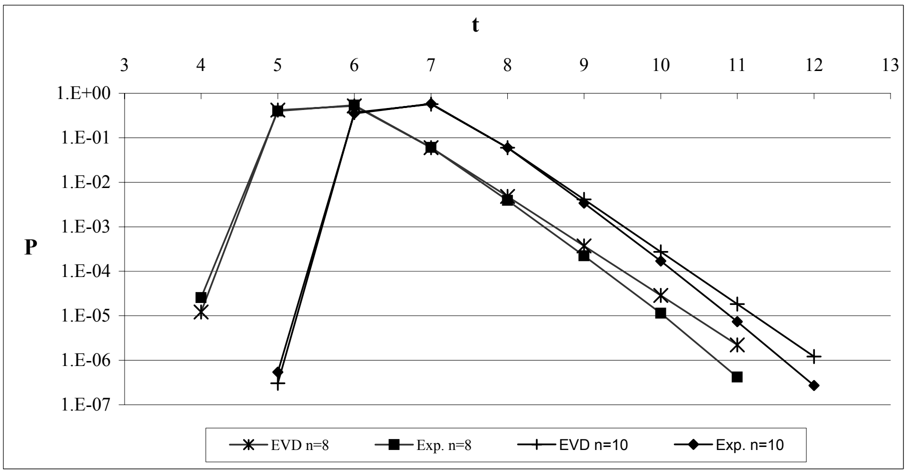

Figure 2. Experimental and theoretical distributions of maximum cardinality $t = \max_{a,b} N(a,b)$ in 8-bit and 10-bit permutations.

which can be solved iteratively. The derivative of $f(x)$ is given by:

$$\ln\left(\frac{x}{\lambda}\right) + \frac{1}{2x}\ . \tag{5.11}$$

Filling in p and using $p \gg \lambda$, we obtain:

$$w \approx \frac{1}{\ln(\frac{p}{\lambda})}\ . \tag{5.12}$$

It follows that if p is much larger than λ, the standard deviation becomes smaller than 1. Since the distribution of the maximum is discrete, this small value of the standard deviation means that the distribution is concentrated at the two integer values near p.

To verify our approximations, we have randomly generated a large number of permutations ranging from 4 to 10 bits and computed the distribution of the maximum cardinality. Starting from 5 bits the typical shape of the extreme value distribution becomes apparent. Figure 2 gives the distributions for 8-bit and 10-bit permutations obtained from our experiments (exp.) and the corresponding extreme value distributions (EVD) with peak and width values given by equations (5.10) and (5.11). The figure illustrates that around the peak the extreme value approximations match the experimental data quite closely. The divergence between the extreme value distribution and the experiments for larger cardinality values, due to the nonlinearity of equation (5.8) becomes less important as n grows.

Table 1 gives the approximated probability distributions for the maximum cardinality over all permutation differentials for $n = 64, 128$ and 256. The values in the table illustrate that the distributions are very narrow. The probability is only large in a single, or two successive values of x.

$n = 64$		$n = 128$		$n = 256$	
t	Pr	t	Pr	t	Pr
26	3.0×10^{-62}	47	1.8×10^{-9}	83	8.0×10^{-56}
27	0.080	48	0.81	84	0.47
28	0.88	49	0.18	85	0.52
29	0.042	50	0.002	86	0.004
30	7.1×10^{-4}	51	2.0×10^{-5}	87	2.5×10^{-5}
31	1.2×10^{-5}	52	1.9×10^{-7}	88	1.4×10^{-7}
32	1.8×10^{-7}	53	1.8×10^{-9}	89	7.9×10^{-10}
33	2.7×10^{-9}	54	1.7×10^{-11}	90	4.4×10^{-12}
34	4.0×10^{-11}	55	1.5×10^{-13}	91	2.4×10^{-14}

Table 1. Distribution of $t = \max_{a,b} N(a,b)$ for n-bit permutations, using equations (5.2) and (5.8) with $y = 2n$.

n	$p(\max_a N)$	$p(\max_{a,b} N)$	n	$p(\max_a N)$	$p(\max_{a,b} N)$
8	3.59	5.95	64	16.60	28.23
12	4.82	8.00	96	22.61	38.75
16	5.95	9.99	128	28.23	48.66
24	8.00	13.37	192	38.75	67.30
32	9.99	16.60	256	48.66	84.94
48	13.37	22.61	384	67.30	118.33

Table 2. Peak value of the maximum cardinality, using equation (5.10) with $y = n$ and $y = 2n$.

Table 2 lists the peak values of the distributions for permutations with typical dimensions that we obtain using these approximations. Table 3 lists the peak values of this distribution for typical block cipher dimensions.

5.3 Maximum EDP and ELP

Theorem 3.7 states that the density function of the EDP of a block cipher differential has a normal shape. Theorem 4.8 does the same for the ELP of a hull. Hence, we apply the results from Section 5.1 to variables with a normal distribution.

Let us first consider the easier case of a variable x with a standard normal distribution. We have:

$$D(x) = \int_{-\infty}^{x} Z(u) du \, . \tag{5.13}$$

This function is closely related to the error function (erf), and can for large x be closely

n	h	$p(\max_{k,a,b} N[k])$	h	$p(\max_{k,a,b} N[k])$
64	56	37.48	128	48.66
96	96	53.44	192	67.30
128	128	67.30	256	84.94
192	128	84.94	256	101.89
256	192	110.17	256	118.33

Table 3. Peak values for the maximum of the fixed-key cardinality over all keys and differentials, using equation (5.10) with $y = 2n + h$.

approximated by [5, 14]:

$$D(x) \approx 1 - \frac{1}{x} Z\,(x) = 1 - \frac{1}{x\sqrt{2\pi}} e^{\frac{-x^2}{2}} \;. \tag{5.14}$$

From this we can derive the following expression for $f(x)$:

$$f(x) = -\ln\left(\frac{1}{x} Z\,(x)\right) = \frac{1}{2}(\ln(2\pi) + x^2) + \ln(x) \;. \tag{5.15}$$

The parameter p_s (subscript s for standard) is the solution of

$$p_s = \sqrt{2\ln(2)y - \ln(2\pi) - 2\ln(p_s)} \;, \tag{5.16}$$

which can be solved for p iteratively, ignoring the rightmost term in the first iteration. The derivative of $f(x)$ is given by:

$$x + \frac{1}{x} \;, \tag{5.17}$$

hence

$$w_s = \frac{p_s}{{p_s}^2 + 1} \approx \frac{1}{p_s} \;. \tag{5.18}$$

If $y > 30$, which is always the case in block ciphers, we obtain that the maximum has an extreme value distribution with $p_s \approx 1.18\sqrt{y}$ and $w_s \approx 1/(1.18\sqrt{y})$. We can find the values of p and w for any normal distribution with mean $\mu(x)$ and standard deviation σ by substituting x by $\frac{x-\mu(x)}{\sigma}$. This gives for the EDP and the ELP:

$$p \;=\; \mu(x) + \sigma p_s \approx 2^{-n}\left(1 + 1.18\sqrt{y}\,2^{(1-h)/2}\right) \;, \tag{5.19}$$

$$w \;=\; \sigma w_s \qquad \approx 2^{-n}\frac{1}{1.18\sqrt{y}}\,2^{(1-h)/2}. \tag{5.20}$$

5.4 Maximum LP

If we accept Assumption 5.2, we can apply the results from Section 5.1.

Assumption 5.2. The joint distributions of a set of LP values $\mathrm{LP}(v_i, u_i)$ can be approximated by assuming that the LP values are statistically independent.

$$\begin{aligned}&\Pr(\mathrm{LP}(v_1,u_1) = \mathrm{LP}_1, \mathrm{LP}(v_2,u_2) = \mathrm{LP}_2, \dots)\\ &\qquad\approx \quad \Pr(\mathrm{LP}(v_1,u_1) = \mathrm{LP}_1) \times \Pr(\mathrm{LP}(v_2,u_2) = \mathrm{LP}_2) \times \dots\end{aligned}$$

We start again with the easier case of $z = x^2$ and x a variable with the standard normal distribution. We have:

$$D(z) = \int_{-\sqrt{z}}^{\sqrt{z}} Z(u)du = 2\int_{-\infty}^{\sqrt{z}} Z(u)du - 1 \,. \tag{5.21}$$

Using (5.14), this gives:

$$D(z) \approx 1 - \sqrt{\frac{2}{\pi z}}\, e^{\frac{-z}{2}} \,, \tag{5.22}$$

yielding

$$f(z) = \frac{1}{2}\left(\ln\left(\frac{\pi}{2}\right) + z + \ln(z)\right) \,. \tag{5.23}$$

p_s is the solution of:

$$p_s = 2\ln(2)y - \ln\left(\frac{\pi}{2}\right) - \ln(p_s) \,. \tag{5.24}$$

The derivative of $f(z)$ is given by:

$$\frac{1}{2}\left(1 + \frac{1}{z}\right) \,. \tag{5.25}$$

As the function $f(z)$ around $(p_s, f(p_s))$ only differs from a linear function by a logarithmic term, the approximation (5.3) is in this case particularly good. For large values of y, the maximum has an extreme value distribution with $p_s \approx 1.38y - \ln(1.38y)$ and $w_s \approx 2$.

We can find the values of p and w for any normal distribution with mean 0 and standard deviation σ by substituting x by $\frac{x}{\sigma}$. This gives for the LP:

$$p = \sigma^2 p_s \approx (1.38y - \ln(1.38y))2^{-n} \,, \tag{5.26}$$

$$w = \sigma^2 w_s \approx 2 \times 2^{-n} \,. \tag{5.27}$$

This results in mean $\approx 2^{-n}(1.38y - \ln(1.38y) + 1)$ and standard deviation $\approx 2.6 \times 2^{-n}$.

Figure 3 gives distributions of the maximum LP for 8-bit and 10-bit permutations, scaled by a factor 2^8, respectively 2^{10}. It shows the distributions obtained from our experiments (exp.) and the corresponding extreme value distributions (EVD) with peak and width values given by equations (5.24) and (5.25). The figure illustrates that the extreme value approximations match the experimental distributions quite closely in

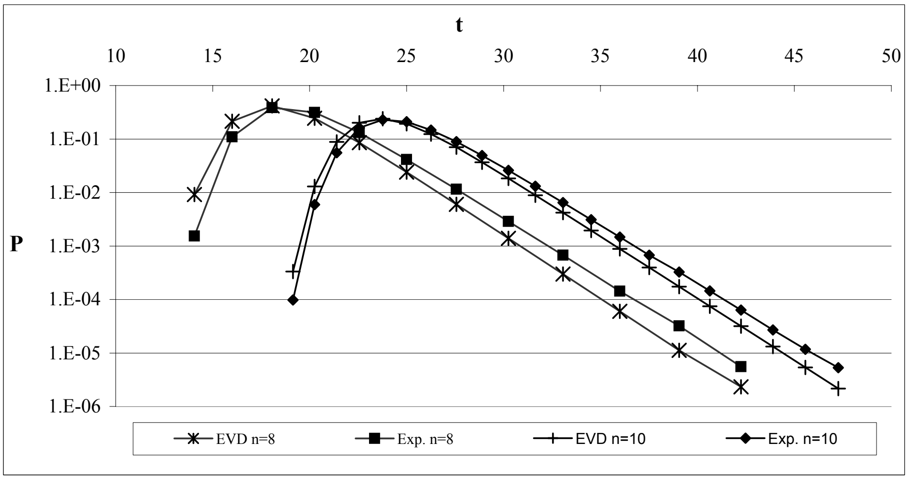

Figure 3. Experimental and theoretical distributions of maximum LP $t = 2^n \max_{a,b} \mathrm{LP}(a,b)$ in 8 and 10-bit permutations.

n	h	$p(\max_{k,v,u} \mathrm{LP}[k])$	h	$p(\max_{k,v,u} \mathrm{LP}[k])$
64	56	249×2^{-64}	128	349×2^{-64}
96	96	393×2^{-96}	192	526×2^{-96}
128	128	526×2^{-128}	256	703×2^{-128}
192	128	703×2^{-192}	256	880×2^{-192}
256	192	969×2^{-256}	256	1057×2^{-256}

Table 4. Peak values of the maximum fixed-key LP for some typical block cipher dimensions, using equation (5.24) with $y = 2n + h$

shape but that they slightly underestimate the maximum LP. This can be explained by the fact that equation (5.21) is based on the continuous normal approximation of the discrete distribution of the LP and that no continuity correction is performed. The figure shows that the deviation becomes smaller as n grows as the distance between the possible LP values shrinks. Table 4 lists the peak values of the maximum fixed-key LP for some typical block cipher dimensions.

5.5 Example

If we apply the results of this section to 128-bit block ciphers with 128-bit keys, we see that the mean of the maximum EDP over all differentials (and ELP over all hulls) is about $2^{-128}(1 + 2^{-59})$ while for any given key, the distribution of the maximum DP is narrowly centered around 96×2^{-128} and that of the maximum LP around 350×2^{-128}.

6 Other difference operations

The bit-wise XOR operation is by far the most common operation used in differential cryptanalysis of block ciphers, but in principle one can choose another group operation for conducting attacks, such as modular integer addition and multiplication. The most characteristic feature of the XOR operation is the fact that all elements have order 2. For other group operations, this is not the case. Our analysis on the cardinality can be extended to other group operations, by taking into account the following observations.

1. The order of the elements in pairs becomes important. The number of ordered pairs with a difference a equals 2^n rather than 2^{n-1}. The distribution of the cardinality $N(a,b)$ over all n-bit to m-bit vector Boolean functions becomes:

$$\Pr\left(N(a,b)=i\right)=\left(2^{-m}\right)^i\left(1-2^{-m}\right)^{2^n-i}\binom{2^n}{i}. \tag{6.1}$$

All numerical approximations of the distribution of the cardinality can be adapted by replacing 2^{n-1} by 2^n. In order to approximate the distribution of the maximum cardinality over all differentials with a given a, we can again invoke Assumption 5.1.

2. Changing the order of the elements in a pair corresponds to changing the sign in the difference. Hence

$$N(a,b)=N(-a,-b),\ \forall\, a,b\,. \tag{6.2}$$

When we approximate the distribution of the maximum cardinality over all differentials, then we have (at most) 2^{m+n-1} independent variables instead of 2^{m+n}.

7 Conclusions

We conclude that while the distribution of the maximum EDP or ELP over a block cipher is centered extremely close to the mean value 2^{-n}, the distribution of the *maximum* DP and LP are significant up to a multiple of 2^{-n}.

In this paper, we have derived approximations for the probability distributions of a number of important parameters over all vector Boolean functions, permutations and block ciphers of given dimensions. For most of these parameters, this is the first time that expressions have been obtained for their distributions. We have shown that their distributions can be approximated by distributions that are well-studied in the field of probability theory such as the normal, Poisson and extreme value distributions. Our experimental verifications suggest that our approximations are very good. We showed that the distributions of fixed-key DP and LP values and EDP and ELP values are quite different.

Acknowledgments. We would like to thank the anonymous reviewers for helpful comments.

References

[1] E. Biham and A. Shamir, *Differential Cryptanalysis of DES-like Cryptosystems*, Journal of Cryptology 4 (1991), pp. 3–72.

[2] J. A. Clark, J. L. Jacob, and S. Stepney, *The design of S-boxes by simulated annealing*, New Generation Computing Journal 23 (2005), pp. 219–231.

[3] J. A. Clark, J. L. Jacob, S. Stepney, S. Maitra, and W. Millan, *Evolving Boolean functions with multiple criteria*. Progress in Cryptology — Indocrypt 2002, Lecture Notes in Computer Science 2551, pp. 246–259. Springer-Verlag, 2002.

[4] J. Daemen, R. Govaerts, and J. Vandewalle, *Correlation matrices*. Fast Software Encryption '94 (B. Preneel, ed.), Lecture Notes in Computer Science 1008, pp. 275–285. Springer-Verlag, 1995.

[5] W. Feller, *An Introduction to Probability Theory and Its Applications*, 1. Wiley & Sons, 1968.

[6] P. Hawkes and L. O'Connor, *XOR and Non-XOR Differential Probabilities*. Advances in Cryptology, Proceedings of Eurocrypt '99 (J. Stern, ed.), Lecture Notes in Computer Science 1592, pp. 272–285. Springer-Verlag, 1999.

[7] P. Hoel, S. Port, and C. Stone, *Introduction to probability theory*. Houghton Mifflin Company.

[8] M. Matsui, *Linear Cryptanalysis Method for DES Cipher*. Advances in Cryptology, Proceedings of Eurocrypt '93 (T. Helleseth, ed.), Lecture Notes in Computer Science 765, pp. 386–397. Springer-Verlag, 1994.

[9] K. Nyberg, *Linear Approximation of Block Ciphers*. Advances in Cryptology, Proceedings of Eurocrypt '94 (A. De Santis, ed.), Lecture Notes in Computer Science 950, pp. 439–444. Springer-Verlag, 1995.

[10] L. O'Connor, *On the Distribution of Characteristics in Bijective Mappings*. Advances in Cryptology, Proceedings of Eurocrypt '93 (T. Helleseth, ed.), Lecture Notes in Computer Science 765, pp. 360–370. Springer-Verlag, 1994.

[11] ______, *Properties of Linear Approximation Tables*. Fast Software Encryption '94 (B. Preneel, ed.), Lecture Notes in Computer Science 1008, pp. 131–136. Springer-Verlag, 1995.

[12] E. Weisstein, *Binomial Distribution*, From MathWorld—A Wolfram Web Resource. http://mathworld.wolfram.com/BinomialDistribution.html.

[13] ______, *Continuity Correction*, From MathWorld—A Wolfram Web Resource. http://mathworld.wolfram.com/ContinuityCorrection.html.

[14] ______, *Erf*, From MathWorld—A Wolfram Web Resource. http://mathworld.wolfram.com/Erf.html.

[15] ______, *Extreme Value Distribution*, From MathWorld—A Wolfram Web Resource. http://mathworld.wolfram.com/ExtremeValueDistribution.html.

[16] ______, *Order Statistic*, From MathWorld—A Wolfram Web Resource. http://mathworld.wolfram.com/OrderStatistic.html.

[17] ______, *Stirling's Approximation*, From MathWorld—A Wolfram Web Resource. http://mathworld.wolfram.com/StirlingsApproximation.html.

A Probability distributions

In this appendix we mention a number of probability distributions that we refer to in the body of the paper. For a more detailed treatment we refer to specialized textbooks such as [5] and [7].

A.1 Binomial distribution

The binomial distribution is a discrete distribution with parameters p and N and is defined as follows:

$$\Pr(X = i) = \binom{\mathrm{N}}{i} p^i (1-p)^{\mathrm{N}-i} \text{ for } 0 \leq i \leq \mathrm{N}\ . \tag{A.1}$$

The mean value of this distribution equals Np, and the variance equals N$p(1-p)$.

A.2 Poisson distribution

The Poisson distribution is a discrete distribution with parameter λ and is defined as follows:

$$\Pr(X = i) = \frac{e^{-\lambda}\lambda^i}{i!} = \mathrm{Poisson}(i;\lambda)\ . \tag{A.2}$$

The mean value and the variance of this distribution are equal to λ. It is well known that a binomial distribution with small p can be closely approximated by a Poisson distribution with $\lambda = np$.

A.3 Normal distribution

A normal distribution is a continuous distribution. For mean $\mu(X)$ and variance σ^2 it has the following density:

$$P_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu(X))^2}{2\sigma^2}} = Z\left(\frac{x-\mu(X)}{\sigma}\right)\ . \tag{A.3}$$

If $\mu(X)$ is 0 and $\sigma = 1$, we speak of the standard normal distribution. It is well known that a binomial distribution with large n can be closely approximated by a normal distribution with mean $\mu(X) = np$ and variance $\sigma^2 = np(1-p)$.

Received 16 November, 2006; revised 26 January, 2007

Author information

Joan Daemen, STMicroelectronics, Belgium.
Email: Joan.Daemen@st.com

Vincent Rijmen, Graz University of Technology, Austria.
Email: Vincent.Rijmen@iaik.tugraz.at