Linear Cryptanalysis Using Low-bias Linear Approximations

Tomer Ashur[1], Daniël Bodden[1], and Orr Dunkelman[2]

[1] Dept. Electrical Engineering, COSIC-imec, KU Leuven, Belgium
[tashur,dbodden]@esat.kuleuven.be
[2] University of Haifa, Haifa, Israel
orrd@cs.haifa.ac.il

Abstract. This paper deals with linear approximations having absolute bias smaller than $2^{-\frac{n}{2}}$ which were previously believed to be unusable for a linear attack. We show how a series of observations which are individually not statistically significant can be used to create a χ^2 distinguisher. This is different from previous works which combined a series of significant observations to reduce the data complexity of a linear attack. We test the distinguisher on a real-world cipher and show that it can be used to improve previous results.

Keywords: Linear cryptanalysis, Multiple linear cryptanalysis, Multi-key linear cryptanalysis, Speck

1 Introduction

Linear cryptanalysis is one of the most powerful cryptanalytic methods. Developed in [26], linear cryptanalysis deals with approximating the non-linear behavior of a cryptographic primitive in a linear way. Through ample research, the understanding of the attack had been improved [24, 27, 28, 30], and it was extended and generalized in many subsequent works [8, 12, 18, 20, 21].

Generally speaking, an adversary using linear cryptanalysis uses a linear approximation linking the input bits, output bits, and key bits with some probability. The quality of the approximation is usually measured through its "bias" or "correlation" and it is evaluated using a large amount of pairs of plaintexts and ciphertexts.

The amount of pairs required for the attack to succeed determines the data complexity and several works analyzed it in order to estimate this quantity [24, 28, 30]. The common wisdom is that for a linear attack using an approximation with bias satisfying $\epsilon = p - \frac{1}{2}$, the data complexity of the order of ϵ^{-2}, which is usually interpreted as the number of pairs the adversary needs to observe. As the number of possible pairs in a block cipher with a block size n is 2^n, this essentially limits the possible approximations to ones with bias $|\epsilon| \geq 2^{-n/2}$, which often puts an upper limit on the number of rounds an approximation can cover. It is worth noting that although the mechanism of linear cryptanalysis is

well understood, the problem of obtaining the exact bias (or a good estimation of it) is still an open one as we discuss later.

One of the extensions proposed for linear cryptanalysis is Multiple Linear cryptanalysis by Kaliski and Robshaw [25]. Originally, the extension was limited to approximations involving the same key bits, but this restriction was later lifted by Biryukov et al. in [8] where it was only required that the set of approximations would be linearly independent. Another progress was made when Hermelin et al. presented Multidimensional Linear cryptanalysis in [21] where the approximations need not be linearly independent. However, the focus of all these works was always on reducing the data complexity of the attack, and all approximations used had biases satisfying $|\epsilon| \geq 2^{-n/2}$. Moreover, it seems that many believe that approximations with bias $|\epsilon| < 2^{-n/2}$ can never be used to attack a block cipher via a linear attack; see e.g., [1, 7, 8, 14, 16, 22, 25, 31]. This belief has reached the point where it is taken as an assumption, and is used to "prove" the resistance of new designs against linear cryptanalysis. In light of the results presented in this paper, such proofs should be reconsidered, especially for lightweight ciphers which often offer only small security margins.

1.1 Our Contributions

In this paper we show that contrary to common belief, approximations with bias $|\epsilon| < 2^{-n/2}$ (which we refer to as "low-bias approximations") **can** be used for a linear attack. This is done by employing the χ^2 distinguisher [31] which can capture the statistical difference between a set of approximations with low biases and a set of random approximations. The implication of this observation is that security arguments based on counting S-boxes and showing an upper bound on the bias of the best approximation should be re-evaluated. Moreover, using low-bias approximations, linear distinguishers can cover additional rounds, thus extend previous attacks.

As a second contribution, we present a set of low-bias linear approximations for Speck32/64, and conjecture that our method can be used to extend the best known distinguisher from 9 to 10 rounds. To confirm this hypothesis, we conduct several experiments and employ statistical tests to show that the results are statistically significant. These experiments serve two purposes as they confirm our theory as well as improve the state of the art cryptanalysis with respect to Speck32/64, which is an important cipher.

2 Preliminaries and Notation

We present the tools and notations used throughout the paper. The space of n-dimensional binary vectors is denoted by $\mathcal{V}_n$. A boolean vector $x \in \mathcal{V}_n$ is composed of a string of n bits $x = (x_{n-1}, \ldots, x_0)$ where x_0 is the least significant bit. A boolean function f is a function taking a vector of size n and outputting a single bit, i.e., $f : \mathcal{V}_n \to \mathcal{V}_1$. For two vectors $x, y \in \mathcal{V}_n$ we define the inner

product $\cdot$ as $x \cdot y = \bigoplus_{i=0}^{n-1} x_i \wedge y_i$ where $\oplus$ stands for addition modulo 2 (XOR) and $\wedge$ is the AND operation between two bits.

A block cipher is a function $E : \mathcal{V}_n \times \mathcal{V}_k \to \mathcal{V}_n$ taking a plaintext P and converting it using a key K to a ciphertext C. An iterative block cipher is a block cipher composed of a relatively simple round function $R(k_i, x_i) = R_{k_i}(x_i)$ that takes a subkey and the round input and returns the round output. The block cipher is then constructed from multiple applications of the round function R with different subkeys obtained through a key-schedule algorithm from the master key K.

A linear approximation ψ for a function $R(k, X)$ is a tuple (α, β) evaluating the equation $\alpha \cdot X \oplus \beta \cdot R(k, X) = 0$ for some given X and k. Each linear approximation is associated with a probability p, which measures the probability that ψ holds for a uniformly chosen X. The bias ϵ_ψ of a linear approximation is defined as $\epsilon_\psi = p - \frac{1}{2}$.

A linear trail is a sequence of m linear approximations $\psi_0, \ldots, \psi_{m-1}$ that covers a set of consecutive rounds $R_0, \ldots, R_{m-1}$ in such a way that ψ_i corresponds to R_i and the output mask of ψ_{i-1} is equal to the input mask of ψ_i. The bias of a linear approximation can be obtained through the Piling Up Lemma and is equal to $2^{m-1} \cdot \prod_{i=0}^{m-1} \epsilon_i$ where ϵ_i is the bias associated with ψ_i. The masks for a linear trail are sometimes written as a tuple (α, β, γ) to indicate the key involvement, and the approximation is then evaluated as $\alpha \cdot X \oplus \beta \cdot R(k, X) \oplus \gamma \cdot k = 0$.

A linear hull with masks (α, β) is the collection of all linear trails with input mask α and output mask β, keeping all intermediate values as free variables. The bias of the hull is the sum of biases for the composing trails. Although often treated as such, the bias of a linear hull is not a fixed value but rather key-dependent, and we elaborate on that in Section 2.1. The term *linear approximation* is used to describe both trails and hulls, and the respective masks will be clear from the context.

In the sequel we denote random variables by boldface, capital letters **X,Y,Z**. A random variable $\mathbf{X}$ is associated with a probability distribution $P_x = (p_0, \ldots, p_\ell)$ where each value p_i is the probability that $\mathbf{X}$ takes the value i, i.e., $\Pr[\mathbf{X} = i] = p_i$. In places where it is clear which random variable is considered, we simply write P instead of P_x. To denote a specific value **X,Y,Z** took, we use $\hat{X}, \hat{Y}, \hat{Z}$. In the sequel, random variables are used to model the fact that the bias is a probability distribution rather than a fixed value (i.e., key dependent). To avoid the complication of estimating the bias through a sample of the data (i.e., through the empirical bias), we will only consider in this paper the case where the entire codebook is used. The extension to cases where only a sample of the codebook is used (e.g., due to a larger block size) can be readily obtained using the same methods as [3, 10, 13].

The normal distribution is an important distribution defined by two parameters, the mean and the variance. We denote by $\mathbf{X} \sim \mathcal{N}(\mu_x, \sigma_x^2)$ the fact that the random variable $\mathbf{X}$ follows a normal distribution with mean μ_x and variance σ_x^2. When $\mu_x = 0$, and $\sigma_x^2 = 1$ we call the resulting distribution the *standard normal distribution*, and say that $\mathbf{X}$ is a standard normal random variable. It is

possible to convert any normal random variable $\mathbf{X} \sim \mathcal{N}(\mu_x, \sigma_x^2)$ into a standard normal variable $\mathbf{Y}$ by setting $\mathbf{Y} = \frac{\mathbf{X}-\mu_x}{\sigma_x}$.

Another important distribution we use is the χ^2 distribution with k degrees of freedom. We write $\mathbf{Z} \sim \chi_k^2$ to say that the random variable $\mathbf{Z}$ follows the χ^2 distribution with k degrees of freedom. It is known that if $\mathbf{X} \sim \mathcal{N}(0,1)$ then $\mathbf{X}^2 \sim \chi_1^2$, and that the sum of m squared standard normal variables follows the χ_m^2 distribution.

In what follows, the goal of the adversary is to distinguish a block cipher E_k from a random permutation. This can be done through *hypothesis testing*. The adversary collects some information that could come from either a random permutation or a block cipher, and uses it to calculate a *test statistic*. We derive the expected behavior of the test statistic when the underlying data comes from a random permutation, and build a *confidence interval* for its distribution. If the test statistic falls inside the interval, we determine that the data came from a random permutation, otherwise, we determine that it came from the block cipher. It is interesting to see that by taking this approach, we make a generic observation, **independent** of any specific application of the method. When executing the attack, the statistical distance, hence the advantage, are determined by the specific block cipher being attacked and the quality of the approximations used. This approach is useful in many real life scenarios when it is hard to obtain a good estimate for the bias for reasons discussed in the next subsection.

2.1 A Note on Bias Estimation and the Required Data Complexity

It is important to note that the problem of obtaining the right value for the bias ϵ is still open. The essence of the problem lies in the fact that although in many cases it is easy to calculate the probability of a specific linear trail directly from the cipher's structure, this bias remains impossible to measure unless the key is already known. On the other hand, the bias of the linear hull, which is easy to measure, is hard to compute theoretically. To better understand this, it is useful to consider the values the bias can take. When a single trail is considered, the magnitude of its bias is fixed, and the key only affects the sign. However, since an adversary has access only to the input and output of the cipher (and not to intermediate rounds), the bias can only be measured over the linear hull, which is composed of an unknown number of trails.

In Matsui's original paper [26], he presented a single trail and used its bias for the attack. This approach worked in the case of DES because the corresponding linear hull consisted only of one significant trail. However, in other ciphers, the fact that the hull may be composed of multiple trails leads to an under- or overestimation of the bias. A counterexample to the case of DES was presented in [4, 5], showing how such an erroneous estimation of the hull's bias through a single trail affects the success probability of the attack.

Another important aspect of bias estimation is the influence of the key. In the case of a linear trail, the magnitude of the bias is known and only the sign is affected by the key. The bias is therefore a random variable that can take only

two values. The actual value is determined by a subset of key bits.[3] However, in the case of a linear hull the bias is the sum of multiple biases coming from the underlying trails. Each of these underlying biases is a random variable with different values coming from different distributions, and resulting in a probability distribution rather than a single value for the hull's bias. Generally, the nature of this distribution is cipher-dependent. For real-life ciphers, the distribution is hard to derive due to the large number of involved trails, and is therefore unknown to the adversary. The probabilistic nature of the bias was discussed in [1]. In practice, many works simply use the mean of the absolute bias in place of the actual distribution for attacks, which is generally insufficient.

In a private communication with Kaisa Nyberg, the authors received a soon to be published manuscript of [11], which constitutes an attempt to address these concerns through an estimation of the ELP. The idea behind their approach is to split the involved linear trails into two groups Q and R representing the dominating trails and the "rest". Once the set of dominating trails is properly accounted for, the "rest" can be modeled as random noise. However, this approach simplifies the problem only slightly. First, it assumes that the adversary can find all dominating trails, which again requires sieving through the space of all possible ones. Second, even if a certain set is already given, it is required that it be exhaustive, which is again hard to verify in practice. Finally, although it can readily be shown that the estimation error of the bias decreases with the number of known trails, this is not reflected in [11], which always models R as random noise, regardless of the set's size.

Interestingly, [11] also mentions (independently of us) that linear approximations with low bias can be used for an attack through the variance of the distribution. However, they discuss this in the context of a key recovery attack using a single approximation. While theirs is an interesting observation, it still relies on several unrealistic assumptions. First, it assumes that sufficient knowledge of the ELP is given, which again implies that all the most dominant trails are known, if not the full distribution. More importantly, it assumes that the ELP is always larger than the bias variance of a random approximation, which allows them to define an interval $[-\Theta, \Theta]$ and claim that keys outside this interval are more likely to be right keys. As can be seen in the example we use in Section 4, this is not always the case, which means that in some cases (a-prioi unknown due to the difficulty to calculate the ELP), keys inside the interval $[-\Theta, \Theta]$ are actually more likely to be the right keys.

A recent paper [29] takes a different approach for obtaining the distribution of biases, by trying to bound them. They explicitly criticize prevailing methods using order statistics, due to the reliance of such methods on assumptions regarding the distribution of the bias. As a result, they question the applicability of formulas formerly used to derive the data complexity. Clearly, [11] and [29] (as well as others) are in disagreement as to some fundamental aspects of linear cryptanalysis. In spite of that, progress in the practical application of linear cryptanalysis is still often based on common wisdom (and hence stands on un-

[3] This is the property exploited when using Matsui's Algorithm 1.

stable ground). For example, papers which follow the evolution of linear trails either manually [2] or using automated tools [9,32] are still published, suggesting that an attack can be leveraged using such trails. Similarly, papers ignoring the key dependency [15] are being published, then improved [33], criticized [10, 23], then salvaged [11].

This discussion makes it apparent that there is still much more work to be done before ciphers' security against linear cryptanalysis is understood. Without taking a stance in the ongoing discussion about bias' distribution, we wish to highlight a different aspect of linear cryptanalysis that was overlooked. In this paper, we show that contrary to what seems to be an almost undisputed axiom, linear approximations with bias smaller than $2^{-n/2}$ **can** be used for linear cryptanalysis. To avoid the minefield that bias estimation seems to be, we take an approach that is based on what is the foundation of statistical cryptanalysis, namely, the behavior of random oracles. Our null hypothesis is that a set of observations came from a random oracle, and we use a statistical test that does not require an alternative hypothesis. In doing so, we avoid the need to obtain information about the bias distribution of the specific cipher under consideration.

This is not to say that obtaining the bias distribution is unimportant. When the distribution is known it allows to calculate the ELP, which can then be used to calculate the data requirement and the advantage of an attack. However, being independent to the bias estimation problem, our approach has merits in real world scenarios, even if the bias is unknown. Since both parameters are not required a-priori to execute the attack, an adversary can simply test the attack offline (e.g., in a lab), establish that they hold enough approximations given the data complexity they tried, then execute it on an online target. For example, a corrupt cipher designer can ensure that one linear approximation out of all 2^{2n} has a low absolute bias but a distinct distribution before releasing the cipher. Such a backdoor will be practically impossible to find due to the difficulty to obtain this distribution, nonetheless to check the distributions for each of the 2^{2n} possible approximations.

3 Using Hypothesis Testing for Linear Cryptanalysis

In this section we present the tools used in linear cryptanalysis and those we use later in Section 4 for distinguishing low-bias linear approximations. We stress that the methods presented in this section are mostly repeating previous results. Viewing linear cryptanalysis as a hypothesis testing problem was already considered in [28] and the χ^2 distinguisher was previously presented in [31] where it was shown to be as efficient as the standard distinguisher. The novelty of the paper starts in the next section where we show that due to the differences between the distinguishers, the χ^2 distinguisher can be used to detect statistical differences between sets of approximations each with a low bias.

Let π be a random permutation of n bits, and let ψ be a linear approximation. We denote by $\mathbf{X}$ a counter for the number of times ψ is satisfied over all 2^n possible inputs. The behavior of $\mathbf{X}$ is given by the following Lemma:

Lemma 1. *For π, ψ, and $\mathbf{X}$ as before*

$$\mathbf{X} \sim \mathcal{N}(2^{n-1}, 2^{n-2}) \tag{1}$$

Proof. We define $X|\epsilon = 2^n \cdot \epsilon$ to be the number of times the linear approximation was satisfied over all possible inputs once the permutation was chosen and ϵ was fixed. Substituting ϵ with [17]'s random variable ϵ yields the random variable $\mathbf{X}$ and completes the proof.

Since we only consider the case where the full codebook is used, $\mathbf{X}$ is obtained by a simple translation of the bias distribution into a distribution on the pairs of plaintexts and ciphertexts. In cases where a sample of the codebook is used, the distribution of $\mathbf{X}$ should be adapted accordingly.

The random variable $\mathbf{X}$ can be converted into a standard normal variable by setting $\frac{\mathbf{X}-2^{n-1}}{2^{n/2-1}}$. Furthermore, knowing that the square of a standard normal variable follows the χ_1^2 distribution, we obtain the following corollary:

Corollary 1. *Let π, ψ, and $\mathbf{X}$ as before, then*

$$\mathbf{Y} = (\frac{\mathbf{X} - 2^{n-1}}{2^{n/2-1}})^2 \sim \chi_1^2 \tag{2}$$

Let E be a family of block ciphers with block size n, and E_k a member of this family characterized by a key k. Further, let ψ be a linear approximation that is biased when applied to a member of E. We define a counter $\hat{T}$ to count the number of times ψ is satisfied after observing 2^n pairs of plaintexts and ciphertexts. As ψ is biased when applied to E_k, so is $\hat{T}$ and the value $\Pr[\mathbf{X} = \hat{T}]$ is small. On the other hand, when applied to a permutation that is not a member of E (i.e., random permutation), $\hat{T}$ follows Lemma 1 and hence $\Pr[\mathbf{X} = \hat{T}]$ is large. The goal of the adversary upon receiving $\hat{T}$ is to decide if it was obtained by applying ψ to a member of E or to a random permutation.

In the classical linear attack, the adversary would normally use $\hat{T}$ to calculate an empirical bias $\hat{\epsilon}$ and compare it to some threshold. For example, an adversary may decide to use $\pm 2^{-n/2}$, meaning that he returns "random permutation" if $-2^{-n/2} \leq \hat{\epsilon} \leq 2^{-n/2}$; "a member of E" otherwise.

This procedure can be viewed as a form of hypothesis testing. The null hypothesis H_0 is that $\hat{T}$ was generated through a random permutation. The alternative hypothesis, namely H_1, is that it was generated using a block cipher. The adversary constructs a confidence interval for H_0, and rejects the null hypothesis if the test statistic falls outside of it. According to [17], ϵ, the bias of a random permutation, has a normal distribution with mean 0 and standard deviation $2^{-n/2-1}$, which means that in the above example, $2^{-n/2}$ is 2 standard deviations away from the mean, yielding a confidence interval of $\alpha = \Phi(2) - \Phi(-2) = 0.95449$. Using this interval means that the distinguisher

Table 1. The linear approximations used to verify the equivalence between the classical linear distinguisher and the χ^2 distinguisher. The difference between the trail bias predicted by [32] and the measured bias is due to the linear hull effect. All masks are reported in hexadecimal notation.

No. rounds	Input mask (left,right)	Output mask (left,right)	Trail bias reported in [32]	Measured bias
6	(000D, 56E0)	(0800, 0800)	2^{-8}	2^{-10}
7	(000D, 56E0)	(2040, 2050)	2^{-10}	2^{-12}
8	(000D, 56E0)	(0083, 80C3)	2^{-13}	2^{-15}
9	(000D, 56E0)	(170B, 130A)	2^{-15}	$2^{-15.98}$

reports $1 - \alpha \approx 4.5\%$ of the values coming from a random permutation incorrectly, and thus, it gives an advantage when the probability to correctly identify the block cipher, namely β, is high enough such that $\beta - (1 - \alpha) > 0$.

Note that this procedure does not require much knowledge about the bias. It is sufficient that the underlying key-dependent bias be large enough, and that the data complexity is sufficient to give a good estimation of the bias for the attack to work.

3.1 The χ^2 Distinguisher

We now present the χ^2 distinguisher and demonstrate that in the classical setting (i.e., for a high-bias approximations) this distinguisher is as efficient as the classical linear distinguisher.

Recalling Corollary 1, we can convert the normal variable $\mathbf{X}$ into a χ_1^2 variable. The $[-2^{-n/2}, 2^{-n/2}]$ interval which was 2 standard deviations from the mean to each direction now becomes a $[0, 4]$ interval. Computing $F_{\chi_1^2}(4) = 0.95449$ where $F_{\chi_1^2}$ is the Cumulative Distribution Function of the χ_1^2 distribution shows that the two intervals cover area of the same size.

We have also verified this equivalence empirically, using versions of Speck32/64 reduced to 6–9 rounds. The linear approximations we used were first published in [32]. However, [32] only reports the bias of a single linear trail found using their automated tools (i.e., they suggest that the bias of a single trail can be used in place of the bias of the respective hull, and ignore the key dependency). We present these approximations in Table 1, and report the average absolute bias of the hull which we have obtained empirically.

In each experiment we used 2^{11} random keys to encrypt 2^{32} plaintexts for 6–9 rounds and applied the respective linear approximations from Table 1. For each key we computed the empirical bias, and used both the classical distinguisher and the χ^2 distinguisher. For the classical distinguisher, we checked how many of the 2^{11} experiments have an absolute bias larger than 2^{-16}. For the χ^2 distinguisher, we checked how many times the test statistic was smaller than 4. Table 2 reports the results, which expectedly show that the two distinguishers are the same.

Table 2. Comparing the two distinguishers: we show that the two distinguishers produce exactly the same results and are in fact equivalent.

No. rounds	Bias	No. Successes for the	
		classical Linear distinguisher	χ^2 distinguisher
6	2^{-10}	2048	2048
7	2^{-12}	2048	2048
8	2^{-15}	1900	1900
9	$2^{-15.98}$	1020	1020

4 Linear Cryptanalysis Using Multiple Low-bias Approximations

In the previous section we presented a distinguisher using the χ^2 distribution in the classical setting. One limitation of the classical normal distinguisher is that it cannot use biases which are too small. This is often used to "prove" the resistance of new designs against linear cryptanalysis by claiming that since no linear approximations with a bias larger than $2^{-n/2}$ exists, no attack can be mounted with a non-negligible success probability.

Recall that Corollary 1 allows to convert a counter $\hat{T}$ following a normal distribution into a χ_1^2 variable. Then, since the sum of m independent χ_1^2 variables is distributed according to χ_m^2, we get the following lemma:

Definition 1. *Let* $\mathbf{X_0}, \ldots, \mathbf{X_{m-1}}$ *be normal random variables with*

$$\mathbf{X_0}, \ldots, \mathbf{X_{m-1}} \sim \mathcal{N}(2^{n-1}, 2^{n-2}).$$

Then a test statistic T *defined as*

$$\mathbf{T} = \sum_{i=0}^{m-1} \left(\frac{\mathbf{X_i} - 2^{n-1}}{2^{n/2-1}}\right)^2 \tag{3}$$

follows the χ_m^2 *distribution.*

Noting that $\mathcal{N}(2^{n-1}, 2^{n-2})$ is exactly the distribution of the counters obtained using random approximations, we can use the χ^2 distinguisher in two scenarios which we soon describe. The intuition in both cases is that although the statistical distance is insufficient to distinguish a single biased distribution from a random one, this distance increases when multiple observations are considered. In other words, we use a series of insignificant observations to create a significant one. This is different from previous works which combined a series of significant observations to reduce the data complexity.

4.1 The Multi-key Setting

The first case we consider is the multi-key scenario. In this setting, data is encrypted using the same family of block ciphers by using multiple independent keys $k_0, \ldots, k_{m-1}$. Under our assumption of using the entire codebook, this

means that each key is used to encrypt 2^n plaintexts, using m different keys. This results in a set of m counters, $\hat{T}_0, \ldots, \hat{T}_{m-1}$, where each counter $\hat{T}_i$ counts the number of times the same linear approximation ψ holds for data encrypted under k_i. The goal of the adversary is to distinguish between this case and a case where the data was obtained using m independent random permutations.

Claim 1 *Let $\pi_0, \ldots, \pi_{m-1}$ be a set of m random permutations, ψ a linear approximation, and $\hat{T}_0, \ldots, \hat{T}_{m-1}$ a set of counters such that $\hat{T}_i$ counts the number of times ψ was satisfied when applied to π_i after using the entire codebook. Then, the test statistic T which is defined as*

$$\mathbf{T} = \sum_{i=0}^{m-1} \left(\frac{\hat{T}_i - 2^{n-1}}{2^{n/2-1}}\right)^2 \tag{4}$$

has a χ^2_m distribution.

The reason $\mathbf{T}$ is modeled as a random variable is that in both cases, the counters $\hat{T}_0, \ldots, \hat{T}_{m-1}$ follow a certain distribution. In the case of a set of random permutations, each value is drawn according to Lemma 1. Similarly, for the case of a block cipher, each counter $\hat{T}_i$ is drawn according to the unknown distribution governing the bias.

With that in mind, we can build a confidence interval of size α for χ^2_m and use it for hypothesis testing. An adversary receives a set of counters $\hat{T}_0, \ldots, \hat{T}_{m-1}$ and needs to decide if they were obtained from a set of random permutations $\pi_0, \ldots, \pi_{m-1}$ or from a family of block ciphers with m different keys, i.e., $E_{k_0}, \ldots, E_{k_{m-1}}$. The adversary builds a test statistic T according to Claim 1 and checks if it falls inside the confidence interval. The null hypothesis H_0 is that the counters were obtained using m random permutations, and therefore $\mathbf{T} \sim \chi^2_m$. The alternative hypothesis H_1 is that they were not. The adversary builds a confidence interval $[a, b]$ of size α for χ^2_m, meaning that if H_0 is true, then $\Pr[a \leq T \leq b] = \alpha$. As is inherent in any attack using confidence intervals (such as the classical linear attack), the rate of false positives is $1 - \alpha$.

The fact that we use the χ^2 distribution has major benefits over the classical distinguisher. Firstly, by squaring, it avoids sign considerations that are inherent to classical linear cryptanalysis. Moreover, since the average bias of a random permutation as well as a biased one is 0, using more approximations in the classical setting only make it more similar to the random case. When using the χ^2 distribution, the mean of the distribution increases with more degrees of freedom (i.e., when more counters are used), while a test statistic coming from a block cipher tends to the second moment of the unknown distribution, i.e., since the mean is 0, to its variance.

As an instructive example, we define a toy cipher, Pitz, and use it to test the distinguisher. In Section 5 we test this distinguisher for a real block cipher.

Example 1. Pitz is an 8-bit block cipher using a balanced Feistel structure. Given two 4-bit inputs x_i and y_i, and a subkey k_i, the round function is

$$(x_{i+1}, y_{i+1}) = R_{k_i}(x, y) = ((x_i \boxplus y_i) \oplus k_i, x_{i+1} \oplus (y_i \ggg 1)) \tag{5}$$

where $\boxplus$ is modular addition, $\oplus$ is bitwise addition, and $\ggg$ is a cyclic shift right. The block cipher is composed of only three rounds, i.e.,

$$(x_3, y_3) = E(k_0, k_1, k_2, x, y) = R_{k_3}(R_{k_2}(R_{k_1}(x, y))).$$

Let us now investigate a linear approximation with input mask $(\mathtt{F_x}, \mathtt{F_x})$ and output mask $(\mathtt{a_x}, \mathtt{0})$. Testing this approximation over all keys, we see that there are 4 possible biases: $\{-2^{-7}, -2^{-5.415}, 2^{-5.415}, 2^{-7}\}$, which consist the support of the distribution. The distribution has respective probabilities $\{\frac{3}{8}, \frac{1}{8}, \frac{1}{8}, \frac{3}{8}\}$ suggesting that the average absolute bias over all keys is $\mathbb{E}[|\epsilon|] = 2^{-6.415}$ with variance $2^{-12.415}$. Note that the traditional belief would set the data complexity to $\mathbb{E}^{-2}[|\epsilon|] = (2^{-6.415})^{-2} = 2^{12.83}$ and would conclude that this approximation cannot be used for a linear attack. Recently, Ashur and Rijmen showed in [5] that knowledge of the full distribution allows to refine the data complexity requirements, but even they argue that at least $(2^{-5.415})^2 = 2^{10.83}$ pairs are needed, which are not available in an 8-bit block cipher.

Indeed, when we build a confidence interval of size $\alpha = 0.95$ and try to apply our distinguisher with $m = 1$, the test statistic never falls outside of it. This is also the case for $m = 2$. However, when $m = 4$, the test statistic falls outside the confidence interval 306 times out of 1024, for $m = 8$ and $m = 16$ the number of times is 905 and 1024, respectively. The probability that a random variable with probability 0.05 would fall outside the test interval more than 305, 904, or 1023 times out of 1024 trials is $\Phi(\frac{305-51.2}{6.97}) \approx \Phi(\frac{904-51.2}{6.97}) \approx \Phi(\frac{1023-51.2}{6.97}) \approx 0$. We see that the more keys we consider, the better the distinguisher performs.

4.2 Using Multiple Linear Approximations

the same arguments in favor of the χ^2 distinguisher that were presented in the previous section can be made for using multiple linear cryptanalysis with low-bias approximations. However, in this case, we need to slightly modify Claim 1 as follows:

Claim 2 *Let π be a random permutation, and let $\psi_0, \ldots, \psi_{m-1}$ be a set of m linear approximation, and $\hat{T}_0, \ldots, \hat{T}_{m-1}$ a set of counters such that $\hat{T}_i$ counts the number of times ψ_i was satisfied when applied to π after using the entire codebook. Then, the test statistic $\mathbf{T}$ which is defined as*

$$\mathbf{T} = \sum_{i=0}^{m-1} (\frac{\hat{T}_i - 2^{n-1}}{2^{n/2-1}})^2 \tag{6}$$

has a χ^2_m distribution.

We can again use Pitz to exemplify this.

Example 2. We now investigate the behavior of multiple linear approximations. We look into several linear approximations, all having the same input mask $(\mathtt{f_x}, \mathtt{f_x})$, and output masks: $(\mathtt{8_x}, \mathtt{0_x})$, $(\mathtt{9_x}, \mathtt{0_x})$,$(\mathtt{a_x}, \mathtt{0_x})$, $(\mathtt{b_x}, \mathtt{0_x})$, $(\mathtt{c_x}, \mathtt{0_x})$, $(\mathtt{d_x}, \mathtt{0_x})$,

$(\mathtt{e_x}, \mathtt{0_x})$, and $(\mathtt{f_x}, \mathtt{0_x})$, with average absolute biases of $2^{-6.0}$, $2^{-5.830}$, $2^{-6.415}$, $2^{-6.415}$, $2^{-6.415}$, $2^{-6.678}$, $2^{-6.093}$, $2^{-5.912}$, respectively. Note that these masks are not linearly independent. For example, $f_x = 8_x \oplus 9_x \oplus e_x$. When applying our distinguisher to these approximations (i.e., $m = 8$), we get that in 1024 experiments, the test statistic falls outside the confidence interval 368 times (36%). The probability that a random variable would fall outside a confidence interval of size 0.95 more than 367 times in 1024 experiments is $\Phi(\frac{367-51.2}{6.97}) \approx 0$.

5 Experimental Verification

In this section we present an experimental verification of our distinguisher. We use SPECK32/64 to present several linear distinguishers for 9 and 10 rounds. We start by presenting SPECK, and then move to verify our results.

5.1 SPECK

SPECK is a family of lightweight block ciphers designed by the NSA in 2013 [6]. A member of the family is denoted by SPECK$2n/mn$, where the block size is $2n$ for $n \in \{16, 24, 32, 48, 64\}$, and the key size is mn for $m \in \{2, 3, 4\}$, depending on the desired security.

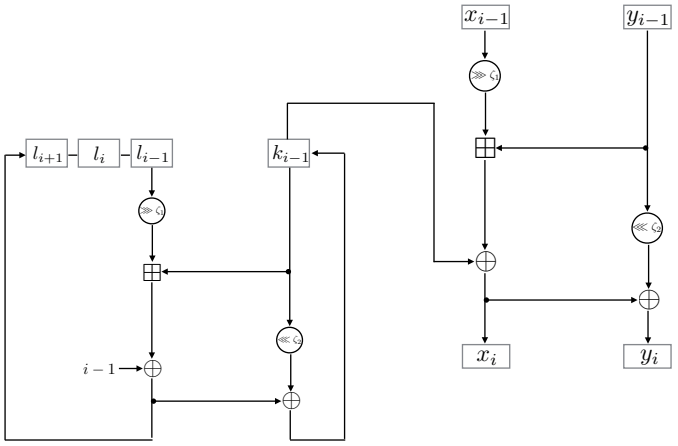

Fig. 1. One round of SPECK

The round function of SPECK receives two words x_i and y_i, and a subkey k_i, all of size n, and outputs two words of size n, x_{i+1} and y_{i+1}, such that

$$(x_{i+1}, y_{i+1}) = F_{k_i}(x_i, y_i) = (f_{k_i}(x_i, y_i), f_{k_i}(x_i, y_i) \oplus (y_i \lll \zeta_1)),$$

Table 3. The 9-round linear approximations used in our experiments.

No. rounds	Trail number	Input mask (left,right)	Output mask (left,right)	Trail bias reported in [32]	Measured bias
9	1	(000D, 56E0)	(170B, 130A)	2^{-15}	$2^{-15.98}$
9	2	(000D, 56E0)	(1D0B, 1B0A)	2^{-15}	$2^{-15.97}$

where $f_{k_i}(\cdot,\cdot)$ is

$$f_{k_i}(x_i, y_i) = ((x_i \ggg \zeta_0) \boxplus y_i) \oplus k_i.$$

The SPECK key schedule algorithm uses the same round function to generate the subkeys. Let $K = (l_{m-2}, ..., l_0, k_0)$ be a master key for SPECK2n, where $l_i, k_0 \in \mathbb{F}_{2^n}$. The sequence of subkeys k_i is generated as

$$k_{i+1} = f_{ct}(l_i, k_i) \oplus (k_i \lll \zeta_1)$$

for

$$l_{i+m-1} = f_{ct}(l_i, k_i),$$

with $ct = i$ the round number starting from 0.

The rotation offsets (ζ_0, ζ_1) are $(7, 2)$ for SPECK32 and $(8, 3)$ for the larger versions. A single round of SPECK with $m = 4$ is depicted in Figure 1. For more details, we refer the interested reader to the original report [6].

In ISC 2015, Yao et al. [32] used an automatic tool to find the longest linear trail for SPECK32/64 using Matsui's Branch-and-Bound framework. As a result, they obtained a 9-round linear trail with bias of 2^{-15} which is guaranteed to be the longest possible trail, and the one with the largest bias among all 9-round trails. This result was later verified using a MILP model developed by Fu et al. in [19]. For differential cryptanalysis, Biryukov et al. presented in [9] a distinguisher also covering 9 rounds with probability 2^{-30}.

We see that the longest distinguishers for SPECK32 cover 9 rounds to-date. In the sequel we present 10 round distinguishers forming, to the best of our knowledge, the best distinguishers for SPECK32, while also validating our results.

5.2 Multi-key Distinguishers for Speck

We start discussing our results with a sanity check. In addition to the 9-round linear trail presented by Yao et al., we found another trail of the same length and the same bias. Both trails are presented in Table 3. We conducted 2048 experiments with each of the trails, using a confidence interval of $\alpha = 0.9$. A success in a single experiment is defined to be "the test statistic falls outside the confidence interval" which, for a random variable, should happen with probability 0.1.[4]

[4]Note that the difference from Table 2 is due to the different size of the confidence interval.

Table 4. Results of the multi-key distinguisher for 9 rounds. The probability reported inside the parentheses is the probability that a binomial random variable will have the reported number of successes or more. We see that the success rate of the distinguisher improves as we consider more approximations obtained using different keys.

m	No. of Experiments	Trail	No. of successes (prob.)
1	2048	1	1069 ($< 2^{-53}$)
	2048	2	1104 ($< 2^{-53}$)
2	1024	1	749 ($< 2^{-53}$)
	1024	2	740 ($< 2^{-53}$)

When using a single key with a single trail, the number of successes in 2048 experiments is 1069 for Trail 1 and 1104 for Trail 2, corresponding to 52.2% of the experiments and 53.9%, respectively.

When setting $m = 2$ and creating a distinguisher based on multiple keys by combining the i key with the $i+1$ key, we get that the number of successes (out of 1024) is 749 for Trail 1 and 740 for Trail 2 corresponding to 73.1% of the experiments and 72.3%, respectively.

For completeness, we model the expected number of successes in a random permutation over N experiments where the success probability is 0.1 (corresponding to the false positive rate) by

$$W \sim \mathcal{B}(N, 0.1).$$

Fixing N, we can compute the probability that W takes a value which is the number of successes x_i or higher, i.e., $p = \Pr[W \geq x_i]$. The results are summarized in Table 4.

We now turn to present our 10-round distinguisher based on multiple keys. Using the 9-round trails as a basis, we extended each of them into 128 10-round trails (i.e., the total number of trails we have is 256, divided into two sets, based on the 9-round distinguisher they were extended from). We first start by applying the distinguisher to each of the approximations independently, and check if it falls inside the confidence interval induced by χ^2_1. The result over $2048 \cdot 256 = 524,288$ experiments is, not surprisingly, 52,215 successes, corresponding to 10%, which is the expected false positive rate. We therefore see that each individual approximation cannot be distinguished from a random one.

When setting $m = 256$ (i.e., each approximation is evaluated against data from 256 different keys), the number of successes over 2048 experiments is 224 corresponding to 10.9% of the experiments. The probability that a random binomial variable with probability 0.1 will have 224 or more success in $N = 2048$ experiments is given by

$$W \sim \mathcal{B}(2048, 0.1)$$
$$\Pr[W \geq 224] = 0.085$$

Table 5. Results of the multi-key distinguisher for 10 rounds. The probability reported inside the parentheses is the probability that a binomial random variable will have the reported number of successes or more. We see that each individual approximation is insufficient for constructing a linear distinguisher, but that a collection of 256 is.

m	No. of Experiments	Trail	No. of successes (Prob.)
1	524,288	1 + 2	52,215 (= 0.838)
256	2048	1 + 2	224 (= 0.085)

The results are summarized in Table 5.

We see that the success rate of the distinguisher improves as we consider more approximations obtained using different keys.

5.3 Multiple Linear Cryptanalysis Using Low-Bias Approximations

We now present a way to use our distinguisher for multiple approximations, where all the approximations have a bias smaller than $2^{-n/2}$. As before, we start with a sanity check using our 9-round linear approximations. As we saw on Section 5.2, when using $\alpha = 0.9$ and $m = 1$, the success rate is 1069 for Trail 1 and 1104 for Trail 2, corresponding to 52.2% of the experiments and 53.9%, respectively.[5] When combining both trails, i.e., setting $m = 2$, the number of successes in 2048 experiments increases to 1429, corresponding to 69.8% of the experiments. The results, as well as the probability to have this number of successes are presented in Table 6.

For a 10-round distinguisher for SPECK32 based on multiple linear approximations we extended each of the 2 trails into 128 10-round trails. Setting $\alpha = 0.9$ and $m = 128$ we get that the number of successes in 2048 experiments is 232 when using Trail 1, and 222 when using Trail 2, corresponding to 11.3% and 10.8%, respectively.

Interestingly, when setting $m = 256$ and using the test statistic over all 256 linear approximations, the obtained result is 223 (10.9%), which is an improvement, but is not as significant as the theory predicts. We conjecture that this behavior is due to the different distributions of the underlying biases for the 9-round linear approximations and leave it for future research.

Instead, we build two test statistics T_1, and T_2, for each group of 10-round linear approximations, and consider the experiments successful if either of the statistics falls outside the confidence interval. We get that when using this test for SPECK32, the number of successes in 2048 experiments is 438 (21.4%), wheres the expected false positive rate in this case is $0.1 + 0.1 - 0.1^2 = 0.19$. A summary of these results is presented in Table 6.

[5] These numbers are the same as in Section 5.2 as they describe the same experiment.

Table 6. Results of a distinguisher using multiple approximations for 9 and 10 rounds. The probability reported inside the parentheses is the probability that a binomial random variable will have the reported number of successes or more.

No. Rounds	m	No. of Experiments	Trail	No. of successes (Prob.)
9	1	2048	1	1069 ($< 2^{-53}$)
		2048	2	1104 ($< 2^{-53}$)
9	2	2048	1+2	1429 ($< 2^{-53}$)
10	1	524,288	1 + 2	52,215 ($= 0.838$)
10	128	2048	1	232 ($= 0.026$)
		2048	2	222 ($= 0.110$)
10	256	2048	1 + 2	223 ($= 0.097$)
10	128+128	2048	1 or 2	438 ($= 0.004$)

6 Conclusion

In this paper we showed how to use a χ^2 distinguisher to detect biases smaller than $2^{-n/2}$ which were previously believed to be unusable. We successfully tested the distinguisher in two settings: the multi-key setting and the multilinear setting. All the results in this paper have been verified experimentally. We showed that in addition to providing a general observation about linear cryptanalysis using low-bias approximations, we can improve the success probability of the previously presented 9-round distinguishing attack in both the multi-key and the multiple approximations scenarios. We were further able to construct several 10-round distinguishers in both settings, which are the longest distinguishers for Speck32/64 to-date.

Since we use a fairly small amount of approximations, future research should be able to further increase the successes probability of the 10-round distinguishers, or extend them to cover even more rounds. In Section 2.1 we discussed related problems in linear cryptanalysis such as getting a good estimation of the bias, and properly deriving the data complexity. We showed that these problems are inherent to all linear attacks, and are independent of our results. In future works we plan to try addressing these related problems. We also plan to refine the way we aggregate counters coming from different distributions. As for a key recovery, while it seems that Matsui's Algorithm 2 can be implemented over the multiple approximations distinguisher as it only uses a single key, it is unclear how a key recovery attack can be implemented in the case of multiple keys. In both cases, more work is required. Naturally, another research direction is to apply our technique to other block ciphers.

References

1. Abdelraheem, M.A., Ågren, M., Beelen, P., Leander, G.: On the Distribution of Linear Biases: Three Instructive Examples. In Safavi-Naini, R., Canetti, R., eds.:

Advances in Cryptology - CRYPTO 2012 - 32nd Annual Cryptology Conference, Santa Barbara, CA, USA, August 19-23, 2012. Proceedings. Volume 7417 of Lecture Notes in Computer Science., Springer (2012) 50–67
2. Alizadeh, J., AlKhzaimi, H., Aref, M.R., Bagheri, N., Gauravaram, P., Kumar, A., Lauridsen, M.M., Sanadhya, S.K.: Cryptanalysis of SIMON Variants with Connections. In Saxena, N., Sadeghi, A., eds.: Radio Frequency Identification: Security and Privacy Issues - 10th International Workshop, RFIDSec 2014, Oxford, UK, July 21-23, 2014, Revised Selected Papers. Volume 8651 of Lecture Notes in Computer Science., Springer (2014) 90–107
3. Ashur, T., Beyne, T., Rijmen, V.: Revisiting the Wrong-Key-Randomization Hypothesis. IACR Cryptology ePrint Archive **2016** (2016) 990
4. Ashur, T., Rijmen, V.: On Linear Hulls and Trails. In Dunkelman, O., Sanadhya, S.K., eds.: Progress in Cryptology - INDOCRYPT 2016 - 17th International Conference on Cryptology in India, Kolkata, India, December 11-14, 2016, Proceedings. Volume 10095 of Lecture Notes in Computer Science. (2016) 269–286
5. Ashur, T., Rijmen, V.: On Linear Hulls and Trails in Simon. IACR Cryptology ePrint Archive **2016** (2016) 88
6. Beaulieu, R., Shors, D., Smith, J., Treatman-Clark, S., Weeks, B., Wingers, L.: The SIMON and SPECK Families of Lightweight Block Ciphers. IACR Cryptology ePrint Archive **2013** (2013) 404
7. Beaulieu, R., Shors, D., Smith, J., Treatman-Clark, S., Weeks, B., Wingers, L.: SIMON and SPECK: Block Ciphers for the Internet of Things. IACR Cryptology ePrint Archive **2015** (2015) 585
8. Biryukov, A., Cannière, C.D., Quisquater, M.: On Multiple Linear Approximations. In Franklin, M.K., ed.: Advances in Cryptology - CRYPTO 2004, 24th Annual International CryptologyConference, Santa Barbara, California, USA, August 15-19, 2004, Proceedings. Volume 3152 of Lecture Notes in Computer Science., Springer (2004) 1–22
9. Biryukov, A., Velichkov, V., Corre, Y.L.: Automatic Search for the Best Trails in ARX: Application to Block Cipher Speck. In Peyrin, T., ed.: Fast Software Encryption - 23rd International Conference, FSE 2016, Bochum, Germany, March 20-23, 2016, Revised Selected Papers. Volume 9783 of Lecture Notes in Computer Science., Springer (2016) 289–310
10. Blondeau, C., Nyberg, K.: Joint Data and Key Distribution of the Linear Cryptanalysis Test Statistic and Its Impact to Data Complexity Estimates of Multiple/Multidimensional Linear and Truncated Differential Attacks. IACR Cryptology ePrint Archive **2015** (2015) 935
11. Blondeau, C., Nyberg, K.: Improved parameter estimates for correlation and capacity deviates in linear cryptanalysis. IACR Trans. Symmetric Cryptol. (to appear) **2016**(2) (2016)
12. Bogdanov, A., Rijmen, V.: Linear hulls with correlation zero and linear cryptanalysis of block ciphers. Des. Codes Cryptography **70**(3) (2014) 369–383
13. Bogdanov, A., Tischhauser, E.: On the Wrong Key Randomisation and Key Equivalence Hypotheses in Matsui's Algorithm 2. In Moriai, S., ed.: Fast Software Encryption - 20th International Workshop, FSE 2013, Singapore, March 11-13, 2013. Revised Selected Papers. Volume 8424 of Lecture Notes in Computer Science., Springer (2013) 19–38
14. Chen, H., Wang, X.: Improved Linear Hull Attack on Round-Reduced Simon with Dynamic Key-guessing Techniques. IACR Cryptology ePrint Archive **2015** (2015) 666

15. Cho, J.Y.: Linear Cryptanalysis of Reduced-Round PRESENT. In Pieprzyk, J., ed.: Topics in Cryptology - CT-RSA 2010, The Cryptographers' Track at the RSA Conference 2010, San Francisco, CA, USA, March 1-5, 2010. Proceedings. Volume 5985 of Lecture Notes in Computer Science., Springer (2010) 302–317
16. Cho, J.Y., Hermelin, M., Nyberg, K.: A New Technique for Multidimensional Linear Cryptanalysis with Applications on Reduced Round Serpent. In Lee, P.J., Cheon, J.H., eds.: Information Security and Cryptology - ICISC 2008, 11th International Conference, Seoul, Korea, December 3-5, 2008, Revised Selected Papers. Volume 5461 of Lecture Notes in Computer Science., Springer (2008) 383–398
17. Daemen, J., Rijmen, V.: Probability distributions of correlation and differentials in block ciphers. J. Mathematical Cryptology **1**(3) (2007) 221–242
18. Dinur, I., Shamir, A.: Cube Attacks on Tweakable Black Box Polynomials. In Joux, A., ed.: Advances in Cryptology - EUROCRYPT 2009, 28th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Cologne, Germany, April 26-30, 2009. Proceedings. Volume 5479 of Lecture Notes in Computer Science., Springer (2009) 278–299
19. Fu, K., Wang, M., Guo, Y., Sun, S., Hu, L.: MILP-Based Automatic Search Algorithms for Differential and Linear Trails for Speck. In Peyrin, T., ed.: Fast Software Encryption - 23rd International Conference, FSE 2016, Bochum, Germany, March 20-23, 2016, Revised Selected Papers. Volume 9783 of Lecture Notes in Computer Science., Springer (2016) 268–288
20. Harpes, C., Kramer, G.G., Massey, J.L.: A Generalization of Linear Cryptanalysis and the Applicability of Matsui's Piling-Up Lemma. In Guillou, L.C., Quisquater, J., eds.: Advances in Cryptology - EUROCRYPT '95, International Conference on the Theory and Application of Cryptographic Techniques, Saint-Malo, France, May 21-25, 1995, Proceeding. Volume 921 of Lecture Notes in Computer Science., Springer (1995) 24–38
21. Hermelin, M., Cho, J.Y., Nyberg, K.: Multidimensional Linear Cryptanalysis of Reduced Round Serpent. In Mu, Y., Susilo, W., Seberry, J., eds.: Information Security and Privacy, 13th Australasian Conference, ACISP 2008, Wollongong, Australia, July 7-9, 2008, Proceedings. Volume 5107 of Lecture Notes in Computer Science., Springer (2008) 203–215
22. Hermelin, M., Nyberg, K.: Linear Cryptanalysis Using Multiple Linear Approximations. IACR Cryptology ePrint Archive **2011** (2011) 93
23. Huang, J., Vaudenay, S., Lai, X., Nyberg, K.: Capacity and Data Complexity in Multidimensional Linear Attack. In Gennaro, R., Robshaw, M., eds.: Advances in Cryptology - CRYPTO 2015 - 35th Annual Cryptology Conference, Santa Barbara, CA, USA, August 16-20, 2015, Proceedings, Part I. Volume 9215 of Lecture Notes in Computer Science., Springer (2015) 141–160
24. Junod, P., Vaudenay, S.: Optimal Key Ranking Procedures in a Statistical Cryptanalysis. In Johansson, T., ed.: Fast Software Encryption, 10th International Workshop, FSE 2003, Lund, Sweden, February 24-26, 2003, Revised Papers. Volume 2887 of Lecture Notes in Computer Science., Springer (2003) 235–246
25. Kaliski, B.S.J., Robshaw, M.J.B.: Linear Cryptanalysis Using Multiple Approximations. In Desmedt, Y., ed.: Advances in Cryptology - CRYPTO '94, 14th Annual International Cryptology Conference, Santa Barbara, California, USA, August 21-25, 1994, Proceedings. Volume 839 of Lecture Notes in Computer Science., Springer (1994) 26–39
26. Matsui, M.: Linear Cryptanalysis Method for DES Cipher. In Helleseth, T., ed.: Advances in Cryptology - EUROCRYPT '93, Workshop on the Theory and

Application of of Cryptographic Techniques, Lofthus, Norway, May 23-27, 1993, Proceedings. Volume 765 of Lecture Notes in Computer Science., Springer (1993) 386–397
27. Nyberg, K.: Linear Approximation of Block Ciphers. In Santis, A.D., ed.: Advances in Cryptology - EUROCRYPT '94, Workshop on the Theory and Application of Cryptographic Techniques, Perugia, Italy, May 9-12, 1994, Proceedings. Volume 950 of Lecture Notes in Computer Science., Springer (1994) 439–444
28. Pascal Junod: On the optimality of linear, differential, and sequential distinguishers. In Biham, E., ed.: Advances in Cryptology - EUROCRYPT 2003, International Conference on the Theory and Applications of Cryptographic Techniques, Warsaw, Poland, May 4-8, 2003, Proceedings. Volume 2656 of Lecture Notes in Computer Science., Springer (2003) 17–32
29. Samajder, S., Sarkar, P.: Another look at normal approximations in cryptanalysis. J. Mathematical Cryptology **10**(2) (2016) 69–99
30. Selçuk, A.A., Biçak, A.: On Probability of Success in Linear and Differential Cryptanalysis. In Cimato, S., Galdi, C., Persiano, G., eds.: Security in Communication Networks, Third International Conference, SCN 2002, Amalfi, Italy, September 11-13, 2002. Revised Papers. Volume 2576 of Lecture Notes in Computer Science., Springer (2002) 174–185
31. Vaudenay, S.: An Experiment on DES Statistical Cryptanalysis. In Gong, L., Stearn, J., eds.: CCS '96, Proceedings of the 3rd ACM Conference on Computer and Communications Security, New Delhi, India, March 14-16, 1996., ACM (1996) 139–147
32. Yao, Y., Zhang, B., Wu, W.: Automatic Search for Linear Trails of the SPECK Family. In Lopez, J., Mitchell, C.J., eds.: Information Security - 18th International Conference, ISC 2015, Trondheim, Norway, September 9-11, 2015, Proceedings. Volume 9290 of Lecture Notes in Computer Science., Springer (2015) 158–176
33. Zheng, L., Zhang, S.: FFT-based multidimensional linear attack on PRESENT using the 2-bit-fixed characteristic. Security and Communication Networks **8**(18) (2015) 3535–3545