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

April 22, 2016

Abstract

The log-likelihood ratio (LLR) test statistic has been proposed in the literature for performing statistical analysis of attacks on block ciphers. A limitation of the LLR test statistic is that its application requires the full knowledge of the corresponding distribution. Another test statistic which has been proposed in the literature does not possess this limitation. The statistical analysis using this test statistic requires *approximating* its distribution by a chi-squared distribution. Problematic issues regarding such approximations have been reported in the literature. This work proposes a new test statistic which offers the following two features. Its application does not require knowledge of the underlying distribution and it is possible to carry out an analysis using this test statistic without using any approximation. This is made possible by applying the theory of martingales to build a Doob martingale satisfying an appropriate Lipschitz condition so that the Azuma-Hoeffding inequality can be applied. Experimental comparison of the data complexity obtained using the new method is made to the data complexity obtained using the chi-squared based method for both simulated distributions and the previously reported distribution for the block cipher SERPENT. In all cases, the data complexity obtained by the new method turns out to be greater. While this may appear to be a drawback of the new method, this is a rigorous bound while the data complexity obtained using the chi-squared method is an approximation where there is little knowledge about the error in the approximation. So, if rigorous bound is desired, then one will have to be satisfied with a more conservative estimate of the data complexity.
Keywords: multiple linear cryptanalyis, LLR statistic, chi-squared statistic.

1 Introduction

Consider the setting of multiple linear cryptanalysis of block ciphers. Statistical analysis proceed by identifying a 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 $\ell > 1$.

$$\begin{aligned} \text{LLR} &= \sum_{\eta=0}^{2^\ell-1} Q_\eta \ln(2^\ell p_\eta); \\ \Lambda &= \sum_{\eta=0}^{2^\ell-1} (Q_\eta/N - 2^{-\ell})^2. \end{aligned}$$

The LLR test statistic arises from the log-likelihood ratio while the distribution of the Λ test statistic can be approximated by a chi-squared distribution. Both these test statistics have some limitations which we mention below.

To apply the LLR test statistic, it is required to have complete 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.

On the other hand, 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. The problem, however, is that the analysis of the chi-squared based method makes several approximations. This issue has been noted in the literature [11, 10, 21] and problems related to such approximations have been analysed in details in [21].

In this work, we propose to perform a statistical analysis without making any approximations. For the LLR test statistic, this has been carried out in [22] using the theory of martingales and the Azuma-Hoeffding inequality. As discussed above, the LLR test statistic has the shortcoming that to apply the method, complete information about $\tilde{p}$ is required. To avoid doing this, one requires a different test statistic. A natural candidate for this is the test statistic Λ. Our analysis of Λ shows that the theory of martingales used in [22] is difficult to apply to this statistic. This leaves us with the problem of obtaining a new test statistic satisfying the following two conditions.

1. An attack based on the test statistic should not require complete information about the joint distribution.
2. It should be possible to carry out the analysis without using any approximations.

To achieve the above, 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. Following the approach outlined in [22], we set up a Doob martingale and show that an appropriate Lipschitz condition holds allowing the application of the Azuma-Hoeffding inequality. The key recovery attack is analysed using the hypothesis testing framework. This allows obtaining a lower bound on the data complexity necessary for attaining specified upper bounds on the probabilities of Type-1 and Type-2 errors. These probabilities are directly related to the success probability and the advantage of an attack.

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. Experiments were done with different values of d. Based on these experiments, we suggest that the value of d be taken to be 1. In most cases, choosing $d = 1$ leads to a lower data complexity compared to choosing $d \neq 1$.

Experiments were done on the block cipher SERPENT using the new test statistic. The data complexities turn out to be higher than that of the Λ test statistic. This may seem like a disadvantage of the new test statistic.

On the other hand, the data complexity obtained using the new method is a rigorous upper bound, while for the Λ method it is an approximate value and the literature does not provide any estimate of the approximation errors.

1.1 Previous and Related Works

Linear cryptanalysis was proposed by Matsui in [17] as an attack on DES and involved a single linear approximation of the cipher. Later, in [18], Matsui used two linear approximations (which were assumed to be independent) to improve the attack. Independently, Kaliski and Robshaw [16] 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 [15] in the context of distinguishing attacks. Further work on distinguishing attacks without the independence assumption was carried out in [1, 14, 2]. Murphy [20] argued that the independence assumption need not be valid.

Junod [13] gave a detailed analysis of Matsui's ranking method [17, 18]. This work introduced the notion of ordered statistics in linear cryptanalysis. This was further developed by Selçuk in [23], 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, 14, 2] was the log-likelihood ratio (LLR). The chi-squared test statistic, another important statistic, was initially used by Handschuh and Gilbert for the cryptanalysis of the SEAL encryption algorithm. Later Johansson and Maximov [12] 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 [10].

2 Background

2.1 Linear Cryptanalysis

Let $E : \{0,1\}^k \times \{0,1\}^n \mapsto \{0,1\}^n$ be a block cipher, and so for each $K \in \{0,1\}^k$, $E_K(\cdot) \stackrel{\Delta}{=} E(K,\cdot)$ 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 cipher 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., $K^{(i)} = k^{(0)} \,||\, \cdots \,||\, k^{(i-1)}$ 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:

$$\langle \Gamma_P^{(i)}, P\rangle \oplus \langle \Gamma_B^{(i)}, B\rangle = \langle \Gamma_K^{(i)}, K^{(r)}\rangle; \quad i = 1, 2, \ldots, \ell;$$

where $\Gamma_P^{(i)}, \Gamma_B^{(i)} \in \{0,1\}^n$ and $\Gamma_{K^{(r)}}^{(i)} \in \{0,1\}^{nr}$ 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

$$L_i \stackrel{\Delta}{=} \langle \Gamma_P^{(i)}, P\rangle \oplus \langle \Gamma_B^{(i)}, B\rangle; \text{ for } i = 1, 2, \ldots, \ell. \tag{1}$$

Inner key bits: Let

$$z_i = \langle \Gamma_K^{(i)}, K^{(r)}\rangle; \quad i = 1, \ldots, \ell.$$

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 [8], 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 to be the following:

$$X = (L_1, \ldots, L_\ell).$$

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 $z \in \{0,1\}^\ell$,

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

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., $\tilde{p} \stackrel{\Delta}{=} \tilde{p}_{0^\ell}$. Write

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

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

For $\kappa \in \{0, 1, \ldots, 2^m - 1\}$, $j = 1, \ldots, N$ and $i = 1, \ldots, \ell$, define

$$\begin{aligned} L_{\kappa,j,i} &= \langle \Gamma_P^{(i)}, P_j \rangle \oplus \langle \Gamma_B^{(i)}, B_{\kappa,j} \rangle; & (3)\\ X_{\kappa,j} &= (L_{\kappa,j,1}, \ldots, L_{\kappa,j,1}); & (4)\\ Q_{\kappa,\eta} &= \#\{j \in \{1,2,\ldots,N\} : X_{\kappa,j} = \eta\}. & (5) \end{aligned}$$

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 the distribution of $Q_{\kappa,\eta}$ is the following:

$$Q_{\kappa,\eta} \sim \begin{cases} \mathrm{Bin}(N, p_z(\eta)) & \text{if } \kappa = \kappa^* \\ \mathrm{Bin}(N, 2^{-\ell}) & \text{if } \kappa \neq \kappa^*. \end{cases}$$

Denote the uniform distribution over the set $\{0,1\}^\ell$ by $p_\$ = (2^{-\ell}, \ldots, 2^{-\ell})$.

2.2 Martingales

We provide a brief description of martingales for for discrete random variables. Further details can be found in standard texts such as [7, 19]. Conditional expectation is defined in the following manner.

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 $\delta > 0$*

$$\Pr[Z_\lambda - Z_0 \geq \delta] \leq e^{-\delta^2/\left(2\sum_{i=1}^{\lambda} \upsilon_i^2\right)}; \quad (6)$$

$$\Pr[Z_\lambda - Z_0 \leq -\delta] \leq e^{-\delta^2/\left(2\sum_{i=1}^{\lambda} \upsilon_i^2\right)}. \quad (7)$$

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.

3 Drawbacks of Previously Proposed Statistics

Two test statistics have been proposed in the literature for performing statistical analysis of attacks on block ciphers. In this section, we briefly review these statistics and point out certain drawbacks.

3.1 The Log-Likelihood Ratio Test Statistics

Recall that $\tilde{p} = (p_0, \ldots, p_{2^\ell - 1})$ and $p_\$ = (2^{-\ell}, \ldots, 2^{-\ell})$. For a fixed κ and $1 \leq j \leq N$, $X_{\kappa,j}$ is given by (4). For $j = 1, \ldots, N$, define

$$Y_{\kappa,j} \quad = \quad \ln\left(p_{X_{\kappa,j}}/2^{-\ell}\right). \tag{8}$$

The LLR random variable is defined to be the following.

$$\mathrm{LLR}_\kappa \quad = \quad \sum_{j=1}^{N} Y_{\kappa,j} = \sum_{j=1}^{N} \ln\left(2^\ell p_{X_{\kappa,j}}\right) = \sum_{\eta=0}^{2^\ell-1} Q_{\kappa,\eta} \ln(2^\ell p_\eta). \tag{9}$$

The LLR test statistic has been used for key recovery attacks as well as distinguishing attacks in several works in the literature [1, 10, 4, 21, 22]. One drawback of this statistics is that to compute it, the full knowledge of $\tilde{p}$ is required. This can be seen from the above two expressions for LLR_κ. 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.

3.2 Chi-Squared Test Statistic

Recall from (5) 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 T_κ in the following manner:

$$T_\kappa \quad = \quad 2^\ell N \sum_{\eta=0}^{2^\ell-1} (Q_{\kappa,\eta}/N - 2^{-\ell})^2. \tag{10}$$

For the correct choice κ^* of the inner key bits, the right hand side of (10) involves $Q_{\kappa^*,\eta}$ whose distribution depends on the correct inner key bits z. Due to the relation $p_{z\oplus\beta}(\eta) = p_z(\eta \oplus \beta)$, the distribution of T_{κ^*}, however, does not depend on z.

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 T_κ approximately follows a chi-squared distribution for both correct and incorrect choices of κ. There is, however, the issue of error in approximation. This issue has been mentioned in the literature [11, 10, 21] and has been analysed in details in [21] where several shortcomings have been pointed out.

A question then arises as to whether it is possible to use this 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 [22] through the application of the theory of martingales. In this section, we explore the applicability of this theory to the test statistic T_κ given by (10).

Recalling the discussion in Section 2.2, an approach to applying the theory of martingales is to build a Doob martingale and obtain a upper bound on the absolute difference in the values of two successive random variables in the martingale sequence. The first part is quite easy to do, but, obtaining a good upper bound seems to be difficult.

For $0 \leq \eta \leq 2^\ell - 1$, define $Y_{\kappa,\eta} = Q_{\kappa,\eta} - N2^{-\ell}$ so that $T_\kappa = \frac{2^\ell}{N} \sum_{\eta=0}^{2^\ell-1} Y_{\kappa,\eta}^2$.

Define $Z_{\kappa,0} = E[T_\kappa] = \mu_0$ and for $1 \leq \eta \leq 2^\ell$, define

$$\begin{aligned} Z_{\kappa,\eta} \quad &= \quad E[T_\kappa \mid Y_{\kappa,0}, Y_{\kappa,1}, \ldots, Y_{\kappa,\eta-1}] \\ &= \quad \frac{2^\ell}{N}\left[\sum_{j=0}^{\eta-1} Y_{\kappa,j}^2 + \sum_{j=\eta}^{2^\ell-1} E[Y_{\kappa,j}^2 \mid Y_{\kappa,0}, \ldots, Y_{\kappa,\eta-1}]\right]; \quad [\text{Since, } E[X \mid X] = X] \end{aligned}$$

It is not difficult to show that the sequence $\{Z_{\kappa,\eta}\}_{\eta=0}^{2^\ell}$ forms a Doob Martingale with respect to $\{Y_{\kappa,\eta}\}_{\eta=0}^{2^\ell-1}$. Further, $T_\kappa = Z_{\kappa,2^\ell}$.

Let us now consider the problem of upper bounding $|Z_{i+1} - Z_i|$. In Section 2.2, the υ-Lipschitz condition was utilised for this purpose. To apply this condition in the present context, we need to set $f(y_0,\ldots,y_{2^\ell-1}) = \sum_{\eta=0}^{2^\ell-1} y_\eta^2$. From this, it is possible to obtain an υ such that f satisfies the υ-Lipschitz condition. Moving from this to a upper bound on $|Z_{i+1} - Z_i|$ requires the crucial condition that the random variables $Y_{\kappa,0},\ldots,Y_{\kappa,2^\ell-1}$ are independent. This condition, however, does not hold. The random variable $Y_{\kappa,\eta}$ is defined from $Q_{\kappa,\eta}$ and so $Y_{\kappa,0},\ldots,Y_{\kappa,2^\ell-1}$ are independent if and only if the random variables $Q_{\kappa,0},\ldots,Q_{\kappa,2^\ell-1}$ are independent. From the definition of Q_η, we have that $\sum_{\eta=0}^{2^\ell-1} Q_{\kappa,\eta} = N$. So, $Q_{\kappa,0},\ldots,Q_{\kappa,2^\ell-1}$ are not independent and so neither are $Y_{\kappa,0},\ldots,Y_{\kappa,2^\ell-1}$. So, the technique of using the Lipschitz condition to upper bound the martingale differences does not work.

Let us now consider the problem of directly trying to obtain a bound for the $\mid Z_{\eta+1} - Z_\eta \mid$. We have,

$$\begin{aligned}
& Z_{\kappa,\eta+1} - Z_{\kappa,\eta} \\
&= \frac{2^\ell}{N}\left[Y_{\kappa,\eta}^2 + \sum_{j=\eta+1}^{2^\ell-1} E[Y_{\kappa,j}^2 \mid Y_{\kappa,0},\ldots,Y_{\kappa,\eta}] - \sum_{j=\eta}^{2^\ell-1} E[Y_{\kappa,j}^2 \mid Y_{\kappa,0},\ldots,Y_{\kappa,\eta-1}]\right] \\
&= \frac{2^\ell}{N}\left[Y_{\kappa,\eta}^2 - E[Y_{\kappa,\eta}^2 \mid Y_{\kappa,0},\ldots,Y_{\kappa,\eta-1}] + \sum_{j=\eta+1}^{2^\ell-1} \left(E[Y_{\kappa,j}^2 \mid Y_{\kappa,0},\ldots,Y_{\kappa,\eta}] - E[Y_{\kappa,j}^2 \mid Y_{\kappa,0},\ldots,Y_{\kappa,\eta-1}]\right)\right].
\end{aligned}$$

Clearly, the involved random variables take values from a finite set and so there is indeed a maximum value for $\mid Z_{\eta+1} - Z_\eta \mid$. Getting a good bound on this maximum value, however, seems to be quite difficult to obtain from the above expansion. Without an appropriate bound, the application of the Azuma-Hoeffding inequality does not provide meaningful results. It is due to this reason that we do not explore this approach any further.

4 A New Test Statistic

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

$$T_\kappa = \sum_{\eta\in\{0,1\}^\ell} \underline{\eta}^d Q_{\kappa,\eta}. \tag{11}$$

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

$$\mu_1 = E[T_\kappa] = \sum_{\eta\in\{0,1\}^\ell} \underline{\eta}^d E[Q_{\kappa,\eta}] = N2^{-\ell} \sum_{\eta\in\{0,1\}^\ell} \underline{\eta}^d; \tag{12}$$

$$\begin{aligned}
\mu_0 &= E[T_{\kappa^*}] \\
&= \sum_{\eta\in\{0,1\}^\ell} \underline{\eta}^d E[Q_{\kappa^*,\eta}] \\
&= \sum_{\eta\in\{0,1\}^\ell} \underline{\eta}^d N(2^{-\ell} + \epsilon_\eta(z)) = \mu_1 + N \sum_{\eta\in\{0,1\}^\ell} \underline{\eta}^d \epsilon_\eta(z) = \mu_1 + N \sum_{\eta\in\{0,1\}^\ell} \underline{\eta}^d \epsilon_\eta.
\end{aligned} \tag{13}$$

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 (11) can be rewritten in the following manner.

$$T_\kappa = \sum_{\eta\in\{0,1\}^\ell} \underline{\eta}^d Q_{\kappa,\eta} = \sum_{j=1}^{N} \underline{X}^d_{\kappa,j}. \tag{14}$$

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.

Define, a sequence of random variables as

$$\begin{aligned} Z_{\kappa,0} &= E[T_\kappa] \\ Z_{\kappa,j} &= E[T_\kappa \mid X_1,\ldots,X_i]; j\geq 1. \end{aligned}$$

Therefore, $Z_N = T_\kappa$. The sequence $Z_{\kappa,0}, Z_{\kappa,1},\ldots Z_{\kappa,N}$ forms a Doob Martingale with respect to $X_{\kappa,1},\ldots,X_{\kappa,N}$.

We would like to apply the Lipschitz condition to upper bound the martingale differences. To this end, let $f(x_1,\ldots,x_N) = x_1^d+\cdots+x_N^d$. Let $x_1,\ldots,x_N,x_i'$ be any $N+1$ elements in $\{0,\ldots,2^\ell-1\}$. Then

$$\begin{aligned} &\left|f(x_1,\ldots,x_{i-1},x_i,x_{i+1},x_N) - f(x_1,\ldots,x_{i-1},x_i',x_{i+1},x_N)\right| \\ &= \left|\sum_{j=1}^{N} x_j^d - \sum_{j=1}^{i-1} x_j^d - x_i'^d - \sum_{j=i+1}^{N} x_j^d\right| \\ &= \mid x_i^d - x_i'^d \mid \leq \max_{x_i,x_i'\in\{0,1\}^\ell} \mid x_i^d - x_i'^d \mid \\ &= (2^\ell-1)^d. \end{aligned}$$

Let $\upsilon = (2^\ell-1)^d$. This implies that $f()$ is υ-Lipschitz. Since $X_1^d,\ldots,X_N^d$ are independent, it follows that $\mid Z_{\kappa,j} - Z_{\kappa,j-1} \mid \leq \upsilon$ for all $j\in\{1,2,\ldots,N\}$. Therefore, for a particular value of κ, one can apply Azuma-Hoeffding inequality on the Doob Martingale $Z_{\kappa,0}, Z_{\kappa,1},\ldots Z_{\kappa,N}$.

Recall that μ_0 is the expectation of T_{κ^*} and μ_1 is the expectation of T_κ for any incorrect choice of κ. The expressions for μ_0 and μ_1 are given by (12) and (13) respectively. 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 3. *Let $0<\alpha,\beta<1$. In Hypothesis Test-1, it is possible to choose t such that for*

$$N \geq \frac{2(2^\ell-1)^{2d}(\sqrt{\ln(1/\alpha)}+\sqrt{\ln(1/\beta)})^2}{\left(\sum_{\eta=0}^{2^\ell-1}\eta^d\epsilon_\eta\right)^2} \tag{15}$$

the probabilities of the Type-1 and Type-2 errors are upper bounded by α and β respectively.

Proof. We provide the proof for the case $\mu_0>\mu_1$ with the other case being similar. The probabilities of Type-1

and Type-2 errors are given by

$$\begin{aligned}
\Pr[\text{Type-I Error}] &= \Pr[T_\kappa \le t \mid H_0 \text{ holds}] = \Pr[T_\kappa - E[T_\kappa] \le t - E[T_\kappa] | H_0 \text{ holds}] \\
&= \Pr[Z_N - Z_0 \le -(\mu_0 - t) | H_0 \text{ holds}] \\
&\le \exp\left(-\frac{(\mu_0 - t)^2}{2N\upsilon^2}\right); \quad \text{[By Azuma-Hoeffding inequality].} \\
\Pr[\text{Type-2 Error}] &= \Pr[T_\kappa > t \mid H_1 \text{ holds}] = \Pr[T_\kappa - E[T_\kappa] > t - E[T_\kappa] \mid H_1 \text{ holds}] \\
&= \Pr[Z_N - Z_0 > (t - \mu_1) \mid H_1 \text{ holds}] \\
&\le \exp\left(-\frac{(t - N\mu_1)^2}{2N\upsilon^2}\right); \quad \text{[By Azuma-Hoeffding inequality].}
\end{aligned}$$

Let,

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

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

$$t = \mu_0 - \upsilon\sqrt{2N\ln(1/\alpha)} \tag{16}$$

$$t = \mu_1 + \upsilon\sqrt{2N\ln(1/\beta)}. \tag{17}$$

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 (15). For any N greater than this value, the probabilities of Type-1 and Type-2 errors will be at most α and β respectively. □

4.1 Success Probability and Expected Advantage

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. By definition, the success probability is $1 - \Pr[\text{Type-1 error}]$. So, if α is an upper bound on the success probability, then $P_S = 1 - \alpha$ is a lower bound on the success probability.

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}. A particular choice of κ is reported as a candidate key if a Type-2 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.

4.2 Attack Procedure

The actual application of the attack will be as follows. Given P_S and a, determine α and β as discussed in Section 4.1; then determine N as given by the right hand side of (15). From α and N determine t as given by (16). 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.

4.3 Choice of d

The theory developed above works for all positive d. The question that arises is what is the appropriate value of d that should be used? There are two factors that need to be kept in mind.

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 (15) 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 (15) goes to infinity. Experiments confirm this behaviour.

Based on the above, we do not consider values of $d < 1$. For integer values of $d \geq 1$, 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 $d = 2$. 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.

5 Experimental Results for SERPENT

We compare the data complexity given by the new test statistic with that of the data complexity of the Λ-test statistic given in [10, Equation (18)] for the block cipher SERPENT.

A reduced round linear cryptanalysis of SERPENT was earlier reported in [6] using a set of linear approximations [5]. This set was later used in [9, 10] to perform multidimensional linear cryptanalysis on SERPENT using the LLR and the Λ-test statistic. To perform their experiments Hermelin et al. (see [9]) used a subset of 64 linear approximations among the list of linear approximations given in [5]. This set 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 $m = 10$.

We note at this point that the total number of linear approximation required to generate the full probability distribution for the correct key is $2^{10} - 1 = 1023$. However, out of these only 64 are given in [5]. To find the full probability distribution for the correct key, two methods were suggested in [9]. We have used the second method.

For our experiments, the value of the probability of success (P_S) was fixed to 0.95. The data complexities for both the new test statistic with $d = 1$ and the chi-squared method were then computed for $a = 1, 2, \ldots, 10$, where "a" denotes advantage of the method. Table 1 summarises the output of the experiment. In the Table, N_Λ and N_X denote the data complexity of the Λ [10, equation 18] and the new test statistic with $d = 1$, respectively. The last column of the Table gives the ratio of the two data complexities. From the Table, it is clear that the Λ test

statistic requires lower data complexity compared to the new method. As a increases, the ratio decreases and for values of $a = 6, \ldots, 10$, N_X is about 8800 to 8300 times more that N_Λ. Another experiment was conducted to

a	N_Λ	N_X	N_X/N_Λ
1	1.11×10^{11}	1.25×10^{6}	88987.47
2	1.43×10^{11}	9.48×10^{6}	15135.63
3	1.71×10^{11}	1.53×10^{7}	11173.77
4	1.96×10^{11}	2.0×10^{7}	9799.09
5	2.19×10^{11}	2.4×10^{7}	9132.67
6	2.41×10^{11}	2.75×10^{7}	8760.68
7	2.62×10^{11}	3.07×10^{7}	8538.75
8	2.83×10^{11}	3.37×10^{7}	8403.21
9	3.03×10^{11}	3.64×10^{7}	8321.65
10	3.23×10^{11}	3.90×10^{7}	8275.83

Table 1: Table showing the comparison of the new test statistic with $d = 1$ and the chi-squared test statistic for SERPENT with a ranging from 1 to 10.

determine the value of d for which the data complexity of the new test statistic is minimum. For this experiment, 100 integral and 99 fractional values of d of the form $1, 2, \ldots, 100$ and $0.01, 0.02, \ldots, 0.99$, respectively, were considered. For each of these values of d and for each $a = 1, 2, \ldots, 10$, the corresponding data complexities were computed for the given distribution of SERPENT.

The value of d corresponding to the minimum data complexity and the corresponding value of the data complexity were then recorded. These are reported in Table 2. As can be seen from the table, in case of SERPENT bith $d = 1$ and $d = 2$ give the minimum data complexity among all possible values of $d > 0$. This table supports our choice of $d = 1$.

a	Minimum Data Complexity	
	Value of d	Data Complexity
1	1, 2	1.11×10^{11}
2	1, 2	1.43×10^{11}
3	1, 2	1.71×10^{11}
4	1, 2	1.96×10^{11}
5	1, 2	2.19×10^{11}
6	1, 2	2.41×10^{11}
7	1, 2	2.62×10^{11}
8	1, 2	2.83×10^{11}
9	1, 2	3.03×10^{11}
10	1, 2	3.23×10^{11}

Table 2: Table showing the minimum data complexity for different values of d for SERPENT with a ranging from 1 to 10.

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 [22] towards a rigorous analysis without using any unfounded approximations. We first considered the chi-squared based test statistic and argued that this test statistic is not amenable to analysis using the theory of martingales used in [22].

To resolve the problem, we introduced a new test statistic using which an attack can be applied without the 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.

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] Baydoin 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.

[6] Baydoin 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.

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

[8] 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`.

[9] 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.

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

[11] Miia Hermelin, Joo Yeon Cho, and Kaisa Nyberg. Statistical Tests for Key Recovery Using Multidimensional Extension of Matsuis Algorithm 1. In *EUROCRYPT*, 2009.

[12] 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.

[13] Pascal Junod. On the Complexity of Matsuis Attack. In *Selected Areas in Cryptography*, pages 199–211. Springer, 2001.

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

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

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

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

[18] 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.

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

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

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

[22] 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`.

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