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

September 21, 2015

Abstract

All statistical analysis of symmetric key attacks use the central limit theorem to approximate the distribution of a sum of random variables using the normal distribution. Expressions for data complexity using such an approach are *inherently approximate*. 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 differetial cryptanalysis, the Chernoff bound is used. For the cases of multiple linear and multiple differential cryptanalysis, the theory of martingales is required. A Doob martingale satisfying the Lipschitz condition is set up so that the Azuma-Hoeffding inequality can be applied. 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 have much wider applicability.

Keywords: block cipher, linear cryptanalysis, differential cryptanalysis, log-likelihood test, order statistics, normal distribution, hypothesis testing, Chernoff bound, Martingales, Lipschitz condition, Azuma-Hoeffding inequality.

1 Introduction

Statistical methods are commonly used for analysing attacks on block ciphers and more generally symmetric key ciphers. There are three basic parameters of interest.

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 above parameters are of interest in a key recovery attack. For a distinguishing attack, the situation is a little different and we consider this later in the paper.

A 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. The reason is that all known statistical methods for analysing attacks rely on the central limit theorem to approximate the distribution of a sum of random variables by the standard normal distribution. From a theoretical point of view

we find this to be unsatisfactory. It would be desirable to carry out the statistical analysis without using any approximation. Further, a recent work [28] takes a detailed look at the error in normal approximations and points out several shortcomings of such an approach.

The major motivation of this work is to derive rigourous 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 rigourous upper bounds on the data complexity.

We make detailed experimental comparisons of the upper bounds that we obtain to the previously best known approximate values of data complexities. 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 about 50 or so, while for multiple differential cryptanalysis, the ratio is between 500 to 2000. This indicates that the upper bounds that we obtain are good. 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 underestimates.

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. Our main observation is that 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 the theory of martingales. We set up a Doob martingale which satisfies an appropriate Lipschitz condition and hence the Azuma-Hoeffding inequality can be applied. 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 applicability and it should be possible to tackle data complexities of other attacks using these techniques.

Previous and related works

Linear Cryptanalysis: This was first proposed by Matsui in [22] to cryptanalyze the block cipher DES. Later Matsui [23] extended this idea by using two linear approximations. In an independent work, Kaliski and Robshaw [18] 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.

In [7], the idea of multiple linear cryptanalysis was further refined. 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 stochastically independent linear approximations. Analysis under the independence assumption was separately done in [17]. Murphy [26] argued that the independence assumption need not be valid.

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

In a later work, Baigneres 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 [17, 3].

Selçuk in [29] 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 [29] are approximate. The order statistics based approach was built upon by Hermelin et al [16]. The authors combined the results obtained in [2, 26, 27, 29] to develop a multilinear cryptanalytic method without the independence assumption.

Differential cryptanalysis: This cryptanalytic method was first proposed by Biham and Shamir in [5]. It was used to successfully cryptanalyze reduced round variants (with up to 15 rounds) of DES using less than 2^{56} operations. Later in [6], 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 [20], truncated differentials [19], cube attack [14], boomerang attack [31], impossible differential cryptanalysis [4] and improbable differential cryptanalysis [30].

The general approach to multiple differential cryptanalysis was considered in [9]. This work considered ℓ differentials having both unequal input and unequal output differences. Later [10] considered ℓ differentials having same input difference but different output differences. 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 [11].

We note that a recent work [28] 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 [29] 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, 8, 12, 13, 21] and also use normal 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. In the later part of the section, we provide relevant details of tail probability inequlities, specifically the Chernoff-Hoeffding bounds for Poisson trials and the Azuma-Hoeffding bounds for martingales.

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 $E : \{0,1\}^k \times \{0,1\}^n \rightarrow \{0,1\}^n$ such that for each $K \in \{0,1\}^k$, the function $E_K(\cdot) \triangleq E(K,\cdot)$ 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 relations, 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_i, C_i), $i = 1, \ldots, N$ 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 distribution 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. 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 $\Gamma^{(i)}_P, \Gamma^{(i)}_B$ and $\Gamma^{(i)}_K$, 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:

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

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

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

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

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

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

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

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

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

$$\begin{aligned} L_{\kappa,j,i} &= \langle \Gamma_P^{(i)}, P_j \rangle \oplus \langle \Gamma_B^{(i)}, B_{\kappa,j} \rangle; \quad (7) \\ X_{\kappa,j} &= (L_{\kappa,j,1}, \ldots, L_{\kappa,j,\ell}). \quad (8) \end{aligned}$$

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.

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

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

From the random variable X_j, we define two random variables p_{X_j} and q_{X_j}. If X_j follows the distribution $\tilde{p}$, then the random variable p_{X_j} takes the value $p(\eta)$ with probability $p(\eta)$ and if X_j follows the distribution $\tilde{q}$, then p_{X_j} takes the value $p(\eta)$ with probability $q(\eta)$. Similarly, for q_{X_j}. For $j = 1, \ldots, N$, define

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

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.

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

The LLR random variable is defined to be the following.

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

Here $Q_\eta = \#\{j : X_j = \eta\}$. The LLR based test statistics for distinguishing between $\tilde{p}$ and $\tilde{q}$ is taken to be the following.

$$T = \frac{\text{LLR}/N - \mu_1}{\sigma_1/\sqrt{N}}. \quad (13)$$

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. For concrete values of N, it is difficult to determine conditions under which the error can be assumed to be less than a pre-specified bound. See [28] for details of this analysis. For the present, we proceed with the normal approximations.

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

$$P_e = \frac{\alpha + \beta}{2}. \tag{14}$$

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, for

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

the expression for P_e is the following.

$$P_e = \Phi(-t) = \Phi\left(-\frac{\sqrt{N}|\mu_0 - \mu_1|}{\sigma_0 + \sigma_1}\right) = \Phi\left(-\frac{\sqrt{N}|D(\tilde{p}||\tilde{q}) + D(\tilde{q}||\tilde{p})|}{\sigma_0 + \sigma_1}\right). \tag{15}$$

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 $\tilde{q}$. The Taylor series expansion involves certain conditions which restricts the applicability of the distinguisher. This has been pointed out in [28].

From the expression for P_e given by (15), it is possible to obtain an expression for the data complexity N required to achieve a desired value of P_e.

$$N = \left(\frac{(\sigma_0 + \sigma_1)\Phi^{-1}(1 - P_e)}{D(\tilde{p}||\tilde{q}) + D(\tilde{q}||\tilde{p})}\right)^2. \tag{16}$$

3 Tail Probabilities

3.1 Chernoff-Hoeffding 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 [25, 24] and are usually referred to as the Chernoff-Hoeffding bounds.

Theorem 1. *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:*

$$\text{For any } \delta > 0,\ \Pr[X \geq (1+\delta)\mu] < \left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right)^{\mu}. \tag{17}$$

$$\text{For any } 0 < \delta < 1,\ \Pr[X \leq (1-\delta)\mu] \leq \left(\frac{e^{-\delta}}{(1-\delta)^{(1-\delta)}}\right)^{\mu}. \tag{18}$$

These bounds can be simplified to the following form.

$$\text{For any } 0 < \delta \leq 1,\ \Pr[X \geq (1+\delta)\mu] \leq e^{-\mu\delta^2/3}. \tag{19}$$

$$\text{For any } 0 < \delta < 1,\ \Pr[X \leq (1-\delta)\mu] \leq e^{-\mu\delta^2/2}. \tag{20}$$

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

$$\text{For any } \delta > 0,\ \Pr[X \geq (1+\delta)\mu] \leq e^{-\delta^2\mu}. \tag{21}$$

$$\text{For any } 0 < \delta < 1,\ \Pr[X \leq (1-\delta)\mu] \leq e^{-\delta^2\mu}. \tag{22}$$

3.2 Martingales

The description of martingales that follows is for discrete random variables. Details can be found in standard texts such as [15, 24]. We start with the definition of conditional expectation.

Definition 1 (Conditional Expectation)**.** *Let X and Y be two random variables such that $E[X] < \infty$. Define*

$$\psi(y) \stackrel{\Delta}{=} E[X|Y = y] = \sum_x x \Pr[X = x|Y = y].$$

Thus, $E[X|Y = y]$ is a function of y. The conditional expectation of X given Y is defined to be $\psi(Y)$ and is written as $\psi(Y) \stackrel{\Delta}{=} E[X|Y]$. So, the conditional expectation of X given Y is a random variable $\psi(Y)$ which is a function of the random variable Y.

The following are several standard properties of conditional expectation.

Proposition 1. *1. $E[E[Y \mid X]] = E[X]$.*

2. If X has a finite expectation and if g is a function such that $Xg(Y)$ has a finite expectation, then $E[Xg(Y) \mid Y] = E[X \mid Y]\, g(Y)$.

3. $E\left[(X - g(Y))^2\right] \geq E\left[(X - E[X \mid Y])^2\right]$ for any pair of random variables X and Y such that X^2 and $g(Y)^2$ have finite expectations.

4. For any function g, such that $g(X)$ has finite expectation, $E[g(X) \mid Y = y] = \sum_x g(x) \Pr[X = x \mid Y = y]$.

5. $\mid E[X \mid Y] \mid \leq E[\mid X \mid\mid Y]$.

6. $E[E[X \mid Y, Z] \mid Y] = E[X \mid Y]$.

7. $E[E[g(X,Y) \mid Z, W] \mid Z] = E[g(X,Y) \mid Z]$.

Definition 2 (Martingale)**.** *A sequence of random variables $Z_1, Z_2, Z_3, \ldots$ is a martingale with respect to another sequence of random variables $Y_1, Y_2, Y_3, \ldots$ if for all $n \geq 1$ the following two conditions hold.*

1. $E[|Z_n|] < \infty$.

2. $E[Z_{n+1}|Y_1, Y_2, \ldots, Y_n] = Z_n$.

If $Z_n = Y_n$ for all $n \geq 1$ then the sequence is a martingale with respect to itself.

The basic Azuma-Hoeffding inequality for martingales is the following.

Theorem 2. *Let, $Z_0, Z_1, Z_2, \ldots$ be a martingale with respect $Y_0, Y_1, Y_2, \ldots$ and suppose that there exists a sequence $\upsilon_1, \upsilon_2, \ldots$ of real numbers such that for all $i \geq 1$, $\mid Z_i - Z_{i-1} \mid \leq \upsilon_i$. Then for any integer $\lambda > 0$ and real $\delta > 0$*

$$\Pr[Z_\lambda - Z_0 \geq \delta] \leq e^{-\delta^2/(2\sum_{i=1}^{\lambda} \upsilon_i^2)}; \quad (23)$$

$$\Pr[Z_\lambda - Z_0 \leq -\delta] \leq e^{-\delta^2/(2\sum_{i=1}^{\lambda} \upsilon_i^2)}. \quad (24)$$

A simple way to construct a martingale is the following. Let $Y_0, Y_1, \ldots Y_\lambda$ be a sequence of random variables and Y is a random variable with $E[\mid Y \mid] < \infty$. Define $Z_i = E[Y \mid Y_0, Y_1, \ldots, Y_i]$ for $i = 0, 1, \ldots, n$. Then using properties of conditional expectation given in Proposition 1, it is easy to see that the following condition holds.

$$E[Z_{i+1} \mid Y_0, Y_1, \ldots, Y_i] = Z_i.$$

So, $\{Z_\lambda\}$ is a martingale with respect to $\{Y_\lambda\}$. A martingale of this type is called a **Doob Martingale**.

To apply the Azuma-Hoeffding inequality, it is required to ensure that the differences $|Z_i - Z_{i-1}|$ are bounded. A general technique for obtaining a Doob martingale with bounded differences is as follows. A function $f(y_1, y_2, \ldots, y_\lambda)$ is said to satisfy the υ-**Lipschitz condition**, if for any i and for any set of values $y_1, y_2, \ldots, y_\lambda$ and y_i',

$$\mid f(y_1, y_2, \ldots, y_{i-1}, y_i, y_{i+1}, \ldots, y_\lambda) - f(y_1, y_2, \ldots, y_{i-1}, y_i', y_{i+1}, \ldots, y_\lambda) \mid \leq \upsilon.$$

That is by changing the value of any single coordinate changes the value of the function by at most υ. Let $Y_1, \ldots, Y_\lambda$ be a finite sequence of random variables and set

$$\begin{aligned} Z_0 &= E[f(Y_1, Y_2, \ldots, Y_\lambda)] \\ Z_i &= E[f(Y_1, Y_2, \ldots, Y_\lambda) \mid Y_1, Y_2, \ldots, Y_i]. \end{aligned}$$

Then $Z_0, Z_1, \ldots, Z_\lambda$ form a Doob martingale with respect to $Y_1, \ldots, Y_\lambda$. Further, if the random variables Y_i's are independent it can be shown that $|Z_i - Z_{i-1}| \leq \upsilon$. The martingale $Z_0, \ldots, Z_\lambda$ satisfies the conditions of Theorem 2 and so the inequality stated in the theorem applies to this martingale.

A special martingale: In our application, the function f will simply be the sum of its arguments. For later convenience, we provide the details of this special case.

Let $Y_1, Y_2, \ldots, Y_\lambda$ be a sequence independent and identically distributed random variables having finite mean μ and suppose that υ is such that for any two elements y and y' in the support of the Y_i's, $\max_{y,y'} |y - y'| = \upsilon$.

Let $Y = f(Y_1, \ldots, Y_\lambda) = \sum_{i=1}^{\lambda} Y_i$. Define a sequence of random variables $Z_0, Z_1, Z_2, \ldots, Z_\lambda$, where $Z_0 = E[Y] = \lambda\mu$ and for all $i \in \{1, 2, \ldots, \lambda\}$, $Z_i = E[Y | Y_1, Y_2, \ldots, Y_i]$. Then the sequence $Z_1, Z_2, \ldots, Z_\lambda$ is a Doob martingale with respect to $Y_1, Y_2, \ldots, Y_\lambda$. Further, using properties of conditional expectation given by Proposition 1, it can be shown that

$$Z_\lambda = E[Y \mid Y_1, \ldots, Y_\lambda] = E[Y_1 + \cdots + Y_\lambda \mid Y_1, \ldots, Y_\lambda] = Y_1 + \cdots + Y_\lambda. \tag{25}$$

For $1 \leq i \leq \lambda$,

$$\left|f(y_1, \ldots, y_{i-1}, y, y_{i+1}, \ldots, y_\lambda) - f(y_1, \ldots, y_{i-1}, y', y_{i+1}, \ldots, y_\lambda)\right| \leq \max_{y,y'} | y - y' | = \upsilon. \tag{26}$$

This shows that the function f is υ-Lipschitz and so $|Z_i - Z_{i-1}| \leq \upsilon$. Then by Theorem 2, for any real $\delta > 0$, we obtain

$$\Pr[Y_1 + \cdots + Y_\lambda - E[Y_1 + \cdots + Y_\lambda] \geq \delta] = \Pr[Z_\lambda - Z_0 \geq \delta] \leq e^{-\delta^2/(2\lambda\upsilon^2)}; \tag{27}$$

$$\Pr[Y_1 + \cdots + Y_\lambda - E[Y_1 + \cdots + Y_\lambda] \leq -\delta] = \Pr[Z_\lambda - Z_0 \leq -\delta] \leq e^{-\delta^2/(2\lambda\upsilon^2)}. \tag{28}$$

4 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 $\mu_1 = 1/2$. Following the hypothesis testing framework, we will be testing the null hypothesis "H_0: κ is correct" versus the alternate hypothesis "H_1: κ is incorrect." The test statistics is $T_\kappa = | X_\kappa - N\mu_1 |$ where $X_\kappa = \sum_{j=1}^{N} L_{\kappa,j}$. Under H_0, $E[X_\kappa] = N\mu_0$ and under H_1, $E[X_\kappa] = N\mu_1$. The decision rule is to reject H_0 if $T_\kappa \leq t$. The actual value of t is to be determined later.

Given the above hypothesis testing setting, the Type-I and Type-II error probabilities can be determined.

Define

$$\delta_0 = (|\mu_0 - \mu_1| - t/N)/\mu_0. \tag{29}$$

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 (19) and (20) of Theorem 1 with this δ_0.

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

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

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 (20) of Theorem 1.

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,

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

The last step follows from (19) of Theorem 1. Let

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

so that we obtain $\Pr[\text{Type-I error}] \leq \alpha$ 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

$$t = N \times |\mu_0 - \mu_1| - \sqrt{3N\mu_0 \ln(1/\alpha)}. \tag{30}$$

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

$$\delta_1 = t/(N\mu_1) \tag{31}$$

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 (21) and (22) of Theorem 1, we obtain

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

Let

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

so that $\Pr[\text{Type-II error}] \leq \beta$. Solving for t in terms of β and using $0 < t/N < |\mu_0 - \mu_1|$ yields

$$t = \sqrt{N\mu_1 \ln\left(\frac{2}{\beta}\right)}. \tag{32}$$

Eliminating t from (30) and (32), we obtain

$$\begin{aligned}
& N \times |\mu_0 - \mu_1| - \sqrt{3N\mu_0 \ln\left(\frac{1}{\alpha}\right)} = \sqrt{N\mu_1 \ln\left(\frac{2}{\beta}\right)} \\
\Rightarrow \quad & \frac{N|c|}{2} - \sqrt{\frac{3N}{2}(1+c)\ln\left(\frac{1}{\alpha}\right)} = \sqrt{\frac{N}{2}\ln\left(\frac{2}{\beta}\right)}; \\
\Rightarrow \quad & N = \frac{2\left(\sqrt{\ln\left(\frac{2}{\beta}\right)} + \sqrt{3\,(1+c)\ln\left(\frac{1}{\alpha}\right)}\right)^2}{c^2}.
\end{aligned} \tag{33}$$

The two expressions for t given by (30) and (32) 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 (33) satisfies both these lower bounds.

Recall that $c = 2(-1)^{z^*}|\epsilon|$. So, depending on the value of z^*, (33) provides two expressions for N, with the expression for $z^* = 1$ being (slightly) greater than the expression for $z^* = 0$. Since the value of z^* will not be known in advance, an upper bound on the data complexity is obtained by choosing $z^* = 1$ and is given by the following expression.

$$N \leq \frac{2\left(\sqrt{\ln\left(\frac{2}{\beta}\right)} + \sqrt{3\left(1+|c|\right)\ln\left(\frac{1}{\alpha}\right)}\right)^2}{c^2}. \tag{34}$$

5 Distinguishers: A Martingale Based Approach

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 $\text{LLR} = Y_1 + \cdots + Y_N$.

We wish to use LLR to design a test of hypothesis to distinguish between $\tilde{p}$ and $\tilde{q}$. The postulated hypotheses are the null hypothesis "H_0: the distribution is $\tilde{p}$" versus the alternate hypothesis "H_1: the distribution is $\tilde{q}$". 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. The test takes the following form.

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

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

The difference to Section 2.3 is that we do not wish to use normal approximations. The analysis of the error probabilities will still require bounds on probabilities and our goal is to obtain these bounds using the Azuma-Hoeffding inequality. For this, it is necessary to define a martingale. The method of doing this is described next. Define

$$\upsilon = \max_{\eta,\eta' \in \{0,\ldots,\nu-1\}} |\ln(p_\eta/q_\eta) - \ln(p_{\eta'}/q_{\eta'})| = \max_{\eta,\eta' \in \{0,\ldots,\nu-1\}} |\ln(p_\eta q_{\eta'}/(p_{\eta'} q_\eta))|. \tag{35}$$

Then for any $y_1, \ldots, y_{i-1}, y_i, y_{i+1}, \ldots, y_N, y_i'$ taking values from the set $\{\ln(p_0/q_0), \ldots, \ln(p_\nu/q_\nu\}$, we have

$$|(y_1 + \cdots + y_{i-1} + y_i + y_{i+1} + \cdots + y_N) - (y_1 + \cdots + y_{i-1} + y_i' + y_{i+1} + \cdots + y_N)| = |y_i - y_i'| < \upsilon.$$

From this it follows that the function $f(y_1, \ldots, y_N) = y_1 + \cdots + y_N$ is υ-Lipschitz. We now build a Doob martingale as described in Section 3. Define

$$\begin{aligned} Z_0 &= E[\text{LLR}] = E[f(Y_1, \ldots, Y_N)] = E[Y_1 + \cdots + Y_N]; \\ Z_j &= E[\text{LLR}|Y_1, \ldots, Y_j] \quad \text{for } j = 1, \ldots, N. \end{aligned}$$

The sequence $Z_0, Z_1, \ldots, Z_N$ forms a Doob martingale with respect to $Y_1, \ldots, Y_N$. Further, since $Y_1, \ldots, Y_N$ are independent and f is υ-Lipschitz, it follows that $|Z_i - Z_{i-1}| \leq \upsilon$. Thus, the Azuma-Hoeffding inequality holds for the martingale $Z_0, \ldots, Z_N$. Note that $Z_N = \text{LLR}$ and $Z_0 = E[\text{LLR}]$.

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 $\mu_0 > \mu_1$:

$$\begin{aligned}
\Pr[\text{Type-I error}] &= \Pr[\text{LLR} \leq t | H_0 \text{ holds}] \\
&= \Pr[Z_N \leq t | H_0 \text{ holds}] \\
&= \Pr[Z_N - Z_0 \leq t - Z_0 | H_0 \text{ holds}] \\
&= \Pr[Z_N - Z_0 \leq -(N\mu_0 - t) | H_0 \text{ holds}] \\
&\leq \exp\left(-\frac{(N\mu_0 - t)^2}{2N\upsilon^2}\right).
\end{aligned}$$

The last inequality follows from (28). Similarly, the probability of Type-II error is computed as follows.

$$\begin{aligned}
\Pr[\text{Type-II error}] &= \Pr[\text{LLR} > t | H_1 \text{ holds}] \\
&= \Pr[Z_N > t | H_1 \text{ holds}] \\
&= \Pr[Z_N - Z_0 > t - Z_0 | H_1 \text{ holds}] \\
&= \Pr[Z_N - Z_0 > t - N\mu_1 | H_1 \text{ holds}] \\
&\leq \exp\left(-\frac{(t - N\mu_1)^2}{2N\upsilon^2}\right).
\end{aligned}$$

The last inequality follows from (27).

Case $\mu_0 < \mu_1$:

$$\begin{aligned}
\Pr[\text{Type-I error}] &= \Pr[\text{LLR} \geq t | H_0 \text{ holds}] \\
&= \Pr[Z_N \geq t | H_0 \text{ holds}] \\
&= \Pr[Z_N - Z_0 \geq t - Z_0 | H_0 \text{ holds}] \\
&= \Pr[Z_N - Z_0 \geq t - N\mu_0 | H_0 \text{ holds}] \\
&\leq \exp\left(-\frac{(t - N\mu_0)^2}{2N\upsilon^2}\right).
\end{aligned}$$

The last inequality follows from (27). Similarly, the probability of Type-II error is computed as follows.

$$\begin{aligned}
\Pr[\text{Type-II error}] &= \Pr[\text{LLR} < t | H_1 \text{ holds}] \\
&= \Pr[Z_N < t | H_1 \text{ holds}] \\
&= \Pr[Z_N - Z_0 < t - Z_0 | H_1 \text{ holds}] \\
&= \Pr[Z_N - Z_0 < -(N\mu_1 - t) | H_1 \text{ holds}] \\
&\leq \exp\left(-\frac{(N\mu_1 - t)^2}{2N\upsilon^2}\right).
\end{aligned}$$

The last inequality follows from (28). Let

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

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

$$P_e = \frac{1}{2}\left(\Pr[\text{Type-I error}] + \Pr[\text{Type-II error}]\right) \leq \frac{\alpha + \beta}{2}.$$

Setting $t = N(\mu_0 + \mu_1)/2$ ensures $\alpha = \beta$ and then we obtain the following upper bound on P_e.

$$P_e \leq \exp\left(-\frac{N(\mu_0-\mu_1)^2}{2\upsilon^2}\right) = \exp\left(-\frac{N(D(\tilde{p}||\tilde{q})+D(\tilde{q}||\tilde{p}))^2}{2\upsilon^2}\right). \quad (36)$$

From the expression for P_e given by (36), the expression for data complexity N for a given value of P_e is obtained to be the following.

$$N = \frac{2\upsilon^2 \ln(1/P_e)}{\left(D(\tilde{p}||\tilde{q})+D(\tilde{q}||\tilde{p})\right)^2}. \quad (37)$$

The expression for N given by (37) does not involve normal approximation. We later compare to the expression for N given by (16) obtained using normal approximation.

6 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

$$\begin{aligned} L_{\kappa,j,i} &= \langle \Gamma_P^{(i)}, P_j \rangle \oplus \langle \Gamma_P^{(i)}, B_{\kappa,j} \rangle; \\ X_{\kappa,j} &= (L_{\kappa,j,1}, L_{\kappa,j,2}, \ldots, L_{\kappa,j,\ell}); \\ Y_{\kappa,z,j} &= \ln\left(p_z(X_{\kappa,j})/2^{-\ell}\right) = \ln\left(2^\ell p_z(X_{\kappa,j})\right). \end{aligned}$$

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 hypothesis testing problem is to test the null hypothesis "H_0: κ is correct" versus the alternate hypothesis "H_1: κ is incorrect." 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 fo all z (see [28] 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). The test statistics is defined to be

$$\mathrm{LLR}_{\kappa,z} = Y_{\kappa,z,1} + \cdots + Y_{\kappa,z,N} = \sum_{\eta \in \{0,1\}^\ell} Q_{\kappa,\eta} \ln(2^\ell p_z(\eta)) \quad (38)$$

where $Q_{\kappa,\eta} = \#\{j : X_{\kappa,j} = \eta\}$. 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.

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

$\mu_0 > \mu_1$: Reject H_0 if $\mathrm{LLR}_{\kappa,z} \leq t$ for all $z \in \{0,1\}^\ell$. Here t is in the range $N\mu_1 < t < N\mu_0$;
$\mu_0 < \mu_1$: Reject H_0 if $\mathrm{LLR}_{\kappa,z} \geq t$ for all $z \in \{0,1\}^\ell$. Here t is in the range $N\mu_0 < t < N\mu_1$.

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 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. As in Section 5, we set up a martingale and use the Azuma-Hoeffding inequality to bound the probabilities of the two types of errors.

Writing $f(y_1, \ldots, y_N) = y_1 + \ldots + y_N$, we have that f is υ-Lipschitz where

$$\upsilon = \max_{\eta,\eta' \in \{0,1\}^\ell} |\ln(2^\ell p_\eta) - \ln(2^\ell p'_\eta)| = \max_{\eta,\eta' \in \{0,1\}^\ell} |\ln(p_\eta / p'_\eta)|. \tag{39}$$

Then $\mathrm{LLR}_{\kappa,z} = f(Y_{\kappa,z,1}, \ldots, Y_{\kappa,z,N})$. As in Section 5, define the following random variables.

$$\begin{aligned} Z_0 &= E[\mathrm{LLR}_{\kappa,z}] = E[f(Y_{\kappa,z,1}, \ldots, Y_{\kappa,z,N})] = E[Y_{\kappa,z,1} + \cdots + Y_{\kappa,z,N}]; \\ Z_j &= E[\mathrm{LLR}_{\kappa,z} | Y_{\kappa,z,1}, \ldots, Y_{\kappa,z,j}] \quad \text{for } j = 1, \ldots, N. \end{aligned}$$

Then $Z_0, \ldots, Z_N$ form a Doob martingale with respect to $Y_{\kappa,z,1}, \ldots, Y_{\kappa,z,N}$ and further by the υ-Lipschitz condition, $|Z_i - Z_{i-1}| \le \upsilon$. So, the Azuma-Hoeffding inequality applies to this martingale. Note that under H_0, $E[\mathrm{LLR}_{\kappa,z}] = N\mu_0$ and under H_1, $E[\mathrm{LLR}_{\kappa,z}] = N\mu_1$. Also, note that $Z_N = \mathrm{LLR}_{\kappa,z}$.

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 $\mu_0 > \mu_1$:

$$\begin{aligned} \Pr[\text{Type-I error}] &= \Pr[\mathrm{LLR}_{\kappa,z} \le t \text{ for all } z | H_0 \text{ holds}] \\ &\le \Pr[\mathrm{LLR}_{\kappa,z^*} \le t | H_0 \text{ holds}] \\ &= \Pr[Z_N \le t | H_0 \text{ holds}] \\ &= \Pr[Z_N - Z_0 \le t - Z_0 | H_0 \text{ holds}] \\ &= \Pr[Z_N - Z_0 \le -(N\mu_0 - t) | H_0 \text{ holds}] \\ &\le \exp\left(-\frac{(N\mu_0 - t)^2}{2N\upsilon^2}\right). \end{aligned}$$

The last inequality follows from (28). Similarly, the probability of Type-II error is computed as follows.

$$\begin{aligned} \Pr[\text{Type-II error}] &= \Pr[\mathrm{LLR}_{\kappa,z} > t \text{ for some } z | H_1 \text{ holds}] \\ &\le \sum_{z \in \{0,1\}^\ell} \Pr[\mathrm{LLR}_{\kappa,z} > t | H_1 \text{ holds}] \\ &\le 2^\ell \left(\Pr[\mathrm{LLR}_{\kappa,z} > t | H_1 \text{ holds}]\right) \\ &\le 2^\ell \left(\Pr[Z_N > t | H_1 \text{ holds}]\right) \\ &\le 2^\ell \left(\Pr[Z_N - Z_0 > t - Z_0 | H_1 \text{ holds}]\right) \\ &\le 2^\ell \left(\Pr[Z_N - Z_0 > t - N\mu_1 | H_1 \text{ holds}]\right) \\ &\le 2^\ell \exp\left(-\frac{(t - N\mu_1)^2}{2N\upsilon^2}\right). \end{aligned}$$

The last inequality follows from (27). Define

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

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

$$t = N\mu_0 - \upsilon\sqrt{2N\ln\left(\frac{1}{\alpha}\right)}. \tag{40}$$

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

$$N > \frac{2\upsilon^2\ln(1/\alpha)}{(\mu_0-\mu_1)^2}. \tag{41}$$

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

$$t = N\mu_1 + \upsilon\sqrt{2N\ln\left(\frac{1}{\beta}\right)} \tag{42}$$

and

$$N > \frac{2\upsilon^2\ln(2^\ell/\beta)}{(\mu_0-\mu_1)^2}. \tag{43}$$

From equation (40) and (42), we get

$$N = 2\upsilon^2\left(\frac{\sqrt{\ln\left(\frac{2^\ell}{\beta}\right)}+\sqrt{\ln\left(\frac{1}{\alpha}\right)}}{D(\tilde{p}\,||\,\tilde{q})+D(\tilde{q}\,||\,\tilde{p})}\right)^2. \tag{44}$$

The expression for N given by (44) satisfies the bounds in (41) and (43).

Case $\mu_0 < \mu_1$:

$$\begin{aligned}
\Pr[\text{Type-I error}] &= \Pr[\text{LLR}_{\kappa,z} \geq t \text{ for all } z|H_0 \text{ holds}]\\
&= \Pr[\text{LLR}_{\kappa,z^*} \geq t|H_0 \text{ holds}]\\
&= \Pr[Z_N \geq t|H_0 \text{ holds}]\\
&= \Pr[Z_N - Z_0 \geq t - Z_0|H_0 \text{ holds}]\\
&= \Pr[Z_N - Z_0 \geq t - N\mu_0|H_0 \text{ holds}]\\
&\leq \exp\left(-\frac{(t-N\mu_0)^2}{2N\upsilon^2}\right).
\end{aligned}$$

The last inequality follows from (28). Similarly, the probability of Type-II error is computed as follows.

$$\begin{aligned}
\Pr[\text{Type-II error}] &= \Pr[\text{LLR}_{\kappa,z} < t \text{ for some } z|H_1 \text{ holds}]\\
&\leq \sum_{z\in\{0,1\}^\ell}\Pr[\text{LLR}_{\kappa,z} < t|H_1 \text{ holds}]\\
&\leq 2^\ell\left(\Pr[\text{LLR}_{\kappa,z} < t|H_1 \text{ holds}]\right)\\
&\leq 2^\ell\left(\Pr[Z_N < t|H_1 \text{ holds}]\right)\\
&\leq 2^\ell\left(\Pr[Z_N - Z_0 > t - Z_0|H_1 \text{ holds}]\right)\\
&\leq 2^\ell\left(\Pr[Z_N - Z_0 > t - N\mu_1|H_1 \text{ holds}]\right)\\
&\leq 2^\ell\exp\left(-\frac{(t-N\mu_1)^2}{2N\upsilon^2}\right).
\end{aligned}$$

The last inequality follows from (28). 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 (44).

7 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, $\Pr[D] = 1/(2^m - 1)$. Let $p_w = 1/(2^m - 1)$. 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.

Let $\mu_0 = p_0$ and $\mu_1 = p_w$. The hypothesis testing framework is applied to test the null hypothesis "H_0: κ is correct" versus the alternate hypothesis "H_1: κ is incorrect." The test statistics is $T_\kappa = |X_\kappa - \mu_1|$, where $X_\kappa = W_{\kappa,1} + \cdots + W_{\kappa,N}$. Under H_0, $E[X_\kappa] = N\mu_0$ and under H_1, $E[X_\kappa] = N\mu_1$. The decision rule is to reject the null hypothesis if $T_\kappa \leq t$ for a suitable threshold t.

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 cannot apply the bounds (21) and (22) of Theorem 1 and instead have to use the bounds (19) and (20) to upper bound this probability.

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 (19) and (20), 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 following expression for N.

$$N = \frac{3\left(\sqrt{\mu_0\ln(1/\alpha)}+\sqrt{\mu_1\ln(2/\beta)}\right)^2}{(\mu_0-\mu_1)^2} = \frac{3\left(\sqrt{p_0\ln(1/\alpha)}+\sqrt{p_w\ln(2/\beta)}\right)^2}{(p_0-p_w)^2} \quad (45)$$

To apply the bounds of Theorem 1, it is required that $0 < \delta_0, \delta_1 < 1$. As in Section 4, 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 (45) satisfies these two lower bounds.

8 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:

$$X_\kappa = \begin{cases} i & \text{if } R_\kappa^{(r)-1}\left(E_{K^{(r)}}(P)\right) \oplus R_\kappa^{(r)-1}\left(E_{K^{(r)}}(P\oplus\delta_0)\right) = \delta_{r-1}^{(i)} \\ 0; & \text{otherwise.} \end{cases} \quad (46)$$

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

$$\Pr[X_\kappa = i] = \begin{cases} p_i & \text{if } \kappa \text{ is the correct choice;} \\ \theta & \text{if } \kappa \text{ is an incorrect choice.} \end{cases} \quad (47)$$

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

$$p_0 = 1 - (p_1 + \cdots + p_\nu); \quad (48)$$

$$\theta_0 = 1 - \nu\theta. \quad (49)$$

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 (2) 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 $Q_{\kappa,\eta} = \#\{j : Y_{\kappa,j} = \eta\}$.

The hypothesis testing framework tests the null hypothesis "H_0: κ is correct" versus the alternate hypothesis "H_1: κ is incorrect." The actual test takes the following form.

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

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

For $y_1, \ldots, y_N$ taking values from the set $\{\ln(p_0/\theta_0), \ldots, \ln(p_\nu/\theta_\nu)\}$, define $f(y_1, \ldots, y_N) = y_1 + \cdots + y_N$. Then f is υ-Lipschitz where

$$\upsilon = \max_{\eta,\eta' \in \{0,\ldots,\nu\}} |\ln((p_\eta \theta'_\eta)/(p'_\eta \theta_\eta))|. \tag{50}$$

Define

$$\begin{aligned} Z_0 &= E\left[\text{LLR}_\kappa\right]; \\ Z_j &= E\left[\text{LLR}_\kappa \mid \text{Y}_{\kappa,1}, \text{Y}_{\kappa,\text{j}}, \ldots, \text{Y}_{\kappa,\text{j}}\right] \text{ for } j \geq 1. \end{aligned}$$

Note that $Z_0 = E\left[\text{LLR}_\kappa\right] = N\mu_0$ under H_0 and $Z_0 = E\left[\text{LLR}_\kappa\right] = N\mu_1$ under H_1. Further, $Z_N = \text{LLR}_\kappa$. Since f is υ-Lipschitz, it follows that $|Z_j - Z_{j-1}| \leq \upsilon$ for $j = 1, \ldots, N$. The sequence $Z_0, \ldots, Z_N$ forms a Doob martingale (with respect to $Y_{\kappa,1}, \ldots, Y_{\kappa,N}$) to which the Azuma-Hoeffding bound can be applied. 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-1 and Type-2 errors are computed as follows:

$$\begin{aligned} \Pr[\text{Type-1 Error}] &= \Pr\left[\text{LLR}_\kappa \leq \text{t} | \text{H}_0 \text{ holds}\right] \\ &= \Pr\left[Z_N - Z_0 \leq -(N\mu_0 - t) | H_0 \text{ holds}\right] \\ &\leq \exp\left(-\frac{(N\mu_0 - t)^2}{2N\upsilon^2}\right); \\ \Pr[\text{Type-2 Error}] &= \Pr\left[\text{LLR}_\kappa > \text{t} | \text{H}_1 \text{ holds}\right] \\ &= \Pr\left[Z_N - Z_0 > t - N\mu_1 | H_1 \text{ holds}\right] \\ &\leq \exp\left(-\frac{(t - N\mu_1)^2}{2N\upsilon^2}\right). \end{aligned}$$

Here the inequalities given by (27) and (28) have been used. Define

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

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.

$$t = N\mu_0 - \upsilon\sqrt{2N \ln\left(\frac{1}{\alpha}\right)}; \tag{51}$$

$$t = N\mu_1 + \upsilon\sqrt{2N \ln\left(\frac{1}{\beta}\right)}. \tag{52}$$

Eliminating t from equations (51) and (52), we get

$$N = 2v^2 \left(\frac{\sqrt{\ln\left(\frac{1}{\beta}\right)} + \sqrt{\ln\left(\frac{1}{\alpha}\right)}}{D(\tilde{p} \,||\, \tilde{\theta}) + D(\tilde{\theta} \,||\, \tilde{p})} \right)^2 . \tag{53}$$

The expression for t given by (51) has to satisfy $N\mu_1 < t$ and the expression for t given by (52) 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 (53).

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 (53).

9 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 get 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

$$\beta = \frac{2^m}{2^m - 1} 2^{-a}. \tag{54}$$

The right hand side can be approximated by 2^{-a} for moderate values of m. It is possible to use (54) 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 (17), 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

$$\beta = \frac{s}{(1+\delta)\,(2^m - 1)} \tag{55}$$

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

$$\Pr\left[W > s\right] < \left(\frac{e^{\frac{s-\mu}{\mu}}}{\left(\frac{s}{\mu}\right)^{\left(\frac{s}{\mu}\right)}} \right)^{\mu} = \frac{e^{s-\mu}\mu^s}{s^s} = P_\beta \text{ (say)}. \tag{56}$$

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 (55), 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).

10 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 9, 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 $a^\star \geq a$? Let $N_{\min}$ denote this minimum required data complexity.

The data complexity expressions that we have obtained earlier provides an expression for N in terms of P_S and a. In other words, this means N plaintext-ciphertext pairs are sufficient to obtain $P_S^\star \geq P_S$ and $a^\star \geq a$. From the definition of $N_{\min}$, it follows that in each case

$$N_{\min} \quad \leq \quad N. \tag{57}$$

So, all the expressions for data complexities that we have obtained are upper bounds on the minimum data complexities required to achieve at least a certain success probability and a certain expected advantage. In particular, we note that our analysis does not involve any approximation (normal or otherwise) and hence these are proper upper bounds. It is in this spirit that we call the obtained upper bounds to be rigorous upper bounds.

The issue of not using any approximations needs a further clarification. The statistical analysis is based upon probabilities of linear and differential relations obtained through an intricate analysis of the structure of the block cipher. Such an analysis may itself involve approximations. Such approximations are not avoided in our approach. Our work only avoids making approximations as part of the statistical analysis itself.

11 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 rigourous 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 as follows.

$$\begin{aligned}\left(\sigma_0^{(L)}\right)^2 &= \sum_{\eta=0}^{2^{\ell}-1} p(\eta)\left(\ln\left(\frac{p(\eta)}{2^{-\ell}}\right)\right)^2 - D(\tilde{p} \,||\, \tilde{p}_{\$})^2;\\ \left(\sigma_1^{(L)}\right)^2 &= \sum_{\eta=0}^{2^{\ell}-1} 2^{-\ell}\left(\ln\left(\frac{2^{-\ell}}{p(\eta)}\right)\right)^2 - D(\tilde{p}_{\$} \,||\, \tilde{p})^2.\end{aligned}$$

For multiple differential cryptanalysis we denote the variances as

$$\begin{aligned}\left(\sigma_0^{(D)}\right)^2 &= \sum_{\eta=0}^{\nu} p(\eta)\left(\ln\left(\frac{p(\eta)}{\theta(\eta)}\right)\right)^2 - D(\tilde{p} \,||\, \tilde{\theta})^2;\\ \left(\sigma_1^{(D)}\right)^2 &= \sum_{\eta=0}^{\nu} \theta(\eta)\left(\ln\left(\frac{\theta(\eta)}{p(\eta)}\right)\right)^2 - D(\tilde{\theta} \,||\, \tilde{p})^2.\end{aligned}$$

Lastly, for the LLR distinguisher we denote the variances as

$$\begin{aligned}\left(\sigma_0^{(\mathrm{Dist})}\right)^2 &= \sum_{\eta=0}^{\nu-1} p(\eta)\left(\ln\left(\frac{p(\eta)}{q(\eta)}\right)\right)^2 - D(\tilde{p} \,||\, \tilde{q})^2;\\ \left(\sigma_1^{(\mathrm{Dist})}\right)^2 &= \sum_{\eta=0}^{\nu-1} q(\eta)\left(\ln\left(\frac{q(\eta)}{p(\eta)}\right)\right)^2 - D(\tilde{q} \,||\, \tilde{p})^2.\end{aligned}$$

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 [29]. The approximate expression for the data complexity of multiple linear cryptanalysis was obtained in [16] while the approximate expression for the data complexity of multiple differential crypanalysis was obtained in [10]. These expressions were obtained using the order statistics based approach. In [28], 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 [28]. 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 (16).

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	$\frac{\left\{\sigma_1^{(L)}\Phi^{-1}\left(1-2^{-\ell-a}\right)+\sigma_0^{(L)}\Phi^{-1}(P_S)\right\}^2}{(D(\tilde{p}\lvert\lvert\tilde{p}_\$)+D(\tilde{p}_\$\lvert\lvert\tilde{p}))^2}$	$\frac{2\upsilon^2\left\{\sqrt{(a+\ell)\ln 2}+\sqrt{\ln(1/(1-P_S))}\right\}^2}{(D(\tilde{p}\lvert\lvert\tilde{q})+D(\tilde{q}\lvert\lvert\tilde{p}))^2}$
Multiple DC	$\frac{\left\{\sigma_1^{(D)}\Phi^{-1}\left(1-2^{-a}\right)+\sigma_0^{(D)}\Phi^{-1}(P_S)\right\}^2}{(D(\tilde{p}\lvert\lvert\tilde{\theta})+D(\tilde{\theta}\lvert\lvert\tilde{p}))^2}$	$\frac{2\upsilon^2\left\{\sqrt{a\ln 2}+\sqrt{\ln(1/(1-P_S))}\right\}^2}{\left(D(\tilde{p}\lvert\lvert\tilde{\theta})+D(\tilde{\theta}\lvert\lvert\tilde{p})\right)^2}$
Distinguisher	$\frac{\left\{\left(\sigma_0^{(\mathrm{Dist})}+\sigma_1^{(\mathrm{Dist})}\right)\Phi^{-1}(1-P_e)\right\}^2}{(D(\tilde{p}\lvert\lvert\tilde{q})+D(\tilde{q}\lvert\lvert\tilde{p}))^2}$	$\frac{2\upsilon^2\ln(1/P_e)}{(D(\tilde{p}\lvert\lvert\tilde{q})+D(\tilde{q}\lvert\lvert\tilde{p}))^2}$

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.

11.1 Experimental Comparison

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 $A = 1, 2, 3, 6$. 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 $\sqrt{A\ln(1/(1-x))}/\Phi^{-1}(x)$. 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 experiments, we took c in the range $(-2^{-40}, 2^{-40})$ and in this range $\sqrt{1-c^2} \approx 1 \approx \sqrt{1+\mid c\mid}$.

Comparing $\sigma_0^{(L)}$ and $\sigma_1^{(L)}$ with $\sqrt{2}\upsilon$: 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 $\sqrt{2}\upsilon$, were then compared by computing the ratios $\sqrt{2}\upsilon/\sigma_0^{(L)}$, $\sqrt{2}\upsilon/\sigma_1^{(L)}$ 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 2 gives the values of $\sqrt{2}\upsilon$, $\sigma_0^{(L)}$ and $\sqrt{2}\upsilon/\sigma_0^{(L)}$.

Comparing $\sigma_0^{(D)}$ and $\sigma_1^{(D)}$ with $\sqrt{2}\upsilon$: This arises in the case of multiple differential cryptanalysis. For the simulation we took $m = 10$ and $\nu = 20$ and again ensured that $\epsilon_\eta \in (-2^{-40}, 2^{-40})$ for all $\eta = 0, 1, \ldots, 20$. Random

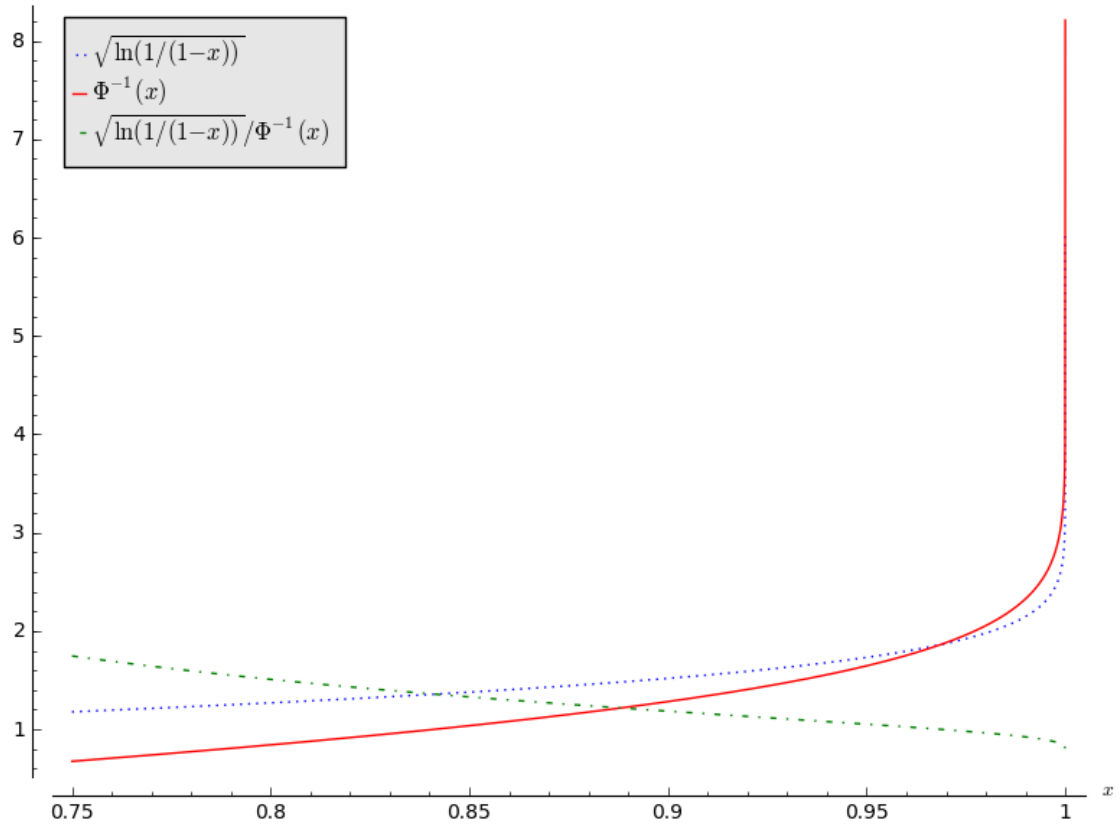

Figure 1: Plots of $\Phi^{-1}(x)$, $\sqrt{\ln(1/(1-x))}$ and $\sqrt{\ln(1/(1-x))}/\Phi^{-1}(x)$.

$\sqrt{2}\upsilon$	$\sigma_0^{(L)}$	$\sqrt{2}\upsilon/\sigma_0^{(L)}$
1.58×10^{-10}	2.54×10^{-11}	6.23
1.03×10^{-10}	2.01×10^{-11}	5.12
1.13×10^{-10}	2.05×10^{-11}	5.52
7.76×10^{-11}	1.38×10^{-11}	5.60
1.19×10^{-10}	2.21×10^{-11}	5.38
2.83×10^{-10}	4.46×10^{-11}	6.33
2.29×10^{-10}	3.71×10^{-11}	6.16
7.69×10^{-11}	1.83×10^{-11}	4.20
1.80×10^{-10}	2.93×10^{-11}	6.13
3.87×10^{-10}	6.32×10^{-11}	6.12

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

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 3 gives the values of $\sqrt{2}\upsilon$, $\sigma_0^{(D)}$ and $\sqrt{2}\upsilon/\sigma_0^{(D)}$.

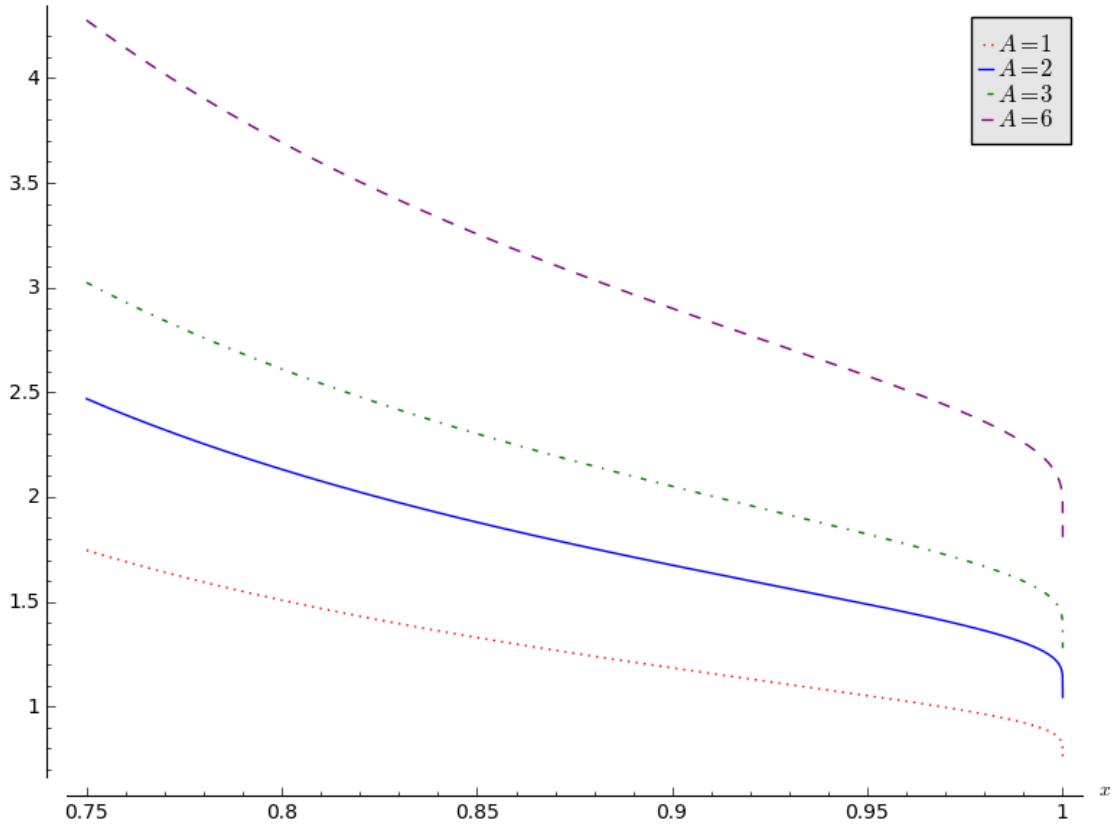

Figure 2: Plots of $\sqrt{A\ln(1/(1-x))}/\Phi^{-1}(x)$ for $A = 1, 2, 3$ and 6.

$\sqrt{2}\upsilon$	$\sigma_0^{(D)}$	$\sqrt{2}\upsilon/\sigma_0^{(D)}$
1.93×10^{-9}	5.36×10^{-11}	35.95
2.48×10^{-9}	8.24×10^{-11}	30.06
2.42×10^{-9}	7.93×10^{-11}	30.45
2.56×10^{-9}	7.35×10^{-11}	34.86
2.38×10^{-9}	6.74×10^{-11}	35.38
2.23×10^{-9}	7.34×10^{-11}	30.33
2.43×10^{-9}	8.28×10^{-11}	29.35
1.86×10^{-9}	4.85×10^{-11}	38.32
2.24×10^{-9}	6.92×10^{-11}	32.32
2.47×10^{-9}	7.22×10^{-11}	34.25

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

Comparing $(\sigma_0^{(\text{Dist})} + \sigma_1^{(\text{Dist})})$ with $\sqrt{2}\upsilon$: 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 $\sqrt{2}\upsilon/(\sigma_0^{(\text{Dist})} + \sigma_1^{(\text{Dist})})$. As before the experiment was repeated 10 times and the observations are listed in Table 4.

$\sqrt{2}\upsilon$	$\sigma_0^{(\text{Dist})} + \sigma_1^{(\text{Dist})}$	$\sqrt{2}\upsilon/(\sigma_0^{(\text{Dist})} + \sigma_1^{(\text{Dist})})$
1.78×10^{-10}	5.60×10^{-11}	3.17
2.19×10^{-10}	6.97×10^{-11}	3.14
7.85×10^{-11}	4.07×10^{-11}	1.93
8.24×10^{-11}	3.55×10^{-11}	2.32
1.43×10^{-10}	4.95×10^{-11}	2.90
7.93×10^{-11}	3.51×10^{-11}	2.26
1.89×10^{-10}	6.11×10^{-11}	3.09
2.25×10^{-10}	7.36×10^{-11}	3.06
7.83×10^{-11}	3.46×10^{-11}	2.26
1.32×10^{-10}	4.54×10^{-11}	2.90

Table 4: Table showing the values of $\sqrt{2}\upsilon$, $(\sigma_0^{(\text{Dist})} + \sigma_1^{(\text{Dist})})$ 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 $m = 10$. 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 $p_w = 1/(2^m - 1)$ 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 1 reports these minimums and maximums. For the case of the LLR distinguisher, it is required that $\alpha = \beta$ and hence for our example, $a = 5$. Since this is a single value of a, we ran the experiment for this value of a 100 times and recorded the minimum and the maximum. The last row of Table 1 reports these values.

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	1703.14	448.57	1345.41	472.68	1474.51	452.98
Multiple LC	43.74	25.42	29.46	17.46	27.20	16.50
LLR Distinguisher	10.13	4.43	8.35	2.87	7.33	2.84

Table 5: 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.

From Table 5 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. Even for multiple differential cryptanalysis, the upper bound is not too much greater than the approximate value. Further, 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 2, 3 and 4.

While 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.

Comparing the two expressions for the upper bounds of 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 (44), we get

$$\begin{aligned}
\upsilon^2 &= \left\{\max\left\{\left|\ln\left(\frac{1+c}{1-c}\right)\right|, \left|\ln\left(\frac{1-c}{1+c}\right)\right|\right\}\right\}^2 \\
&= \left(\ln\left(\frac{1+c}{1-c}\right)\right)^2; \\
\mu_0 &= \frac{1}{2}\left[\ln(1-c^2) + \ln\left(\frac{1+c}{1-c}\right)\right]; \\
\mu_1 &= \frac{1}{2}\ln(1-c^2); \text{ and} \\
N &= \frac{8\{\sqrt{(a+1)\ln 2} + \sqrt{\ln(1/(1-P_S))}\}^2}{c^2}.
\end{aligned}$$

This needs to be compared with the expression obtained using the Chernoff bound, and 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 ratio x/y as z, the data complexity obtained using Chernoff bound as N_C and the data complexity obtained using Azuma-Hoeffding inequality as N_{AH}. Then,

$$\begin{aligned}
N_{AH} - N_C &= \frac{2y^2}{c^2}\{4(z+1)^2 - (z+\sqrt{3})^2\} \\
&= \frac{2y^2}{c^2}\{3z^2 + 1 + 2z(4-\sqrt{3})\} \\
&> 0; \quad \text{[Since, } x \text{ and } y \text{ are greater than zero].}
\end{aligned}$$

Thus, we have $N_{AH} > N_C$, which means that the data complexity obtained directly using the Chernoff 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 (53), we get

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

For single differential cryptanalysis, an analytical comparison is complicated. Hence, we opted for an experimental comparison of the two data complexities. For the experiment, we again took three values of $P_S = 1-2^{-5}, 1-2^{-7}$ and $1-2^{-10}$ and for each value of P_S, a was varied from 1 to 100. Also, as before, we took the size of the target sub-key m to be 10. and $p_0 = p_w + c$, where $p_w = 1/(2^m - 1)$ and c is randomly chosen from $(-2^{-40}, 2^{-40})$.

The ratio N_{AH}/N_C was computed. It was seen that in all simulations, N_{AH}/N_C is greater than 1. This indicates that the Chernoff bound based data complexity gives a better upper bound in case of single differential cryptanalysis.

12 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. Experiments 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.

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 Baigneres, 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] 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.

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

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

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

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

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

[10] 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.

[11] 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.

[12] 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.

[13] Joan Daemen and Vincent Rijmen. Probability Distributions of Correlation and Differentials in Block Ciphers. *Journal of Mathematical Cryptology JMC*, 1(3):221–242, 2007.

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

[15] Geoffrey Grimmett and David Stirzaker. *Probability and Random Processes*. Oxford university press, 2001.

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

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

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

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

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

[21] 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.

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

[23] 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.

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

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

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

[27] 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.

[28] Subhabrata Samajder and Palash Sarkar. Another look at normal approximations in cryptanalysis. Cryptology ePrint Archive, Report 2015/679, 2015. `http://eprint.iacr.org/`.

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

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

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