Rigorous Upper Bounds on Data Complexities of Block Cipher Cryptanalysis

Subhabrata Samajder and Palash Sarkar
Applied Statistics Unit
Indian Statistical Institute
203, B.T.Road, Kolkata, India - 700108.
{subhabrata_r,palash}@isical.ac.in

Abstract

Statistical analysis of symmetric key attacks aims to obtain an expression for the data complexity which is the number of plaintext-ciphertext pairs needed to achieve the parameters of the attack. Existing statistical analyses invariably use some kind of approximation, the most common being the approximation of the distribution of a sum of random variables by a normal distribution. Such an approach leads to expressions for data complexities which are *inherently approximate*. Prior works do not provide any analysis of the error involved in such approximations. In contrast, this paper takes a rigorous approach to analysing attacks on block ciphers. In particular, no approximations are used. Expressions for upper bounds on the data complexities of several basic and advanced attacks are obtained. The analysis is based on the hypothesis testing framework. Probabilities of Type-I and Type-II errors are upper bounded using standard tail inequalities. In the cases of single linear and differential cryptanalysis, we use the Chernoff bound. For the cases of multiple linear and multiple differential cryptanalysis, Hoeffding bounds are used. This allows bounding the error probabilities and obtaining expressions for data complexities. We believe that our method provides important results for the attacks considered here and more generally, the techniques that we develop should have much wider applicability.
AMS Classifications: 94A60, 11T71, 68P25, 62P99
Keywords: block cipher, linear cryptanalysis, differential cryptanalysis, log-likelihood ratio test, hypothesis testing, Chernoff bound, Hoeffding's inequality.

1 Introduction

Statistical methods are commonly used for analysing attacks on block ciphers and more generally symmetric key ciphers. For an attack that aims at recovering a portion of the secret key, there are three basic parameters of interest. (For a distinguishing attack, the situation is a little different and we consider this later.)

1. The success probability P_S, i.e., the probability that the correct key will be recovered by the attack.
2. The advantage a such that the number of false alarms is a fraction 2^{-a} of the number of possible values of the sub-key which is the target of the attack.
3. The data complexity N which is the number of plaintext-ciphertext pairs required to achieve at least a pre-specified success probability and at least a pre-specified advantage.

The main goal of any statistical analysis of an attack is to be able to express the data complexity N in terms of P_S and a. All the known methods for doing this, however, provide only approximate expressions for N without deriving bounds on the approximation errors.

1.1 Our Contributions

The major motivation of this work is to derive rigorous upper bounds on the data complexity in terms of P_S and a. In particular, we do not use any approximation in the statistical analysis[1]. To show that this can indeed be done, we consider five basic cryptanalytic scenarios. These are single linear cryptanalysis; single differential cryptanalysis; multiple linear cryptanalysis; multiple differential cryptanalysis; and the task of distinguishing between two probability distributions. In each case, we show that it is indeed possible to obtain rigorous upper bounds on the data complexity.

The theoretical work is supported by several computations. For the block cipher SERPENT, we use the joint distribution of multiple linear approximations [15] to compute the approximate data complexity given by the analysis in [19] and also the upper bound on data complexity obtained in this work. The ratio of these two values turn out to be between 43 and 63. We further make detailed experimental comparisons of the upper bounds that we obtain to the previously best known approximate values of data complexities using simulated joint distributions. For the cases of single linear cryptanalysis, single differential cryptanalysis and distinguisher, the ratio of the upper bound to the approximate expression is around 10 or smaller. For multiple linear cryptanalysis, the ratio is between 4 and 200. These indicate that the upper bounds that we obtain are not too far away from the approximate values obtained earlier. From a practical point of view, we think it is better to use the upper bound to measure the strength of a cipher, since it may turn out that the approximate data complexities are actually underestimates.

For multiple differential cryptanalysis, however, the upper bound turns out to be much larger than the approximate estimate obtained earlier. The reason for this could be one or both of the following: the approximate value is an underestimate or, the upper bound is an overestimate. Deciding the exact reason requires more work.

The data complexity expressions that we obtain are valid for all values of the success probability P_S and advantage a. So, for example, these expressions can be evaluated to obtain data complexities for $P_S = 0.1$. Such an attack has a 10% chance of being successful and from a cryptanalytic point of view would be considered a valid attack. Similarly, even lower values of P_S can be considered. However, in earlier works on multiple linear cryptanalysis [19] and multiple differential cryptanalysis [11], for the data complexity expressions to be valid, the condition $P_S > 0.5$ is required. This is mentioned in [19] without any explanation. In [11], this condition is not even mentioned, though, evaluating the expression given there with $P_S = 0.5$ leads to meaningless values of the data complexity. The condition $P_S > 0.5$ is a consequence of using the normal approximation and we refer to [32] for more details on this issue.

The hypothesis testing based approach is used to analyse the attacks. This requires obtaining the probabilities of Type-I and Type-II errors. In the approximate analysis, the normal approximations are used to conveniently handle these probabilities. We use a different approach. The Type-I and Type-II error probabilities are essentially tail probabilities for a sum of some random variables. There are known rigorous methods for handling such tail probabilities, though, to the best of our knowledge, these methods have not been applied to the hypothesis testing setting.

For the cases of single linear and single differential cryptanalysis, it is required to bound the tail probabilities of a sum of independent Bernoulli distributed random variables. The usual method for handling this is to use the Chernoff bound. Using the Chernoff bound to upper bound the Type-I and Type-II error probabilities quite nicely leads to an expression for the data complexity.

In the cases of multiple linear or multiple differential cryptanalysis, the test statistic is no longer a sum of Bernoulli distributed random variables. As a result, the Chernoff bound does not apply. To tackle these cases, we take recourse to Hoeffding's inequality. This inequality allows us to bound the required tail probabilities to obtain upper bounds on the Type-I and Type-II error probabilities. The case of distinguisher is tackled similarly.

The importance of our work is twofold. On the one hand, we bring an amount of rigour to the statistical treatment of basic block cipher cryptanalysis. More generally, the techniques that we apply have broad appli-

[1]Note that the structural analysis of a block cipher itself usually involves approximations. Our work does not address this issue.

cability and it should be possible to tackle data complexities of other attacks using these techniques. From a practical point of view, our computations confirm that the upper bounds that we obtain are greater than the approximate data complexities reported earlier. Since it is not known whether the approximate values are under or overestimates, we think it is better to use the upper bounds.

1.2 Bounds on Data Complexity

We separately discuss the issue for key recovery attacks and distinguishing attacks.

Case of key recovery attacks: Let $N_{\min}(P_S, a)$ be the minimum amount of data required to achieve success probability at least P_S and advantage at least a, where the minimum is over all possible methods of statistical analysis. Any particular method of statistical analysis provides an expression for the data complexity that is required if the method is followed. Considering a statistical analysis as an algorithm $\mathcal{A}$, let $N_{\mathcal{A}}(P_S, a)$ denote the data complexity expression obtained using $\mathcal{A}$ to obtain success probability at least P_S and advantage at least a. Clearly $N_{\mathcal{A}}(P_S, a)$ is an upper bound on $N_{\min}(P_S, a)$. It is also a lower bound in the sense that at least $N_{\mathcal{A}}(P_S, a)$ amount of data will be required to achieve the parameters P_S and a *if the method* $\mathcal{A}$ *is followed*.

A bound $N_{\mathcal{A}}(P_S, a)$ obtained using a statistical method $\mathcal{A}$ is useful to a cryptanalyst. It tells the cryptanalyst that this amount of data is *sufficient* to attain success probability at least P_S and advantage at least a. Put another way, a upper bound tells a cryptanalyst that no more data is required to achieve the attack parameters.

From a cipher designer's point of view, a data complexity expression of the type $N_{\mathcal{A}}(P_S, a)$ is also useful. It tells the designer that if method $\mathcal{A}$ is followed, then at least $N_{\mathcal{A}}(P_S, a)$ amount of data is required to attain the parameters P_S and a. This provides useful information in quantifying the resistance of the cipher against a particular type of attack. This is particularly important if $\mathcal{A}$ is the best known method for carrying out the statistical analysis. It would be even more useful to a cipher designer to obtain $N_{\min}(P_S, a)$. Unfortunately, to the best of our knowledge, there is no work in the literature which provides this information.

Case of distinguishing attacks: A distinguishing attack proceeds as a test of hypothesis to distinguish between two different probability distributions. In this case, the data complexity is considered to be a function of the error probability which is defined to be half the sum of the probabilities of Type-I and Type-II errors. Let $N_{\min}(P_e)$ be the minimum amount of data required to ensure that the error probability is at most P_e, where the minimum is over all possible methods of statistical analysis. For a particular statistical method $\mathcal{A}$, let $N_{\mathcal{A}}(P_e)$ be the data complexity required to ensure error probability at most P_e. Similar to the case of key recovery attacks, $N_{\mathcal{A}}(P_e)$ is an upper bound on $N_{\min}(P_e)$ and at least $N_{\mathcal{A}}(P_e)$ amount of data is required to ensure error probability at most P_e if the method $\mathcal{A}$ is followed. Also, the usefulness of $N_{\mathcal{A}}(P_e)$ to a cryptanalyst and to a cipher designer remains the same as in the case of key recovery attacks.

An asymptotic expression for $N_{\min}(P_e)$ has been described in [4]. The expression is given in terms of the Chernoff information which involves taking an infimum over all real numbers in $(0, 1)$. Consequently, the resulting expression cannot be computed and [4] provides *approximations*.

To the best of our knowledge, all previously proposed statistical methods either for key recovery attacks or for distinguishing attacks use *approximations* to obtain expressions for data complexity without detailed analysis of the approximation errors. Consequently, the obtained data complexities cannot be considered to be either lower or upper bounds. The present work provides upper bounds on the data complexities and we write *rigorous* upper bound to emphasise that no approximations are used in our analysis.

1.3 How Good are the Bounds?

The bounds on data complexity that we obtain crucially depend on the bounds for tail probabilities that we use. We have used the Chernoff and the Hoeffding bounds. These are general bounds which apply to sums of

independent random variables. This leads to the question of whether better bounds are known and whether these can be applied to the current context?

The theory of large deviations is concerned with the probability of rare events and so tail probabilities can be handled by this theory. It can be shown that the tail probability is upper bounded by an exponential in N times a function called the rate function. This rate function is the Legendre transform of the moment generating function of the corresponding random variable. In theory, it is indeed possible to express the tail probabilities in terms of the rate function. However, this does not automatically provide meaningful bounds for the data complexity. There are several difficulties involved. For a more detailed discussion of these difficulties, we refer to [33].

1.4 Previous and related works

Linear Cryptanalysis: This was first proposed by Matsui in [26] to cryptanalyze the block cipher DES. Later Matsui [27] extended this idea by using two linear approximations. In an independent work, Kaliski and Robshaw [22] extended Matsui's attack involving single linear approximation to ℓ (≥ 1) linear approximations. Their result, however, was restrictive as it is required for all ℓ linear approximations to have the same plaintext and ciphertext bits though the key bits could be different.

Biryukov et al [8] further refined the idea of multiple linear cryptanalysis. The authors considered ℓ linear approximations without any assumption on their structure. This, though, also had a restriction. The analysis was valid only for ℓ independent linear approximations. Analysis under the independence assumption was separately done Junod and Vaudenay [21]. Murphy [30] argued that the independence assumption need not be valid.

In a later work, Baignères et al [2] used the log-likelihood ratio (LLR) statistic to build an optimal distinguisher between two distributions. This result did not require the independence assumption. The theme of obtaining optimal distinguishers was also investigated in [20, 3].

Selçuk in [34] proposed an order statistics based ranking methodology for analysing single linear and differential cryptanalysis. The paper provided expressions for the data complexity of these attacks. The order statistics based approach uses a well known theorem from statistics to approximate the distribution of an order statistics using the normal distribution. Consequently, the data complexities obtained in [34] are approximate. The order statistics based approach was built upon by Hermelin et al [19]. The authors combined the results obtained in [2, 30, 31, 34] to develop a multilinear cryptanalytic method without the independence assumption.

Differential cryptanalysis: This cryptanalytic method was first proposed by Biham and Shamir in [6]. It was used to successfully cryptanalyze reduced round variants (with up to 15 rounds) of DES using less than 2^{56} operations. Later in [7], the authors further improved their attack by considering several differentials having the same output difference. Over time, several variants of differential cryptanalysis have been proposed. These include higher order differentials [24], truncated differentials [23], cube attack [16], boomerang attack [36], impossible differential cryptanalysis [5] and improbable differential cryptanalysis [35].

The general approach to multiple differential cryptanalysis was considered in [10]. This work considered ℓ differentials having both unequal input and unequal output differences. The case of ℓ differentials having same input difference but different output differences was analysed in details in [11]. The order statistics based framework was used to derive an expression for the data complexity. A general study of data complexity and success probability of statistical attacks was carried out in [12].

We note that a recent work [32] performs a concrete analysis of normal approximations used in symmetric key cryptanalysis using the Berry-Esséen theorem. In particular, the work critiques the order statistics based approach advocated by Selçuk [34] and points out several shortcomings. More generally, the entire approach of using normal approximations (without consideration of the error) is questioned.

A related line of work is based on the key dependent behaviour of linear and differential characteristics [1, 9, 13, 25] and use approximations. The techniques introduced in this paper should also be applicable to this setting

and can form the basis for future work.

2 Background

In this section, we provide the background for the work. The section starts with a brief background on block cipher cryptanalysis (to the extent necessary for understanding this paper) with emphasis on linear cryptanalysis. Next we provide some details about the important log-likelihood ratio (LLR) test statistics. Appendix A provides relevant details of tail probability inequalities, specifically the Chernoff bound for Poisson trials and the Hoeffding bounds.

2.1 Background for Block Cipher Cryptanalysis

The description of block cipher cryptanalysis given here is tailored towards linear cryptanalysis. Differential cryptanalysis is separately considered later.

A block cipher is a function $E : \{0,1\}^k \times \{0,1\}^n \to \{0,1\}^n$ such that for each $K \in \{0,1\}^k$, the function $E_K(\cdot) \stackrel{\Delta}{=} E(K,\cdot)$ is a bijection from $\{0,1\}^n$ to itself. Here K is the secret key. The n-bit input to the block cipher is called the plaintext and the n-bit output of the block cipher is called the ciphertext.

Practical constructions of block ciphers have an iterated structure consisting of several rounds. Each round consists of applying a round function parameterised by a round key. The round functions are bijections of $\{0,1\}^n$. An expansion function, called the key scheduling algorithm, is applied to the secret key to obtain round keys. Let the round keys be $k^{(0)}, k^{(1)}, \ldots$, and denote the round functions as $R^{(0)}_{k^{(0)}}, R^{(1)}_{k^{(1)}}, \ldots$. Further, denote by $K^{(i)}$ the concatenation of the first i round keys, i.e., $K^{(i)} = k^{(0)}||\cdots||k^{(i-1)}$; and let $E^{(i)}_{K^{(i)}}$ denote the composition of the first i round functions, i.e.,

$$\begin{aligned} E^{(0)}_{K^{(0)}} &= R^{(0)}_{k^{(0)}}; \\ E^{(i)}_{K^{(i)}} &= R^{(i-1)}_{k^{(i-1)}} \circ \cdots \circ R^{(0)}_{k^{(0)}} = R^{(i-1)}_{k^{(i-1)}} \circ E^{(i-1)}_{K^{(i-1)}}, \quad i \geq 1. \end{aligned}$$

A block cipher may have many rounds and a reduced round cryptanalysis may target only a few of these rounds. Suppose that an attack targets $r+1$ rounds. For a plaintext P, let C be the output after $r+1$ rounds and B be the output after r rounds. So, $B = E^{(r)}_{K^{(r)}}(P)$ and $C = R^{(r)}_{k^{(r)}}(B)$.

Relations between plaintext and the input to the last round: The basic step in block cipher cryptanalysis is to perform a detailed analysis of the structure of a block cipher. Such a study reveals one or more possible relations between the following quantities: a plaintext P; the input to the last round B; and possibly $K^{(r)}$. Such relations can be in the form of a linear function or in the form of a differential as we explain later. Usually, such a relation holds only with some probability. The probability is taken over the uniform random choice of P. If there are more than one relation, then it is required to consider the joint distribution of the probabilities that these relations hold. Obtaining relations and their possibly joint distribution is a non-trivial task which requires a great deal of experience and ingenuity. These relations form the bedrock on which a statistical analysis of an attack can be carried out.

Target sub-key: A single relation between P and B will usually involve only a subset of the bits of B. If several (or multiple) relations between P and B are known, it is required to consider the subset of the bits of B which cover all the relations. Obtaining these bits from C will require a partial decryption of the last round. Such a partial decryption will involve a subset of the bits of secret key (or of the last round key). Obtaining the correct values of these key bits is the goal of the attack and these bits will be called the target sub-key. The size of the target sub-key in bits will be denoted by m. So, m key bits are sufficient to partially decrypt C to obtain

the bits of B which are involved in any of the relation between P and B. There are 2^m possible choices of the target sub-key bits out of which one is correct and all others are incorrect. The goal is to pick out the correct key.

Setting of an attack: Suppose there are N plaintext-ciphertext pairs (P_j, C_j), $j = 1, \ldots, N$ which have been generated using the correct key and are available. For each choice κ of the last round key bits, it is possible to invert C_j to obtain the relevant bits of $B_{\kappa,j}$. The relevant bits are those which are required to evaluate the relations discovered in the prior analysis of the block cipher. Note that $B_{\kappa,j}$ depends on κ even though C_j may not. If κ is the correct choice for the target sub-key, then C_j indeed depends on κ, otherwise C_j has no relation to κ.

Given P_j and the relevant bits of $B_{\kappa,j}$ it is possible to evaluate all the known relations. From the results of these evaluations, a test statistic T_κ is defined. Since there are a total of 2^m possible values of κ, there are also 2^m random variables T_κ. These random variables are assumed to be independent and the distributions of these random variables depend on whether κ is correct or incorrect. It is also assumed that the distributions of T_κ for incorrect κ are identical. This assumption was considered in [17]. For an attack to be possible, it is required to obtain the two possible distributions of T_κ – one when κ is the correct choice and the other when κ is an incorrect choice.

2.2 Linear Cryptanalysis

Assume that the analysis of the structure of the block cipher provides $\ell \geq 1$ linear approximations. These are given by masks $\Gamma_P^{(i)}, \Gamma_B^{(i)}$ and $\Gamma_K^{(i)}$, for $i = 1, \ldots, \ell$. The subscript P denotes plaintext mask; the subscript B denotes mask after r rounds; and the subscript K denotes the mask for $K^{(r)}$. So, $\Gamma_P^{(i)}$ and $\Gamma_B^{(i)}$ are in $\{0,1\}^n$ and $\Gamma_K^{(i)}$ is in $\{0,1\}^{nr}$. If $\ell > 1$, then the attack is called multiple linear cryptanalysis and if $\ell = 1$, we will call the attack single linear cryptanalysis, or simply, linear cryptanalysis. Define

$$L_i = \langle \Gamma_P^{(i)}, P \rangle \oplus \langle \Gamma_B^{(i)}, B \rangle; \text{ for } i = 1, \ldots, \ell. \tag{1}$$

Inner key bits: For a fixed but unknown key $K^{(r)}$, the quantity $z_i = \langle \Gamma_K^{(i)}, K^{(r)} \rangle$ is a single unknown bit. Denote by $z = (z_1, \ldots, z_\ell)$ the collection of the ℓ bits arising in this manner. The key masks $\Gamma_K^{(1)}, \ldots, \Gamma_K^{(\ell)}$ are known. So, z is determined only by the unknown key $K^{(r)}$. The bits represented by z are called the inner key bits. The key $K^{(r)}$ is unknown but, fixed and so there is no randomness in $K^{(r)}$. Correspondingly, z is also unknown but fixed and there is no randomness in z.

Consider a uniform random choice of P. The round functions are deterministic bijections and so the uniform distribution on P induces a uniform distribution on B. Each L_i is a random variable which can take the values 0 or 1. The randomness of L_i arises solely from the randomness of P. Define the random variable X to be the following:

$$X \quad = \quad (L_1, \ldots, L_\ell). \tag{2}$$

So, X is distributed over $\{0,1\}^\ell$ and its distribution is determined by the distribution of the L_i's which in turn is determined by the distribution of P.

A single linear approximation is of the form

$$L_i = \langle \Gamma_K^{(i)}, K^{(r)} \rangle = z_i. \tag{3}$$

Note that we are not assuming any randomness over the key $K^{(r)}$ and the bits z_i's have no randomness even though they are unknown. So, the distribution of $L_i \oplus z_i$ is determined completely by the distribution of L_i.

Joint distribution parameterised by inner key bits: A linear approximation of the type given by (3) holds with some probability over the uniform random choice of P. The random variables $L_1, \ldots, L_\ell$ are not necessarily independent. The joint distribution of these variables is given as follows: For $z = (z_1, \ldots, z_\ell)$, and $\eta = (\eta_1, \ldots, \eta_\ell) \in \{0,1\}^\ell$, define

$$p_z(\eta) \quad = \quad \Pr[L_1 = \eta_1 \oplus z_1, \ldots, L_\ell = \eta_\ell \oplus z_\ell] = \frac{1}{2^\ell} + \epsilon_\eta(z) \tag{4}$$

where $-1/2^\ell \leq \epsilon_\eta(z) \leq 1 - 1/2^\ell$.

The vector $\tilde{p}_z \stackrel{\Delta}{=} (p_z(0), \ldots, p_z(2^\ell - 1))$ is a probability distribution, where the integers $\{0, \ldots, 2^\ell - 1\}$ are identified with the set $\{0,1\}^\ell$. For each choice of z, we obtain a different distribution. These distributions are, however, related to each other. Suppose $z' = z \oplus \beta$ for some $\beta \in \{0,1\}^\ell$. Then it is easy to verify that $\epsilon_\eta(z') = \epsilon_{\eta \oplus \beta}(z)$. It follows that

$$p_{z \oplus \beta}(\eta) \quad = \quad p_z(\eta \oplus \beta). \tag{5}$$

Let $\tilde{p}$ be the probability distribution $\tilde{p} \stackrel{\Delta}{=} \tilde{p}_{0^\ell}$ and under the usual identification of $\{0,1\}^\ell$ and the integers in $\{0, \ldots, 2^\ell - 1\}$, write

$$\tilde{p} \quad = \quad (p_0, \ldots, p_{2^\ell - 1}) \tag{6}$$

so that for $\eta \in \{0,1\}^\ell$, $p_\eta \stackrel{\Delta}{=} p(\eta) = 1/2^\ell + \epsilon_\eta$.

Notation: There are N plaintext-ciphertext pairs (P_j, C_j) for $j = 1, \ldots, N$. For a choice κ of the target subkey, the C_j's are partially decrypted to obtain the relevant bits of $B_{\kappa,j}$. For $\kappa \in \{0, \ldots, 2^m - 1\}$, $j = 1, \ldots, N$ and $i = 1, \ldots, \ell$, define

$$L_{\kappa,j,i} \quad = \quad \langle \Gamma_P^{(i)}, P_j \rangle \oplus \langle \Gamma_B^{(i)}, B_{\kappa,j} \rangle; \tag{7}$$

$$X_{\kappa,j} \quad = \quad (L_{\kappa,j,1}, \ldots, L_{\kappa,j,\ell}). \tag{8}$$

2.3 LLR Statistics

Let $\tilde{p} = (p_0, \ldots, p_{\nu-1})$ and $\tilde{q} = (q_0, \ldots, q_{\nu-1})$ be two probability distributions over a finite alphabet of size $\nu > 0$. The Kullback-Leibler divergence between $\tilde{p}$ and $\tilde{q}$ is defined as follows.

$$D(\tilde{p} || \tilde{q}) \quad = \quad \sum_{\eta=0}^{\nu-1} p_\eta \ln(p_\eta / q_\eta). \tag{9}$$

The problem of distinguishing between the two distributions is the following. Let $X_1, \ldots, X_N$ be a sequence of independent and identically distributed random variables taking values from the set $\{0, \ldots, \nu - 1\}$. It is known that all the X_i's follow one of the distributions $\tilde{p}$ or $\tilde{q}$, but, which one is not known.

The goal is to formulate a test of hypothesis to distinguish between these two distributions. This test takes the form where the null hypothesis "H_0: the distribution is $\tilde{p}$" is tested against the alternate hypothesis "H_1: the distribution is $\tilde{q}$".

Note that $\tilde{p}$ is a probability distribution on $\{0, \ldots, \nu - 1\}$ and the probability at a point $\eta \in \{0, \ldots, \nu - 1\}$ is written as p_η. For $1 \leq j \leq N$, the random variable X_j takes values from the set $\{0, \ldots, \nu - 1\}$. So, the derived random variable p_{X_j} is well defined. One may set $W_j = p_{X_j}$. The possible values of W_j are $p_0, p_1, \ldots, p_{\nu-1}$. If X_j follows $\tilde{p}$, then for $\eta \in \{0, \ldots, \nu - 1\}$, $\Pr[W_j = p_\eta] = \Pr[X_j = \eta] = p_\eta$; if X_j follows another distribution $\tilde{q}$, then $\Pr[W_j = p_\eta] = \Pr[X_j = \eta] = q_\eta$.

For $j = 1, \ldots, N$, define

$$Y_j \;=\; \ln\left(p_{X_j}/q_{X_j}\right). \tag{10}$$

Let μ_0 and σ_0^2 be the mean and variance of Y_j under hypothesis H_0. Similarly, let μ_1 and σ_1^2 be the mean and variance of Y_j under hypothesis H_1. Then the expressions for μ_0, μ_1, σ_0^2 and σ_1^2 can be computed to be the following.

$$\left.\begin{array}{ll} \mu_0 = D(\tilde{p} \,||\, \tilde{q}); & \mu_1 = -D(\tilde{q} \,||\, \tilde{p}); \\ \sigma_0^2 = \sum_{\eta=0}^{\nu-1} p(\eta)\left(\ln\left(\frac{p(\eta)}{q(\eta)}\right)\right)^2 - \mu_0^2; & \sigma_1^2 = \sum_{\eta=0}^{\nu-1} q(\eta)\left(\ln\left(\frac{q(\eta)}{p(\eta)}\right)\right)^2 - \mu_1^2. \end{array}\right\} \tag{11}$$

The LLR random variable is defined to be the following.

$$\text{LLR} \;=\; \sum_{j=1}^{N} Y_j = \sum_{j=1}^{N} \ln\left(p_{X_j}/q_{X_j}\right) = \sum_{\eta=0}^{\nu-1} Q_\eta \ln(p_\eta/q_\eta). \tag{12}$$

Here $Q_\eta = \#\{j : X_j = \eta\}$. Following the method described in [2], it is possible to define a test of hypothesis to distinguish between the two distributions $\tilde{p}$ and $\tilde{q}$ using *approximately*

$$N \;=\; \left(\frac{(\sigma_0 + \sigma_1)\Phi^{-1}(1 - P_e)}{D(\tilde{p}||\tilde{q}) + D(\tilde{q}||\tilde{p})}\right)^2 \tag{13}$$

plaintext-ciphertext pairs, where P_e is half the sum of the probabilities of Type-I and Type-II errors and Φ is the standard normal distribution function. More details are given in the appendix.

3 Single Linear Approximation

In this section, we consider the case of a single linear approximation. Let $P_1, \ldots, P_N$ be N independent and uniformly distributed plaintexts. For simplicity, in this section, we will write L instead of L_1 and $L_{\kappa,j}$ instead of $L_{\kappa,j,1}$. Since there is a single linear approximation, the joint distribution $\tilde{p}$ reduces to simply a probability value $p = \Pr[L_{\kappa,j} = 0] \neq 1/2$ when κ is the correct choice. For an incorrect choice of κ, it is conventional to assume that $\Pr[L_{\kappa,j} = 0] = 1/2$. For the correct choice of κ, $L_{\kappa,j}$ follows $\mathsf{Bernoulli}(p)$ for all j, where $p = 1/2 + \epsilon = 1/2 \pm |\epsilon|$. The appropriate sign is determined by the correct value of the inner key bit z^* and we can write $p = 1/2 + (-1)^{z^*}|\epsilon|$. Under the wrong key hypothesis, for an incorrect choice of κ, $L_{\kappa,j}$ follows $\mathsf{Bernoulli}(1/2)$ for all j.

Let $c = 2(p - 1/2) = 2(-1)^{z^*}|\epsilon|$ and define $\mu_0 = p = (1 + c)/2$ and $\mu_1 = 1/2$. The hypothesis testing framework will be used. The test statistics is $T_\kappa = |X_\kappa - N\mu_1|$ where $X_\kappa = \sum_{j=1}^{N} L_{\kappa,j}$. Consider the following test of hypothesis:

Hypothesis Test-1 (single linear cryptanalysis):
H_0: "κ is correct" versus H_1: "κ is incorrect."
Decision rule: Reject H_0 if $T_\kappa \leq t$.

Proposition 1. *Let $0 < \alpha, \beta < 1$. In Hypothesis Test-1, the value of t can be chosen such that for*

$$N \;\geq\; \frac{2\left(\sqrt{\ln\left(\frac{2}{\beta}\right)} + \sqrt{3\,(1 + |c|)\ln\left(\frac{1}{\alpha}\right)}\right)^2}{c^2} \tag{14}$$

the probabilities of the Type-I and Type-II errors are upper bounded by α and β respectively.

Proof. The requirement is to show the bound on N given the values of α and β. As is usual in the hypothesis testing framework, we will obtain two equations, one relating α, t and N and another relating β, t and N. Eliminating t variable between these two equations will provide the expression for N in terms of α and β.

Note that under H_0, $E[X_\kappa] = N\mu_0$ and under H_1, $E[X_\kappa] = N\mu_1$. Define $\delta_0 = (|\mu_0 - \mu_1| - t/N)/\mu_0$.

The decision threshold t will be chosen to satisfy $0 < t/N < |\mu_0 - \mu_1|$. For t in this range, we have $0 < \delta_0 < |\mu_0 - \mu_1|/\mu_0 < 1$. So, it is possible to apply the Chernoff bound (specifically (40) and (41) of Theorem 7) with this δ_0.

First suppose $\mu_0 > \mu_1$. Then $\delta_0 = (\mu_0 - \mu_1 - t/N)/\mu_0$ and so $(1-\delta_0)\mu_0 = \mu_1 + t/N$.

$$\begin{aligned}
\Pr[\text{Type-I Error}] &= \Pr[T_\kappa \le t | H_0 \text{ holds}] = \Pr[-t \le X_\kappa - N\mu_1 \le t | H_0 \text{ holds}] \\
&\le \Pr[X_\kappa - N\mu_1 \le t | H_0 \text{ holds}] = \Pr[X_\kappa \le t + N\mu_1 | H_0 \text{ holds}] \\
&= \Pr[X_\kappa \le (1-\delta_0)N\mu_0 | H_0 \text{ holds}] \\
&\le \exp\left(-N\mu_0\delta_0^2/2\right) \le \exp\left(-N\mu_0\delta_0^2/3\right).
\end{aligned}$$

Recall that X_κ is the sum $L_{\kappa,1} + \cdots + L_{\kappa,N}$ and under H_0, each $L_{\kappa,j}$ follows $\mathsf{Bernoulli}(p)$. So, the last step of the above calculation follows from the Chernoff bound (Equation (41) in the appendix).

Now suppose that $\mu_1 > \mu_0$. (Note that since $p \neq 1/2$, the case $\mu_0 = \mu_1$ does not occur.) Then $\delta_0 = (\mu_1 - \mu_0 - t/N)/\mu_0$ and so $(1+\delta_0)\mu_0 = \mu_1 - t/N$. In this case,

$$\begin{aligned}
\Pr[\text{Type-I Error}] &= \Pr[T_\kappa \le t | H_0 \text{ holds}] = \Pr[-t \le X_\kappa - N\mu_1 \le t | H_0 \text{ holds}] \\
&\le \Pr[X_\kappa \ge (1+\delta_0)N\mu_0 | H_0 \text{ holds}] \le \exp\left(-N\mu_0\delta_0^2/3\right)
\end{aligned}$$

The last step follows from the Chernoff bound (Equation (40) in the appendix). The actual bound used in this case is different from that used for the case of $\mu_0 > \mu_1$.

A relation involving α and N is obtained by enforcing

$$\alpha = \exp\left(-N\mu_0\delta_0^2/3\right).$$

This shows that $\Pr[\text{Type-I Error}] \le \alpha$ irrespective of the values of μ_0 and μ_1. From the expressions for α and δ_0 and using the fact that $0 < t/N < |\mu_0 - \mu_1|$ we obtain

$$t = N \times |\mu_0 - \mu_1| - \sqrt{3N\mu_0 \ln(1/\alpha)}. \tag{15}$$

The probability of Type-II error is given by,

$$\begin{aligned}
\Pr[\text{Type-II Error}] &= \Pr\left[T_\kappa > t \,|H_1 \text{ holds}\right] = \Pr\left[|X_\kappa - N\mu_1| > t \,|H_1 \text{ holds}\right] \\
&= \Pr\left[X_\kappa > t + N\mu_1 \,|H_1 \text{ holds}\right] + \Pr\left[X_\kappa < -t + N\mu_1 \,|H_1 \text{ holds}\right].
\end{aligned}$$

Let

$$\delta_1 = t/(N\mu_1) \tag{16}$$

so that $t/N + \mu_1 = (1+\delta_1)\mu_1$ and $-t/N + \mu_1 = (1-\delta_1)\mu_1$. The analysis of Type-I error shows that $0 < t/N < |\mu_0 - \mu_1|$ from which it follows that $0 < \delta_1 < 1$. Using (42) and (43) of Theorem 7, we obtain

$$\Pr[\text{Type-II Error}] \le 2\exp\left(-N\mu_1\delta_1^2\right).$$

A relation involving β and N is obtained by enforcing

$$\beta = 2\exp\left(-N\mu_1\delta_1^2\right) = 2\exp\left(-t^2/(N\mu_1)\right).$$

This shows that Pr[Type-II Error] $\leq \beta$. Solving for t in terms of β and using $0 < t/N < |\mu_0 - \mu_1|$ yields

$$t = \sqrt{N\mu_1 \ln\left(\frac{2}{\beta}\right)}. \tag{17}$$

Eliminating t from (15) and (17), we obtain

$$N = \frac{2\left(\sqrt{\ln\left(\frac{2}{\beta}\right)} + \sqrt{3\,(1+c)\ln\left(\frac{1}{\alpha}\right)}\right)^2}{c^2}. \tag{18}$$

The two expressions for t given by (15) and (17) combined with the condition $0 < t/N < |\mu_0 - \mu_1|$ gives rise to two lower bounds on N. It is easy to check that the expression for N given by (18) satisfies both these lower bounds.

Recall that $c = 2(-1)^{z^*}|\epsilon|$. So, depending on the value of z^*, (18) provides two expressions for N, with the expression for $z^* = 1$ being (slightly) greater than the expression for $z^* = 0$. Taking $z^* = 1$ provides the expression on the right hand side of (14). So, for any N greater than this value, the probabilities of Type-I and Type-II errors are upper bounded by α and β respectively.

□

4 Multiple Linear Cryptanalysis

We assume the setting and notation explained in Sections 2.1 and 2.2. There are $\ell \geq 1$ linear approximations, κ denotes the choice of the target sub-key and z denotes the choice of the inner key bits. There are N plaintext-ciphertext pairs $(P_1, C_1), \ldots, (P_N, C_N)$. For a choice κ of the target sub-key; a choice $z = (z_1, \ldots, z_\ell)$ of the inner key bit; $j \in \{1, \ldots, N\}$; and $1 \leq i \leq \ell$, define

$$L_{\kappa,j,i} = \langle \Gamma_P^{(i)}, P_j \rangle \oplus \langle \Gamma_P^{(i)}, B_{\kappa,j} \rangle; \quad X_{\kappa,j} = (L_{\kappa,j,1}, L_{\kappa,j,2}, \ldots, L_{\kappa,j,\ell});$$

$$Y_{\kappa,z,j} = \ln\left(p_z(X_{\kappa,j})/2^{-\ell}\right) = \ln\left(2^\ell p_z(X_{\kappa,j})\right).$$

Suppose z is the correct choice of the inner key bits. For a particular choice of κ, the random variables $X_{\kappa,z,1}, \ldots, X_{\kappa,z,N}$ are independent and these variables follow either the distribution $\tilde{p}_z$ or the distribution $\tilde{q} = (2^{-\ell}, \ldots, 2^{-\ell})$ according as κ is the correct choice or κ is an incorrect choice.

The test statistic is defined to be

$$\mathrm{LLR}_{\kappa,z} \;=\; Y_{\kappa,z,1} + \cdots + Y_{\kappa,z,N} = \sum_{\eta \in \{0,1\}^\ell} Q_{\kappa,\eta} \ln(2^\ell p_z(\eta)) \tag{19}$$

where $Q_{\kappa,\eta} = \#\{j : X_{\kappa,j} = \eta\}$. Consider the following test of hypothesis:

Hypothesis Test-2 (multiple linear cryptanalysis):
H_0: "κ is correct" versus H_1: "κ is incorrect."
Decision rule:
Case $\mu_0 > \mu_1$: Reject H_0 if $\mathrm{LLR}_{\kappa,z} \leq t, \forall z \in \{0,1\}^\ell$; where $t \in (N\mu_1, N\mu_0)$;
Case $\mu_0 < \mu_1$: Reject H_0 if $\mathrm{LLR}_{\kappa,z} \geq t, \forall z \in \{0,1\}^\ell$; where $t \in (N\mu_0, N\mu_1)$.

Proposition 2. *Let $0 < \alpha, \beta < 1$. In Hypothesis Test-2, it is possible to choose t such that for*

$$N \geq \frac{\upsilon^2\{\sqrt{\ln(2^\ell/\beta)} + \sqrt{\ln(1/\alpha)}\}^2}{2(D(\tilde{p} \,||\, \tilde{q}) + D(\tilde{q} \,||\, \tilde{p}))^2}. \tag{20}$$

the probabilities of the Type-I and Type-II errors are upper bounded by α and β respectively. Here

$$\upsilon = \max_{\eta \in \{0,1\}^\ell} \ln(2^\ell p_\eta) - \min_{\eta \in \{0,1\}^\ell} \ln(2^\ell p_\eta) = \ln\left(\frac{\max_\eta p_\eta}{\min_\eta p_\eta}\right).$$

Proof. Under H_0, each $Y_{\kappa,z,j}$ has mean $\mu_0 = D(\tilde{p}_z||\tilde{q})$ while under H_1, each $Y_{\kappa,z,j}$ has mean $\mu_1 = -D(\tilde{q}||\tilde{p}_z)$. It is not difficult to prove that μ_0 and μ_1 have the same value for all z (see [32] for a proof) and so we simply write $\mu_0 = D(\tilde{p}||\tilde{q})$ and $\mu_1 = -D(\tilde{q}||\tilde{p})$, where $\tilde{p} = (p_0, \ldots, p_{2^\ell-1})$ as defined in (6).

We now proceed to analyse the probabilities of Type-I and Type-II errors and derive expressions for the data complexity. While doing this, we avoid using normal approximations. We use Hoeffding's inequalities (see Appendix A.2) to bound the probabilities of the two types of errors.

Recall that for a fixed value of κ and z, the random variables $\mathrm{LLR}_{\kappa,z,j}$ $(j = 1, \ldots, N)$ are independently and identically distributed with each random variables taking values from the set $\{\ln(2^\ell p_0), \ldots, \ln(2^\ell p_{2^\ell-1})\}$. This, implies that for a fixed value of κ and z,

$$\upsilon_{\min} = \min_{\eta \in \{0,1\}^\ell} \ln(2^\ell p_\eta) \leq \mathrm{LLR}_{\kappa,z,j} \leq \max_{\eta \in \{0,1\}^\ell} \ln(2^\ell p_\eta) = \upsilon_{\max};$$

for all $j = 1, \ldots, N$. Let, $\upsilon = \upsilon_{\max} - \upsilon_{\min}$. Therefore one can use Hoeffding bounds (see Appendix A) on the sum of independent and identically distributed random variables $\mathrm{LLR}_{\kappa,z} = \sum_{j=1}^N \mathrm{LLR}_{\kappa,z,j}$; where $D_N = \sum_{j=1}^N (\upsilon_{\max} - \upsilon_{\min})^2 = N\upsilon^2$.

We now turn to bounding the error probabilities and obtaining expression for the data complexity. This is done separately for the two cases depending on the relative values of μ_0 and μ_1. Let z^* be the correct choice of the inner key bits.

Case $\mu_0 > \mu_1$: In this case for $t \in (N\mu_1, N\mu_0)$, to be determined later, the null hypothesis is rejected if $\mathrm{LLR}_{\kappa,z} \leq t$ for all $z \in \{0,1\}^\ell$. Then,

$$\begin{aligned}
\Pr[\text{Type-I Error}] &= \Pr[\mathrm{LLR}_{\kappa,z} \leq t \text{ for all } z | H_0 \text{ holds}] \leq \Pr[\mathrm{LLR}_{\kappa,z^*} \leq t | H_0 \text{ holds}] \\
&= \Pr[\mathrm{LLR}_{\kappa,z^*} N\mu_0 \leq -(N\mu_0 - t) | H_0 \text{ holds}] \leq \exp\left(-\frac{2(N\mu_0 - t)^2}{N\upsilon^2}\right).
\end{aligned}$$

The last inequality follows from the Hoeffding's inequality (see (45) of the appendix). Similarly, the probability of Type-II error is computed as follows.

$$\begin{aligned}
\Pr[\text{Type-II Error}] &= \Pr[\mathrm{LLR}_{\kappa,z} > t \text{ for some } z | H_1 \text{ holds}] \leq \sum_{z \in \{0,1\}^\ell} \Pr[\mathrm{LLR}_{\kappa,z} > t | H_1 \text{ holds}] \\
&= 2^\ell \cdot \Pr[\mathrm{LLR}_{\kappa,z} > t | H_1 \text{ holds}] = 2^\ell \cdot \Pr[\mathrm{LLR}_{\kappa,z} - N\mu_1 > t - N\mu_1 \mid H_1 \text{ holds}] \\
&\leq 2^\ell \exp\left(-\frac{2(t - N\mu_1)^2}{N\upsilon^2}\right).
\end{aligned}$$

The last inequality follows from the Hoeffding's inequality (see (44) of the appendix). Define

$$\alpha = \exp\left(-\frac{2(N\mu_0 - t)^2}{N\upsilon^2}\right); \quad \beta = 2^\ell \exp\left(-\frac{2(t - N\mu_1)^2}{2N\upsilon^2}\right).$$

Then Pr[Type-I Error] $\leq \alpha$ and Pr[Type-II Error] $\leq \beta$. The expression for α gives two values for t. Using the upper bound on t, i.e., $t < N\mu_0$, the expression for t has to be

$$\sqrt{2}t = \sqrt{2}N\mu_0 - \upsilon\sqrt{N\ln(1/\alpha)}. \tag{21}$$

The lower bound on t, i.e., $N\mu_1 < t$ provides the following lower bound on N.

$$N > \frac{\upsilon^2 \ln(1/\alpha)}{2(\mu_0 - \mu_1)^2}. \tag{22}$$

Similarly, the expression for β leads to two values for t and again using the range for t, we obtain

$$\sqrt{2}t = \sqrt{2}N\mu_1 + \upsilon\sqrt{N\ln(2^\ell/\beta)} \tag{23}$$

and

$$N > \frac{\upsilon^2 \ln(2^\ell/\beta)}{2(\mu_0 - \mu_1)^2}. \tag{24}$$

From (21) and (23), we obtain the expression on the right hand side of (20). The expression for N given by (20) satisfies the bounds in (22) and (24).

Case $\mu_0 < \mu_1$: In this case for $t \in (N\mu_0, N\mu_1)$, to be determined later, the null hypothesis is rejected if $\mathrm{LLR}_{\kappa,z} > t$ for all $z \in \{0,1\}^\ell$. Then,

$$\begin{aligned}\Pr[\text{Type-I Error}] &= \Pr[\mathrm{LLR}_{\kappa,z} \geq t \text{ for all } z | H_0 \text{ holds}] \leq \Pr[\mathrm{LLR}_{\kappa,z^*} \geq t | H_0 \text{ holds}] \\ &= \Pr[\mathrm{LLR}_{\kappa,z^*} - N\mu_0 \geq t - N\mu_0 | H_0 \text{ holds}] \leq \exp\left(-\frac{2(t-N\mu_0)^2}{N\upsilon^2}\right).\end{aligned}$$

The last inequality follows from the Hoeffding's inequality (see (44) of the appendix). Similarly, the probability of Type-II error is computed as follows.

$$\begin{aligned}\Pr[\text{Type-II Error}] &= \Pr[\mathrm{LLR}_{\kappa,z} < t \text{ for some } z | H_1 \text{ holds}] \leq \sum_{z\in\{0,1\}^\ell} \Pr[\mathrm{LLR}_{\kappa,z} < t | H_1 \text{ holds}] \\ &= 2^\ell \cdot \Pr[\mathrm{LLR}_{\kappa,z} - N\mu_1 < -(N\mu_1 - t) | H_1 \text{ holds}] \leq 2^\ell \exp\left(-\frac{2(t-N\mu_1)^2}{N\upsilon^2}\right).\end{aligned}$$

The last inequality follows from the Hoeffding's inequality (see (45) of the appendix). Further analysis of this case in the manner similar to that done for $\mu_1 < \mu_0$ shows that the expression for N in this case is also given by (20). □

Algorithmically, the test is performed in the following manner. Consider $\mu_0 > \mu_1$, the case for $\mu_0 < \mu_1$ being similar. Initialise a set $\mathcal{L}$ to be the empty set. For each κ and z, if $\mathrm{LLR}_{\kappa,z} > t$, then $\mathcal{L} \leftarrow \mathcal{L} \cup \{\kappa\}$. At the end, $\mathcal{L}$ contains the list of candidate keys.

We consider the time required for computing $\mathrm{LLR}_{\kappa,z}$ for all values of κ and z. For a fixed κ, the values of $Q_{\kappa,\eta}$ for all $\eta \in \{0,1\}^\ell$ can be computed in $O(\ell N)$ time. Given these $Q_{\kappa,\eta}$'s, for any z, the value of $\mathrm{LLR}_{\kappa,z}$ can be computed in $O(2^\ell)$ additional time; for a fixed κ, given the values of $Q_{\kappa,\eta}$'s, the values of $\mathrm{LLR}_{\kappa,z}$ for all $z \in \{0,1\}^\ell$ can be computed in $O(2^{2\ell})$ additional time. Thus, the values of $\mathrm{LLR}_{\kappa,z}$ for all $\kappa \in \{0,1\}^m$ and for all $z \in \{0,1\}^\ell$ can be computed in $O(2^m(\ell N + 2^{2\ell}))$ time.

5 Single Differential Cryptanalysis

Let the n-bit strings $\delta_0, \delta_1, \ldots, \delta_r$ with $\delta_0 \neq 0$, be the input differences to the rounds of an $r+1$-round block cipher. Let P be a plaintext and set $P' = P \oplus \delta_0$. Let, $B^{(0)} = P, B^{(1)}, \ldots, B^{(r)}$ denote the inputs to round number $0, \ldots, r$ respectively, i.e., $B^{(i+1)} = R^{(i)}_{k^{(i)}}(B^{(i)})$ corresponding to the plaintext P. Further, let $B^{(0)\prime} = P', B^{(1)\prime}, \ldots, B^{(r)\prime}$ be the inputs to round numbers $0, \ldots, r$ respectively corresponding to the plaintext P'. Then $A = \wedge_{i=0}^{r}(B^{(i)} \oplus B^{(i)\prime} = \delta_i)$ denotes the event that the differential characteristic $\delta_0 \rightarrow \delta_1 \rightarrow \cdots \rightarrow \delta_r$ occurs. Suppose that for the correct key K, $\Pr[A] = p$. Notice that as in the case of linear cryptanalysis the randomness also comes from the uniform random choice of P.

As in Section 2.2, we assume that guessing m bits of the key allows the partial decryption of C to obtain $B^{(r)}$. These m bits will constitute the target sub-key and the goal will be to obtain the correct value of the sub-key. Further, as done previously, we will denote a choice of the target sub-key by κ.

Let, D denote the event $B^{(r)} \oplus B^{(r)\prime} = \delta_r$. Further, let $\Pr[D|\overline{A}] = p'$ and $p_0 = p + (1-p)p'$. Then for the correct choice κ of the target sub-key $\Pr[D] = p_0$. Since δ_0 is not the zero string, $P \neq P'$. This further implies that $B^{(i)} \neq B^{(i)\prime}$ for $i = 1, \ldots, r$ since each round function is a bijection. For incorrect choices of κ, it is assumed that $B^{(r)}$ and $B^{(r)\prime}$ correspond to uniform sampling without replacement of two n-bit strings from $\{0,1\}^n$. Hence, for incorrect of κ, $\Pr[D] = 1/(2^n - 1)$. Let $p_w = 1/(2^n - 1)$. In general $p_0 > p_w$ and we will be proceeding with this assumption. The analysis for the case $p_0 < p_w$ is similar.

Consider N plaintext pairs $(P_1, P_1'), \ldots, (P_N, P_N')$ with $P_j \oplus P_j' = \delta_0$ and their corresponding ciphertexts $(C_1, C_1'), \ldots, (C_N, C_N')$. For a choice κ of the target sub-key, the attacker obtains $(B^{(r)}_{\kappa,1}, B^{(r)\prime}_{\kappa,1}), \ldots, (B^{(r)}_{\kappa,N}, B^{(r)\prime}_{\kappa,N})$ by partially decrypting $(C_1, C_1'), \ldots, (C_N, C_N')$ respectively. So, for $j = 1, \ldots, N$, it is possible to determine whether the condition $B^{(r)}_{\kappa,j} \oplus B^{(r)\prime}_{\kappa,j} = \delta_r$ holds.

For a choice κ of the target sub-key, define the binary valued random variables $W_{\kappa,1}, \ldots, W_{\kappa,N}$ as follows: $W_{\kappa,j} = 1$ if $B^{(r)}_{\kappa,j} \oplus B^{(r)}_{\kappa,j} = \delta_r$; and $W_{\kappa,j} = 0$ otherwise. If κ is the correct choice, then $\Pr[W_{\kappa,j} = 1] = p_0$ and if κ is an incorrect choice, then $\Pr[W_{\kappa,j} = 1] = p_w$ for all j.

The test statistics is $T_\kappa = |X_\kappa - \mu_1|$. Consider the following test of hypothesis:

Hypothesis Test-3 (single differential cryptanalysis):

H_0: "κ is correct" versus H_1: "κ is incorrect."

Decision rule: Reject H_0 if $T_\kappa \leq t$.

Proposition 3. *Let $0 < \alpha, \beta < 1$. In Hypothesis Test-3 it is possible to choose t such that for*

$$N \geq \frac{3\left(\sqrt{p_0 \ln(1/\alpha)} + \sqrt{p_w \ln(2/\beta)}\right)^2}{(p_0 - p_w)^2}. \tag{25}$$

the probabilities of the Type-I and Type-II errors are upper bounded by α and β respectively.

Proof. Let $\mu_0 = p_0$ and $\mu_1 = p_w$. where $X_\kappa = W_{\kappa,1} + \cdots + W_{\kappa,N}$. Under H_0, $E[X_\kappa] = N\mu_0$ and under H_1, $E[X_\kappa] = N\mu_1$.

This setting is almost the same as that for single linear cryptanalysis, the only differences being the facts that $\mu_1 = p_w$ is not in general $1/2$ and the inner key bit z is absent. As a result of μ_1 not being equal to $1/2$, for analysing the Type-II error probability we have to apply slightly different forms of the Chernoff bounds.

The expressions for $\delta_0, \delta_1, \alpha$ and the expression for t in terms of α are obtained as in the case of single linear cryptanalysis to be the following:

$$\begin{aligned}
\delta_0 &= (|\mu_0 - \mu_1| - t/N)/\mu_0; \\
\delta_1 &= t/(N\mu_1); \\
\alpha &= \exp(-(N\mu_0\delta_0^2)/3); \\
t &= N \times |\mu_0 - \mu_1| - \sqrt{3N\mu_0 \ln(1/\alpha)}.
\end{aligned}$$

Due to the use of the bounds (40) and (41), the expression for β changes as does the expression for t in terms of β.

$$\begin{aligned} \beta &= 2\exp\left(-N\mu_1\delta_1^2/3\right); \\ t &= \sqrt{3N\mu_1\ln(2/\beta)}. \end{aligned}$$

Equating the two expressions for t provides the expression on the right hand side of (25).

To apply the Chernoff bound (see Theorem 7), it is required that $0 < \delta_0, \delta_1 < 1$. As in Section 3, having $0 < t/N < |\mu_0 - \mu_1|$ ensures that the conditions on δ_0 and δ_1 hold. The bound on t leads to two lower bounds on N and the expression for N given by (25) satisfies these two lower bounds.

□

6 Multiple Differential Cryptanalysis

Here we consider a version of the multiple differential cryptanalysis, where the attacker uses ν r-round differentials all having the same input difference. Suppose that the ν r-round differentials for a block cipher are given by n-bit strings δ_0 and $\delta_r^{(1)}, \ldots, \delta_r^{(\nu)}$; where δ_0 denotes the input difference and $\delta_r^{(i)}$ denotes the i^{th} output difference. Each of the $\delta_r^{(i)}$'s must be non-zero n-bit strings and so $\nu \leq 2^n - 1$. As in the case of linear cryptanalysis, consider an m-bit target sub-key for some $m \leq n$. Guessing the value of this sub-key allows the inversion of the $(r+1)$-th round. For a uniform random plaintext P, and a choice κ of the target sub-key, define a random variable X_κ as follows:

$$X_\kappa = \begin{cases} i & \text{if } R_\kappa^{(r)-1}\left(E_{K^{(r)}}(P)\right) \oplus R_\kappa^{(r)-1}\left(E_{K^{(r)}}(P \oplus \delta_0)\right) = \delta_{r-1}^{(i)} \\ 0; & \text{otherwise.} \end{cases} \tag{26}$$

For $1 \leq i \leq \nu$, let p_i and θ be such that

$$\Pr[X_\kappa = i] = \begin{cases} p_i & \text{if } \kappa \text{ is the correct choice;} \\ \theta & \text{if } \kappa \text{ is an incorrect choice.} \end{cases} \tag{27}$$

Under the wrong key assumption, $\theta = 1/(2^n - 1)$. Further, define

$$p_0 = 1 - (p_1 + \cdots + p_\nu); \tag{28}$$

$$\theta_0 = 1 - \nu\theta. \tag{29}$$

Then both $\tilde{p} = (p_0, p_1, \ldots, p_\nu)$ and $\tilde{\theta} = (\theta_0, \theta, \ldots, \theta)$ are proper probability distributions. For the correct choice of κ, p_0 is the probability that none of the ν differentials hold. Similarly, for an incorrect choice of κ, θ_0 is the probability that none of the ν differentials hold. The random variable X_κ follows $\tilde{p}$ if κ is the correct choice and X_κ follows $\tilde{\theta}$ if κ is an incorrect choice.

Define another random variable $Y_\kappa = \ln\left(\frac{p_{X_\kappa}}{\theta_{X_\kappa}}\right)$. Let $\mu_0 = E[Y_\kappa]$ if X_κ follows $\tilde{p}$ (i.e., κ is the correct choice) and let $\mu_1 = E[Y_\kappa]$ if X_κ follows $\tilde{\theta}$ (i.e., κ is an incorrect choice). Then, $\mu_0 = D(\tilde{p} \,||\, \tilde{\theta})$ and $\mu_1 = D(\tilde{p} \,||\, \tilde{\theta})$.

Consider the N plaintext-ciphertext pairs $(P_1, C_1), \ldots, (P_N, C_N)$. For a choice κ of the target sub-key and $j = 1, \ldots, N$, let $X_{\kappa,j}$ be the random variable given by (26) corresponding to (P_j, C_j) and let $Y_\kappa = \ln\left(\frac{p_{X_{\kappa,j}}}{\theta_{X_{\kappa,j}}}\right)$. The test statistics is defined to be the following.

$$\text{LLR}_\kappa = \sum_{j=1}^{N} Y_{\kappa,j} = \sum_{\eta \in \{0,\ldots,\nu\}} Q_{\kappa,\eta} \ln(p_\eta/\theta_\eta)$$

where $Q_{\kappa,\eta} = \#\{j : Y_{\kappa,j} = \eta\}$. Consider the following test of hypothesis:

Hypothesis Test-4 (multiple differential cryptanalysis):

H_0: "κ is correct" versus H_1: "κ is incorrect."

Decision rule:

Case $\mu_0 > \mu_1$: Reject H_0 if LLR $\leq t$ where $t \in (N\mu_1, N\mu_0)$;

Case $\mu_0 < \mu_1$: Reject H_0 if LLR $\geq t$ where $t \in (N\mu_0, N\mu_1)$.

Proposition 4. *Let $0 < \alpha, \beta < 1$ and N be such that*

$$N \geq \frac{\upsilon^2\{\sqrt{\ln(1/\beta)} + \sqrt{\ln(1/\alpha)}\}^2}{2(D(\tilde{p} \,||\, \tilde{\theta}) + D(\tilde{\theta} \,||\, \tilde{p}))^2}. \tag{30}$$

Then the probabilities of the Type-I and Type-II errors in Hypothesis Test-4 are upper bounded by α and β respectively. Here,

$$\upsilon = \max_{\eta \in \{0,\ldots,\nu\}} \ln(p_\eta/\theta_\eta) - \min_{\eta \in \{0,\ldots,\nu\}} \ln(p_\eta/\theta_\eta).$$

Proof. Under H_0, $E[\text{LLR}] = N\mu_0$ while under H_1, $E[\text{LLR}] = N\mu_1$.

Here $Y_{\kappa,1}, \ldots, Y_{\kappa,N}$ are independently and identically distributed random variables taking values from the set $\{\ln(p_0/\theta_0), \ldots, \ln(p_\nu/\theta_\nu)\}$. Then, for a fixed κ

$$\upsilon_{\min} = \min_{\eta \in \{0,1,\ldots,\nu\}} \ln(p_\eta/\theta_\eta) \leq Y_{\kappa,j} \leq \max_{\eta \in \{0,1,\ldots,\nu\}} \ln(p_\eta/\theta_\eta) = \upsilon_{\max};$$

for all $j = 1, \ldots, N$. Let, $\upsilon = \upsilon_{\max} - \upsilon_{\min}$. Therefore, Hoeffding bounds can be applied on the sum of independently and identically distributed random variables $\text{LLR}_\kappa = \sum_{j=1}^{N} Y_{\kappa,j}$; where $D_N = N\upsilon^2$.

The error analysis is carried out separately in the two cases $\mu_0 > \mu_1$ and $\mu_0 < \mu_1$.

Case $\mu_0 > \mu_1$: In this case, $N\mu_1 < t < N\mu_0$. The probabilities of Type-I and Type-II errors are computed as follows:

$$\begin{aligned}
\Pr[\text{Type-I Error}] &= \Pr[\text{LLR}_\kappa \leq \text{t} | \text{H}_0 \text{ holds}] = \Pr[\text{LLR}_\kappa - \text{N}\mu_0 \leq -(\text{N}\mu_0 - \text{t}) | \text{H}_0 \text{ holds}] \\
&\leq \exp\left(-\frac{2(N\mu_0 - t)^2}{N\upsilon^2}\right); \\
\Pr[\text{Type-II Error}] &= \Pr[\text{LLR}_\kappa > \text{t} | \text{H}_1 \text{ holds}] = \Pr[\text{LLR}_\kappa - \text{N}\mu_1 > \text{t} - \text{N}\mu_1 | \text{H}_1 \text{ holds}] \\
&\leq \exp\left(-\frac{2(t - N\mu_1)^2}{N\upsilon^2}\right).
\end{aligned}$$

Here the inequalities given by (45) and (44) have been used. Define

$$\alpha = \exp\left(-\frac{2(N\mu_0 - t)^2}{N\upsilon^2}\right); \quad \beta = \exp\left(-\frac{2(t - N\mu_1)^2}{N\upsilon^2}\right).$$

The equation for α gives two values of t. The range for t eliminates one of the values. Similarly, the equation for β gives two values of t where one of the values is eliminated using the range for t. The two allowed values of t are the following.

$$\sqrt{2}t = \sqrt{2}N\mu_0 - \upsilon\sqrt{N\ln(1/\alpha)}; \tag{31}$$

$$\sqrt{2}t = \sqrt{2}N\mu_1 + \upsilon\sqrt{N\ln(1/\beta)}. \tag{32}$$

Eliminating t from equations (31) and (32), we obtain the expression given by the right hand side of (30). The expression for t given by (31) has to satisfy $N\mu_1 < t$ and the expression for t given by (32) has to satisfy $t < N\mu_0$. These give rise to two lower bounds on t both of which are satisfied by the expression for N given by (30).

Case $\mu_0 < \mu_1$: The analysis of this case is similar and leads to an expression for N which is the same as that given by (30).

□

7 Relating Advantage to Type-II Error Probability

The size of the target sub-key is m bits and there is one correct choice and the rest are incorrect choices. The hypothesis test is carried out independently for each choice κ of the target sub-key. Every time a Type-II error occurs, an incorrect choice gets labelled as a candidate key.

In the previous analyses, we have assumed β to be an upper bound on the probability of Type-II error. For the present, let us assume that β is indeed the actual probability of Type-II error. In the next section, we will consider the situation when β is an upper bound.

Since the probability of Type-II error is β, the expected number of incorrect keys which gets labelled as a candidate key is $\beta(2^m - 1)$. An attack is said to have an a-bit advantage if the size of the list of candidate keys produced by the attack is 2^{m-a}. Equating $(2^m - 1)\beta = 2^{m-a}$, we have that for an attack with a-bit expected advantage

$$\beta \;=\; \left(\frac{2^m}{2^m - 1}\right) 2^{-a}. \tag{33}$$

The right hand side can be approximated by 2^{-a} for moderate values of m. It is possible to use (33) to substitute $2^m/(2^m - 1) \times 2^{-a}$ for β in all the expressions for data complexities that have been obtained previously. This allows the data complexities to be expressed in terms of the expected advantage a.

While relating the expected advantage to β is sufficient for most purposes, it is possible to say more. One can upper bound the probability that the size of the list of false alarms exceeds a certain threshold. This is done as follows.

For each incorrect choice κ of the target sub-key, define W_κ to be a random variable which takes the value 1 if a Type-II error occurs for this choice of κ; and it takes the value 0 otherwise. Then the random variables W_κ's are independent Bernoulli distributed random variables having probability of success β. Let

$$W = \sum_{\kappa \text{ incorrect}} W_\kappa$$

and let $\mu = E[W] = \beta(2^m - 1)$. Using the Chernoff bound (38), we have that for any $\delta > 0$,

$$\Pr\left[W > (1+\delta)\,\mu\right] < \left(\frac{e^\delta}{(1+\delta)^{(1+\delta)}}\right)^\mu.$$

Define s such that $s = (1+\delta)\,\mu$ which combined with $\mu = \beta(2^m - 1)$ gives

$$\beta = \frac{s}{(1+\delta)\,(2^m - 1)}. \tag{34}$$

Using $s = (1+\delta)\,\mu$, we have

$$\Pr\left[W > s\right] < \left(\frac{e^{\frac{s-\mu}{\mu}}}{\left(\frac{s}{\mu}\right)^{\left(\frac{s}{\mu}\right)}}\right)^\mu = \frac{e^{s-\mu}\mu^s}{s^s} = P_\beta \text{ (say)}. \tag{35}$$

It is now possible to say that the probability that the list of false alarms exceeds s is at most P_β. Since μ is fixed, fixing P_β fixes s and then the relation $s = (1+\delta)\mu$ also fixes δ. Using (34), β can be expressed in terms of s and δ. Substituting this expression for β in the data complexities obtained earlier provides expressions for data complexities in terms s and P_β (and the Type-I error probability).

8 Distinguishers

Consider the problem of distinguishing between the probability distributions $\tilde{p}$ and $\tilde{q}$ over the set $\{0, \ldots, \nu-1\}$. Let, as in Section 2.3, $X_1, \ldots, X_N$ be independent and identically distributed random variables following either $\tilde{p}$ or $\tilde{q}$ but, which one is not known. As before, let $Y_j = \ln(p_{X_j}/q_{X_j})$ for $j = 1, \ldots, N$ and $\text{LLR} = Y_1 + \cdots + Y_N$.

Consider the log-likelihood ratio (LLR) based test statistics to design a test of hypothesis to distinguish between $\tilde{p}$ and $\tilde{q}$.

Hypothesis Test-5 (distinguisher):

H_0: "the distribution is $\tilde{p}$" versus H_1: "the distribution is $\tilde{q}$."

Decision rule:

Case $\mu_0 > \mu_1$: Reject H_0 if $\text{LLR} \leq t$ where $t \in (\mu_1, \mu_0)$;

Case $\mu_0 < \mu_1$: Reject H_0 if $\text{LLR} \geq t$ where $t \in (\mu_0, \mu_1)$.

Proposition 5. *Let $0 < P_e < 1$. In Hypothesis Test-5, it is possible to choose t such that for*

$$N \geq \frac{\upsilon^2 \ln(1/P_e)}{2(D(\tilde{p}||\tilde{q}) + D(\tilde{q}||\tilde{p}))^2}. \tag{36}$$

the Type-I and Type-II error probabilities satisfy

$$\Pr[\text{Type-I error}] + \Pr[\text{Type-II error}] \leq 2P_e.$$

Here,

$$\upsilon = \max_{\eta \in \{0,\ldots,\nu-1\}} \ln(p_\eta/q_\eta) - \min_{\eta \in \{0,\ldots,\nu-1\}} \ln(p_\eta/q_\eta).$$

Proof. Under H_0, Y_j has mean μ_0 and variance σ_0^2; while under H_1, Y_j has mean μ_1 and variance σ_1^2. The expressions for $\mu_0, \mu_1, \sigma_0^2, \sigma_1^2$ are given by (11). In the present case, we will not have any use for the variances. Under H_0, $E[\text{LLR}] = N\mu_0$ while under H_1, $E[\text{LLR}] = N\mu_1$. Also note that for the independently and identically distributed random variables $Y_1, \ldots, Y_N$, each

$$\upsilon_{\min} = \min_{\eta \in \{0,1,\ldots,\nu-1\}} \ln(p_\eta/q_\eta) \leq Y_j \leq \max_{\eta \in \{0,1,\ldots,\nu-1\}} \ln(p_\eta/q_\eta) = \upsilon_{\max}.$$

Let, $\upsilon = \upsilon_{\max} - \upsilon_{\min}$. Therefore, Hoeffding bounds can be applied on the sum of independently and identically distributed random variables $\text{LLR} = \sum_{\text{j=1}}^{\text{N}} \text{Y}_\text{j}$; where $D_N = N\upsilon^2$.

We now consider the probabilities of Type-I and Type-II errors. Since the form of the test is determined by the relative values of μ_0 and μ_1, the analysis is also done separately.

Case $\mu_0 > \mu_1$:

$$\begin{aligned} \Pr[\text{Type-I Error}] &= \Pr[\text{LLR} \leq t | H_0 \text{ holds}] = \Pr[\text{LLR} - N\mu_0 \leq -(N\mu_0 - t) | H_0 \text{ holds}] \\ &\leq \exp\left(-\frac{2(N\mu_0 - t)^2}{N\upsilon^2}\right). \end{aligned}$$

The last inequality follows from Hoeffding's inequality (see (45)). Similarly, the probability of Type-II error is computed as follows.

$$\begin{aligned} \Pr[\text{Type-II Error}] &= \Pr[\text{LLR} > t | H_1 \text{ holds}] = \Pr[\text{LLR} - N\mu_1 > t - N\mu_1 | H_1 \text{ holds}] \\ &\leq \exp\left(-\frac{2(t - N\mu_1)^2}{N\upsilon^2}\right). \end{aligned}$$

The last inequality follows from Hoeffding's inequality (see (44)).

Case $\mu_0 < \mu_1$:

$$\begin{aligned}\Pr[\text{Type-I Error}] &= \Pr[\text{LLR} \geq t | H_0 \text{ holds}] = \Pr[\text{LLR} - N\mu_0 \geq t - N\mu_0 | H_0 \text{ holds}] \\ &\leq \exp\left(-\frac{2(t-N\mu_0)^2}{N\upsilon^2}\right).\end{aligned}$$

The last inequality follows from Hoeffding's inequality (see (44)). Similarly, the probability of Type-II error is computed as follows.

$$\begin{aligned}\Pr[\text{Type-II Error}] &= \Pr[\text{LLR} < t | H_1 \text{ holds}] = \Pr[\text{LLR} - N\mu_1 < -(t - N\mu_1) | H_1 \text{ holds}] \\ &\leq \exp\left(-\frac{2(N\mu_1 - t)^2}{N\upsilon^2}\right).\end{aligned}$$

The last inequality follows from Hoeffding's inequality (see (45)). Let

$$\alpha = \exp\left(-\frac{2(N\mu_0 - t)^2}{N\upsilon^2}\right); \quad \beta = \exp\left(-\frac{2(t - N\mu_1)^2}{N\upsilon^2}\right).$$

These expressions are upper bounds on the probabilities of Type-I and Type-II errors respectively irrespective of whether $\mu_0 > \mu_1$ or $\mu_0 < \mu_1$.

The quantities α and β are determined by N. To obtain a relation between P_e and N, we set

$$P_e = \frac{\alpha + \beta}{2}.$$

Then it follows that $(\Pr[\text{Type-I Error}] + \Pr[\text{Type-II Error}]) \leq 2P_e$. Setting $t = N(\mu_0 + \mu_1)/2$ ensures $\alpha = \beta$ and then we obtain the following.

$$P_e = \exp\left(-\frac{2N(\mu_0 - \mu_1)^2}{\upsilon^2}\right) = \exp\left(-\frac{2N(D(\tilde{p}||\tilde{q}) + D(\tilde{q}||\tilde{p}))^2}{\upsilon^2}\right). \tag{37}$$

From the expression for P_e given by (37), the expression for N is given by the right hand side of (36). From this the statement of the result follows.

□

9 Upper Bounds

In the previous sections, we have obtained expressions for data complexities. These expressions are in terms of upper bounds on the probabilities of Type-I and Type-II errors.

Let $\alpha^\star$ and $\beta^\star$ be the actual probabilities of Type-I and Type-II errors respectively and further, let α and β be upper bounds on $\alpha^\star$ and $\beta^\star$ respectively. The success probability is $P_S^\star$ which by definition is $1 - \alpha^*$. Letting $P_S = 1 - \alpha$, we have, $P_S^\star \geq P_S$. Setting P_S to a pre-specified value ensures that the actual probability of success $P_S^\star$ is at least this value.

Following the discussion in Section 7, the probability of Type-II error can be related to the expected advantage of an attack. Let $a^\star$ be such that $2^{-a^\star} \times 2^m/(2^m - 1) = \beta^\star$. Also, define $a = -\lg \beta$ so that $\beta = 2^{-a}$. Then

$$2^{-a} = \beta \geq \beta^\star = 2^{-a^\star} \times 2^m/(2^m - 1) \geq 2^{-a^\star}$$

which shows that $a^\star \geq a$. So, fixing a to a pre-specified value ensures that the actual advantage is at least this value.

Using $P_s = 1 - \alpha$ and $\beta = 2^{-a}$ all the expressions for the data complexities obtained earlier can be written in terms of P_S and a.

The main question about data complexity that a cryptanalyst is interested in is the following. For a pre-specified value of P_S and a, what is the minimum number of plaintext-ciphertext pairs which ensures that $P_S^\star \geq P_S$ and $a^\star \geq a$? Following the discussion in Section 1.2, $N_{\min}(P_S, a)$ denotes this minimum required data complexity.

The data complexity expressions that we have obtained for the key recovery attacks earlier provide expressions for N in terms of P_S and a which can be written as $N(P_S, a)$. In other words, this means $N(P_S, a)$ plaintext-ciphertext pairs are sufficient to obtain $P_S^\star \geq P_S$ and $a^\star \geq a$. Again from the discussion in Section 1.2, we have $N_{\min}(P_S, a) \leq N(P_S, a)$ for all the cases of key recovery attacks. Similarly, for the case of distinguishing attacks $N_{\min}(P_e) \leq N(P_e)$. We record these in the following theorem.

Theorem 6. *1. For key recovery attacks using a single linear approximation based on Hypothesis Test-1,*

$$N_{\min}(P_S, a) \leq \frac{2\left\{\sqrt{(a+1)\ln 2} + \sqrt{3\,(1+|c|)\ln\left(1/(1-P_S)\right)}\right\}^2}{c^2}.$$

2. For key recovery attacks using multiple linear approximations based on Hypothesis Test-2,

$$N_{\min}(P_S, a) \leq \frac{\upsilon^2\left\{\sqrt{(a+\ell)\ln 2} + \sqrt{\ln\left(1/(1-P_S)\right)}\right\}^2}{2\left(D(\tilde{p}\,||\,\tilde{q}) + D(\tilde{q}\,||\,\tilde{p})\right)^2}.$$

3. For key recovery attacks using a single differential based on Hypothesis Test-3,

$$N_{\min}(P_S, a) \leq \frac{3\left\{\sqrt{p_w(a+1)\ln 2} + \sqrt{p_0 \ln(1/(1-P_S))}\right\}^2}{(p_0 - p_w)^2}.$$

4. For key recovery attacks using multiple differentials based on Hypothesis Test-4,

$$N_{\min}(P_S, a) \leq \frac{\upsilon^2\left\{\sqrt{a\ln 2} + \sqrt{\ln\left(1/(1-P_S)\right)}\right\}^2}{2\left(D(\tilde{p}\,||\,\tilde{\theta}) + D(\tilde{\theta}\,||\,\tilde{p})\right)^2}.$$

5. For distinguishing attacks based on Hypothesis Test-5,

$$N_{\min}(P_e) \leq \frac{\upsilon^2 \ln\left(1/P_e\right)}{2(D(\tilde{p}||\tilde{q}) + D(\tilde{q}||\tilde{p}))^2}.$$

10 Comparison

Previous works have obtained expressions for data complexities of the various attacks considered in this paper. The analyses have been based on using the central limit theorem to approximate the distribution of the sum of some random variables using the normal distribution. In this work, we have not used any approximation in our analysis. It is of interest to compare the rigorous upper bounds on data complexities that we have obtained with the expressions for data complexities using normal approximations.

We start by making a theoretical comparison of the various expressions. To facilitate the comparison, we introduce some notation to denote the expressions for the variances that arise in the different cases.

Let $\tilde{p}_{\$} \stackrel{\Delta}{=} (2^{-\ell}, \ldots, 2^{-\ell})$ be the uniform probability distribution over $\{0,1\}^{\ell}$. The variances in case of multiple linear cryptanalysis will be denoted by $(\sigma_0^{(L)})^2$ and $(\sigma_1^{(L)})^2$ (see [32] for further details). For multiple differential cryptanalysis we denote the variances by $(\sigma_0^{(D)})^2$ and $(\sigma_1^{(D)})^2$ (see [32] for further details). Lastly, for the LLR distinguisher we denote the variances by $(\sigma_0^{(\text{Dist})})^2$ and $(\sigma_1^{(\text{Dist})})^2$ (see [32] for further details) The expressions are all similar and our use of different notation is only for the sake of convenience in comparison.

Table 1 compares the expressions for the approximate data complexities that exist in the literature to the corresponding upper bounds on the data complexities obtained in this paper. For single linear and single differential cryptanalysis, the approximate expressions for data complexities were originally obtained in [34]. The approximate expression for the data complexity of multiple linear cryptanalysis was obtained in [19] while the approximate expression for the data complexity of multiple differential crypanalysis was obtained in [11]. These expressions were obtained using the order statistics based approach. In [32], the hypothesis testing framework was used to analyse data complexities. The actual forms of the approximate expressions for the data complexities listed in Table 1 are from [32]. For the case of distinguisher, the original analysis based on normal approximation was done in [2]. This was recapitulated in Section 2.3 and the approximate expression for the data complexity listed in Table 1 is given by (13).

The main observation from Table 1 is that in each case, the denominator of the approximate expression is the same as that of the upper bound. So, the difference between the approximate expression and the upper bound arises from the difference in the numerator. An analytical comparison of the numerators is infeasible. So, we perform an experimental comparison.

Attack Type	**Approximate Data Complexities**	**Upper Bounds**
Single LC	$\frac{\left\{\Phi^{-1}\left(1-2^{-a-1}\right)+\sqrt{1-c^2}\Phi^{-1}(P_S)\right\}^2}{c^2}$	$\frac{2\left\{\sqrt{(a+1)\ln 2}+\sqrt{3(1+\lvert c\rvert)\ln(1/(1-P_S))}\right\}^2}{c^2}$
Single DC	$\frac{\left\{\sqrt{p_w(1-p_w)}\Phi^{-1}\left(1-2^{-a}\right)+\sqrt{p_0(1-p_0)}\Phi^{-1}(P_S)\right\}^2}{(p_0-p_w)^2}$	$\frac{3\left\{\sqrt{p_w(a+1)\ln 2}+\sqrt{p_0\ln(1/(1-P_S))}\right\}^2}{(p_0-p_w)^2}$
Multiple LC	$\frac{\left\{\sigma_1^{(L)}\Phi^{-1}\left(1-2^{-\ell-a}\right)+\sigma_0^{(L)}\Phi^{-1}(P_S)\right\}^2}{(D(\bar{p}\lvert\lvert\bar{p}_{\$})+D(\bar{p}_{\$}\lvert\lvert\bar{p}))^2}$	$\frac{\upsilon^2\left\{\sqrt{(a+\ell)\ln 2}+\sqrt{\ln(1/(1-P_S))}\right\}^2}{2(D(\bar{p}\lvert\lvert\bar{q})+D(\bar{q}\lvert\lvert\bar{p}))^2}$
Multiple DC	$\frac{\left\{\sigma_1^{(D)}\Phi^{-1}\left(1-2^{-a}\right)+\sigma_0^{(D)}\Phi^{-1}(P_S)\right\}^2}{(D(\bar{p}\lvert\lvert\tilde{\theta})+D(\tilde{\theta}\lvert\lvert\bar{p}))^2}$	$\frac{\upsilon^2\left\{\sqrt{a\ln 2}+\sqrt{\ln(1/(1-P_S))}\right\}^2}{2\left(D(\bar{p}\lvert\lvert\tilde{\theta})+D(\tilde{\theta}\lvert\lvert\bar{p})\right)^2}$
Distinguisher	$\frac{\left\{\left(\sigma_0^{(\text{Dist})}+\sigma_1^{(\text{Dist})}\right)\Phi^{-1}(1-P_e)\right\}^2}{(D(\bar{p}\lvert\lvert\bar{q})+D(\bar{q}\lvert\lvert\bar{p}))^2}$	$\frac{\upsilon^2\ln(1/P_e)}{2(D(\bar{p}\lvert\lvert\bar{q})+D(\bar{q}\lvert\lvert\bar{p}))^2}$

Table 1: Table giving the upper bound on the data complexities along with the existing data complexities. Here LC denotes linear cryptanalysis and DC denotes differential cryptanalysis.

10.1 Comparison for SERPENT

This section compares the approximate data complexity of multiple linear cryptanalysis with the upper bound for the block cipher SERPENT. Collard et al [14] had presented reduced round linear cryptanalysis of the block cipher SERPENT using a set of linear approximations [15]. This set was later used in [18, 19]. The experiments conducted by Hermelin et al [18] made use of one subset of 64 linear approximations among the set given in [15]. It was found that this subset can be generated from 10 linear approximations, which they called the basis linear approximations. Table 2 of [18] lists these 10 linear approximations. These linear approximations can be used to recover 10 bits of the first round key. Thus, we have $\ell = 10$ and $m = 10$.

Notice that in order to generate the full joint distribution it is required to get the biases for all the $2^{10} - 1 = 1023$ non-zero linear approximations, generated from the 10 basis linear approximations. Since, only 64 out of these 1023 linear approximations were given in [15], the authors of [18, 19] used two different techniques to generate the full distribution. We have used the second method. Following [19] the value of P_S was fixed to 0.95. Table 2, summarises the output of the experiment for $a = 1, \ldots, 10$. In the table, N_{LLR} denotes the data complexity given by Equation (38) of [19] and N_{Upp} denotes the upper bound for multiple linear cryptanalysis given in Theorem 6. From the table, it follows that the upper bound on the data complexity is about 43 to 63 times that of the approximate value.

a	N_{LLR}	N_{Upp}	N_{Upp}/N_{LLR}
1	4.48×10^6	1.95×10^8	43.60
2	4.95×10^6	2.22×10^8	44.84
3	5.35×10^6	2.50×10^8	46.72
4	5.72×10^6	2.80×10^8	48.84
5	6.09×10^6	3.11×10^8	51.08
6	6.44×10^6	3.44×10^8	53.39
7	6.79×10^6	3.79×10^8	55.73
8	7.14×10^6	4.15×10^8	58.11
9	7.49×10^6	4.53×10^8	60.51
10	7.83×10^6	4.93×10^8	62.93

Table 2: Table showing comparison between N_{LLR} and N_{Upp} for the block cipher Serpent

10.2 Comparisons Using Simulated Joint Distributions

The approximate expressions contain terms of the type $\Phi^{-1}(x)$ and the corresponding term in the upper bound is $\sqrt{A \ln(1/(1-x))}$ for $A = 1, 2, 3, 6$. (For $x = P_S$ this can be seen directly; the other x's are $1 - 2^{-a-1}$, $1 - 2^{-a}$, $1 - 2^{-\ell-a}$ and $1 - P_e$ and the corresponding values of $1/(1-x)$ are 2^{a+1}, 2^a, $2^{\ell+a}$ and $1/P_e$ respectively.) These terms do not depend on the probability distributions $\tilde{p}$ or $\tilde{q}$.

Comparing $\Phi^{-1}(x)$ with $\sqrt{A \ln(1/(1-x))}$: For x varying from $1 - 2^{-2}$ to $1 - 2^{-100}$, Figure 1 shows the plots of $\Phi^{-1}(x)$, $\sqrt{\ln(1/(1-x))}$ and $\sqrt{\ln(1/(1-x))}/\Phi^{-1}(x)$. This shows that for the given range of x, the ratio $\sqrt{\ln(1/(1-x))}/\Phi^{-1}(x)$ is between 1 and 2. For $A = 2, 3$ or 6, the ratio increases by $\sqrt{A}$. Figure 2 shows the plots for the ratio $\sqrt{A \ln(1/(1-x))}/\Phi^{-1}(x)$ for $A = 1, 2, 3$ and 6.

From these plots we can infer that the difference in the approximate data complexities and the upper bounds arising due to the difference in $\Phi^{-1}(x)$ and $\sqrt{A \ln(1/(1-x))}$ is only by a small constant.

Comparisons of components depending on actual distributions: Some of the components in the numerators of the expressions given in Table 1 depend on the actual distributions $\tilde{p}$ and $\tilde{q}$. Performing these comparisons require simulating appropriate distributions. Below, we mention the actual simulations that were done and the corresponding results.

Comparing $1 - c^2$ and $1 + \mid c \mid$: Clearly, $1 - c^2 < 1 + \mid c \mid$. For our computations, we took c in the range $(-2^{-40}, 2^{-40})$ and in this range $\sqrt{1 - c^2} \approx 1 \approx \sqrt{1 + \mid c \mid}$.

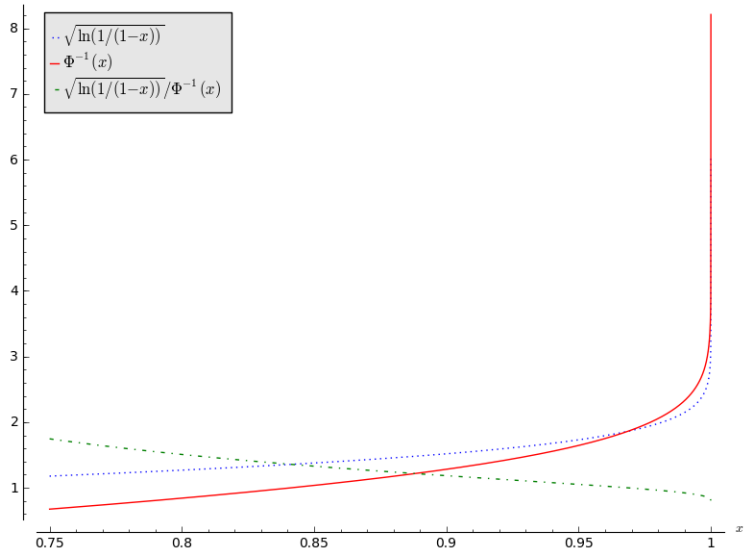

Figure 1: Plots of $\Phi^{-1}(x)$, $\sqrt{\ln(1/(1-x))}$ and $\sqrt{\ln(1/(1-x))}/\Phi^{-1}(x)$.

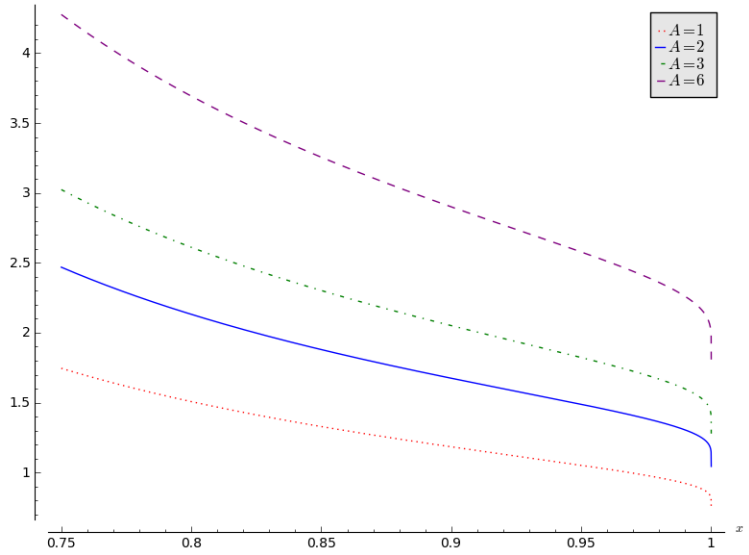

Figure 2: Plots of $\sqrt{A\ln(1/(1-x))}/\Phi^{-1}(x)$ for $A = 1, 2, 3$ and 6.

Comparing $\sigma_0^{(L)}$ and $\sigma_1^{(L)}$ with $\upsilon/\sqrt{2}$: This arises in the case of multiple linear cryptanalysis. For simulating the distributions, we took $\ell = 5$ and randomly selected the probabilities of $\tilde{p}$ in such a way that for all $\eta = 0, 1, \ldots, 2^5 - 1$, $\epsilon_\eta \in (-2^{-40}, 2^{-40})$. The values $\sigma_0^{(L)}$, $\sigma_1^{(L)}$ and $\upsilon/\sqrt{2}$, were then compared by computing the ratios $\upsilon/(\sqrt{2}\sigma_0^{(L)})$, $\upsilon/(\sqrt{2}\sigma_1^{(L)})$ and $\sigma_0^{(L)}/\sigma_1^{(L)}$. This experiment was repeated 10 times.

It was observed that the ratio $\sigma_0^{(L)}/\sigma_1^{(L)} \approx 1$ and also the ratio $\sqrt{2}\upsilon/\sigma_0^{(L)} \approx \sqrt{2}\upsilon/\sigma_1^{(L)}$. Table 3 gives the values of $\upsilon/\sqrt{2}$, $\sigma_0^{(L)}$ and $\sqrt{2}\upsilon/\sigma_0^{(L)}$.

$\upsilon/\sqrt{2}$	$\sigma_0^{(L)}$	$\upsilon/(\sqrt{2}\sigma_0^{(L)})$
5.98×10^{-9}	6.35×10^{-10}	9.42
3.54×10^{-9}	5.67×10^{-10}	6.25
2.04×10^{-9}	5.44×10^{-10}	3.76
4.18×10^{-8}	2.62×10^{-9}	15.92
1.19×10^{-8}	8.85×10^{-10}	13.41
1.69×10^{-8}	1.15×10^{-9}	14.70
6.06×10^{-9}	6.32×10^{-10}	9.60
1.31×10^{-8}	9.50×10^{-10}	13.83
1.52×10^{-8}	1.05×10^{-9}	14.49
1.16×10^{-8}	8.74×10^{-10}	13.27

Table 3: Table showing the values of $\upsilon/\sqrt{2}$, $\sigma_0^{(L)}$ and $\sqrt{2}\upsilon/\sigma_0^{(L)}$.

Comparing $\sigma_0^{(D)}$ and $\sigma_1^{(D)}$ with $\upsilon/\sqrt{2}$: This arises in the case of multiple differential cryptanalysis. For the simulation we took $n = 32, m = 10$ and $\nu = 20$ and again ensured that $\epsilon_\eta \in (-2^{-40}, 2^{-40})$ for all $\eta = 0, 1, \ldots, 20$. Random distributions were generated using these parameters like multiple linear cryptanalysis, The ratios $\sqrt{2}\upsilon/\sigma_0^{(D)}$, $\sqrt{2}\upsilon/\sigma_1^{(D)}$ and $\sigma_0^{(D)}/\sigma_1^{(D)}$ were considered. The experiment was also repeated 10 times.

As before the result showed that the ratio $\sqrt{2}\upsilon/\sigma_0^{(D)} \approx \sqrt{2}\upsilon/\sigma_1^{(D)}$ and $\sigma_0^{(D)}/\sigma_1^{(D)} \approx 1$. Table 4 gives the values of $\upsilon/\sqrt{2}$, $\sigma_0^{(D)}$ and $\sqrt{2}\upsilon/\sigma_0^{(D)}$.

$\upsilon/\sqrt{2}$	$\sigma_0^{(D)}$	$\upsilon/(\sqrt{2}\sigma_0^{(D)})$
0.0071	1.56×10^{-7}	32174.65
0.0070	1.51×10^{-7}	32578.80
0.0070	1.51×10^{-7}	32891.29
0.0066	1.60×10^{-7}	28959.21
0.0074	1.44×10^{-7}	36168.05
0.0076	1.62×10^{-7}	32985.94
0.0077	1.72×10^{-7}	31684.23
0.0071	1.44×10^{-7}	34608.71
0.0073	1.53×10^{-7}	33980.48
0.0074	1.50×10^{-7}	34872.68

Table 4: Table showing the values of $\upsilon/\sqrt{2}$, $\sigma_0^{(D)}$ and $\sqrt{2}\upsilon/\sigma_0^{(D)}$.

The experiment clearly shows that value of $\sigma_0^{(D)}$ is quite small compared to $\upsilon/\sqrt{2}$. Reason being, for $M = 40$ and $n = 32$, their difference $M - n = 8$ is quite small. We explain this more clearly. For the distributions considered, we have for all $\eta \neq 0$, $p_\eta = q_\eta + \epsilon_\eta$, where $q_\eta = 1/(2^n - 1) \approx 2^{-n}$ and $\epsilon_\eta \in (-2^{-M}, 2^{-M})$. This implies,

$$1 - 2^{-(M-n)} < \frac{p_\eta}{q_\eta} \approx 1 + \frac{\epsilon_\eta}{2^{-n}} < 1 + 2^{-(M-n)}.$$

Therefore, we have

$$1 - 2^{-(M-n)} < \upsilon_{\min}; \quad \text{and } \upsilon_{\max} < 1 + 2^{-(M-n)},$$

which implies that υ is upper bounded by $2^{-(M-n-1)}$, i.e., $0 \leq \upsilon < 2^{-(M-n-1)}$. Therefore, υ is small if $2^{-(M-n-1)}$ is small. In the present case we have $M = 40$ and $n = 32$, which implies that $2^{-(M-n-1)} = 2^{-7}$. Similarly, for multiple linear cryptanalysis υ is upper bounded by $2^{-(M-\ell-1)}$. Previously we had taken $\ell = 10$, this makes $2^{-(M-\ell-1)} = 2^{-29}$. This, somewhat explains the reason as to why the value $\upsilon/\sqrt{2}$ is closer to $\sigma_0^{(L)}$ in case of multiple linear cryptanalysis than compared to $\sigma_0^{(D)}$ for the multiple differential cryptanalysis.

Comparing $(\sigma_0^{(\text{Dist})} + \sigma_1^{(\text{Dist})})$ with $\upsilon/\sqrt{2}$: This is relevant for the distinguisher. The distinguisher is defined for arbitrary probability distributions $\tilde{p}$ and $\tilde{q}$. For the experimental comparison, we applied the distinguisher to the context of multiple linear cryptanalysis. Here, as before, we chose $\ell = 5$ and ϵ_η in the same range as that of multiple linear cryptanalysis. Unlike the previous cases, here it is required to compute $\upsilon/(\sqrt{2}(\sigma_0^{\sqrt{2}(\text{Dist})} + \sigma_1^{(\text{Dist})}))$. As before the experiment was repeated 10 times and the observations are listed in Table 5.

$\upsilon/\sqrt{2}$	$\sigma_0^{(\text{Dist})} + \sigma_1^{(\text{Dist})}$	$\sqrt{2}\upsilon/(\sigma_0^{(\text{Dist})} + \sigma_1^{(\text{Dist})})$
1.33×10^{-9}	5.43×10^{-10}	2.45
2.14×10^{-8}	1.40×10^{-9}	15.28
4.11×10^{-9}	5.81×10^{-10}	7.08
1.86×10^{-9}	5.45×10^{-10}	3.41
4.35×10^{-9}	5.94×10^{-10}	7.33
4.34×10^{-9}	5.76×10^{-10}	7.55
1.83×10^{-8}	1.22×10^{-9}	14.96
1.32×10^{-9}	5.48×10^{-10}	2.40
1.98×10^{-8}	1.31×10^{-9}	15.16
8.10×10^{-9}	7.13×10^{-10}	11.35

Table 5: Table showing the values of $\upsilon/\sqrt{2}$, $(\sigma_0^{(\text{Dist})} + \sigma_1^{(\text{Dist})})$ and $\sqrt{2}\upsilon/(\sigma_0^{(\text{Dist})} + \sigma_1^{(\text{Dist})})$.

Overall comparison of approximate data complexities with the upper bounds: The size of the target sub-key was taken to be $m = 10$ bits and the block size $n = 32$. For single linear cryptanalysis, we chose c randomly in the range $(-2^{-40}, 2^{-40})$. For single differential cryptanalysis, it was assumed that $p_0 = p_w + c$, where $p_w = 1/(2^n - 1)$ and c was chosen randomly from $(-2^{-40}, 2^{-40})$. In the cases of multiple linear cryptanalysis and the LLR distinguisher we took $\ell = 5$ and for multiple differential cryptanalysis we took $\nu = 20$. In all three cases, the ϵ_η's were randomly chosen from $(-2^{-40}, 2^{-40})$.

As is normally the case, the success probability P_S was fixed to a constant. We have used three different success probabilities, namely, $P_S = 1 - 2^{-5}$, $1 - 2^{-7}$ and $1 - 2^{-10}$. The advantage was varied from $a = 2$ to 100 for all cases other than the LLR distinguisher. For each value of a, the ratio of the upper bound on the data complexity to the approximate data complexity was computed and the minimum and maximum of these values were recorded. The rows of Table 6 reports these minimums and maximums. For the case of the LLR distinguisher, it is required that $\alpha = \beta$ and hence for our example, $a = 5, 7$ and 10. Since we get a single value of a, we ran the experiment for this value of a 100 times for each value of a and recorded the minimum and the maximum. The last row of Table 6 reports these values.

From Table 6 it can be observed that other than the case of multiple differential cryptanalysis, the upper bound is not significantly larger than the approximate data complexity. For multiple differential cryptanalysis, the upper bound is significantly greater than the approximate value. To a large extent, the higher value of the upper bound is explained by the differences in the values of υ and the variances as reported in Tables 3, 4 and 5.

Type of Attack	$P_S = 1 - 2^{-5}$		$P_S = 1 - 2^{-7}$		$P_S = 1 - 2^{-10}$	
	Maximum	Minimum	Maximum	Minimum	Maximum	Minimum
Single LC	6.02	1.70	5.21	1.73	4.63	1.76
Single DC	5.09	1.89	4.17	1.84	3.50	1.80
Multiple DC	2.30×10^9	3.05×10^8	2.63×10^9	1.70×10^8	1.90×10^9	2.54×10^8
Multiple LC	200.55	4.43	197.75	4.43	199.06	4.53
LLR Distinguisher	2.55	1.01	2.17	0.86	1.82	0.77

Table 6: Table giving the maximum and minimum values of the ratios of the upper bound to the approximate data complexity for each row of Table 1.

For the cases where the approximate data complexities and the upper bounds are close, our conclusion is that it is perhaps better to use the upper bounds as the data complexities of the corresponding attacks. While this will push up the data requirement to some extent, it is based on rigorous analysis and is certain to hold in all cases. For multiple differential cryptanalysis, the gap in the approximate and upper bound on data complexity is fairly large so that no clear conclusion can be drawn. This gap could be due to the approximate value being a significant underestimate or due to the fact that the upper bound is an overestimate. At this point of time, we are unable to determine the exact reason. More work is necessary to settle this point.

10.3 Comparing the Two Upper Bounds for Single Linear and Differential Cryptanalysis

Note that in our analysis we get two upper bounds on data complexity of single linear cryptanalysis – one obtained directly using the Chernoff bound and another by putting $\ell = 1$ in the expression for data complexity of multiple linear cryptanalysis. Putting $\ell = 1$ in equation (20), we get

$$\begin{aligned}
\upsilon^2 &= [\max\{\ln(1+c), \ln(1-c)\} - \max\{\ln(1-c), \ln(1+c)\}]^2 = \left(\ln\left(\frac{1+c}{1-c}\right)\right)^2; \\
\mu_0 &= \frac{1}{2}\left[\ln(1-c^2) + c\ln\left(\frac{1+c}{1-c}\right)\right]; \\
\mu_1 &= \frac{1}{2}\ln(1-c^2); \text{ and} \\
N &= \frac{2\{\sqrt{(a+1)\ln 2} + \sqrt{\ln(1/(1-P_S))}\}^2}{c^2}.
\end{aligned}$$

This needs to be compared with the expression obtained using the Chernoff bound, i.e.,

$$N = \frac{2\{\sqrt{(a+1)\ln 2} + \sqrt{3(1+\mid c \mid)\ln(1/(1-P_S))}\}^2}{c^2}.$$

Let us call $\sqrt{(a+1)\ln 2}$ as x, $\sqrt{\ln(1/(1-P_S))}$ as y, the data complexity obtained using Chernoff bound as N_C and the data complexity obtained using Hoeffding bounds as N_H. Then,

$$\begin{aligned}
N_H - N_C &= \frac{2}{c^2}\{(x+y)^2 - (x + \sqrt{3(1+\mid c \mid)}y)^2\} \\
&= -\frac{2}{c^2}\{(2+3\mid c \mid)y^2 + 2(\sqrt{3(1+\mid c \mid)} - 1)xy\} \\
&< 0; \quad \text{[Since, } x \text{ and } y \text{ are greater than zero, and } \sqrt{3(1+\mid c \mid)} > 1\text{]}.
\end{aligned}$$

Thus, we have $N_H < N_C$, which means that the data complexity obtained using the Hoeffding bound gives a better upper bound in case of single linear cryptanalysis.

Similarly, one obtains two upper bounds on the data complexity of single differential cryptanalysis. Putting $\nu = 1$ in the right hand side of (30), we get

$$\begin{aligned}
\tilde{p} &= (1-p_0, p_0); \quad \tilde{\theta} = (1-p_w, p_w); \\
\upsilon^2 &= [\max\{\ln(p_0/p_w), \ln((1-p_0)/(1-p_w))\} - \min\{\ln(p_0/p_w), \ln((1-p_0)/(1-p_w))\}]^2; \\
&= \left(\ln\left(\frac{p_0(1-p_w)}{p_w(1-p_0)}\right)\right)^2 \\
D(\tilde{p} \,||\, \tilde{\theta}) &= (1-p_0)\ln\left(\frac{1-p_0}{1-p_w}\right) + p_0 \ln\left(\frac{p_0}{p_w}\right) \\
&= \ln\left(\frac{1-p_0}{1-p_w}\right) + p_0 \ln\left(\frac{p_0(1-p_w)}{p_w(1-p_0)}\right); \\
D(\tilde{\theta} \,||\, \tilde{p}) &= \ln\left(\frac{1-p_w}{1-p_0}\right) - p_w \ln\left(\frac{p_0(1-p_w)}{p_w(1-p_0)}\right); \\
(D(\tilde{p} \,||\, \tilde{\theta}) + D(\tilde{\theta} \,||\, \tilde{p}))^2 &= (p_0-p_w)^2 \left(\ln\left(\frac{p_0(1-p_w)}{p_w(1-p_0)}\right)\right)^2 = (p_0-p_w)^2\upsilon^2; \text{ and} \\
N_H &= \frac{\{\sqrt{a\ln 2} + \sqrt{\ln(1/(1-P_S))}\}^2}{2(p_0-p_w)^2}.
\end{aligned}$$

This needs to be compared with the expression obtained using the Chernoff bound, i.e.,

$$N_C = \frac{3\{\sqrt{p_w(a+1)\ln 2} + \sqrt{p_0 \ln(1/(1-P_S))}\}^2}{(p_0-p_w)^2}.$$

Then,

$$N_H - N_C = \frac{\left\{(x+y)^2 - 6(\sqrt{p_w}x + \sqrt{p_0}y)^2\right\}}{2(p_0-p_w)^2} = \frac{\left\{(1-6p_w)x^2 + (1-6p_0)y^2 + 2(1-6\sqrt{p_0 p_w})xy\right\}}{2(p_0-p_w)^2}$$

Now, $1-6p_w \geq 0$, implies $p_w \leq 1/6$ or in other words, $n \geq 3$. Recall, that n denotes the block size. Therefore, it is safe to assumed that $p_w \leq 1/6$. Similarly, it is also safe to assume that $p_0 \leq 1/6$. Then, these two assumption gives $1 - 6\sqrt{p_0 p_w} \geq 0$. Thus, we have

$$N_H - N_C \geq 0,$$

or in other words, $N_H \geq N_C$. Therefore, the data complexity obtained using Chernoff bounds gives a better upper bound in case of single differential cryptanalysis.

11 Conclusion

The paper obtains rigorous upper bounds on the data complexities of linear and differential cryptanalysis. No use is made of the central limit theorem to approximate the distribution of a sum of random variables using the normal distribution. Computations show that the obtained upper bounds are not too far away from previously obtained approximate data complexities. Due to the rigorous nature of our analysis, we believe that this approach may be adopted in the future to analyse other techniques for cryptanalysis.

The statistical techniques that have been used for obtaining the upper bounds are fairly standard, though, to the best of our knowledge they have not been used in this context earlier. We, however, make no claims that the bounds that we obtain cannot be improved. In fact, one of the goals of our work is to stimulate interest in rigorous statistical analysis of attacks on block ciphers. Hopefully, the community will further explore this direction of research since we believe that if something is worth doing, then it is worth doing properly.

References

[1] Mohamed Ahmed Abdelraheem, Martin Ågren, Peter Beelen, and Gregor Leander. On the Distribution of Linear Biases: Three Instructive Examples. In *Advances in Cryptology–CRYPTO 2012*, pages 50–67. Springer, 2012.

[2] Thomas Baignères, Pascal Junod, and Serge Vaudenay. How Far Can We Go Beyond Linear Cryptanalysis? In *Advances in Cryptology–ASIACRYPT 2004*, pages 432–450. Springer, 2004.

[3] Thomas Baignères, Pouyan Sepehrdad, and Serge Vaudenay. Distinguishing Distributions Using Chernoff Information. In *Provable Security*, pages 144–165. Springer, 2010.

[4] Thomas Baignères and Serge Vaudenay. The complexity of distinguishing distributions (invited talk). In Reihaneh Safavi-Naini, editor, *Information Theoretic Security, Third International Conference, ICITS 2008, Calgary, Canada, August 10-13, 2008, Proceedings*, volume 5155 of *Lecture Notes in Computer Science*, pages 210–222. Springer, 2008.

[5] Eli Biham, Alex Biryukov, and Adi Shamir. Cryptanalysis of Skipjack Reduced to 31 Rounds Using Impossible Differentials. In *Advances in Cryptology–Eurocrypt99*, pages 12–23. Springer, 1999.

[6] Eli Biham and Adi Shamir. Differential Cryptanalysis of DES-like Cryptosystems. In *Advances in Cryptology–CRYPTO'90*, pages 2–21. Springer, 1990.

[7] Eli Biham and Adi Shamir. Differential Cryptanalysis of DES-like Cryptosystems. *Journal of CRYPTOLOGY*, 4(1):3–72, 1991.

[8] Alex Biryukov, Christophe De Cannière, and Michaël Quisquater. On Multiple Linear Approximations. In *Advances in Cryptology–CRYPTO 2004*, pages 1–22. Springer, 2004.

[9] Céline Blondeau, Andrey Bogdanov, and Gregor Leander. Bounds in Shallows and in Miseries. In *Advances in Cryptology–CRYPTO 2013*, pages 204–221. Springer, 2013.

[10] Céline Blondeau and Benoît Gérard. Multiple Differential Cryptanalysis: Theory and Practice. In *Fast Software Encryption*, pages 35–54. Springer, 2011.

[11] Céline Blondeau, Benoît Gérard, and Kaisa Nyberg. Multiple Differential Cryptanalysis using LLR and χ^2 Statistics. In *Security and Cryptography for Networks*, pages 343–360. Springer, 2012.

[12] Céline Blondeau, Benoît Gérard, and Jean-Pierre Tillich. Accurate Estimates of the Data Complexity and Success Probability for Various Cryptanalyses. *Designs, Codes and Cryptography*, 59(1-3):3–34, 2011.

[13] Andrey Bogdanov and Elmar Tischhauser. On the Wrong Key Randomisation and Key Equivalence Hypotheses in Matsuis Algorithm 2. In *Fast Software Encryption*, pages 19–38. Springer, 2014.

[14] Baudoin Collard, François-Xavier Standaert, and Jean-Jacques Quisquater. Experiments on the multiple linear cryptanalysis of reduced round serpent. In *Fast Software Encryption*, pages 382–397. Springer, 2008.

[15] Baydoin Collard, François-Xavier Standaert, and Jean-Jacques Quisquater. 2008. `http://www.dice.ucl.ac.be/fstandae/PUBLIS/50b.zip`.

[16] Itai Dinur and Adi Shamir. Cube Attacks on Tweakable Black Box Polynomials. *Advances in Cryptology–EUROCRYPT 2009*, pages 278–299, 2009.

[17] Carlo Harpes, Gerhard G. Kramer, and James L. Massey. A Generalization of Linear Cryptanalysis and the Applicability of Matsui's Piling-Up Lemma. In Louis C. Guillou and Jean-Jacques Quisquater, editors, *Advances in Cryptology - EUROCRYPT '95, International Conference on the Theory and Application of Cryptographic Techniques, Saint-Malo, France, May 21-25, 1995, Proceeding*, volume 921 of *Lecture Notes in Computer Science*, pages 24–38. Springer, 1995.

[18] Miia Hermelin, Joo Yeon Cho, and Kaisa Nyberg. Multidimensional Linear Cryptanalysis of Reduced Round Serpent. In *Information Security and Privacy*, pages 203–215. Springer, 2008.

[19] Miia Hermelin, Joo Yeon Cho, and Kaisa Nyberg. Multidimensional Extension of Matsuis Algorithm 2. In *Fast Software Encryption*, pages 209–227. Springer, 2009.

[20] Pascal Junod. On the Optimality of Linear, Differential, and Sequential Distinguishers. In *Advances in Cryptology–EUROCRYPT 2003*, pages 17–32. Springer, 2003.

[21] Pascal Junod and Serge Vaudenay. Optimal Key Ranking Procedures in a Statistical Cryptanalysis. In *Fast Software Encryption*, pages 235–246. Springer, 2003.

[22] Burton S Kaliski Jr and Matthew JB Robshaw. Linear Cryptanalysis Using Multiple Approximations. In *Advances in Cryptology–Crypto94*, pages 26–39. Springer, 1994.

[23] Lars R Knudsen. Truncated and Higher Order Differentials. In *Fast Software Encryption*, pages 196–211. Springer, 1995.

[24] Xuejia Lai. Higher order derivatives and differential cryptanalysis. In *Communications and Cryptography*, pages 227–233. Springer, 1994.

[25] Gregor Leander. On linear hulls, statistical saturation attacks, present and a cryptanalysis of puffin. In *Advances in Cryptology–EUROCRYPT 2011*, pages 303–322. Springer, 2011.

[26] Mitsuru Matsui. Linear Cryptanalysis Method for DES Cipher. In *Advances in Cryptology–EUROCRYPT'93*, pages 386–397. Springer, 1993.

[27] Mitsuru Matsui. The First Experimental Cryptanalysis of the Data Encryption Standard. In Y. G. Desmedt, editor, *Advances in Cryptology–Crypto94*, pages 1–11. Springer, 1994.

[28] Michael Mitzenmacher and Eli Upfal. *Probability and computing: Randomized algorithms and probabilistic analysis*. Cambridge University Press, 2005.

[29] Rajeev Motwani and Prabhakar Raghavan. *Randomized Algorithms*. Chapman & Hall/CRC, 2010.

[30] Sean Murphy. The Independence of Linear Approximations in Symmetric Cryptanalysis. *Information Theory, IEEE Transactions on*, 52(12):5510–5518, 2006.

[31] Kaisa Nyberg and Miia Hermelin. Multidimensional walsh transform and a characterization of bent functions. In *Proceedings of the 2007 IEEE Information Theory Workshop on Information Theory for Wireless Networks*, pages 83–86, 2007.

[32] Subhabrata Samajder and Palash Sarkar. Another Look at Normal Approximations in Cryptanalysis. *Journal of Mathematical Cryptology*, 2016. DOI: 10.1515/jmc-2016-0006.

[33] Subhabrata Samajder and Palash Sarkar. Can large deviation theory be used for estimating data complexity? Cryptology ePrint Archive, Report 2016/465, 2016. `http://eprint.iacr.org/`.

[34] Ali Aydın Selçuk. On Probability of Success in Linear and Differential Cryptanalysis. *Journal of Cryptology*, 21(1):131–147, 2008.

[35] Cihangir Tezcan. The Improbable Differential Attack: Cryptanalysis of Reduced Round CLEFIA. In *Progress in Cryptology-INDOCRYPT 2010*, pages 197–209. Springer, 2010.

[36] David Wagner. The Boomerang Attack. In *Fast Software Encryption*, pages 156–170. Springer, 1999.

A Concentration Inequalities

A.1 Chernoff Bounds

We briefly recall some results on tail probabilities of sums of Poisson trials that will be used later. These results can be found in standard texts such as [29, 28] and are usually referred to as the Chernoff bounds.

Theorem 7. *Let $X_1, X_2, \ldots, X_\lambda$ be a sequence of independent Poisson trials such that for $1 \leq i \leq \lambda$, $\Pr[X_i = 1] = p_i$. Then for $X = \sum_{i=1}^{\lambda} X_i$ and $\mu = E[X] = \sum_{i=1}^{\lambda} p_i$ the following bounds hold:*

$$\text{For any } \delta > 0,\ \Pr[X \geq (1+\delta)\mu] < \left(\frac{e^{-\delta}}{(1+\delta)^{(1+\delta)}}\right)^{\mu}. \tag{38}$$

$$\text{For any } 0 < \delta < 1,\ \Pr[X \leq (1-\delta)\mu] \leq \left(\frac{e^{-\delta}}{(1-\delta)^{(1-\delta)}}\right)^{\mu}. \tag{39}$$

These bounds can be simplified to the following form.

$$\text{For any } 0 < \delta \leq 1,\ \Pr[X \geq (1+\delta)\mu] \leq e^{-\mu\delta^2/3}. \tag{40}$$

$$\text{For any } 0 < \delta < 1,\ \Pr[X \leq (1-\delta)\mu] \leq e^{-\mu\delta^2/2}. \tag{41}$$

Further, if $p_i = 1/2$ for $i = 1, \ldots, \lambda$, then the following stronger bounds hold.

$$\text{For any } \delta > 0,\ \Pr[X \geq (1+\delta)\mu] \leq e^{-\delta^2\mu}. \tag{42}$$

$$\text{For any } 0 < \delta < 1,\ \Pr[X \leq (1-\delta)\mu] \leq e^{-\delta^2\mu}. \tag{43}$$

A.2 Hoeffding Inequality

we briefly recall Hoeffding's inequality for sum of independent random variables. The result can be found in standard texts such as [28].

Theorem 8 (Hoeffding Inequality)**.** *Let, $X_1, X_2, \ldots, X_\lambda$ be a finite sequence of independent random variables, such that for all $i = 1, \ldots, \lambda$, there exists real numbers $a_i, b_i \in \mathbb{R}$, with $a_i < b_i$ and $a_i \leq X_i \leq b_i$. Let $X = \sum_{i=1}^{\lambda} X_i$. Then for any positive $t > 0$,*

$$\Pr[X - E[X] \geq t] \leq \exp\left(-\frac{2t^2}{D_\lambda}\right) \tag{44}$$

$$\Pr[X - E[X] \leq -t] \leq \exp\left(-\frac{2t^2}{D_\lambda}\right) \tag{45}$$

$$\Pr[|X - E[X]| \geq t] \leq 2\exp\left(-\frac{2t^2}{D_\lambda}\right); \tag{46}$$

where $D_\lambda = \sum_{i=1}^{\lambda} (b_i - a_i)^2$.

B Data Complexity of Distinguisher Using Normal Approximation

The LLR based test statistics for distinguishing between $\tilde{p}$ and $\tilde{q}$ is taken to be the following.

$$T = \frac{\text{LLR}/N - \mu_1}{\sigma_1/\sqrt{N}}. \tag{47}$$

The following two asymptotic assumptions are usually made.

1. If the X_j's follow $\tilde{q}$, then for sufficiently large N, T approximately follows the standard normal distribution $\Phi(0,1)$.
2. On the other hand, if the X_j's follow $\tilde{p}$, then T is rewritten as follows.

$$T = \frac{\sigma_0}{\sigma_1} Z + \frac{\sqrt{N}(\mu_0 - \mu_1)}{\sigma_1}$$

where $Z = \dfrac{\text{LLR}/N - \mu_0}{\sigma_0/\sqrt{N}}$. For sufficiently large N, Z approximately follows the standard normal distribution $\Phi(0,1)$.

Both the above assumptions involve an error term. The error can be bounded above using the Berry-Esséen theorem. See [32] for details of this analysis.

The form of the test is determined by the relative values of μ_0 and μ_1.

$\mu_0 > \mu_1$: Reject H_0 if $T \leq t$ where t is in the range $\mu_1 < t < \mu_0$;
$\mu_0 < \mu_1$: Reject H_0 if $T \geq t$ where t is in the range $\mu_0 < t < \mu_1$.

Let α and β be the probabilities of Type-I and Type-II errors respectively. Define

$$P_e = \frac{\alpha + \beta}{2}. \tag{48}$$

The goal is to choose a value of t for which $\alpha = \beta$ holds. The analysis of α and β is done as follows. First suppose $\mu_0 > \mu_1$.

$$\begin{aligned} \alpha = \Pr[\text{Type-I Error}] &= \Pr[T \leq t | H_0 \text{ holds}] = \Phi\left(\frac{\sigma_1 t}{\sigma_0} - \frac{\sqrt{N}(\mu_0 - \mu_1)}{\sigma_0}\right); \\ \beta = \Pr[\text{Type-II Error}] &= \Pr[T > t | H_1 \text{ holds}] = 1 - \Phi(t) = \Phi(-t). \end{aligned}$$

In this case, $t = \sqrt{N}(\mu_0 - \mu_1)/(\sigma_0 + \sigma_1)$ ensures that $\alpha = \beta$.

Now suppose that $\mu_0 < \mu_1$. Proceeding as above shows that choosing $t = \sqrt{N}(\mu_1 - \mu_0)/(\sigma_0 + \sigma_1)$ ensures $\alpha = \beta$. So, irrespective of the relative values of μ_0 and μ_1, for

$$t = \frac{\sqrt{N}|\mu_0 - \mu_1|}{\sigma_0 + \sigma_1}$$

the expression for P_e is the following.

$$P_e = \Phi(-t) = \Phi\left(-\frac{\sqrt{N}|\mu_0 - \mu_1|}{\sigma_0 + \sigma_1}\right) = \Phi\left(-\frac{\sqrt{N}|D(\tilde{p}||\tilde{q}) + D(\tilde{q}||\tilde{p})|}{\sigma_0 + \sigma_1}\right). \tag{49}$$

In [2], a second order Taylor series expansion of ln term was used in the expression for the Kullback-Leibler divergence. This resulted in the expression for P_e simplifying to $P_e = \Phi(-\sqrt{N}C(\tilde{p},\tilde{q})/2)$, where $C(\tilde{p},\tilde{q})$ is defined to be the capacity between the two probability distributions $\tilde{p}$ and $\tilde{q}$.

From the expression for P_e given by (49), it is possible to obtain an expression for the data complexity N required to achieve a desired value of P_e.

$$N = \left(\frac{(\sigma_0 + \sigma_1)\Phi^{-1}(1 - P_e)}{D(\tilde{p}||\tilde{q}) + D(\tilde{q}||\tilde{p})}\right)^2. \tag{50}$$