A New Test Statistic for Key Recovery Attacks Using Multiple Linear Approximations

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

January 19, 2017

Abstract

The log-likelihood ratio (LLR) and the chi-squared distribution based test statistics have been proposed in the literature for performing statistical analysis of key recovery attacks on block ciphers. A limitation of the LLR test statistic is that its application requires the full knowledge of the corresponding distribution. Previous work using the chi-squared approach required *approximating* the distribution of the relevant test statistic by chi-squared and normal distributions. Problematic issues regarding such approximations have been reported in the literature. Perhaps more importantly, both the LLR and the chi-squared based methods are applicable only if the success probability P_S is greater than 0.5. On the other hand, an attack with success probability less than 0.5 is also of considerable interest. This work proposes a new test statistic for key recovery attacks which has the following features. Its application does not require the full knowledge of the underlying distribution; it is possible to carry out an analysis using this test statistic without using any approximations; the method applies for all values of the success probability. The statistical analysis of the new test statistic follows the hypothesis testing framework and uses Hoeffding's inequalities to bound the probabilities of Type-I and Type-II errors.

keywords: multiple linear cryptanalyis, LLR statistic, chi-squared statistic, Hoeffding inequality.

1 Introduction

Consider the setting of multiple linear cryptanalysis of block ciphers. Statistical analyses of such attacks proceed by identifying a suitable test statistic. In purely statistical terms, the setting is as follows. Let $X_1, \ldots, X_N$ be independent and identically distributed random variables taking values from the set $\{0,1\}^\ell$. The distribution of the X_j's is either a distribution $\tilde{p} = (p_0, \ldots, p_{2^\ell-1})$ or it is the uniform distribution on $\{0,1\}^\ell$. For $\eta \in \{0,1\}^\ell$, let Q_η be the random variable which counts the number of j's such that $X_j = \eta$. The following test statistics have been used in the literature on block cipher cryptanalysis. Assume $\ell > 1$.

$$\mathrm{LLR} = \sum_{\eta=0}^{2^\ell-1} Q_\eta \ln(2^\ell p_\eta); \quad \Lambda = 2^\ell N \sum_{\eta=0}^{2^\ell-1} (Q_\eta/N - 2^{-\ell})^2.$$

The LLR test statistic arises from the log-likelihood ratio while the distribution of Λ can be approximated by a chi-squared distribution. By the chi-squared test statistic, we will mean Λ. Approximate expressions for data complexities of key recovery attacks using the LLR and the chi-squared test statistics have been obtained in [14]. Both the LLR and the chi-squared test statistics have some limitations which are mentioned below.

Knowledge of the distribution: To apply the LLR test statistic, it is required to have *full* knowledge of the probability distribution $\tilde{p}$. In many situations, this information may be difficult to obtain. The distribution $\tilde{p}$ is uncovered by a detailed analysis of the block cipher and for $\ell > 1$, obtaining the full distribution $\tilde{p}$ may not be possible. In such situations, it is not possible to apply the LLR test statistic.

To apply the chi-squared test statistic, the knowledge of $\tilde{p}$ is not required. The analysis needs to only unearth the expected value of the test statistic which is one of the factors that determine the number of plaintext-ciphertext pairs required to mount the attack. So, to apply an analysis based on the chi-squared test statistic, the requirement from the analysis of the block cipher is substantially lower than that required from the LLR test statistic.

Approximation issues: For both the LLR and the chi-squared test statistics, the analysis in [14] approximates the corresponding distributions by normal. This involves an error in approximation which has not been studied in details. For the chi-squared based test statistic, this issue has been briefly noted in the literature [15, 14]. For detailed analysis of problems arising from normal approximations we refer to [27].

Works only for high success probability: The success probability of a key recovery attack is the probability that the target sub-key is indeed recovered by the algorithm. While it is good to have high success probabilities, from a cryptanalytic point of view, low success probabilities are also meaningful. For example, an attack with success probability 0.1 has a 10% chance of success. Such an attack should be considered to be a valid attack. It is helpful for a cryptanalyst to determine the amount of plaintext-ciphertext pairs required to achieve a certain success probability. The LLR and the chi-squared based approaches have serious limitations with respect to this requirement. Both the approaches are applicable only for success probabilities greater than 0.5. So, if one wishes to obtain an estimate of data complexity for an attack with 10% chance of success, then there is nothing in the literature which allows doing this.

Our Contributions

In this work, we propose to perform a statistical analysis which overcomes the previously mentioned limitations. This requires a suitable test statistic.

Our first choice is the chi-squared test statistic. For this, we considered the possibility of performing an analysis without making any approximations. We follow the hypothesis testing framework. An approach for avoiding approximations in this framework has been outlined in [26]. The idea is to apply the Hoeffding bounds to upper bound the probabilities of Type-I and Type-II errors. This requires expressing the test statistic as a sum of *independent* random variables. Unfortunately, for the chi-squared test statistic, this does not seem to be possible.

Since neither the LLR nor the chi-squared test statistics seem to apply, we propose a new test statistic. For $\eta \in \{0,1\}^\ell$, let $\underline{\eta}$ denote the integer whose binary representation is η. Let d be a positive real number. We propose the test statistic $T = \sum_{\eta \in \{0,1\}^\ell} \underline{\eta}^d Q_\eta$. The computation of this statistic does not require information about $\tilde{p}$. Let μ_0 (resp. μ_1) be the expectation of T when the X_j's follow $\tilde{p}$ (resp. the uniform distribution). If $\mu_0 \neq \mu_1$, then T can be used to carry out a key recovery attack. The requirement from the analysis of the internal structure of the block cipher is to obtain (an estimate of) μ_0. Given the value of μ_0, it is possible to obtain an expression for the data complexity (i.e., the number of plaintext-ciphertext pairs) required to attain the parameters of a successful attack.

The statistical analysis that we perform does not require us to make any approximations. It is possible to express T as a sum of independent random variables. So, the Hoeffding bounds can be used to bound the probabilities of the Type-I and Type-II errors.

The theoretical analysis holds for any positive d. The question that arises is what value of d should be used in practice. An important point to keep in mind is that for the chosen value of d, it should be possible to estimate the value of μ_0. Based on experiments, we suggest that the value of d should be taken to be 1.

We have evaluated the obtained bound using known linear approximations for the block cipher SERPENT. For success probabilities at most 0.5, there is no prior result in the literature to which we can compare. For success probabilities greater than 0.5, the values of the bounds turn out to be higher than the approximate values obtained using the chi-squared test statistic.

Note that the minimum required data complexity to achieve a certain value of the success probability is not known. Our method provides an upper bound on this minimum data complexity while the chi-squared method provides an approximate value where the error in approximation is not known. So, the data complexity of the chi-squared method cannot be taken to be the correct value and then the bound obtained by our method is criticised for being an over-estimate. It is possible that the chi-squared method grossly under-estimates the minimum required data complexity.

A work of independent interest would be to simulate an attack on some particular (toy) cipher to determine the required data complexity and then compare it to the bound that we obtain and the approximate value obtained from the chi-squared method. If done in a comprehensive manner this could be an interesting exercise, but, one that we feel is not directly related to the contribution of the paper. We have obtained a theoretical bound which holds for all ciphers. This in itself should be of some intrinsic interest. Note that the upper bound on the data complexity obtained in this paper depends only on the value of μ_0, provided such an estimate exists and that the estimate is an *accurate* one.

While we do not make any claims that the bound is tight, we do note that carrying out an experiment for one particular cipher will not establish the bound to be loose. There may be ciphers for which the bound is loose while there could be other ciphers for which the bound is tight. Further, it is difficult to extrapolate results on data complexity obtained from simulation of an attack on some toy cipher to results about much higher data complexity on more complex and real-life ciphers. More work is required to establish the tightness (or not) of the bounds obtained here. We refer to [28] for a discussion on this issue. Another equally important issue is to be able to propose another test statistic which shares the advantages of the one that we use and for which it is possible to obtain lower values of the data complexity.

Previous and Related Work

Linear cryptanalysis was proposed by Matsui in [21] as an attack on DES and involved a single linear approximation of the cipher. Later, in [22], Matsui used two linear approximations (which were assumed to be independent) to improve the attack. Independently, Kaliski and Robshaw [20] extended Matsui's attack involving single linear approximations to multiple linear cryptanalysis using $\ell \geq 1$ independent linear approximations. The approximations that were considered had certain restrictions. It was assumed that the ℓ linear approximations have a common data mask (i.e., plaintext and ciphertext mask) but different key masks.

In [3], Biryukov et al. gave a more general method for multiple linear cryptanalysis without any assumption on the corresponding linear approximations. Their analysis, though, still assumed the linear approximations to be independent. Analysis under the independence assumption was also done independently by Junod and Vaudenay in [19] in the context of distinguishing attacks. Further work on distinguishing attacks without the independence assumption was carried out in [1, 18, 2, 10]. Murphy [24] argued that the independence assumption need not be valid.

Junod [17] gave a detailed analysis of Matsui's ranking method [21, 22]. This work introduced the notion of ordered statistics in linear cryptanalysis. This was further developed by Selçuk in [29], where he used a well known asymptotic result from the theory of ordered statistic to arrive at the expression for success probability for both single linear and differential cryptanalysis.

The test statistic used in [1, 18, 2] was the log-likelihood ratio (LLR). The chi-squared test statistic was initially used by Handschuh and Gilbert [11] for the cryptanalysis of the SEAL encryption algorithm. Later Johansson and Maximov [16] gave an explicit analysis of the success and the error probabilities in the context of their attack on the stream cipher Scream. The idea of Selçuk's order statistics based approach has been combined with the LLR and the chi-squared test statistics to obtain expressions for data complexities of multiple linear cryptanalysis [14].

A related line of work considered the situation where the correlation for a linear approximation depends on the key. This line of research originates from the work of Daemen and Rijmen [9] and was explicitly put in the context of linear cryptanalysis in [6] for single linear cryptanalysis and in [5] for multi-dimensional linear cryptanalysis.

In this paper, we have not considered the issue of key-dependent correlations. The problems with the use of normal approximations for linear cryptanalysis without key-dependent behaviour (reported in [27]) also extend to the case of key-dependent correlations. Further, there are several additional subtleties which need to be properly handled. Carefully analysing the setting of key-dependent correlations without approximations requires a separate comprehensive treatment. The goal of the present paper, on the other hand, is primarily to show how several limitations of previous statistical methods for analysing multiple linear cryptanalysis can be overcome. We believe that the usefulness of this contribution can be assessed independently of the issue of key-dependent correlations.

2 Multiple Linear Cryptanalysis

Let $E : \{0,1\}^k \times \{0,1\}^n \mapsto \{0,1\}^n$ be a block cipher, and so for each $K \in \{0,1\}^k$, $E_K(\cdot) \triangleq E(K, \cdot)$ is a bijection from $\{0,1\}^n$ to itself. Here, K is called the secret key, the n-bit input to E_K is called the plaintext and the n-bit output of E_K is called the ciphertext.

Usual constructions of block ciphers involve a simple round function parameterised by a round key which is iterated over several rounds. The round keys are produced by applying an expansion function, called the key scheduling algorithm, to the secret key K. Denote the round keys by $k^{(0)}, k^{(1)}, \ldots$ and round functions by $R^{(0)}_{k^{(0)}}, R^{(1)}_{k^{(1)}}, \ldots$. Also, let $K^{(i)}$ denote the concatenation of the first i round keys, i.e., $K^{(i)} = k^{(0)} \mid\mid \cdots \mid\mid k^{(i-1)}$ and $E^{(i)}_{K^{(i)}}$ denote the composition of the first i round functions, i.e.,

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

Suppose that an attack targets $r+1$ rounds. For a plaintext P, we denote by B the output after r rounds, i.e, $B = E^{(r)}_{K^{(r)}}(P)$ and we denote by C the output after $r+1$ rounds, i.e., $C = E^{(r+1)}_{K^{(r+1)}}(P) = R^{(r)}_{k^{(r)}}(B)$.

Block cipher cryptanalysis starts off with a detailed analysis of the block cipher. This results in one or possibly more relations between the plaintext P, the input to the last round B and possibly the expanded key $K^{(r)}$. In case of linear cryptanalysis these relations are linear in nature and are of the following form:

$$\langle \Gamma^{(i)}_P, P \rangle \oplus \langle \Gamma^{(i)}_B, B \rangle = \langle \Gamma^{(i)}_K, K^{(r)} \rangle; \quad i = 1, 2, \ldots, \ell;$$

where $\Gamma^{(i)}_P, \Gamma^{(i)}_B \in \{0,1\}^n$ and $\Gamma^{(i)}_{K^{(r)}} \in \{0,1\}^{nr}$ denotes the plaintext mask, the mask to the input of the last round and the key mask respectively. A linear relation of the above form is called a linear approximation of the block cipher. Such linear approximations usually hold with some probability which is taken over the uniform random choices of the plaintext P. Obtaining such relations and their joint distribution is not a trivial task and requires a lot of ingenuity and experience. They form the basis on which the statistical analysis of block ciphers are built. If $\ell > 1$, the attack is called a multiple linear cryptanalysis and if $\ell = 1$, we call the attack single linear cryptanalysis, or simply, linear cryptanalysis. Define $L_i \triangleq \langle \Gamma^{(i)}_P, P \rangle \oplus \langle \Gamma^{(i)}_B, B \rangle;$ for $i = 1, 2, \ldots, \ell$.

Inner key bits: Let $z_i = \langle \Gamma^{(i)}_K, K^{(r)} \rangle; \quad i = 1, \ldots, \ell$. Note that for a fixed but unknown key $K^{(r)}$, z_i represents a single unknown bit. Denote by $z = (z_1, \ldots, z_\ell)$ the collection of the bits arising in this manner. Since, the ℓ

key masks $\Gamma_K^{(1)}, \ldots, \Gamma_K^{(\ell)}$ are known, the tuple z is determined only by the unknown but fixed $K^{(r)}$. Hence, there is no randomness either of $K^{(r)}$ or z. The bits of z are called the inner key bits.

Target sub-key bits: Any linear relation of the form above, between P and B, usually involves only a subset of the bits of B. When $\ell > 1$, several (or multiple) relations between P and B are known. In such cases, it is required to consider the subset of the bits of B which covers all the relations. In order to obtain these bits from the ciphertext C it is required to partially decrypt C by one round. This involves a subset of the bits of the last round key $k^{(r)}$. We call this the target sub-key. The goal of linear cryptanalysis is then to find the correct value of the target sub-key using the ℓ linear approximations and their joint distributions. We denote the number of bits in the target sub-key by m. In other words, these m key bits are sufficient to partially decrypt C by one round and obtain the bits of B involved in any of the ℓ linear approximations. Notice that there are 2^m possible choices of the target sub-key out of which only one is correct. The purpose of the attack is to identify the correct key. For convenience of notation, we will denote the correct choice of the target sub-key as κ^*.

Setting of the attack: The block cipher is instantiated with an unknown, but, fixed key. It is assumed that N independent and uniform random plaintexts are chosen and the corresponding ciphertexts under fixed key are obtained. Denote the plaintext-ciphertext pairs as $(P_j, C_j); j = 1, 2, \ldots, N$. For each choice κ of the target sub-key, it is possible for the attacker to partially decrypt each C_j by one round to obtain $B_{\kappa,j}; j = 1, 2, \ldots, N$. Note that $B_{\kappa,j}$ is dependent on κ even though C_j may not. Clearly, if the choice of κ is correct, then the C_j's depend on κ. On the other hand, for an incorrect choice of κ, C_j has no relation with κ.

Statistical analysis proceeds by defining a test statistic T_κ for each choice κ of the target sub-key. This provides 2^m random variables of the type T_κ. The distribution of T_κ depends on whether κ is the correct choice or, it is an incorrect choice. Under the usual wrong key hypothesis [12], it is assumed that the distributions of all the T_κ's for incorrect choices of κ's are the same.

Suppose that the plaintext P is uniformly distributed. Since, each round function is a bijection, the uniform distribution of P also induces a uniform distribution on B. By definition, L_i is a binary random variable taking values from the set $\{0, 1\}$. Also from the discussion above it is clear that the source of randomness of L_i comes from the randomness of P. Define the random variable X as $X = (L_1, \ldots, L_\ell)$. Then X is a random variable distributed over $\{0, 1\}^\ell$.

Joint distribution parameterised by inner key bits: The distribution of the random variable $X = (L_1, \ldots, L_\ell)$ is the following. For $\eta \in \{0, 1\}^\ell$ and $z \in \{0, 1\}^\ell$,

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

where $-1/2^\ell \leq \epsilon_\eta(z) \leq 1 - 1/2^\ell$. Denote by $\tilde{p}_z = (p_z(0), p_z(1), \ldots, p_z(2^\ell - 1))$ the corresponding probability distribution, where the integers $\{0, 1, \ldots, 2^\ell - 1\}$ are identified with the set $\{0, 1\}^\ell$. For each choice of z, we obtain a different but related distribution. Let, $z' = z \oplus \beta$ for some $\beta \in \{0, 1\}^\ell$, then it is easy to verify that $\epsilon_\eta(z') = \epsilon_{\eta\oplus\beta}(z)$, which implies that $p_{z\oplus\beta}(\eta) = p_z(\eta \oplus \beta)$. Let, $\tilde{p}$ denote the probability distribution $\tilde{p}_{0^\ell}$, i.e., $\tilde{p} \triangleq \tilde{p}_{0^\ell}$. Write $\tilde{p} = (p_0, \ldots, p_{2^\ell-1})$, so that for all $\eta \in \{0, 1\}^\ell$, $p_\eta \triangleq p(\eta) = 1/2^\ell + \epsilon_\eta$.

For $\kappa \in \{0, 1, \ldots, 2^m - 1\}$, $j = 1, \ldots, N$ and $i = 1, \ldots, \ell$, define $L_{\kappa,j,i} = \langle \Gamma_P^{(i)}, P_j \rangle \oplus \langle \Gamma_B^{(i)}, B_{\kappa,j} \rangle$; $X_{\kappa,j} = (L_{\kappa,j,1}, \ldots, L_{\kappa,j,\ell})$; and

$$Q_{\kappa,\eta} \quad = \quad \#\{j \in \{1, 2, \ldots, N\} : X_{\kappa,j} = \eta\}. \tag{2}$$

Note that $Q_{\kappa,\eta}$ is the number of times η appears among the random variables $X_{\kappa,1}, \ldots, X_{\kappa,N}$. Suppose z is the correct choice of the inner key bits. Then for the correct choice of the target sub-key (i.e., $\kappa = \kappa^*$) the random variable $Q_{\kappa,\eta}$ follows $\mathrm{Bin}(N, p_z(\eta))$, whereas for the incorrect choice of the target sub-key (i.e., $\kappa \neq \kappa^*$) the random variable $Q_{\kappa,\eta}$ follows $\mathrm{Bin}(N, 2^{-\ell})$. Denote the uniform distribution over the set $\{0, 1\}^\ell$ by $p_\$ = (2^{-\ell}, \ldots, 2^{-\ell})$.

Success probability and advantage of an attack: Two important parameters which are relevant to a key recovery attack are the success probability and the (expected) advantage. The success probability is the probability that the correct value of the target sub-key is recovered in the attack. The advantage of an attack is a, if a fraction 2^{-a} of all possible 2^m values of the target sub-key are reported as candidate values. So, for an attack with advantage a, the size of the list of candidate keys is 2^{m-a}.

3 Drawbacks of Previously Proposed Statistics

As mentioned in the introduction, two test statistics have been proposed earlier [14] for performing statistical analysis of key recovery attacks on block ciphers. In this section, we briefly review these statistics and point out certain drawbacks.

Log-likelihood ratio test statistic: The LLR test statistic has been used for key recovery attacks as well as distinguishing attacks in several works in the literature [1, 14, 4, 26]. One drawback of this statistics is that to compute it, the full knowledge of $\tilde{p}$ is required. This is evident from the expression of the LLR test statistic. In many situations, such complete knowledge of the joint distribution of the multiple linear approximations may not be available. In such cases, it will not be possible to compute the value of LLR_κ.

The analysis in [14] provides an expression for the data complexity in terms of the success probability and the advantage. This expression is stated to be valid only for success probability greater than 0.5.

Chi-squared test statistic: Recall from (2) that for a choice κ of the target sub-key and for $\eta \in \{0,1\}^\ell$, $Q_{\kappa,\eta}$ is the number of times η occurs among the random variables $X_{\kappa,1}, \ldots, X_{\kappa,N}$. Define a test statistic Λ_κ in the following manner:

$$\Lambda_\kappa \quad = \quad 2^\ell N \sum_{\eta=0}^{2^\ell - 1} (Q_{\kappa,\eta}/N - 2^{-\ell})^2. \tag{3}$$

For the correct choice κ^* of the target sub-key bits, the right hand side of (3) involves $Q_{\kappa^*,\eta}$ whose distribution depends on the inner key bits z. Due to the relation $p_{z\oplus\beta}(\eta) = p_z(\eta \oplus \beta)$, the distribution of Λ_{κ^*}, however, does not depend on z.

To apply the chi-squared test statistic, it is not required to know the full distribution of the underlying probability distribution. Statistical analysis using this test statistic has been carried out in [14] in the following manner. The distribution of $Q_{\kappa,\eta}$ follows a binomial for both correct and incorrect choices of κ. The binomial can be approximated using a normal distribution and then the distribution of Λ_κ approximately follows a chi-squared distribution for both correct and incorrect choices of κ. There is, however, the issue of error in approximation which has not been properly analysed. This issue of error in approximation has been briefly mentioned in [15, 14, 27] and has been analysed in details in [27] where several shortcomings have been pointed out.

The data complexity for the chi-squared test statistic was given by Hermelin et al. in [14]. It was shown that for "large" values of a and $P_S > 0.5$, the data complexity, which we denote by N_Λ, is approximately

$$N_\Lambda = \frac{2\sqrt{2^\ell - 1}\Phi^{-1}(1 - 2^{-a}) + 4\left(\Phi^{-1}(2P_S - 1)\right)^2}{C(\tilde{p})}; \tag{4}$$

where $C(\tilde{p}) = \sum_{\eta=0}^{2^\ell-1}(p_\eta - 2^{-\ell})/2^{-\ell}$.

A reduced round linear cryptanalysis of SERPENT was earlier reported in [8] using a set of linear approximations [7]. Out of these, a subset of 64 linear approximations was later used in [13, 14] to perform multidimensional linear cryptanalysis on SERPENT using the LLR and the chi-squared test statistics. It happens so that this subset can be generated by 10 linear approximations called the basis linear approximations and can be used to recover 10 bits of the last round key. Thus, for this particular experiment, $\ell = 10$ and $m = 10$.

It was pointed out in [27], that for a χ^2 approximation of the distribution of the test statistic Λ to be valid, the corresponding distributions under both the null and the alternate hypotheses need to satisfy the following two conditions for all $\eta \in \{0,1\}^\ell$: $\mid p_\eta(1-p_\eta)-q_\eta(1-q_\eta) \mid < p_\eta(1-p_\eta)$; and $\frac{p_\eta(1-p_\eta)-q_\eta(1-q_\eta)}{p_\eta(1-p_\eta)} \approx 0$. We checked whether these conditions hold for the linear approximations of the reduced round block cipher SERPENT reported in [7]. The total number of linear approximation required to generate the full probability distribution for the correct key is $2^{10} - 1 = 1023$. Out of these, only 64 are given in [7]. To find the full probability distribution for the correct key, two methods were suggested in [13]. We have used the second method, where the correlations of the remaining $1023 - 64 = 959$ approximations are assumed to be zero. The Walsh transform method of [25] was then used on these approximations to get the joint distribution.

For the joint distribution of the reduced round SERPENT it was found that for all η, $\mid p_\eta(1-p_\eta)-q_\eta(1-q_\eta) \mid$ is indeed less than $p_\eta(1-p_\eta)$. The maximum value of the ratio $\mid (p_\eta(1-p_\eta)-q_\eta(1-q_\eta)) \mid /(p_\eta(1-p_\eta))$ is 0.0049. So, the χ^2 approximation is valid provided that the value 0.0049 is *assumed* to be sufficiently close to zero. The effect of this assumption on the final expression for the data complexity is not known. This is one of the several approximations that is required to obtain the chi-squared based data complexity expression. We refer to [27] for more details.

A question then arises as to whether it is possible to use the chi-squared test statistic to obtain an expression for the data complexity *without* using any approximation. Such an approach has been shown to be successful for the LLR test statistic [26] through the application of the Hoeffding bounds. This requires expressing the test statistic as a sum of independent random variables. However, Λ_κ is the sum of 2^ℓ random variables where these individual random variables are determined by $Q_{\kappa,\eta}$, $\eta \in \{0,1\}^\ell$. The $Q_{\kappa,\eta}$'s are dependent as $\sum_{\eta\in\{0,1\}^\ell} Q_{\kappa,\eta} = N$. So, the Hoeffding bound does not apply directly. Further, there does not seem any other way to write Λ_κ as the sum of independent random variables.

4 A New Test Statistic

Let d be a positive integer and consider the following test statistic.

$$T_\kappa = \sum_{\eta\in\{0,1\}^\ell} \underline{\eta}^d Q_{\kappa,\eta}. \tag{5}$$

Let μ_0 be the expectation of T_κ for the correct choice of κ and let μ_1 be the expectation of T_κ for an incorrect choice of κ. Then

$$\mu_1 = E[T_\kappa] = \sum_{\eta\in\{0,1\}^\ell} \underline{\eta}^d E[Q_{\kappa,\eta}] = N2^{-\ell} \sum_{\eta\in\{0,1\}^\ell} \underline{\eta}^d; \tag{6}$$

$$\mu_0 = E[T_{\kappa^*}] = \sum_{\eta\in\{0,1\}^\ell} \underline{\eta}^d E[Q_{\kappa^*,\eta}] = \mu_1 + N \sum_{\eta\in\{0,1\}^\ell} \underline{\eta}^d \epsilon_\eta. \tag{7}$$

So, $\mu_0 - \mu_1 = N\sum_{\eta\in 0,1^\ell} \underline{\eta}^d \epsilon_\eta$. One can now aim to design a statistical analysis which attempts to recover κ^* by exploiting the difference in the two expectations. While doing this, we would like to avoid making any approximations. We next show how both of these aims can be achieved.

Recall that for a fixed κ, the random variables $X_{\kappa,1},\ldots,X_{\kappa,N}$ are independent. The test statistic given by (5) can be rewritten in the following manner.

$$T_\kappa = \sum_{\eta\in\{0,1\}^\ell} \underline{\eta}^d Q_{\kappa,\eta} = \sum_{j=1}^{N} \underline{X}^d_{\kappa,j}. \tag{8}$$

This enables writing T_κ as the sum of independent random variables. The computation of T_κ can be done in $O(N)$ time using any one of the two expressions. This computation does not require the knowledge of the ϵ_η's.

Consider the following test of hypothesis:

Hypothesis Test-1:

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

Decision rule:

Case $\mu_0 > \mu_1$: Reject H_0 if $T_\kappa \leq t, \forall z \in \{0,1\}^\ell$; where $t \in (\mu_1, \mu_0)$;

Case $\mu_0 < \mu_1$: Reject H_0 if $T_\kappa \geq t, \forall z \in \{0,1\}^\ell$; where $t \in (\mu_0, \mu_1)$.

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

$$N \geq \frac{(2^\ell - 1)^{2d}(\sqrt{\ln(1/\alpha)} + \sqrt{\ln(1/\beta)})^2}{2\left(\sum_{\underline{\eta}=0}^{2^\ell-1} \underline{\eta}^d \epsilon_\eta\right)^2} \tag{9}$$

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

The proof follows by applying Hoeffding's bound (see Appendix A) to upper bound the probabilities of the type-I and type-II errors, and thereafter eliminating the threshold parameter t. The proof is given in Appendix B.

Let $\mu_1' = 2^{-\ell} \sum_{\eta \in \{0,1\}^n} \underline{\eta}^d$ and $\mu_0' = \sum_{\eta \in \{0,1\}^n} \underline{\eta}^d (2^{-\ell} + \epsilon_\eta)$. Then $\mu_0' - \mu_1' = \sum_{\eta \in \{0,1\}^n} \underline{\eta}^d \epsilon_\eta$ and so (9) can be written as

$$N \geq \frac{(2^\ell - 1)^{2d}(\sqrt{\ln(1/\alpha)} + \sqrt{\ln(1/\beta)})^2}{2\left(\mu_0' - \mu_1'\right)^2}.$$

Thus, although (9) suggests that it is necessary to know all the ϵ_η's to get a lower bound of N, it is actually not the case. It suffices to have a good estimate of μ_0' which is just the expected value of the random variable $\underline{X}^d_{\kappa^*,1}$. (Note that $\underline{X}^d_{\kappa^*,1}, \ldots, \underline{X}^d_{\kappa^*,N}$ are identically distributed.)

Relating to success probability and expected advantage: By definition, the success probability is $1 - \Pr[\text{Type-I error}]$. So, if α is an upper bound on the probability of the type-I error, then $P_S = 1 - \alpha$ is a lower bound on the success probability.

An incorrect value of κ is reported as a candidate key if a Type-II error occurs. Since there are a total of $2^m - 1$ incorrect values of the target sub-key, the expected number of wrong values reported as candidate keys is $\beta(2^m - 1)$. Equating to 2^{m-a} gives $\beta = 2^{-a} \times 2^m/(2^m - 1)$.

In the expression for the data complexity N, we may replace α by $1 - P_S$ and β by $2^{-a} \times 2^m/(2^m - 1)$. This provides an expression for the data complexity required to attain success probability at least P_S and advantage at least a.

Nature of the bound: Proposition 1 shows a lower bound on the data complexity required for ensuring a certain minimum success probability and a certain minimum advantage. This lower bound is with respect to Hypothesis Test-1 which in particular means that the test statistic T_κ is used. We note, on the other hand, that there is a possibility of using some other test statistic for which the required data complexity is lower. This means that taken over all possible test statistics, the data complexity expression in Proposition 1 is actually an upper bound on the minimum data complexity that is required to achieve given values of success probability and advantage.

Attack procedure: The actual application of the attack will be as follows. Given P_S and a, determine α $(= 1 - P_S)$ and β $(= 2^{-a} \times 2^m/(2^m - 1))$; then determine N as given by the right hand side of (9). From α and N determine t (given by (13) or (14) of Appendix B). Once t is determined, Hypothesis Test-1 can be performed. Suppose that $\mu_0 > \mu_1$, the other case being similar. Initialise a list $\mathcal{L}$ to be empty. For each choice κ of the target sub-key, compute T_κ; if $T_\kappa > t$, append κ to $\mathcal{L}$. At the end, $\mathcal{L}$ contains the set of candidate keys.

The above procedure does not require knowledge of $\tilde{p}$ to apply the test. Only the knowledge of μ_0 is required to obtain an estimate of the data complexity N.

Choice of d: The theory described above works for all positive d. We suggest the use of $d = 1$. The rationale behind such a choice is given in Appendix C

5 Experimental Results for SERPENT

We compare the bound on the data complexity given by (9) to that of the approximate data complexity of the Λ-test statistic given by [14, Equation (18)] and reproduced in (4) for the reduced round block cipher SERPENT. The distribution used for all the computations in this section is the one discussed in Section 3. The comparison presented in this section has been broadly classified into two groups, one where P_S has been fixed to 0.95 and the other where experiments have been conducted for different values of P_S.

Fixed P_S: For this experiment, the value of P_S was fixed to 0.95. The bound given by (9) with $d = 1$ and the approximate value given by (4) were then computed for $a = 1, 2, \ldots, 10$. Table 1 summarises the output of the experiment. The last column of the Table gives the ratio of the two data complexities. From the Table, it is clear that approximate estimate obtained from the Λ test statistic is lower than the upper bound obtained from the new method.

We note that the minimum data complexity required to achieve success probability 0.95 and advantage a is not known. While N_X is an upper bound, N_Λ is an approximation where the error in approximation is not known. At present, it is not possible to say anything more than this.

a	N_X (9)	N_Λ (4)	N_X/N_Λ
1	2.79×10^{10}	1.25×10^{6}	22246.87
2	3.59×10^{10}	9.48×10^{6}	3783.91
3	4.27×10^{10}	1.53×10^{7}	2793.44
4	4.89×10^{10}	2.0×10^{7}	2449.77
5	5.47×10^{10}	2.4×10^{7}	2283.17
6	6.03×10^{10}	2.75×10^{7}	2190.17
7	6.56×10^{10}	3.07×10^{7}	2134.69
8	7.08×10^{10}	3.37×10^{7}	2100.80
9	7.58×10^{10}	3.64×10^{7}	2080.41
10	8.07×10^{10}	3.90×10^{7}	2068.96

Table 1: Values of N_X and N_Λ for the joint distribution of SERPENT with a ranging from 1 to 10 and $P_S = 0.95$.

P_S	N_X (9)	N_Λ (4)
0.10	2.03×10^{10}	n.a.
0.20	2.31×10^{10}	n.a.
0.30	2.56×10^{10}	n.a.
0.40	2.81×10^{10}	n.a.
0.50	3.08×10^{10}	n.a.
0.60	3.37×10^{10}	2.33×10^{7}
0.70	3.71×10^{10}	2.28×10^{7}
0.80	4.16×10^{10}	2.28×10^{7}
0.90	4.84×10^{10}	2.33×10^{7}

Table 2: Values of N_X and N_Λ for the joint distribution of SERPENT with $P_S = 0.1, 0.2, \ldots, 0.9$ and $a = 5$. In the table, n.a. denotes "not applicable."

Varying P_S: We computed the value of N_X for different values of $P_S = 0.1, 0.2, \ldots, 0.9$ for the same joint distribution of SERPENT. For this experiment we fixed $a = 5$. Table 2 reports the results of the experiment. From the table, it can be seen that the data complexity N_X increases as P_S increases, which is what one would expect. But, the data complexity N_Λ first increases then decreases even for $P_S > 0.5$. This anomalous behaviour is due to the approximations used in deriving the expression for N_Λ.

6 Conclusion

The paper considered the problem of statistical analysis of attacks on block ciphers in the situation where the LLR test statistic cannot be applied. The other aspect considered was to follow the approach in [26] towards a rigorous analysis without using any approximations. We first considered the chi-squared based test statistic and argued that this test statistic is not amenable to analysis using our approach.

To resolve the problem, we introduced a new test statistic using which an attack can be applied without the full knowledge of the underlying probability distribution. Also, the resulting statistical framework can be analysed rigorously without making any approximations. The obtained expression for data complexity was compared to the *approximate* expression for data complexity for the chi-squared test statistic using known linear approximations for the block cipher SERPENT. As expected, the data complexity of the new test statistic turns out to be higher. This shows that if one wishes to follow a rigorous approach, then one would have to be satisfied with a conservative estimate of the data complexity.

An important aspect of our analysis is that it allows obtaining estimates of the data complexity for all possible values of the success probability. This is in contrast to previous work which required the success probability to be greater than half.

References

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

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

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

[4] 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.

[5] Céline Blondeau and Kaisa Nyberg. Joint Data and Key Distribution of Simple, Multiple, and Multidimensional Linear Cryptanalysis Test Statistic and its Impact to Data Complexity. *Designs, Codes and Cryptography*, pages 1–31, 2016. DOI: 10.1007/s10623-016-0268-6, ISSN: 1573-7586.

[6] Andrey Bogdanov and Elmar Tischhauser. On the wrong key randomisation and key equivalence hypotheses in matsui's algorithm 2. In *Fast Software Encryption - 20th International Workshop, FSE 2013, Singapore, March 11-13, 2013. Revised Selected Papers*, pages 19–38, 2013. http://dx.doi.org/10.1007/978-3-662-43933-3_2.

[7] Baudoin Collard, François-Xavier Standaert, and Jean-Jacques Quisquater. 2008. http://www.dice.ucl.ac.be/fstandae/PUBLIS/50b.zip, accessed on 30^{th} July, 2014.

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

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

[10] Benoît Gérard and Jean-Pierre Tillich. On linear cryptanalysis with many linear approximations. In *IMA International Conference on Cryptography and Coding*, pages 112–132. Springer, 2009.

[11] Helena Handschuh and Henri Gilbert. χ^2 Cryptanalysis of the SEAL Encryption Algorithm. In *Fast Software Encryption*, pages 1–12. Springer, 1997.

[12] Carlo Harpes, Gerhard G. Kramer, and James L. Massey. A Generalization of Linear Cryptanalysis and the Applicability of Matsui's Piling-Up Lemma. In *Advances in Cryptology - EUROCRYPT '95, International Conference on the Theory and Application of Cryptographic Techniques, Saint-Malo, France, May 21-25, 1995, Proceeding*, pages 24–38, 1995. `http://link.springer.de/link/service/series/0558/bibs/0921/09210024.htm`.

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

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

[15] Miia Hermelin, Joo Yeon Cho, and Kaisa Nyberg. Statistical Tests for Key Recovery Using Multidimensional Extension of Matsui's Algorithm 1. In Helena Handschuh, Stefan Lucks, Bart Preneel, and Phillip Rogaway, editors, *Symmetric Cryptography*, number 09031 in Dagstuhl Seminar Proceedings, Dagstuhl, Germany, 2009. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, Germany. Available at `http://drops.dagstuhl.de/opus/volltexte/2009/1954`, ISSN: 1862-4405.

[16] Thomas Johansson and Alexander Maximov. A Linear Distinguishing Attack on Scream. In *Proceedings 2003 IEEE International Symposium on Information Theory*, pages 164–164. IEEE, 2003.

[17] Pascal Junod. On the Complexity of Matsui's Attack. In *Selected Areas in Cryptography*, pages 199–211. Springer, 2001.

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

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

[20] Burton S Kaliski Jr and Matthew JB Robshaw. Linear Cryptanalysis Using Multiple Approximations. In *Advances in Cryptology–Crypto'94*, pages 26–39. Springer, 1994.

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

[22] Mitsuru Matsui. The First Experimental Cryptanalysis of the Data Encryption Standard. In Y. G. Desmedt, editor, *Advances in Cryptology–Crypto'94*, pages 1–11. Springer, 1994.

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

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

[25] 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.

[26] Subhabrata Samajder and Palash Sarkar. Rigorous Upper Bounds on Data Complexities of Block Cipher Cryptanalysis. *IACR Cryptology ePrint Archive*, 2015:916, 2015. `http://eprint.iacr.org/2015/916`.

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

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

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

A Hoeffding Inequality

We briefly recall Hoeffding's inequality for sum of independent random variables. The result can be found in standard texts such as [23].

Theorem 2 (Hoeffding Inequality). *Let, $X_1, X_2, \ldots, X_\lambda$ be a finite sequence of independent random variables, such that for all $i = 1, \ldots, \lambda$, there exists real numbers $a_i, b_i \in \mathbb{R}$, with $a_i < b_i$ and $a_i \leq X_i \leq b_i$. Let $X = \sum_{i=1}^{\lambda} X_i$. Then for any positive $t > 0$,*

$$\begin{aligned} \Pr[X - E[X] \geq t] &\leq \exp\left(-\frac{2t^2}{D_\lambda}\right) && (10) \\ \Pr[X - E[X] \leq -t] &\leq \exp\left(-\frac{2t^2}{D_\lambda}\right) && (11) \\ \Pr[\mid X - E[X] \mid \geq t] &\leq 2\exp\left(-\frac{2t^2}{D_\lambda}\right); && (12) \end{aligned}$$

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

B Proof of Propositon 1

We provide the proof for the case $\mu_0 > \mu_1$ with the other case being similar. Recall that $\underline{X}^d_{\kappa,1}, \ldots, \underline{X}^d_{\kappa,N}$ are N independently and identically distributed random variables such that for all $j = 1, \ldots, N$

$$\upsilon_{\min} = 0 \leq \underline{X}^d_{\kappa,j} \leq (2^\ell - 1)^d = \upsilon_{\max}.$$

Let, $\upsilon = \upsilon_{\max} - \upsilon_{\min} = (2^\ell - 1)^d$ Thus Hoeffding bounds (see Section A) can be used on the sum of independently and identically distributed random variables $T_\kappa = \sum_{j=1}^{N} \underline{X}^d_{\kappa,j}$; where $D_N = N\upsilon^2$.

The probabilities of Type-I and Type-II errors are then given by

$$\begin{aligned}
\Pr[\text{Type-I Error}] &= \Pr[T_\kappa \leq t \mid H_0 \text{ holds}] = \Pr[T_\kappa - \mu_0 \leq -(\mu_0 - t) | H_0 \text{ holds}] \\
&\leq \exp\left(-\frac{2(\mu_0 - t)^2}{N\upsilon^2}\right); \quad [\text{By 11}]. \\
\Pr[\text{Type-II Error}] &= \Pr[T_\kappa > t \mid H_1 \text{ holds}] = \Pr[T_\kappa - \mu_1 > t - \mu_1] \mid H_1 \text{ holds}] \\
&\leq \exp\left(-\frac{2(t - \mu_1)^2}{N\upsilon^2}\right); \quad [\text{By 10}].
\end{aligned}$$

Let,

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

Then, using the fact that $\mu_1 < t < \mu_0$, we get

$$\sqrt{2}t = \sqrt{2}\mu_0 - \upsilon\sqrt{2N\ln(1/\alpha)} \quad (13)$$

$$\sqrt{2}t = \sqrt{2}\mu_1 + \upsilon\sqrt{N\ln(1/\beta)}. \quad (14)$$

Eliminating t from the above two equations and using the expressions for μ_0, μ_1 and υ, we get the expression given by the right hand side of (9). For any N greater than this value, the probabilities of Type-I and Type-II errors will be at most α and β respectively. □

C Choice of d

There are two factors that need to be kept in mind while while choosing a appropriate value of d.

1. The value of d has an effect on the data complexity. So, one should try to choose a value of d which minimises the data complexity.

2. For the chosen value of d, it should be possible to obtain an estimate of μ_0 through the analysis of the block cipher.

Regarding the first point, there does not seem to be a way to formally prove that one particular value of d will minimise the data complexity. Instead, we provide intuitive explanations and experimental evidence.

The statistic $T_\kappa = \sum_{j=1}^{N} \underline{X}_{\kappa,j}^d$. As d goes to zero, $X_{\kappa,j}^d$ goes to 1 and so the effect of $X_{\kappa,j}$ diminishes. Further, as $d \to 0$, $(2^\ell - 1)^d \to 1$ and $\underline{\eta}^d \to 1$ for all $\eta \in \{0,1\}^\ell$. So, the numerator of the data complexity expression given by (9) goes to a constant and the denominator goes to $\sum_{\eta \in \{0,1\}^\ell} \epsilon_\eta$. By definition, the later sum is 0. So, as $d \to 0$, the data complexity expression given by (9) goes to infinity. Experiments confirm this behaviour.

Based on the above, we do not consider values of $d < 1$. For values of $d = 1, \ldots, 100$, we have run experiments with the known linear approximations of SERPENT and have observed that the minimum data complexity is attained for $d = 1$ and $d = 2$. The values are shown in Table 3. To decide between these two values, we consider the second point mentioned above. Intuitively, it is easier to obtain the value of μ_0 for $d = 1$ than for $d = 2$. So, we suggest using $d = 1$ for defining the test statistic T_κ.

Negative values of d: Most of the theory that has been developed also works for negative values of d. The only problem is that for $\underline{\eta} = 0$, the value of $\underline{\eta}^d$ is undefined. This defect can be rectified by defining T_κ to be $\sum_{j=1}^{N}(1 + \underline{X}_{\kappa,j})^d$. Working out the details of this test statistic leads to $\upsilon = |2^{\ell d} - 1|$ and $|\mu_0 - \mu_1| = \sum_{\eta \in \{0,1\}^\ell}(1 + \underline{\eta})^d \epsilon_\eta$. The value of υ does not depend on the sign of d. Suppose $d > 0$, then the value of $|\mu_0 - \mu_1|$ with d is greater than the value of $|\mu_0 - \mu_1|$ with $-d$. As a result, the data complexity with d is lesser compared to the data complexity for $-d$. Due to this reason, we have not considered negative values of d.

a	Minimum Data Complexity	
	Value of d	Data Complexity
1	1, 2	2.79×10^{10}
2	1, 2	3.59×10^{10}
3	1, 2	4.27×10^{10}
4	1, 2	4.89×10^{10}
5	1, 2	5.47×10^{10}
6	1, 2	6.03×10^{10}
7	1, 2	6.56×10^{10}
8	1, 2	7.08×10^{10}
9	1, 2	7.58×10^{10}
10	1, 2	8.07×10^{10}

Table 3: Table showing the minimum data complexity over different values of d for the linear approximations of SERPENT with a ranging from 1 to 10.