Structural and Statistical Analysis of Multidimensional Linear Approximations of Random Functions and Permutations

Tomer Ashur, Mohsin Khan, and Kaisa Nyberg

Abstract—**The goal of this paper is to investigate linear approximations of random functions and permutations. Our motivation is twofold. First, before the distinguishability of a practical cipher from an ideal one can be analysed, the cryptanalyst must have an accurate understanding of the statistical behaviour of the ideal cipher. Secondly, this issue has been neglected both in old and in more recent studies, particularly when multiple linear approximations are being used simultaneously. Traditional models have been based on the average behaviour and simplified using other assumptions such as independence of the linear approximations. Multidimensional cryptanalysis was introduced to avoid making artificial assumptions about statistical independence of linear approximations. On the other hand, it has the drawback of including many trivial approximations that do not contribute to the attack but just cause a waste of time and memory. We show for the first time in this paper that the trivial approximations reduce the degree of freedom of the related χ^2 distribution. Previously, the affine linear cryptanalysis was proposed to allow removing trivial approximations and, at the same time, admitting a solid statistical model. In this paper, we identify another type of multidimensional linear approximation, called Davies-Meyer approximation, which has similar advantages, and present full statistical models for both the affine and the Davies-Meyer type of multidimensional linear approximations. The new models given in this paper are realistic, accurate and easy to use. They are backed up by standard statistical tools such as Pearson's χ^2 test and finite population correction and demonstrated to work accurately using practical examples.**

Index Terms—**Cryptography, block ciphers, linear cryptanalysis, random Boolean functions, statistical distributions.**

I. Introduction

A. Modelling Linear Key-Recovery Attacks

LINEAR cryptanalysis is a statistical method used for distinguishing a block cipher from a random permutation and can be extended to key-recovery attacks of practical block ciphers. It makes use of the nonrandom behaviour of certain linear approximations of the cipher. Linear approximations are single-bit values obtained by exclusive-or summation of certain input bits and output bits over some rounds of the block cipher.

In the setting of linear key-recovery attacks, the traditional heuristic assumption is that a keyed iterated block cipher becomes a pseudorandom function or permutation if some of its rounds are replaced by encryption using a wrong key. On the other hand, if the key is correct, then the data is computed from the cipher. Distinguishing between these two cases using statistical tests requires statistical models of the test statistic. For a recent overview of the existing models, we refer to [1]. Such statistical models are always based on trade-offs between accuracy and feasibility. The traditional approach has been to state some unproven assumptions, called as *wrong-key hypothesis* and *right-key hypothesis*, which are desired to capture the statistical behaviour, but still simple enough to allow feasible computation of the model.

In all existing studies, the wrong-key hypothesis in linear cryptanalysis, as well as in other statistical attacks, is based on a statistical model of the family of random permutations, when the target cipher is a block cipher, or a model of the family of random functions in some other cases such as stream ciphers. Then the main effort in the cryptanalytic attack is focused on identifying and demonstrating evidence of nonrandom behaviour in the target cipher. In linear cryptanalysis, the problem is to find bit combinations that are either strongly biased, or equal to zero for all keys. The known search algorithms for finding suitable biased linear approximations are based on Matsui's seminal work [2], where biased linear approximations were found by identifying one or more strong linear approximation trails that the linear approximations is composed of. The right-key hypothesis is then derived from a statistical model that captures the probability distributions and their parameters of the linear approximations in the case of the cipher.

The success probability and the data complexity of the attack are then estimated based on statistical distinguishing between the probability distributions in the right-key case and the wrong-key case. Since the wrong-key case is typically modelled using linear approximations of randomly and uniformly selected permutations, it is clear that a proper understanding of the random behaviour has an essential role in statistical cryptanalysis.

Manuscript received March 1, 2020; revised March 25, 2021; accepted November 8, 2021. The work of Tomer Ashur was supported by an FWO Post-Doctoral Fellowship under Grant 12ZH420N. *(Corresponding author: Kaisa Nyberg.)*

Tomer Ashur is with the imec—COSIC, KU Leuven, 3000 Leuven, Belgium, and also with the Department of Mathematics and Computer Science, TU Eindhoven, 5612 AZ Eindhoven, The Netherlands (e-mail: t.ashur@tue.nl).

Mohsin Khan was with the Department of Computer Science, University of Helsinki, 00560 Helsinki, Finland. He is now with the Ericsson Research, Ericsson AB, 16480 Stockholm, Sweden (e-mail: mohsin.a.khan@ericsson.com).

Kaisa Nyberg, retired, was with the Department of Computer Science, Aalto University, 00076 Espoo, Finland. She resides in 00100 Helsinki, Finland (e-mail: kaisa.nyberg@aalto.fi).

Communicated by S. Fehr, Associate Editor for Cryptography.

Digital Object Identifier 10.1109/TIT.2021.3128618

Along the history of the linear cryptanalysis, the wrong-key hypothesis has taken different forms, and the main contributions are rather scattered in the literature. The first goal of this paper is to give a concise presentation of the behaviour of random functions and permutations under linear cryptanalysis. Our second goal is to present a new and more realistic model of the wrong-key hypothesis for the multidimensional linear cryptanalysis. The statistical behaviour of a multidimensional linear approximation appears to depend significantly on its structure.

B. Existing Wrong-Key Models in Linear Cryptanalysis

The understanding about the statistical behaviour of linear approximations of random functions and permutations has developed a lot during the times. In early works, correlations of linear approximations of random permutations were estimated to be negligible and equal to their expected value, zero. While it was understood already in 1994 by O'Connor [3] that the correlations of linear approximations vary within the random permutations, it was not until in 2006 this fact was examined in more detail by Daemen and Rijmen [4]. They considered the probability distribution of correlations of linear approximations both for random functions and random permutations and showed that both distributions behave similarly and can be approximated using normal distributions with the same parameters with the only distinction that the interval of the discrete distribution of correlations can have only even values for permutations.

These advanced models of linear approximations of random functions and permutations led to the observation that if a linear approximation of a cipher has correlation equal to zero for all keys, then it is not random and can be distinguished from random [5]. Conversely, this means that under the traditional hypothesis, according to which correlations of linear approximations of random permutations are equal to zero, even a truly randomly selected permutation will be falsely identified as nonrandom, because the correlations of their linear approximations are usually nonzero. This example illustrates how important it is to state the wrong-key assumption accurately.

The wrong-key model of [4] was extended by Bogdanov and Tischhauser [6] by integrating data sampling into it. While being an important opening to key-dependent models, it had two main drawbacks. First, the right-key model was still based on the assumption that the correlation of the linear approximation is equally large (in absolute value) for all keys. Secondly, the plaintexts were assumed to be drawn with replacement. While giving realistic estimates for small sample sizes, this approach lead to significant deviations from the true behaviour when the sample size approaches the full codebook. These two drawbacks of that model were highlighted by the counterintuitive phenomenon that the success probability is not always an increasing function of the data-complexity. The underlying problems were corrected by a new model given in [7] for a single linear approximation based on a single dominant trail. A more detailed study of the conditions for this counterintuitive phenomenon of nonmonotonicity was given in [8].

With the goal of making the linear distinguishers more powerful, several authors have proposed to use multiple linear approximations simultaneously. In the early models, the wrong-key hypothesis was always based on the assumption that in the wrong-key case, the expected correlations, that is, the correlations of linear approximations computed for the full codebook of the cipher behave as on average, that is, are equal to zero [9], see also [10]. Recently, key-dependency has been integrated to the models both in the wrong key and right key cases [7], [11], [12] by adopting a simplifying assumption that the correlations of any set of multiple linear approximations are independent when considered over the set of all permutations. In a subsequent version [1] of [11], this assumption was stated only for correlations of linearly independent linear approximations of random permutations. Whether this means a true theoretical improvement is not known.

In general, not much is known about the statistical independence of correlations considered as random variables over the key space. Only correlations of components of balanced functions are known to be independent trivially as they are always constants, that is, equal to zero. A multidimensional linear approximation of a permutation is not in general a balanced function. Hence the correlations of its components may not be equal to zero and may have statistical dependencies.

The assumption about independence of correlations was needed to derive statistical distributions for the sum of the squared correlations of the linear approximations. More specifically, the independence assumption has been used for expressing the variance of the sum of squared correlations as the sum of the variances of the squared correlations of the individual linear approximations. In this paper, it will be shown that, for certain sets of linear approximations, this result can be achieved without the independence assumption.

C. Our Contributions

We start by deriving exact formulas for the mean and variance of the capacity of multinomially distributed variables and make the observation that the variance of the capacity is additive, that is, it can be expressed as the sum of the variances of the capacities of the individual variables in the case when the expected distribution is uniform. This corresponds to the case of the expected value distribution of a random function.

We continue by revisiting the distributions of correlations of single linear approximations of random functions and random permutations. Adding to the results of [4] we observe that a linear approximation of a random function is a random Boolean function, while this is not the case if the random functions are restricted to permutations. We give the discrete probability distribution of the correlation of a linear approximation of a random permutation in terms of a hypergeometric distribution.

While multidimensional linear approximations of some functions can be modelled using the multinomial distribution, this is never the case for a multidimensional linear approximation of permutations. Even in case of a single variable, the hypergeometric distribution must be used instead of the binomial distribution. We leave it an open question whether the multivariate hypergeometric distribution might give a feasible approach in this case, and instead, use continuous approximations of the probability distributions to model the statistical behaviour of the capacity of a multidimensional linear approximation of a random permutation. This leads us to the study of the χ^2 distribution.

In many practical applications of multidimensional linear cryptanalysis, the linear space of linear approximations contains many trivial approximations that have correlation zero for any permutation. Their impact has been ignored in previous works and the degree of freedom of the χ^2 distribution is taken equal to $2^t - 1$ where t is the dimension of the multidimensional linear approximation, see e.g. [7]. We prove that in the presence of trivial approximations, the degree of freedom is strictly less than $2^t - 1$. Moreover, we conjecture the correct value of the degree of freedom and present experimental evidence to support this conjecture. We also identify a new type of multidimensional linear approximation, which we call the Davies-Meyer approximation, and which is characterised by the property of not containing any trivial linear approximations.

Having found a realistic solution to the problem of how to model wrong-key behaviour for multidimensional linear cryptanalysis, we apply the same approach for the recently presented variant of linear cryptanalysis, named as affine multidimensional cryptanalysis [13]. Preliminary versions of these results appeared in [14].

Affine subsets of linear approximations naturally arise in many ciphers. As an example we analyse SIMON32/64, which is a Feistel cipher that employs bitwise AND operation as the only nonlinear component of the round function. In this case, the affine spaces originate from the linear spaces comprising the four linear approximations of the AND operation that have non-zero correlations. Using our χ^2 model of affine linear approximations of randomly selected permutations, we experimentally identify nonrandom behaviour of 2-dimensional affine sets of linear approximations over 13-18 rounds and a 6-dimensional affine set over 18 rounds of SIMON32/64. These experiments used full codebook of data and 2^{13} randomly selected keys. We also performed similar experiments with less than the full codebook of data.

D. Outline

The standard definitions of linear cryptanalysis are recalled and the mean and variance of capacity are computed for a general multinomial distribution in Section II, where we also recall the related discrete probability distributions and their continuous approximations. In Section III, the distributions of correlations of single linear approximations are revisited. The new contributions of the structure and probability distributions of multidimensional linear approximations are presented in Section IV and applied to affine sets of approximations in Section V. Then we enhance these statistical models by integrating random sampling without replacement to them in Section VI. To perform randomness analysis of linear approximations of SIMON32/64, we define the randomness test in Section VII and present the results in Section VIII. The conclusions are drawn in Section IX.

II. Capacity of Vectorial Boolean Functions

A. Correlation and Capacity

Let F be a function from S to $\mathbb{F}_2^t$, where S is a finite set and $\mathbb{F}_2^t$ is a vector space over $\mathbb{F}_2$ of dimension t. We focus on two ways of defining F. First, we can just give the (indexed) set of the values $F(x)$, $x \in S$. The second way of defining F is to give t Boolean functions $f_1, \ldots, f_t$, that is, t coordinate functions of F, and their values $f_i(x)$, $x \in S$, $i = 1, \ldots, t$. Given $\beta = (\beta_1, \ldots, \beta_t) \in \mathbb{F}_2^t$, we denote by $\beta \cdot F$ the linear combination of the coordinate functions of $F = (f_1, \ldots, f_t)$ determined by β, that is,

$$\beta \cdot F = \beta_1 f_1 + \ldots + \beta_t f_t,$$

and say that the Boolean function $\beta \cdot F$ is a component of F.

Functions are in general imbalanced, that is, all values in the image space are not taken equally often. Related to the two ways of defining F, we have two ways of measuring the imbalance of F. First, we can consider the uniformity of its value distribution. Given $\eta \in \mathbb{F}_2^t$ let us denote by p_η its probability, that is,

$$p_\eta = |S|^{-1} |\{x \in S \mid F(x) = \eta\}|.$$

Then the imbalance of this distribution is measured using the capacity

$$\mathrm{Cap}(F) = 2^t \sum_{\eta \in \mathbb{F}_2^t} (p_\eta - 2^{-t})^2. \tag{1}$$

Secondly, we can consider the imbalance of its components using correlations. Let f be a Boolean function from S to $\mathbb{F}_2$. Then its correlation $\mathrm{cor}(f)$ is given by

$$\mathrm{cor}(f) = |S|^{-1}(|\{x \in S | f(x) = 0\}| - |\{x \in S | f(x) = 1\}|). \tag{2}$$

It is well-known, see e.g. [15], [16], that these two approaches to measuring imbalance are related due to the following equality,

$$p_\eta = 2^{-t} \sum_{\beta \in \mathbb{F}_2^t} (-1)^{\beta \cdot \eta} \mathrm{cor}(\beta \cdot F), \text{ for all } \eta \in \mathbb{F}_2^t, \tag{3}$$

or equivalently, by the Walsh-Hadamard transform,

$$\mathrm{cor}(\beta \cdot F) = \sum_{\eta \in \mathbb{F}_2^t} (-1)^{\beta \cdot \eta} p_\eta, \text{ for all } \beta \in \mathbb{F}_2^t. \tag{4}$$

Then we can express $\mathrm{Cap}(F)$ also as

$$\mathrm{Cap}(F) = \sum_{\beta \in \mathbb{F}_2^t, \beta \neq 0} \mathrm{cor}(\beta \cdot F)^2. \tag{5}$$

In particular, a random function $F : S \to \mathbb{F}_2^t$ can be generated either by selecting its t coordinate functions randomly and independently, or by picking its values $F(x)$ randomly and independently from $\mathbb{F}_2^t$. The value distribution of a random function F follows a multinomial distribution. By (5) the expected value of the capacity of the value distribution of a random function is the sum of the expected values of the squared correlations taken over the non-trivial components of F. We are also interested to determine the variance of the capacity for random functions. The problem is not trivial, since we can neither assume all nonzero components of F to be independent, nor to have independent correlations. Nevertheless, in the next subsection we give a result, see Corollary 1, which shows that, based solely on the properties of the multinomial distribution of the values of a random function F, the variance of its capacity is obtained as the sum of the equal variances of the squared correlations of its nonzero components.

B. Capacity Related to Multinomial Distributed Variables

In the preceding section, the notion of capacity was defined as a measure of the uniformity of the value distribution of the function. More generally, we can define capacity for any finite set of non-negative values. Let $z_1, \ldots, z_k$ be non-negative real numbers and denote by m their sum. Then we define their capacity as

$$\frac{k}{m^2}\sum_{\eta=1}^{k}\left(z_\eta - \frac{m}{k}\right)^2. \tag{6}$$

This quantity is related to the Euclidean distance of the probability distribution from the uniform distribution and also called as the squared Euclidean imbalance. By substituting $z_\eta = |\{\, x \in S \mid F(x) = \eta \,\}|$ and $k = 2^t$ to (6), we have $m = |S|$ and we get the capacity of F as defined by (1). Next we determine the mean and variance of the capacity for stochastic variables that follow a general multinomial distribution.

Let $z_1, \ldots, z_k$ be the outcomes of a set of k, $k \geq 2$, stochastic variables that follow a multinomial distribution with probabilities $p_1, \ldots, p_k$ and let us denote the number of trials by m. Then $z_1 + \cdots + z_k = m$. Let us denote by C the capacity of $z_1, \ldots, z_k$ as given by (6). Then C is also an outcome of a stochastic variable. The proof of the following result is given in Appendix A.

Theorem 1: Let C be the capacity of multinomially distributed variables and let the parameters of the multinomial distribution be $p_1, \ldots, p_k$ and m. Then

$$\begin{aligned}\mathrm{Exp}(C) &= \frac{k-1}{m} + \frac{(m-1)k}{m}\sum_{\eta=1}^{k}(p_\eta - \frac{1}{k})^2\\ \mathrm{Var}(C) &= \frac{(m-1)k^2}{m^3}\left((4m-8)P_3-(4m-6)P_2^2+2P_2\right),\end{aligned}$$

where

$$P_2 = \sum_{\eta=1}^{k} p_\eta^2 \quad \text{and} \quad P_3 = \sum_{\eta=1}^{k} p_\eta^3.$$

Note that in the expression of the expected capacity we have

$$k\sum_{\eta=1}^{k}\left(p_\eta - \frac{1}{k}\right)^2 = kP_2 - 1,$$

which is the capacity of the values $p_1, \ldots, p_k$.

If $p_\eta = \frac{1}{k}$, for all $\eta = 1, \ldots, k$, then $P_2 = 1/k$ and $P_2^2 = P_3 = 1/k^2$, and the mean and variance of the capacity of multinomially distributed variables are given by the following corollary.

Corollary 1: Let C be the capacity of a multinomially distributed variable with distribution parameters $p_\eta = \frac{1}{k}$, for all $\eta = 1, \ldots, k$, and m. Then

$$\begin{aligned}\mathrm{Exp}(C) &= \frac{k-1}{m}\\ \mathrm{Var}(C) &= \frac{2(k-1)(m-1)}{m^3}.\end{aligned}$$

C. Standard Probability Distributions

The normal distribution is denoted by $\mathcal{N}(\mu, \sigma^2)$, where μ is the mean and σ^2 is the variance. In case $\mu = 0$ and $\sigma^2 = 1$ this distribution is called the standard normal distribution.

The binomial distribution is the multinomial distribution with $k = 2$ and is denoted by $\mathcal{B}(m, p)$, where $p = p_1$ and $1 - p = p_2$. The mean and variance of this distribution are mp and $mp(1-p)$, respectively. The binomial distribution corresponds to random sampling with replacement from a set S of size $M = |S|$, where we have two types of elements, denoted by 0 and 1. If the sampling is without replacement then the number of outcomes of type 0 in m experiments follows the hypergeometric distribution $\mathcal{HG}(M, K, m)$, where K is the number of elements of type 0 in the entire S. The mean and variance of the hypergeometric distribution are

$$m\frac{K}{M} = mp \text{ and } m\frac{K}{M}\frac{M-K}{M}\frac{M-m}{M-1} = mp(1-p)\frac{M-m}{M-1},$$

respectively, where we denoted by p the probability of outcomes of type 0 in the entire set S, that is, $p = \frac{K}{M}$. The variances of the binomial and hypergeometric distributions differ by a factor, whose close estimate

$$B = \frac{M-m}{M}, \tag{7}$$

is called the finite population correction coefficient [17]. For sufficiently large S, both distributions can be approximated by the normal distribution $\mathcal{N}(\mu, \sigma^2)$, where μ is the mean and σ^2 is the variance, as follows:

$$\mathcal{B}(m, p) \approx \mathcal{N}(mp, mp(1-p)) \tag{8}$$

$$\mathcal{HG}(M, K, m) \approx \mathcal{N}(mp, mp(1-p)B). \tag{9}$$

The general (noncentral) chi-squared distribution with ℓ degrees of freedom and noncentrality parameter δ is denoted by $\chi^2_\ell(\delta)$. It is defined as the probability distribution of the sum of squares of ℓ independent random variables that follow the normal distribution $\mathcal{N}(\mu_i, 1)$. Then $\delta = \sum_{i=1}^{\ell} \mu_i$. The mean of the $\chi^2_\ell(\delta)$ distribution is $\ell + \delta$ and its variance is $2(\ell + 2\delta)$. If $\delta = 0$ then the distribution is called central and is denoted by χ^2_ℓ.

Another setting that gives rise to a chi-square distribution is the one of the multinomial distribution. Using the same notation as in Subsection II-B we set

$$T = \sum_{\eta=1}^{k}\frac{\left(z_\eta - \frac{m}{k}\right)^2}{\frac{m}{k}}. \tag{10}$$

Then T is the test statistic of the well-known Pearson's chi-squared test and it is known to follow the $\chi^2_{k-1}(\delta)$ distribution, where

$$\delta = \sum_{\eta=1}^{k}\frac{\left(mp_\eta - \frac{m}{k}\right)^2}{\frac{m}{k}}, \tag{11}$$

see, e.g., [18]. Note that in this setting the number of degrees of freedom is the number of variables that are free to vary, that is $k-1$, the size of the domain of the multinomial distribution minus one, due to the constraint $z_1 + \ldots + z_k = m$. If there are other constraints, then the number of degrees of freedom may be further reduced. For example, s additional linearly independent linear constrains on the values z_η will further reduce the number of degrees of freedom to $k - 1 - s$.

By the expression (6) of C we have $T = mC$. Hence the χ^2 distribution of T can be used to give a continuous approximation of the discrete probability distribution of C. For example, we can compare the mean and variance of C given by Corollary 1 in the case where $p_\eta = 1/k$ for all $\eta = 1, \ldots, k$ with the ones obtained from the χ^2 distribution of T. We can see that the means are identical, while the variances differ by a negligible term $2(k-1)/m^3$.

The multinomial distribution and the related Pearson's χ^2 distribution apply to the case when the values z_η are obtained by drawing samples of m elements from S with replacement. If sampling is without replacement then the multivariate hypergeometric distribution shall be used instead of the multinomial distribution. Then the statistic T given in (10) must be multiplied by the inverse of the finite population correction coeffifient to get a χ^2-distributed variable [17]. We state this result for further reference as follows.

Lemma 1: Let T be given by (10) where the values of variables z_{eta}, $\eta = 1, \ldots, k$ are obtained by sampling m elements from S without replacement and the initial probabilities p_η are as defined in the setting of the multinomial distribution. Then the variable

$$B^{-1}T,$$

where B is given by (7), approximately follows $\chi^2_{k-1}(\delta)$ distribution, where δ is given by (11).

III. Probability Distribution of a Single Linear Approximation of a Random Function and Permutation

We denote by $\mathbb{F}_2^n$ the linear space over the field $\mathbb{F}_2 = \{0, 1\}$ with addition denoted by '$+$' and inner product denoted by '$\cdot$'. Let f be a Boolean function from $\mathbb{F}_2^n$ to $\mathbb{F}_2$. Given an element $a \in \mathbb{F}_2^n$ the Boolean function defined as

$$x \mapsto f(x) + a \cdot x$$

is called a linear approximation of f. We first derive the distributions of linear approximations of random Boolean functions and random balanced Boolean functions. They are essentially the same as those given by Daemen and Rijmen in [4]. In this section, we will complete their work by giving the exact distributions in both cases.

A. Zeroes of Linear Approximations

Let f be a Boolean function from $\mathbb{F}_2^n$ to $\mathbb{F}_2$. We say that $x \in \mathbb{F}_2^n$ is a zero of f if $f(x) = 0$. To determine the correlation of a linear approximation of f, let us first determine the number of its zeroes.

Lemma 2: Let $a \in \mathbb{F}_2^n$, $a \neq 0$, and f be a Boolean function from $\mathbb{F}_2^n$ to $\mathbb{F}_2$. Let N_0 be the number of the zeroes of f. Then the number of zeroes of the linear approximation $g(x) = f(x) + a \cdot x$ is equal to

$$\begin{aligned} &|\{x \in \mathbb{F}_2^n \mid f(x) = 0,\, a \cdot x = 0\}| \\ &+|\{x \in \mathbb{F}_2^n \mid f(x) = 1,\, a \cdot x = 1\}| \\ &= 2^{n-1} - N_0 + 2\upsilon, \end{aligned}$$

where we denoted

$$\upsilon = |\{x \in \mathbb{F}_2^n \mid f(x) = 0,\, a \cdot x = 0\}| \qquad (12)$$

Proof: Since the nonzero linear function $x \mapsto a \cdot x$ is balanced, we have

$$|\{x \in \mathbb{F}_2^n \mid f(x) = 1,\, a \cdot x = 1\}| = 2^{n-1} - (N_0 - \upsilon).$$

Adding υ to both sides of this equation gives what is claimed. ■

The following lemma gives the distribution of υ.

Lemma 3: Let Boolean function f over $\mathbb{F}_2^n$ be chosen randomly and equiprobably from the set of all Boolean functions with a fixed number N_0 of zeroes. Let $a \in \mathbb{F}_2^n$ be nonzero and fixed. Then υ defined by (12) follows the hypergeometric distribution $\mathcal{HG}(2^n, 2^{n-1}, N_0)$.

Proof: Given a fixed balanced linear function $a \cdot x$, the N_0 zeroes of f are chosen by choosing υ zeroes among the 2^{n-1} zeroes of $a \cdot x$ and $N_0 - \upsilon$ zeroes among the 2^{n-1} inputs x such that $a \cdot x = 1$. ■

B. Random Boolean Function

The number of zeroes of a Boolean function selected randomly and equiprobably from the set of all Boolean functions of n variables follows the binomial distribution $\mathcal{B}(2^n, \frac{1}{2})$.

Theorem 2: Let f be selected randomly and equiprobably from the set of all Boolean functions of n variables. Then for any fixed $a \in \mathbb{F}_2^n$ the number of zeroes of the linear approximation $a \cdot x + f(x)$ follows a binomial distribution $\mathcal{B}(2^n, \frac{1}{2})$.

Proof: For any fixed Boolean function g, the mapping, which maps a Boolean function f to the function $f + g$, is a bijection in the set of all Boolean functions of n variables. Then if f is chosen uniformly at random from this set then so is $f + g$. In particular, the distribution of the number of zeroes of $f + g$ follows the same distribution as the number of zeroes of f. ■

For an alternative proof that computes the distribution of the zeroes of the linear approximation based on Lemma 3, see Appendix B.

Now we apply Corollary 1 for $k = 2$ to get the following result.

Corollary 2: Let $a \in \mathbb{F}_2^n$ be fixed. The distribution of a correlation c of a linear approximation $a \cdot x + f(x)$ of a Boolean function f that is drawn randomly and equiprobably from the set of all Boolean functions of n variables has the following parameters:

$$\begin{aligned} \mathrm{Exp}(c) &= 0 \\ \mathrm{Var}(c) &= \mathrm{Exp}(c^2) = 2^{-n} \\ \mathrm{Var}(c^2) &= 2^{1-2n} - 2^{1-3n}. \end{aligned}$$

Proof: When $k = 2$ we have $C = c^2$ by (5) and we can apply Corollary 1 with $m = 2^n$ to get $\mathrm{Exp}(c^2) = 2^{-n}$

and $\mathrm{Var}(c^2) = 2(2^n - 1)2^{-3n}$. Further, by Theorem 2 we have that $\mathrm{Exp}(2^n c) = 0$. Hence $\mathrm{Exp}(c) = 0$ and $\mathrm{Var}(c) = \mathrm{Exp}(c^2) - \mathrm{Exp}(c)^2 = 2^{-n}$. ■

C. Balanced Random Boolean Function

A Boolean function over $\mathbb{F}_2^n$ is said to be balanced if its number of zeroes is equal to 2^{n-1}. It is well known that a vectorial Boolean function is a permutation if and only if all its components, that is, nonzero linear combinations of its coordinate functions are balanced.

From Lemma 2 and Lemma 3 we get the following result.

Theorem 3: Let f be selected randomly and equiprobably from the set of all balanced Boolean functions of n variables. Then for any fixed $a \in \mathbb{F}_2^n$, $a \neq 0$, the number of zeroes of the linear approximation $f(x) + a \cdot x$ is an even integer 2υ where $\upsilon \sim \mathcal{H}G(2^n, 2^{n-1}, 2^{n-1})$.

Corollary 3: The distribution of a correlation $c = \mathrm{cor}(g)$ of a linear approximation $g(x) = a \cdot x + f(x)$ of a balanced Boolean function f drawn randomly and equiprobably from the set of all balanced Boolean functions of n variables has the following parameters:

$$\mathrm{Exp}(c) = 0 \text{ and } \mathrm{Var}(c) = \mathrm{Exp}(c^2) = \frac{1}{2^n - 1}.$$

Proof: We have $c = 2^{2-n}\upsilon - 1$, where

$$\mathrm{Exp}(\upsilon) = 2^{n-2} \text{ and } \mathrm{Var}(\upsilon) = \frac{(2^{n-2})^2}{2^n - 1}.$$

■

D. Random Vectorial Boolean Function and Permutation

In the context of linear cryptanalysis, a linear approximation of a vectorial Boolean function $F : \mathbb{F}_2^n \to \mathbb{F}_2^s$ is identified with a Boolean function defined as

$$x \mapsto a \cdot x + b \cdot F(x), \ x \in \mathbb{F}_2^n,$$

where $a \in \mathbb{F}_2^n$ and $b \in \mathbb{F}_2^s$, $b \neq 0$.

Since a single component $b \cdot F(x)$, $b \neq 0$, of a random vectorial Boolean function is a random Boolean function, it follows that the number of zeroes of a linear approximation of a random vectorial Boolean function is binomially distributed as given by Theorem 2.

For permutations, the nonzero component functions $b \cdot F(x)$ are balanced Boolean functions. Therefore, the distribution of the zeroes of a single linear approximation of a permutation drawn uniformly at random among all permutations is given by Theorem 3.

IV. Multidimensional Linear Approximations of Permutations

A. Multidimensional Linear Approximation as a Vectorial Boolean Function

Let $F : \mathbb{F}_2^n \to \mathbb{F}_2^s$ be a vectorial Boolean function. A multidimensional linear approximation Λ is a vectorial Boolean function such that the components of Λ form a linear subspace of the linear space

$$\{\, g : \mathbb{F}_2^n \to \mathbb{F}_2 \,|\, g(x) = a \cdot x + b \cdot F(x) \,\}$$

of all linear approximations of F. Let us denote this subspace by L and its dimension by t. Let us fix a basis $\lambda_1, \ldots, \lambda_t$ of L, and give notations of the basic elements as

$$\lambda_i(x) = a_i \cdot x + b_i \cdot F(x), \text{ for } i = 1, \ldots, t. \tag{13}$$

Then $\Lambda : \mathbb{F}_2^n \to \mathbb{F}_2^t$ is given by $\lambda_1, \ldots, \lambda_t$ as its coordinate functions. Given $\beta = (\beta_1, \ldots, \beta_t) \in \mathbb{F}_2^t$ the component $\beta \cdot \Lambda$ of Λ has a unique representation as a linear approximation of F of the form $a \cdot x + b \cdot F(x)$ as follows:

$$\begin{aligned} \beta \cdot \Lambda(x) &= \beta_1\lambda_1(x) + \ldots + \beta_t\lambda_t(x) \\ &= \left(\sum_{i=1}^{t} \beta_i a_i\right) \cdot x + \left(\sum_{i=1}^{t} \beta_i b_i\right) \cdot F(x). \end{aligned} \tag{14}$$

In the rest of this paper we identify a linear approximation $g(x) = a \cdot x + b \cdot F(x)$ with the pair $(a, b) \in \mathbb{F}_2^n \times \mathbb{F}_2^s$, and call it the mask pair of g. Here the element $a \in \mathbb{F}_2^n$ is called the input mask and the element $b \in \mathbb{F}_2^s$ the output mask. We also denote $\mathrm{cor}(g)$ by $\mathrm{cor}(a, b)$. Also the linear subspace L spanned by $\lambda_1, \ldots, \lambda_t$ of the space of all linear approximations is identified with a linear subspace of $\mathbb{F}_2^n \times \mathbb{F}_2^s$ spanned by the mask pairs $(a_1, b_1), \ldots, (a_t, b_t)$ given by (13). We will use L also to denote this subspace and call it the mask space of Λ.

By (1), (5), and (14) the capacity of Λ is then given as

$$\begin{aligned} \mathrm{Cap}(\Lambda) &= \sum_{(a,b)\in L, (a,b)\neq(0,0)} \mathrm{cor}(a, b)^2 \\ &= \sum_{\beta\in\mathbb{F}_2^t, \beta\neq 0} \mathrm{cor}(\beta \cdot \Lambda)^2 \\ &= 2^t \sum_{\eta\in\mathbb{F}_2^t} (p_\eta - 2^{-t})^2, \end{aligned} \tag{15}$$

One known consequence of this result is that the value distribution of a multidimensional linear approximation is uniform if and only if the correlations of all its non-zero linear approximations are equal to zero.

B. Structure of Multidimensional Linear Approximation of Permutation

In this section we determine the structural properties of a multidimensional approximation of a permutation F. For example, F is an encryption function of a block cipher, or some rounds of a block cipher with a fixed key, or F is just any bijective function of bit strings.

A multidimensional linear approximation Λ of a permutation F may contain nonzero linear approximations with mask pairs of the form $(a, 0)$ or $(0, b)$. Such linear approximations are called trivial, because they have fixed correlations equal to zero for any permutation F. Next we examine their effect on the distribution of the capacity $\mathrm{Cap}(\Lambda)$. Let us denote by U the linear subspace of the multidimensional approximation consisting of the approximations of the form $(a, 0)$ and let u be its dimension. Similarly, we denote by V the subspace of the masks of the form $(0, b)$ and by v its dimension. Then $U \cap V = \{(0, 0)\}$. Often such spaces span the whole multidimensional approximation, that is, all masks are of the form (a, b), where $(a, 0) \in U$ and $(0, b) \in V$. Then the multidimensional approximation is said to have independent input and output masks [19]. But in general, there may exist a linear subspace W of L such that, if $(a, b) \in W$ and $(a, b) \neq (0, 0)$, then $a \neq 0$ and $b \neq 0$. Then $U \cap W = V \cap W = \{(0, 0)\}$ and the mask space L of the multidimensional approximation Λ can be written as a direct sum

$$L = U \oplus V \oplus W. \qquad (16)$$

Mask pairs of the type comprising W do not have independent input and output masks. We will show later in Subsection IV-D that they are actually connected by a one-to-one correspondece.

Let us denote by Λ_1, Λ_2 and Λ_3 the multidimensional linear approximations determined by the mask sets U, V and W, respectively. Then the values of Λ_1 are u-bit vectors, the values of Λ_2 are v-bit vectors, and the values of Λ_3 are $(t-u-v)$-bit vectors, and $\Lambda = (\Lambda_1, \Lambda_2, \Lambda_3)$. Since all linear approximations in U and V are balanced, the value distributions of Λ_1 and Λ_2 are uniform. Considering this property for Λ_1 we get 2^u constraints for the value distribution of Λ as follows

$$\sum_{\zeta,\nu} \Pr(\Lambda(x) = (\xi, \zeta, \nu)) = \Pr(\Lambda_1(x) = \xi) = 2^{-u},$$

for all u-bit vectors ξ. From these constraints $2^u - 1$ are independent, since

$$\sum_{\xi} \Pr(\Lambda_1(x) = \xi) = 1.$$

Similarly, by the uniformity of the value distribution of Λ_2, we get the following constraints from which $2^v - 1$ are independent because

$$\sum_{\xi,\nu} \Pr(\Lambda(x) = (\xi, \zeta, \nu)) = \Pr(\Lambda_2(x) = \zeta) = 2^{-v},$$

for all v-bit vectors ζ.

We conclude that the number of degrees of freedom of the probability distribution of the values of a multidimensional linear approximation Λ of a permutation, as considered above, is bounded from above by

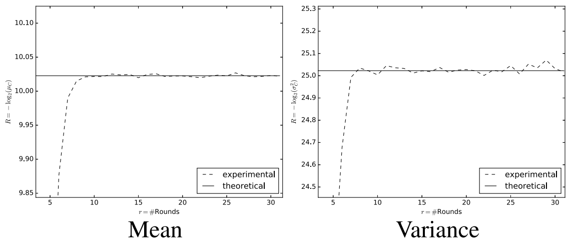
$$2^t - 1 - (2^u - 1) - (2^v - 1) = 2^t - 2^u - 2^v + 1.$$

Let us now consider Λ and the probabilities p_η of its t-bit values $\eta = (\xi, \zeta, \nu)$ as stochastic variables over the space of all equiprobable permutations. We apply Pearson's χ^2 test and compute the test variable as

$$T(\Lambda) = 2^n \sum_{\eta} \frac{\left(p_\eta - 2^{-t}\right)^2}{2^{-t}} = 2^n \mathrm{Cap}(\Lambda). \qquad (17)$$

Then $T(\Lambda)$ follows the χ^2 distribution. By Corollary 3, for linear approximations of randomly and equiprobably drawn permutations, the expected value of correlations $\mathrm{cor}(a, b)$, with $(a, b) \neq 0$, is equal to zero, also of those correlations where $a \neq 0$ and $b \neq 0$. On the other hand, $\mathrm{cor}(a, b) = 1$ for $a = b = 0$. Hence by (3), the expected value of each p_η is equal to 2^{-t}. Thus we have proved the following result.

Theorem 4: Let the multidimensional linear approximation have dimension t and the linear subspaces of trivial masks of the form $(a, 0)$ and $(0, b)$ have dimensions u and v, respectively. Then for a permutation chosen randomly and equiprobably from the set of all permutations from $\mathbb{F}_2^n$ to $\mathbb{F}_2^n$ the capacity of this multidimensional linear approximation follows, when multiplied by the factor 2^n, the central χ^2 distribution with at most $2^t - 2^u - 2^v + 1$ degrees of freedom.



Mean Variance

Fig. 1. Mean and variance of capacity of multidimensional linear approximation of dimension 7. Output masks spanned by bit: 9.

Motivated by this result, we conjecture that the value distribution of a multidimensional linear approximation of a randomly and equiprobably chosen permutation with mask subspaces U and V of dimensions u and v, respectively, has the maximum degree of freedom, that is, $2^t - 2^u - 2^v + 1$.

Conjecture 1: For a permutation from $\mathbb{F}_2^n$ to $\mathbb{F}_2^n$ drawn uniformly at random, the capacity of a multidimensional linear approximation with dimension t and the linear subspaces of trivial masks with dimensions u and v follows, when multiplied by 2^n, the χ^2 distribution with $2^t - 2^u - 2^v + 1$ degrees of freedom.

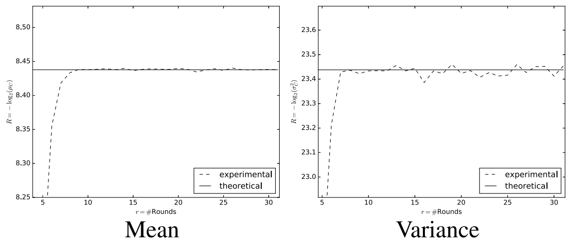
C. Experiments

We performed experiments to check the validity of Conjecture 1 in different dimensions. In our simulations of a random permutation, we used the iterated block cipher SMALLPRESENT-[4] with a varying number of rounds. This cipher has 31 rounds in total and the block size is 16 bits [20]. The state bits at input and output to each round are numbered from 0 to 15 from right to left.

For each fixed number of rounds of SMALLPRESENT-[4] varying from 0 to 31, the distribution of the capacity of the multidimensional linear approximation is computed over 2^{14} keys. Then the mean and the variance of the capacity is computed. The multidimensional linear approximation is of the form $U \oplus V$ where both U and V have nonzero bits in positions 5, 6, 9, 10, 11, 13, 14, 15. Six typical examples are depicted in Figures 1 – 6. In all six examples U is spanned by bits in positions 9, 10, 11, 13, 14, 15, and has dimension equal to 6, while the dimension of V varies from 1 to 6.

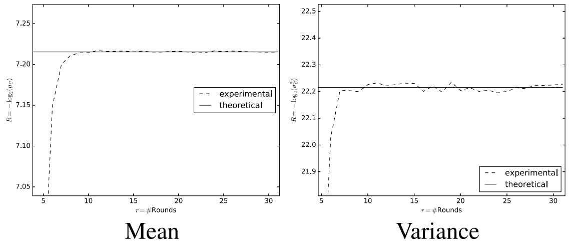
In each figure, the negatives of the base 2 exponents, that is $-\log_2$, of the mean and variance of the capacity are plotted as the number of rounds increases, and compared with the hypothetical value given by Conjecture 1 which is depicted using a horizontal line. We see that the results of the experiments support Conjecture 1 perfectly.

We also computed a number of experimental probability distributions of capacities for a random permutation instantiated by 20 rounds of SMALLPRESENT-[4]. One typical example of such probability distribution is plotted in Figure 7 for a multidimensional linear approximation with mask space



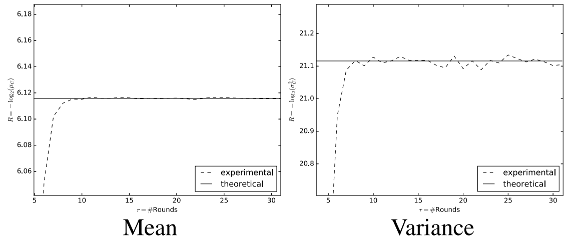
Mean Variance

Fig. 2. Mean and variance of capacity of multidimensional linear approximation of dimension 8. Output masks spanned by bits: $9, 10$.



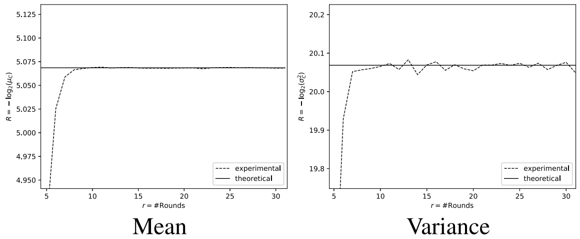
Mean Variance

Fig. 3. Mean and variance of capacity of multidimensional linear approximation of dimension 9. Output masks spanned by bits: $9, 10, 11$.



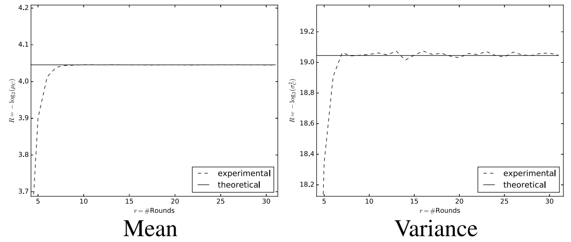
Mean Variance

Fig. 4. Mean and variance of capacity of multidimensional linear approximation of dimension 10. Output masks spanned by bits: $9, 10, 11, 13$.



Mean Variance

Fig. 5. Mean and variance of capacity of multidimensional linear approximation of dimension 11. Output masks spanned by bits: $9, 10, 11, 13, 14$.



Mean Variance

Fig. 6. Mean and variance of capacity of multidimensional linear approximation of dimension 12. Output masks spanned by bits: $9, 10, 11, 13, 14, 15$.

$U \oplus V$ where U and V are of dimension 4 and spanned by bits in positions 5, 6, 9, and 10. The resulting plot is compared with the χ^2 distribution with $2^8 - 2^4 - 2^4 + 1 = 225$ degrees of freedom plotted as a solid curve given by Conjecture 1.

D. Special Case $u = v = 0$

Let us start by defining a special type of multidimensional linear approximation, which we call a Davies-Meyer approximation for reasons to be explained in this subsection.

Definition 1: A multidimensional linear approximation Λ is called a Davies-Meyer approximation if given any linearly independent set of mask pairs (a_i, b_i), $i = 1, \ldots, t$, in the mask space L of Λ, the input masks a_i, $i = 1, \ldots, t$, are linearly independent and the output masks b_i, $i = 1, \ldots, t$, are linearly independent.

An equivalent formulation of this definition can be given as follows.

Theorem 5: A multidimensional linear approximation of a permutation is a Davies-Meyer approximation if and only if it does not contain any nonzero trivial approximations.

Proof: By definition, a Davies-Meyer approximation does not contain any nonzero approximation that has either input or output mask equal to zero, since a zero element cannot be included in a set of linearly independent elements. It remains to show that if the mask space L of a multidimensional linear approximation Λ does not contain any trivial approximations, then Λ must be a Davies-Meyer approximation.

Let us suppose the contrary, that is, L does not contain trivial approximations, but is not a Davies-Meyer approximation. Then L has a basis (a_i, b_i), $i = 1, \ldots, t$, where either a_i, $i = 1, \ldots, t$, are linearly dependent or b_i, $i = 1, \ldots, t$, are linearly dependent. Without loss of generality, we assume that the masks b_i, $i = 1, \ldots, t$, are linearly dependent. Then there is a non-empty subset of masks b_{i_j}, $j = 1, \ldots, k$, such that

$$\sum_{j=1}^{k} b_{i_j} = 0.$$

Since (a_i, b_i), $i = 1, \ldots, t$, is the basis of L, the linear approximations (a_i, b_i), $i = 1, \ldots, t$, are linearly independent, and therefore it must be the case that

$$\sum_{j=1}^{k} (a_{i_j}, b_{i_j}) = (\sum_{j=1}^{k} a_{i_j}, \sum_{j=1}^{k} b_{i_j}) = (\sum_{j=1}^{k} a_{i_j}, 0) \neq (0, 0).$$

This can happen only if

$$\sum_{j=1}^{k} a_{i_j} \neq 0.$$

Then L contains a nonzero mask pair of the form $(a, 0)$, which contradicts the assumption. ■

By this theorem, the multidimensional approximation Λ_3 determined by the mask set W in the presentation (16) of L as $L = U \oplus V \oplus W$ is a Davies-Meyer approximation. Moreover, the theorem can be restated by saying that Λ is a Davies-Meyer approximation if and only if $U = V = \{(0, 0)\}$.

Given a Davies-Meyer approximation, we can define a linear bijection D from the linear space $\mathrm{span}(a_1, \ldots, a_t)$, spanned

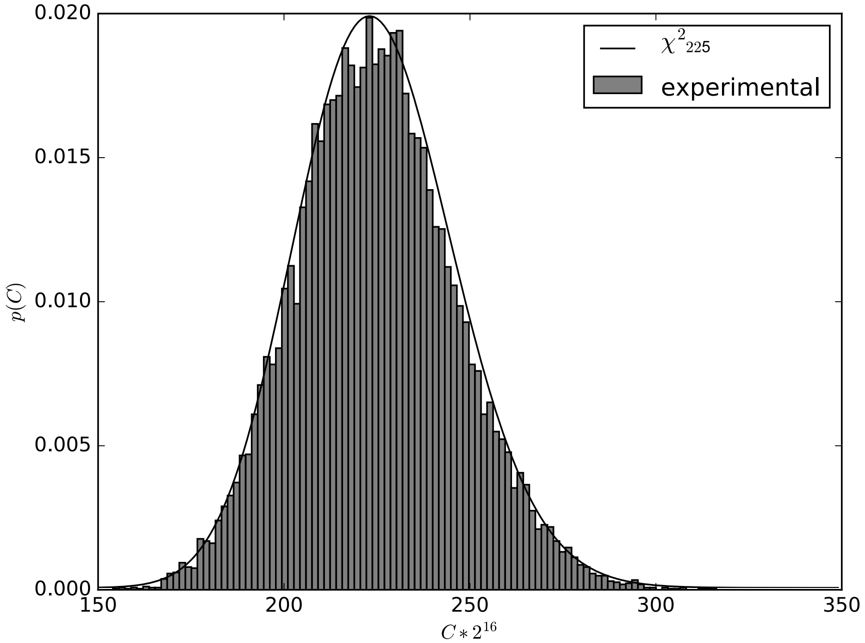


Fig. 7. Experimental probability distribution of capacity C multiplied by 2^{16} of a multidimensional linear approximation of dimension 8.

by the input masks, to the linear space $\mathrm{span}(b_1,\ldots,b_t)$, spanned by the output masks of mask pairs in L, by setting

$$D(a_i)=b_i,, i=1,\ldots,t.$$

By definition, $t\le n$. Then we extend D to a bijective linear mapping from $\mathbb{F}_2^n$ to $\mathbb{F}_2^n$ and denote it by $\bar{D}$. Then a linear approximation $(a,b)\in L$ can be expressed as

$$\begin{aligned} a\cdot x+b\cdot F(x) &= a\cdot x+a\cdot(\bar{D}^\top\circ F)(x)\\ &= a\cdot(x+(\bar{D}^\top\circ F)(x)), \qquad (18)\end{aligned}$$

where $\bar{D}^\top$ is the transpose of $\bar{D}$. If F is chosen randomly and equiprobably from the set of all permutations, then the same holds for the permutation $P=\bar{D}^\top\circ F$. We observe that the function of the form

$$x\mapsto x+P(x)$$

is the Davies-Meyer construction [21], which is known to give a pseudorandom function (more accurately, a family of pseudorandom functions) from $\mathbb{F}_2^n$ to $\mathbb{F}_2^n$ if P is a truly random permutation, that is, chosen randomly and equiprobably among all permutations $\mathbb{F}_2^n$ to $\mathbb{F}_2^n$ [22]. By (18) the linear approximations in L form a linear subspace of the components of a Davies-Meyer function, and hence the Davies-Meyer approximation Λ of F is a pseudorandom function from $\mathbb{F}_2^n$ to $\mathbb{F}_2^t$ if F is a truly random permutation.

In Subsection VII-A we define a test for distinguishing a permutation (cipher) from a truly random permutation. In the theory of cryptography, analogous tests are also used to distinguish a function from a truly random function. Specifically, a pseudorandom function is defined by the property that there is no efficient test that can be used to distinguish between a pseudorandom function and a truly random function with a larger than a negligble distinguishing advantage [23].

This means that any probability distribution computed from the values of a Davies-Meyer approximation Λ over a truly random permutation F can be replaced by the corresponding distribution computed for a truly random function. Recalling that the multinomial distribution of the capacity of a truly random function from $\mathbb{F}_2^n$ to $\mathbb{F}_2^t$ can be approximated by the χ_t^2 distribution, see Section II-C, we can state the following result.

Theorem 6: Conjecture 1 holds for the Davies-Meyer approximation.

This property will be used later in the statistical analysis of a Davies-Meyer approximation, see Theorem 11.

To illustrate a probability distribution of a Davies-Meyer approximation we depict the distribution of capacity over 2^{14} random keys in Figure 8. The capacity is computed for the 6-dimensional linear approximation over 20 rounds of SMALLPRESENT-[4] spanned by mask pairs (e_9,e_9), (e_{10},e_{10}), (e_{11},e_{11}), (e_{13},e_{13}), (e_{14},e_{14}), and (e_{15},e_{15}), where we denoted by e_j the bit vector with a single 1-bit in position j. The solid curve depicts the χ_{63}^2 distribution.

E. Multidimensional Linear Approximation of Serpent

The block cipher Serpent [10] was one of the first ciphers analysed using the multidimensional linear cryptanalysis. The multidimensional approximation Λ for Serpent was built by taking the linear space spanned by a linearly independent set of m strong base approximations of the form $(a_1,b),\ldots,(a_m,b)$ all with the same output mask b. Then L is of the form $U\oplus V$, where $u=m$ and $v=1$. Moreover, the Davies-Meyer part W was non-existent. It means that all the linear combinations of the base approximations involving an even number of base approximations had output mask equal to zero, and hence, correlation zero. In the cryptanalysis, all $2^{m+1}-1$

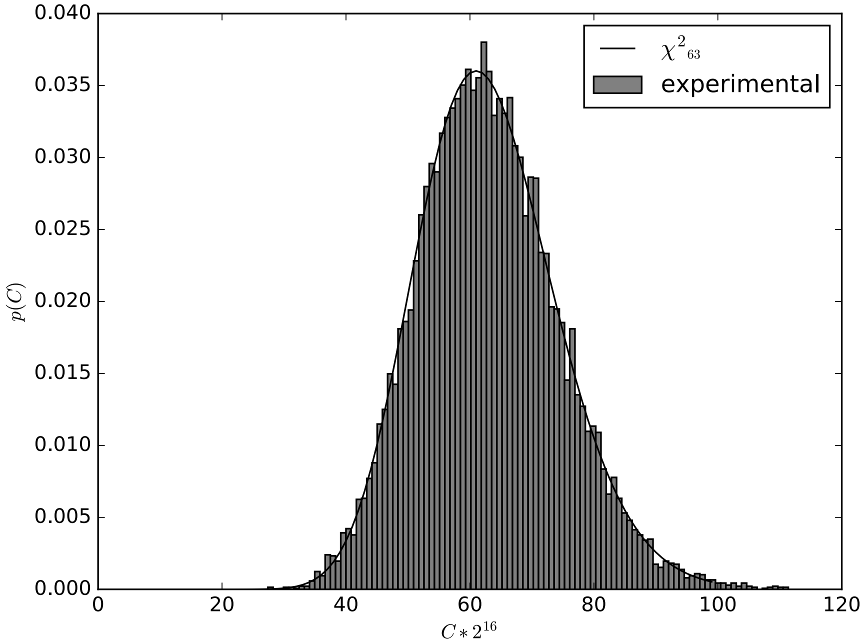


Fig. 8. Probability distribution of $2^{16}C$ for capacity C of a 6-dimensional linear approximation with no mask pairs of the form $(a, 0)$ or $(0, b)$, $a \neq 0$, $b \neq 0$.

non-zero approximations were involved including those $2^m - 1$ of the form $(a, 0)$ with correlation zero. It was mentioned that such approximations can be ignored in the computation of the empirical correlation. Nevertheless, they cannot be ignored when the degree of freedom of the sampled χ^2 statistic is determined as will be explained in Subsection VI-C.

Recently it was proposed by Nyberg to remove the subspace of trivial linear approximations and consider only the remaining set that forms an affine subspace [13]. Let us apply this idea to the multidimensional approximation of Serpent discussed above. Take the $m-1$ dimensional subspace spanned by masks $(a_2 \oplus a_1, 0), \ldots, (a_m \oplus a_1, 0)$ and denote it by H. Then the affine subspace $(a_1, b) + H$ is only a half of the size of the original linear space and still contains all m strong base approximations. Moreover, for each key, the capacity of the affine set of approximations is exactly the same as the capacity of the original set, while the degrees of freedom of the χ^2 statistic is reduced by one half.

To conclude this section let us mention that the structure of multidimensional linear approximations must be taken in consideration also for non-bijective functions. Then only the mask pairs of the form $(a, 0)$ are trivial with mean and variance of the correlation equal to zero. For example, if in the above example the block cipher Serpent is replaced by some non-bijective function but the same set of linear approximations are used, then removing the trivial approximations leads to the same affine set of approximations.

Next we study the distribution of the capacity for an affine set of linear approximations of a random permutation. Further in Subsection VI-D, we will recall the sampled χ^2 statistic from [13] with the following essential improvements: randomisation over the key and sampling without replacement. The compound probability distribution is then given by the integration of the probability distribution of the capacity into the probability distribution of the sampled χ^2 statistic.

V. Capacity of an Affine Set of Approximations

A. Constructing Affine Set of Approximations

The approach for constructing an affine set of linear approximations which does not contain trivial approximations but has a statistical model without artificial independence assumptions, was proposed by Nyberg [13]. Such a set can be constructed, for example, by taking an affine subspace of input masks and an affine subspace of output masks to get a set of the form

$$A = (a_0 + U') \times (b_0 + V') = (a_0, b_0) + (U' \times V'),$$

where the dimensions of U' and V' are positive, $a_0 \notin U'$ and $b_0 \notin V'$. We denote

$$U = \{(a, 0) \mid a \in U'\} \text{ and } V = \{(0, b) \mid b \in V'\}. \qquad (19)$$

Then the smallest linear space that contains A is

$$U \oplus V \oplus \{(0, 0), (a_0, b_0)\} = (U \oplus V) \cup A,$$

that is, the space W in the expression (16) has dimension one. But using the multidimensional linear approximation defined by this set of masks instead of using only the set A would add all trivial linear approximations from U and V to this set and reduce the strength of the attack. To avoid wasting attack resources, such as memory and time, we want to exclude the linear approximations with masks in $U \oplus V$.

More generally, let us consider such a statistic $T(A)$ for any affine set of the form $A = (a_0, b_0) + H$ where H is a linear subspace of masks and $(a_0, b_0) \notin H$. Moreover, we assume that A does not contain trivial masks. Let Λ be the multidimensional linear approximation defined by the linear space of masks $L = \{(0,0),(a_0,b_0)\} \oplus H$. Let Λ' the multidimensional linear approximation defined by H and $\Lambda' = U \oplus V \oplus W$ be its presentation in the form (16). We define the affine test statistic as follows

$$T(A) = 2^n \sum_{(a,b)\in A} \mathrm{cor}(a,b)^2 = T(\Lambda) - T(\Lambda'). \qquad (20)$$

We denote the dimension of Λ by t. Hence we can express Λ as $\Lambda = (f_0, \Lambda')$, where f_0 is the Boolean function $f_0(x) = a_0 \cdot x + b_0 \cdot E(x)$. Then the values of Λ are given as (ν, η), where ν is a bit and η is a $(t-1)$-bit vector.

Since the correlations of the linear approximations are not independent, we cannot examine the distribution of $T(A)$ directly from its expression as a sum of squared correlations. We can, however, do this if instead we express $T(A)$ in terms of value distribution $p_{(\nu,\eta)}$ of Λ as given by the following lemma.

Lemma 4: In the setting defined above, we have

$$T(A) \;=\; 2^n 2^{t-1} \sum_{\eta\in\mathbb{F}_2^{t-1}} (p_{(1,\eta)} - p_{(0,\eta)})^2. \qquad (21)$$

Proof: By applying (17) to $T(\Lambda')$ we obtain

$$\begin{aligned}
&2^n 2^{t-1} \sum_{\eta} (p_{(1,\eta)} - p_{(0,\eta)})^2 + T(\Lambda') \\
&= 2^n 2^{t-1} \sum_{\eta} \left(p_{(1,\eta)} - 2^{-t}) - (p_{(0,\eta)} - 2^{-t}\right)^2 \\
&+ 2^n 2^{t-1} \sum_{\eta} \left(p_{(1,\eta)} + p_{(0,\eta)} - 2^{-(t-1)}\right)^2 \\
&= 2^n 2^{t-1} \sum_{\eta} \left(p_{(1,\eta)} - 2^{-t}) - (p_{(0,\eta)} - 2^{-t}\right)^2 \\
&+ 2^n 2^{t-1} \sum_{\eta} \left(p_{(1,\eta)} - 2^{-t}) + (p_{(0,\eta)} - 2^{-t}\right)^2 \\
&= 2^n 2^{t} \sum_{\eta\in\mathbb{F}_2^{t-1}, \delta\in\mathbb{F}_2} \left(p_{(\delta,\eta)} - 2^{-t}\right)^2.
\end{aligned}$$

By replacing the summation index $(\delta,\eta) \in \mathbb{F}_2 \times \mathbb{F}_2^{t-1}$ by $\eta \in \mathbb{F}_2^t$ we get the expression of $T(\Lambda)$ given by (17). Then the claim follows from (20). ■

B. Distribution of the Statistic $T(A)$ for a Random Permutation

To compute $T(A)$ according to (21) for a permutation, all n-bit inputs x are distributed to 2^{t-1} categories according to the value η of $\Lambda'(x)$. Further, within each category the inputs x are divided into two subsets according to their value $f_0(x)$. The resulting value in category η is the difference of the sizes of its two subsets.

Since the expected probability distribution of the values (ν,η) of Λ over all permutations is uniform, the expected value of the differences $p_{(1,\eta)} - p_{(0,\eta)}$ is zero. Hence we propose to use Pearson's χ^2 test for the values obtained in this way in 2^{t-1} categories. The related χ^2 test statistic is $T(A)$.

To determine the number of degrees of freedom of $T(A)$, we observe that, taken together, the 2^{t-1} variables $p_{(1,\eta)} + p_{(0,\eta)}$ and the 2^{t-1} variables $p_{(1,\eta)} - p_{(0,\eta)}$, where η is a $t-1$-bit vector, uniquely determine the value distribution of Λ with probabilities $p_{\nu,\eta}$, where ν is a bit and η is a $t-1$-bit vector, which by Conjecture 1 has $2^t - 2^u - 2^v + 1$ free variables. Since the masks in $U \oplus V$ (if any) belong also to the multidimensional linear approximation Λ', the value distribution of Λ' has $2^{t-1} - 2^u - 2^v + 1$ free variables, also by Conjecture 1. Since $T(A) + T(\Lambda') = T(\Lambda)$, it follows that $T(A)$ must have at least 2^{t-1} degrees of freedom. On the other hand, by its expression (21) $T(A)$ has at most 2^{t-1} degrees of freedom, and hence exactly 2^{t-1} degrees of freedom.

We conclude that under Theorem 4 and Conjecture 1 for random permutations, $T(A)$ is χ^2 distributed with 2^{t-1} degrees of freedom and summarise the result as follows.

Theorem 7: Let $A = (a_0,b_0) + H$ be an affine subspace of linear approximations of a random permutation such that it does not contain any trivial linear approximations and assume that the multidimensional linear approximations defined by the linear spaces H and $L = \{(0,0),(a_0,b_0)\} \oplus H$ satisfy Conjecture 1. Then the statistic

$$T(A) = 2^n \mathrm{Cap}(A) = 2^n \sum_{(a,b)\in A} \mathrm{cor}(a,b)^2$$

follows χ^2 distribution with $|A|$ degrees of freedom.

VI. Data Sampling for Approximations of Random Permutations

A linear attack can be seen as composed of two parts, first, finding an approximation with good correlation and secondly, detecting this correlation in a collection of input-output pairs. When viewed this way, linear cryptanalysis is mainly a parameter estimation problem and the influence of data sampling is only on the second part. The distribution of the correlation over the keys is determined by the structure of the block cipher. Undersampling introduces an error to this parameter estimation problem. The empirical correlation is therefore a random variable in the key and the choice of the sample of plaintexts.

In Sections IV and V we presented the probability distributions of correlations and capacities computed over the full input domain of random functions and permutations. The goal of this section is to integrate a random variate data sample of fixed size into these probability distributions.

A. Sampling With or Without Replacement for a Random Permutation

In many studies on linear cryptanalysis, known plaintext-ciphertext pairs are assumed to be drawn randomly and independently, which implies sampling with replacement. It has been argued that sampling without replacement implies chosen plaintext and contradicts the essence of linear cryptanalysis of being a "known plaintext attack".

On the other hand, it has been acknowledged that duplicated plaintext-ciphertext pairs do not give new information, for which reason experimental cryptanalysis of practical ciphers typically use non-repeating plaintexts. For example, in the first experimental cryptanalysis on the DES cipher, Matsui generated the plaintexts as distinct powers of a primitive element in a 64-bit field [2].

Considering practical applications, the raw data obtained from the cipher is rarely non-repeating and may have too many duplicates for being random looking. Therefore, it requires preparations before it can be used for statistical analysis. Given two models, one requiring data input that looks like a randomly generated sample with replacement and another one without duplicates, the latter is arguably more practical to achieve. It takes $\mathcal{O}(N)$ memory and time to clear a raw sample of N plaintexts from duplicates and also gives a unique value for the size of the clean sample.

Today, linear cryptanalysis is most commonly used for estimating how many rounds of an iterative block cipher it takes until any reasonable linear attack requires the full codebook of known plaintext-ciphertext pairs. Achieving the whole codebook using sampling with replacement introduces unnecessary uncertainty to the model that can be avoided if sampling is without replacement. Based on the reasons given above, we will consider only sampling without replacement in this paper. In particular, for the experiments given in Section VIII that deal with sample sizes equal or close to full codebook, analysis with distinct plaintexts gives more accurate results.

Whether sampling is with or without replacement has also implications to the statistical models of wrong-key behaviour. The classical wrong-key assumptions commonly use the idea that if the key is wrong then the values of a bitwise linear approximation follow the binomial distribution with probability 1/2. This leads to a normal distribution $\mathcal{N}(0, 1/N)$ of the empirical correlation, where N is the size of the sample drawn with replacement. As long as the cipher has a linear approximation such that its correlation has only a small number of values as the key varies, or the average correlation over the keys is different from zero, then this wrong-key model is reasonable. But the modern block ciphers have been designed not to have such linear approximations. In particular, the correlations typically vary a lot with the key and have average value equal to zero leading to the same distribution $\mathcal{N}(0, 1/N)$ of the empirical correlation in the right-case and the wrong-key case. It follows that the early models of linear cryptanalysis, e.g. [24], [25], hardly apply to modern block ciphers.

The more advanced models of linear correlations of block ciphers consider randomisation over the key and the data sample. Bogdanov and Tischauser were the first to present a wrong-key model of the empirical correlation and gave the distribution $\mathcal{N}(0, 1/N + 2^{-n})$, where n is the block size and N is the size of the sample assuming sampling with replacement [4]. Later Blondeau and Nyberg showed that if sampling is without replacement, then the wrong-key distribution is $\mathcal{N}(0, 1/N)$. While this distribution is the same as in the classical case without key randomisation, the setting is different and the corresponding right-key model allows building a distinguisher [7].

In this section, we present probability distributions of the capacity considered over a random permutation and random sampling without replacement for a single linear approximation, a multidimensional linear approximation including Davies-Meyer approximation as a special case, and an affine set of approximations. In all cases, we first derive the distribution for an arbitrary fixed permutation by randomisation over the data sample only. Then by using the results from Sections IV and V we present the compound probability distributions of the capacity over a random permutation and a random sample.

B. Sampling Without Replacement for a Single Linear Approximation

Given a mask pair (a, b), where $b \neq 0$, and a data sample S of input-output pairs $(x, F(x))$ of size N drawn for a random function $F : \mathbb{F}_2^n \rightarrow \mathbb{F}_2^s$, let us denote by $\widehat{w}(a, b)$ the number of inputs x, for which $(x, F(x) \in S$ and the linear approximation $a \cdot x + b \cdot F(x)$ takes the value zero. Let $w(a, b)$ be the number of zeroes of $a \cdot x + b \cdot F(x)$ over all inputs $x \in \mathbb{F}_2^n$. Then

$$\widehat{w}(a, b) \sim \mathcal{HG}(2^n, w(a, b), N).$$

For a truly random F, we know by Theorem 2 that

$$w = w(a, b) \sim \mathcal{B}(2^n, 1/2).$$

Then the distribution of $\widehat{w}(a, b)$ taken over a truly random function and a random data sample S of size N has the following probability distribution

$$\begin{aligned} \Pr(\widehat{w}(a, b) = k) &= \sum_{w=0}^{2^n} \left(\frac{1}{2}\right)^{2^n} \binom{2^n}{w} \frac{\binom{w}{k}\binom{2^n-w}{N-k}}{\binom{2^n}{N}} \\ &= (\frac{1}{2})^N \binom{N}{k}. \end{aligned}$$

Hence $\widehat{w}(a, b) \sim \mathcal{B}(N, 1/2)$.

Let us denote by $\widehat{\mathrm{cor}}(a, b)$ the sampled correlation, that is,

$$\widehat{\mathrm{cor}}(a, b) = \frac{1}{N}(2\widehat{w}(a, b) - N).$$

By normal approximation (8), we obtain the following result.

Theorem 8: Let $\widehat{\mathrm{cor}}(a, b)$, where $b \neq 0$, be the sampled correlation of the linear approximation (a, b) of a function from $\mathbb{F}_2^n$ to $\mathbb{F}_2^s$. Then the probability distribution of $\widehat{\mathrm{cor}}(a, b)$ taken over a truly random function and a random data sample of size N of distinct plaintexts, where $N \leq 2^n$, can be approximated by the normal distribution $\mathcal{N}(0, 1/N)$.

To prove the corresponding result for a random permutation we use the normal approximation from the beginning.

Theorem 9: Let $\widehat{\mathrm{cor}}(a, b)$, where $b \neq 0$, be the sampled correlation of a linear approximation of a permutation from $\mathbb{F}_2^n$ to $\mathbb{F}_2^n$. Then the probability distribution of $\widehat{\mathrm{cor}}(a, b)$ considered over a truly random permutation and a random data sample of size N of distinct plaintexts, where $N \leq 2^n$, can be approximated by the normal distribution $\mathcal{N}(0, 1/N)$.

Proof: Given a linear approximation (a, b), $b \neq 0$, and a data sample S of size N drawn for a permutation $E : \mathbb{F}_2^n \rightarrow \mathbb{F}_2^n$, the sampled correlation is $\widehat{\mathrm{cor}}(a, b) = \frac{1}{N}(2\widehat{w}(a, b) - N)$ where $\widehat{w}(a, b) \sim \mathcal{HG}(2^n, w(a, b), N)$. Then by normal approximation (8),

$$\widehat{\mathrm{cor}}(a, b) \sim \mathcal{N}\left(\mathrm{cor}(a, b), \frac{B}{N}(1 - \mathrm{cor}(a, b)^2)\right),$$

where $\mathrm{cor}(a,b) = 2^{-n}(2w(a,b) - 2^n)$ and $B = (2^n - N)/2^n$. By Theorem 3, Corollary 3, and using the normal approximation of the hypergeometric distribution (8), we have

$$\mathrm{cor}(a,b) \sim \mathcal{N}(0, 2^{-n}).$$

Then the distribution of $\widehat{\mathrm{cor}}(a,b)$ taken over a random permutation and a random sample of size N is approximately normal with mean $\mathrm{Exp}\,(\mathrm{cor}(a,b)) = 0$ and variance equal to

$$\mathrm{Var}(\mathrm{cor}(a,b)) + \mathrm{Exp}(\mathrm{Var}(\widehat{\mathrm{cor}}(a,b))) = 2^{-n} + \frac{B}{N} - \frac{B}{N}2^{-n} \approx \frac{1}{N}.$$

■

We get another view of this result by observing that a linear approximation of a random permutation F can be expressed as a linear approximation of a pseudorandom function as follows

$$a \cdot x + b \cdot F(x) = (a+b) \cdot x + b \cdot (x + F(x)),$$

see Subsection IV-D, and then applying Theorem 8.

C. Sampling Without Replacement for a Multidimensional Linear Approximation

Let us now recall the sampled test statistic of a multidimensional linear approximation Λ. It is obtained by taking (17) and replacing 2^n by N and correlations $\mathrm{cor}\,(a,b)$ by sampled correlations $\widehat{\mathrm{cor}}\,(a,b)$ as follows

$$T_N(\Lambda) = N \sum_{(a,b)\in L,(a,b)\neq 0} \widehat{\mathrm{cor}}\,(a,b)^2\,. \tag{22}$$

Let us first derive the probability distribution of $T_N(\Lambda)$ for an arbitrary fixed key and randomly chosen sample of distinct plaintexts. The corresponding probability distribution for $T_N(\Lambda)$ is given by the following result originally stated in [7]. The proof given in [7] assumed independent hypergeometric distributions. In [1] the validity of this result was questioned due to the artificial assumption of independence. Therefore another proof will be given here by applying the standard statistical argument of *finite population correction* to the χ^2 distributed variable as given by Lemma 1. In our context, 2^n is the size of the population and N is the size of the random sample of distinct elements from that population.

Theorem 10: Let Λ be a multidimensional linear approximation of dimension t applied to a permutation from $\mathbb{F}_2^n$ to $\mathbb{F}_2^n$. Let $T_N(\Lambda)$ be the statistic defined by (22) computed over a random sample of size N of distinct plaintexts. Then $B^{-1}T_N(\Lambda)$ follows non-central χ^2 distribution with $2^t - 1$ degrees of freedom and non-centrality parameter $B^{-1}N\mathrm{Cap}(\Lambda)$, where B is as defined by (7).

Proof: We denote by $\widehat{p}_\eta$ the sampled probabilities of the distribution of the t-bit values $\eta \in \Lambda$ computed for a sample of size N of inputs x. We apply (15) to this distribution to write $T_N(\Lambda)$ as follows

$$T_N(\Lambda) = N2^t \sum_\eta (\widehat{p}_\eta - 2^{-t})^2 = \sum_\eta \frac{(N\widehat{p}_\eta - N2^{-t})^2}{N2^{-t}}. \tag{23}$$

Then $T_N(\Lambda)$ is Pearson's χ^2-test statistic with $2^t - 1$ degrees of freedom. Since the sample is without replacement we apply Lemma 1 and get that $B^{-1}T_N(\Lambda)$ is non-centrally χ^2 distributed and has expected value equal to $2^t - 1 + \delta$, where δ is the non-centrality parameter. Then the expected value of $T_N(\Lambda)$ is equal to $B(2^t - 1) + B\delta$. To determine δ we compute the expected value of $T_N(\Lambda)$ directly. Expanding the expression (23) we get

$$\begin{aligned} T_N(\Lambda) &= \sum_\eta \frac{(N\widehat{p}_\eta - Np_\eta)^2}{N2^{-t}} &(24)\\ &+ N2^t \sum_\eta (p_\eta - 2^{-t})^2 &(25)\\ &+ N2^{t+1} \sum_\eta p_\eta(\widehat{p}_\eta - p_\eta), &(26) \end{aligned}$$

where p_η is the probability of the t-bit value η in the image space of Λ. Note that in the expansion (24-26) the term $N2^{t+1}\sum_\eta(\widehat{p}_\eta - p_\eta)$ was omitted because it is equal to zero. Now expression (24) is Pearson's χ^2-test statistic with $2^t - 1$ degrees of freedom by using the standard approximation $Np_\eta \approx N2^{-t}$ in the denominator. Moreover it is central, since for each η the expected value of $\widehat{p}_\eta$ is equal to p_η. Since the sampling is without replacement, we get that the expected value of (24) is equal to $B(2^t - 1)$. The expression (25) is constant and equal to $N\mathrm{Cap}(\Lambda)$, and the expected value of (26) is equal to zero. Solving δ from the equation

$$B(2^t - 1) + B\delta = B(2^t - 1) + N\mathrm{Cap}(\Lambda)$$

gives the non-centrality parameter as claimed. ■

As the sample size N grows, and gets equal to 2^n, the sampled statistic $T_N(\Lambda)$ gets equal to the statistic $T(\Lambda)$. In general, the χ^2-variables computed for the entire input space may not have the same number of degrees of freedom as we saw in Subsection IV-B, which complicates the analysis of the compound distribution of the statistic $T_N(\Lambda)$ considered over a random permutation and a random sample of size N. In the case, where Λ does not contain any trivial approximations, that is, Λ is a Davies-Meyer approximation, the distribution of $T(\Lambda)$ also has $t - 1$ degrees of freedom. Moreover, by Theorem 6 the distribution given by Conjecture 1 holds and we get the following result. The proof is similar to the proof of Theorem 13 in the next subsection and is omitted here.

Theorem 11: Let Λ be a Davies-Meyer approximation applied to a random permutation from $\mathbb{F}_2^n$ to $\mathbb{F}_2^n$. Let $T_N(\Lambda)$ be the statistic defined by (22) computed for a sample of size N of distinct plaintexts and considered as a random variable over a truly random permutation and a random sample of size N. Then the mean of $T_N(\Lambda)$ is $|\Lambda| - 1$ and the variance is $2\,(|\Lambda| - 1)$.

D. Sampling Without Replacement for an Affine Approximation

Given an affine subspace A of linear approximations defined by two multidimensional linear approximations Λ and Λ' of dimensions t and $t-1$ respectively, we define the sampled test statistic $T_N(A)$ analogously to (20) as follows

$$T_N(A) = N \sum_{(a,b)\in A} \widehat{\mathrm{cor}}\,(a,b)^2 = T_N(\Lambda) - T_N(\Lambda'). \tag{27}$$

By repeating the derivations of Section V, but now for the sampled statistic $T_N(A)$ and using Theorem 10 we get the following result.

Theorem 12: Let A be an affine set of linear approximations applied to a permutation from $\mathbb{F}_2^n$ to $\mathbb{F}_2^n$ and assume it does not contain trivial approximations. Let $T_N(A)$ be the statistic defined by (27) computed for a random sample of size N of distinct plaintexts. Then $B^{-1}T_N(A)$ follows the noncentral χ^2 distribution with $|A|$ degrees of freedom and noncentrality parameter $B^{-1}N\mathrm{Cap}(A)$, where B is as defined by (7).

The noncentrality parameter $B^{-1}N\mathrm{Cap}(A)$ of the distribution of $T_N(A)$ depends on the permutation. If the permutation is truly random, the distribution of $T(A) = 2^n\mathrm{Cap}(A)$ is given by Theorem 7 under the assumption that Conjecture 1 holds. We get the following result.

Theorem 13: Let A be an affine set of linear approximations applied to a permutation from $\mathbb{F}_2^n$ to $\mathbb{F}_2^n$ and let us assume that A does not contain trivial approximations and Conjecture 1 holds. Let $T_N(A)$ be the statistic defined by (27) computed for a sample of size N of distinct plaintexts and considered as a random variable over a random permutation and a random sample of size N. Then the mean of $T_N(A)$ is $|A|$ and the variance is $2|A|$.

Proof: Let us denote $|A|$ by ℓ. Using the non-central χ^2 distribution of $B^{-1}T_N(A)$ for a fixed permutation with capacity $\mathrm{Cap}(A)$ given by Theorem 12, we get that the mean of $T_N(A)$ is equal to

$$B\ell + N\mathrm{Cap}(A) = B\ell + N2^{-n}T(A). \tag{28}$$

By taking the mean over random permutations, we get the mean ℓ as claimed.

Similarly, by Theorem 12, we get that the variance of $T_N(A)$ is equal to

$$B^2\left(2\ell + 4B^{-1}N\mathrm{Cap}(A)\right). \tag{29}$$

Then the total variance over random permutation is computed as the sum of the mean of (29) and the variance of (28) to get

$$\begin{aligned} &B^2\left(2\ell + 4B^{-1}N2^{-n}\ell\right) + \left(N2^{-n}\right)^2 \cdot 2\ell \\ &= 2B^2\ell + 4B(1-B)\ell + 2(1-B)^2\ell = 2\ell. \end{aligned}$$

■

Based on these considerations one can argue that, when considered as a random variable over a random permutation and a random sample of N distinct plaintexts, the test statistic $T_N(A)$ follows the $\chi^2_{|A|}$ distribution.

We have seen that constructions of multidimensional and affine linear approximations that do not contain any trivial approximations have a simple and clear theory for random permutations. Also for those approximations that contain trivial approximations it is quite straightforward to derive the mean and the variance of the sampled statistic. For permutations originating from ciphers the theory is not that clear. The least one can say is that linear approximations of block ciphers have the same trivial linear approximations as a random permutation. The problem of trivial approximations was observed also in [10] where it was recommended to exclude them in the computation of the empirical correlation. While this helps in speeding up the cryptanalysis, the problem of accuracy still remains. In the case of [10] the trivial linear approximations could have been easily excluded by considering the related affine set as discussed in Subsection IV-E.

VII. Evaluating Nonrandomness of Linear Approximations

In this section, we apply the statistical models of linear approximations of a random permutation and present a test to evaluate non-randomness of linear approximations of a block cipher based on the observed capacities in large experiments.

A. Randomness Test

We consider the set of permutations from $\mathbb{F}_2^n$ to $\mathbb{F}_2^n$ and a permutation drawn from this set. The null hypothesis of the test is defined as follows.

Hypothesis 1 (Null Hypothesis): The permutation is a truly random permutation, that is, it has been drawn uniformly at random from the set of all permutations.

The alternative hypothesis is then defined as follows.

Hypothesis 2 (Alternative Hypothesis): The permutation is not a truly random permutation.

The test is performed by computing the test statistic T for the given permutation. For example, $T = T_N(A)$ defined by (27). Given a threshold τ, the null hypothesis is accepted if $T \leq \tau$ and the alternative hypothesis is accepted if $T > \tau$.

To determine the threshold, we first set the significance level α and then use the probability distribution of T over a randomly and uniformly selected permutation to determine the threshold τ_α in such a way that

$$\Pr(\text{Alternative Hypothesis is accepted}|\text{Permutation is truly random}) = \alpha.$$

The following probability

$$\Pr(\text{Alternative Hypothesis is accepted}|\text{Permutation is not truly random})$$

is called the success probability. It depends on τ_α and we denote it by $P_S(\alpha)$.

The performance of the test for a given significance level α can then be quantified using the distinguishing advantage defined as the absolute value of the difference

$$P_S(\alpha) - \alpha.$$

The distinguishing advantage is a value between 0 and 1. While for practical attacks this value is usually closer to 1, non-negligible deviation from 0 can be considered to reflect nonrandomness of the permutation.

When used for key recovery it is assumed that, if the test statistic is computed from data obtained from the cipher using a wrong key, then the Null Hypothesis holds. Based on this assumption, the significance level α can be interpreted as the fraction of wrong keys falsely accepted as potential right keys. If the total number of key candidates to be tested is 2^k then the number of key candidates accepted by the test as correct is $\alpha 2^k = 2^{k-d}$, where $d = -\log(\alpha)$. This is akin to saying that the number of correct key bits recovered using the distinguisher is d. The levels $\alpha = 0.25,\ 0.125,\ 0.0625$ used in this paper correspond to $d = 2,\ 3,\ 4$, respectively.

B. How to Determine the Success Probability

At each significance level, the success probability depends on the probability distribution of the test statistic considered over the cipher keys and over the data obtained from the cipher. Previously, there have been attempts to estimate the distribution of the capacity of a multidimensional linear approximation of a block cipher [1], [12]. The former paper determines the mean and the variance of the probability distribution which is then assumed to be normal, while the latter paper computes an approximation of the probability density function in simulations.

In this paper, we take a different approach. We use rough estimates of the correlations only to support the search for strong linear approximations, but do not use these estimates to model the cipher. For certain selected values of α, we determine the success probabilities $P_S(\alpha)$ and distinguishing advantages experimentally and see how many rounds the distinguishers can cover before the distinguishing advantage becomes negligible, that is, $\alpha \approx P_S(\alpha)$.

VIII. Nonrandomness of Affine Distinguishers of Simon

A. Description of Simon

Simon is a family of lightweight block ciphers designed by the US National Security Agency (NSA) and published in 2013 [26]. The Simon$2n/mn$ family of lightweight block ciphers has 10 members differing in their block and key sizes. All members of the family have a Feistel structure with round function R employing a non-linear function f. In each round i, R receives two n-bit input words X_i and Y_i, and outputs two n-bit words X_{i+1} and Y_{i+1}. The round function uses three operations in $\mathbb{F}_2^n$: bitwise addition (exclusive-or; XOR), bitwise multiplication (AND), and a left circular shift by j positions, which we denote by '$\oplus$', '&', and '$\lll j$', respectively. Note that the meaning of '$\oplus$' within this section differs from the one used elsewhere in this paper, see Section IV. The internal non-linear function f is defined as:

$$f\left(X_i\right) = \left[\left(X_i \lll 1\right) \& \left(X_i \lll 8\right)\right] \oplus \left(X_i \lll 2\right).$$

The output of the round function R on an input block $X_i || Y_i$ is:

$$R_i\left(X_i, Y_i\right) = \left(Y_i \oplus f\left(X_i\right) \oplus k_i, X_i\right),$$

where i is the round number and k_i is the round key. The entire cipher is a composition of round functions $R_{r-1} \circ R_{r-2} \circ \ldots \circ R_0\left(X_0, Y_0\right)$. The structure of the round function of Simon is depicted in Figure 9.

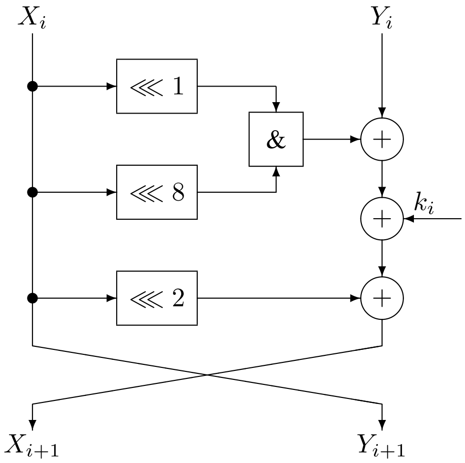


Fig. 9. One round of Simon.

B. Building Linear Approximations of Simon*32/64*

We start by determining the linear approximations over one round of Simon. Referring to Figure 9 let us denote by a the mask on the left input data half X_i and by b the mask on the right input data half Y_i. To have nonzero correlation, we then must have b as the mask on the left output data half X_{i+1}.

The bitwise AND-function maps two bits x and y to the single bit $x\&y$. All four linear combinations of the bits x and y have nonzero correlation with $x\&y$, all with the same absolute value 2^{-1}. Hence the 2-bit to 1-bit AND-function is a bent function and its four linear approximations with nonzero output mask form an affine set with capacity equal to 1.

To capture the linear approximations of the bitwise AND-function over one round, let us assume now that b is a vector with a single 1-bit. If b has a single 1-bit, only a single AND operation is activated. The four linear approximations of the AND function induce four masks on the right half Y_{i+1} which form the two-dimensional affine space

$$a \oplus b_{\ggg 2} \oplus \operatorname{span}\left\{b_{\ggg 1}, b_{\ggg 8}\right\}, \qquad (30)$$

where we denoted by $b_{\ggg j}$ the right circular shift by j positions of the bit string b. Since any Feistel cipher (without the final swap) is its own inverse, the same reasoning works also backwards. Given a mask $b||c$ on the output data $X_{i+1}||Y_{i+1}$, where b is assumed to have a single 1-bit, then the input masks on the data $X_i||Y_i$ that may give nonzero correlations are of the form $a||b$, where a is one of the four vectors of the two-dimensional affine set

$$c \oplus b_{\ggg 2} \oplus \operatorname{span}\left\{b_{\ggg 1}, b_{\ggg 8}\right\}. \qquad (31)$$

Each non-zero bit of the mask b as described above, potentially induces a two-dimensional affine subspace of masks. Later in this section we will see an example of a mask b that has two non-zero bits which together generate a four-dimensional affine space of masks each relating to a linear approximation over one round with a correlation that has absolute value 2^{-2}.

Since it becomes soon impractical to iterate this method over more rounds by keeping track of all linear approximations, we focus on the so-called core approximations which are linear approximations that use at each round only one linear approximation with input mask $a||b$ and output mask $b||(a \oplus b_{\ggg 2})$. This was also the approach used in [27] to build the core approximation trails for SIMON. Then the trail correlations are computed based on how many times the linear approximation $x\&y = 0$ was used.

For a linear approximation of an n-bit block cipher to be useful in cryptanalysis, its average squared correlation should differ from 2^{-n}, that is, the corresponding value for a random permutation. The trail correlations are commonly used to give lower bounds to the average squared correlation. Only if such a lower bound is larger than 2^{-n} it is useful for distinguishing from the random case.

The absolute values of the trail correlations for SIMON32/64 used in [27] were far less than 2^{-16}. Nevertheless, referring to the method by Biryukov *et al.* of multiple linear cryptanalysis [25], it was argued that by collecting sufficiently many of such linear approximations that have squared correlations with known lower bounds, however small they are, one can accumulate the lower bound of the sum of the average squared correlations of the multiple linear approximations to exceed 2^{-32} to get data complexity less than 2^{32}. The problem is that these early methods of linear cryptanalysis make the assumption that for each linear approximation to be used, the cryptanalyst has obtained a good estimate of the absolute value of the correlation, and this value is the same for all keys. This assumption is not satisfied by modern block ciphers such as SIMON whose linear approximations have a large number of trails and correlations that vary a lot with the key.

In this paper, we revisit the constructions of sets of linear approximations of SIMON32/64 from [27], and compute estimates of the correlations experimentally over large sets of keys. As observed, a core trail can be extended at the beginning and at the end by using all four approximations to build multiple linear approximations, which give naturally rise to affine sets of linear approximations.

Using the statistical distribution given in Theorem 13 of the capacity of an affine set of linear approximations for a random permutation and the randomness test presented in Section VII we evaluate the deviation of the behaviour of the block cipher from random behaviour by extensive experiments.

C. Full Codebook Randomness Evaluation of SIMON*32/64*

In this section we build examples of affine sets of linear approximations which cover up to 18 rounds of SIMON32/64 and experimentally evaluate their distinguishing advantages. We begin by describing the core approximation trail we will use for these examples and experimentally evaluate its nonrandomness starting from 13 rounds. The core trail is built starting with the input mask $4000_x||0001_x$ to round 1 and by following its propagation through the rounds as described in Subsection VIII-B. Note that output mask from round i is the input mask to round $i + 1$.

By counting the total number of non-zero bits in the right halves of the output masks from all rounds of the trail, we can evaluate the trail correlation. We see that after 10 rounds the absolute value of the trail correlation drops below 2^{-16}.

TABLE I

CORE TRAIL FOR SIMON32/64

round	output mask	round	output mask
0	$4000_x\|\|0001_x$	9	$0111_x\|\|0004_x$
1	$0001_x\|\|0000_x$	10	$0004_x\|\|0110_x$
2	$0000_x\|\|0001_x$	11	$0110_x\|\|0040_x$
3	$0001_x\|\|4000_x$	12	$0040_x\|\|0100_x$
4	$4000_x\|\|1001_x$	13	$0100_x\|\|0000_x$
5	$1001_x\|\|0400_x$	14	$0000_x\|\|0100_x$
6	$0400_x\|\|1101_x$	15	$0100_x\|\|0040_x$
7	$1101_x\|\|4040_x$	16	$0040_x\|\|0110_x$
8	$4040_x\|\|0111_x$	17	$0110_x\|\|0004_x$
9	$0111_x\|\|0004_x$	18	$0004_x\|\|0111_x$

TABLE II

EXPERIMENTAL SUCCESS PROBABILITIES AND AVERAGE SQUARED CORRELATIONS OF SINGLE LINEAR APPROXIMATIONS DERIVED FROM THE CORE TRAIL WITH INPUT MASK $4000_x||0001_x$. EXPERIMENTS USED 2^{13} KEYS

No.	output	av. squared	succ. probability $P_S(\alpha); \alpha =$		
rounds	mask	correlation	0.25	0.125	0.0625
13	$0100_x\|\|0000_x$	$2^{-29.873}$	0.543	0.422	0.330
14	$0000_x\|\|0100_x$	$2^{-29.873}$	0.543	0.422	0.330
15	$0100_x\|\|0040_x$	$2^{-31.131}$	0.371	0.243	0.161
16	$0040_x\|\|0110_x$	$2^{-31.739}$	0.287	0.159	0.091
17	$0110_x\|\|0004_x$	$2^{-31.975}$	0.255	0.129	0.066
18	$0004_x\|\|0111_x$	$2^{-32.024}$	0.250	0.127	0.059

Nevertheless, the true correlation will stay above this value for at least 16 rounds as we will see next in Table II, which gives the results of the experimental evaluation of the core linear approximation over 13-18 rounds.

The next step is to use an affine subspace of linear approximations. We take the output mask $0001_x||0000_x$ from round 1 of the core trail and determine all four input masks to round 1 that have non-zero correlation. According to (31) they are $4000_x||0001_x$, $C000_x||0001_x$, $4100_x||0001_x$, and $C100_x||0001_x$. This is a two-dimensional affine space

$$4000_x||0001_x \oplus \text{span}\,\{\, 8000_x||0000_x,\ 0100_x||0000_x \,\}\,.$$

With these input masks and a single output mask as given for the core trail we have an affine set of input-output mask pairs of dimension 2. In Table III we give the experimental results of these affine linear approximations over 13-18 rounds using the test described in Section VII. The expected capacity in the random case is 2^{30}. Compared to the corresponding results for the single approximations given in Table II, there is a significant improvement in success probabilities, in all significance levels, up to 17 rounds, after which point also the capacity becomes very close to the random.

In an attempt to analyse nonrandomness of 18 rounds of SIMON32/64 we use another trail constructed from the core trail. We add one more round to the core trail in the beginning and take all four masks. To apply (31) we take $b||c = 4000_x||0001_x$, to get the four input masks $a||b$, where

$$a \in 1001_x \oplus \text{span}\{0040_x,\, 2000_x\}.$$

To obtain the output masks, we start with the output mask $a||b = 0040_x||0110_x$ from round 16, and observe that b has two active bits. Then by (30) the output masks are of the

TABLE III

EXPERIMENTAL SUCCESS PROBABILITIES AND AVERAGE CAPACITIES FOR A 2-DIMENSIONAL AFFINE SUBSPACE A OF LINEAR APPROXIMATIONS WITH FOUR INPUT MASKS: $4000_x||0001_x$, $C000_x||0001_x$, $4100_x||0001_x$, $C100_x||0001_x$ AND ONE OUTPUT MASK. EXPERIMENTS USED 2^{13} KEYS

No.	output	average	succ. probability $P_S(\alpha); \alpha =$		
rounds	mask	capacity	0.25	0.125	0.0625
13	$0100_x\|\|0000_x$	$2^{-27.887}$	0.664	0.56	0.483
14	$0000_x\|\|0100_x$	$2^{-27.887}$	0.663	0.56	0.483
15	$0100_x\|\|0040_x$	$2^{-29.133}$	0.47	0.340	0.254
16	$0040_x\|\|0110_x$	$2^{-29.736}$	0.332	0.193	0.115
17	$0110_x\|\|0004_x$	$2^{-29.968}$	0.262	0.131	0.068
18	$0004_x\|\|0111_x$	$2^{-30.000}$	0.253	0.127	0.064

form $b||c$, where one value for c is $a \oplus b_{\ggg 2} = 0004_x$. All the 16 masks are obtained by adding this value to the masks in the linear space spanned by the right circular shifts by 1 and 8 of the two single-bit components 0100_x and 0010_x of b. That is,

$$c \in 0004_x \oplus \operatorname{span}\{0001_x,\, 0008_x,\, 0080_x,\, 1000_x\}.$$

The resulting affine space of linear masks is of dimension 6. Again, by running experiments with a randomly selected set of 2^{13} keys we get the average capapcity $2^{-25.99}$ which is quite close to the expected capacity of 2^{-26} in the random case. Nevertheless, when applying the randomness test, we get success probabilities $P_S(0.25) = 0.260$, $P_S(0.125) = 0.137$ and $P_S(0.0625) = 0.072$. Thus the 6-dimensional 18-round distinguisher performs better than the 1-dimensional 18-round distinguisher in Table II or the 2-dimensional 18-round distinguisher in Table III.

On the other hand, we can see that the 18-round affine set of approximations does not preserve all the distinguishing power of the 16-round core approximation, see Table II, even if it was constructed by extending this 16-round approximation by one round up and one round down by taking all masks in the input and output that have non-zero correlations. For example, $P_S(0.125) = 0.159$ for the 16-round approximation, while $P_S(0.125) = 0.137$ for the extended 18-round affine approximation.

D. Compliance With the Model of Multidimensional Linear Cryptanalysis

We used this experimental setting also for verifying Conjecture 1 and computed the capacity of the 6-dimensional linear subspace of approximations related to the 6-dimensional affine subspace of approximations of SIMON32/64 constructed in the previous subsection. The dimensions of the spaces are $t = 6$, $u = 4$ and $v = 2$. According to Conjecture 1 the expected capacity in the random case is equal to $2^{-26.508}$. When evaluating it for the cipher by computing the average capacity over 2^{13} keys we get $2^{-26.506}$, which is convincingly close to the conjectured value.

E. Randomness Testing of SIMON *Using Less Than Full Codebook*

The experiments used for this section are identical to those of Subsection VIII-C in all but the data complexity. We use exactly the same approximations as for Table III as well as the same 2^{13} keys. While Table III used the full codebook of 2^{32} plaintext-ciphertext pairs for distinguishing, we use 2^{30} pairs for Table IV and 2^{28} pairs for Table V. By Theorem 13 the expected capacity in the random case is 2^{-28} for $N = 2^{30}$, and 2^{-26} for $N = 2^{28}$.

TABLE IV

SUCCESS PROBABILITIES AND AVERAGE CAPACITIES FOR A 2-DIMENSIONAL AFFINE SUBSPACE USING A PARTIAL CODEBOOK OF 2^{30} DATA. INPUT MASKS: $4000_x||0001_x$, $C000_x||0001_x$, $4100_x||0001_x$, $C100_x||0001_x$. EXPERIMENTS USED 2^{13} KEYS

No.	output	average	succ. probability $P_S(\alpha); \alpha =$		
rounds	mask	capacity	0.25	0.125	0.0625
13	$0100_x\|\|0000_x$	$2^{-27.127}$	0.475	0.342	0.253
14	$0000_x\|\|0100_x$	$2^{-27.127}$	0.475	0.342	0.253
15	$0100_x\|\|0040_x$	$2^{-27.734}$	0.334	0.193	0.117
16	$0040_x\|\|0110_x$	$2^{-27.921}$	0.272	0.144	0.078
17	$0110_x\|\|0004_x$	$2^{-27.983}$	0.257	0.127	0.069
18	$0004_x\|\|0111_x$	$2^{-27.9993}$	0.255	0.122	0.060

TABLE V

SUCCESS PROBABILITIES AND AVERAGE CAPACITIES FOR A 2-DIMENSIONAL AFFINE SUBSPACE USING A PARTIAL CODEBOOK OF 2^{28} DATA. INPUT MASKS: $4000_x||0001_x$, $C000_x||0001_x$, $4100_x||0001_x$, $C100_x||0001_x$. EXPERIMENTS USED 2^{13} KEYS

No.	output	average	succ. probability $P_S(\alpha); \alpha =$		
rounds	mask	capacity	0.25	0.125	0.0625
13	$0100_x\|\|0000_x$	$2^{-25.716}$	0.344	0.205	0.122
14	$0000_x\|\|0100_x$	$2^{-25.716}$	0.344	0.205	0.122
15	$0100_x\|\|0040_x$	$2^{-25.923}$	0.280	0.144	0.076
16	$0040_x\|\|0110_x$	$2^{-25.993}$	0.252	0.124	0.063
17	$0110_x\|\|0004_x$	$2^{-25.986}$	0.254	0.127	0.067
18	$0004_x\|\|0111_x$	$2^{-25.991}$	0.256	0.127	0.065

The results of experiments given in Tables IV–V reveal that distinguishing from random is still possible also with less than the full codebook. Comparison with Table III shows that, as expected, the success probability is lower in rounds 13–16 and becomes closer to the random behaviour already at 17 rounds when 2^{30} pairs of data is used and even earlier, at 16 rounds, when 2^{28} is used. This is due to the smaller sample which leads to a larger *sample error* and thus to bigger variance for the test statistic.

IX. CONCLUSION

In this paper we presented a model which captures the statistical behaviour of the capacity of multidimensional linear approximations computed for a permutation and a sample of plaintext, when the permutation and the sample of distinct plaintext of fixed size are selected uniformly at random. The additivity of the variances of squared correlations is achieved without any assumptions of statistical independence based only on standard statistical tools such as Pearson's χ^2 test and the finite population correction coefficient.

We showed for the first time that the degree of freedom of the related χ^2 distribution over the distribution depends on the structure of the multidimensional linear approximation and that it can be significantly smaller than assumed in previous works due to the existence of trivial approximations. We identify two types of sets of multiple linear approximations, the Davies-Meyer approximation and the affine approximation which do not have trivial approximations. Such types of approximations offer the most efficient χ^2-based linear attacks due to the optimal number of degrees of freedom. When selecting sets of strong multiple linear approximations for actual ciphers such structures are recommended for consideration if possible. As the first example, we mentioned the first multidimensional linear cryptanalysis on Serpent where restricting to an affine set of approximations could potentially improve the attack.

The second example consists of experimental evaluation of certain affine multidimensional linear approximations on block cipher SIMON32/64. Using our statistical model for random permutations we present a simple test to evaluate randomness experimentally. We were able to identify nonrandom behaviour of round-reduced SIMON32/64 up to 18 rounds. It remains, however, an open question whether such affine multidimensional linear approximations on SIMON32/64 and its larger versions can potentially lead to efficient key-recovery attacks.

The best linear key-recovery attack on SIMON32/64 is given in [28]. It makes use of the 13-round linear hull identified in [27] and adds 5 rounds before and after this distinguisher to extend the attack over to 23 rounds. It starts from the input to round 2 of the core trail, see Table I and ends after round 14 thus covering 13 rounds of the cipher. It has been chosen carefully in such a way that the input and output masks of the linear hull have only one active bit each which allows very efficient key guessing techniques.

The distinguishing advantage of the 13-round linear approximation used in [28] corresponds to the one given in Table II, that is, it can recover $d = 4$ bits of the secret key with experimentally determined success probability $P_S(0.0625) = 0.330$. Our 2-dimensional affine distinguishers given in Table III can improve the success probability to $P_S(0.0625) = 0.483$ for 13 rounds, or alternatively, increase the number of rounds to 15 with a slight decrease of the success probability to $P_S(0.0625) = 0.254$. Our distinguishers, however, have several active bits in the input and output making efficient key search a challenging task which is left for future work.

APPENDIX A
PROOF OF THEOREM 1

Lemma 5: In the setting of Theorem 1, let $z_1, \dots, z_k$ be k multinomially distributed variables. Then the variables and some of their first powers and products, where $\eta, \zeta, \xi, \iota \in \{1, \dots, k\}$, have the following expected values.

$$\begin{aligned}
\mathrm{Exp}(z_\eta) &= mp_\eta \\
\mathrm{Exp}(z_\eta z_\zeta) &= m(m-1)p_\eta p_\zeta \\
\mathrm{Exp}(z_\eta^2) &= mp_\eta + m(m-1)p_\eta^2 \\
\mathrm{Exp}(z_\eta^2 z_\zeta) &= m(m-1)p_\eta p_\zeta + m(m-1)(m-2)p_\eta^2 p_\zeta \\
\mathrm{Exp}(z_\eta^3) &= mp_\eta + 3m(m-1)p_\eta^2 \\
&+ m(m-1)(m-2)p_\eta^3 \\
\mathrm{Exp}(z_\eta z_\zeta z_\xi z_\iota) &= m(m-1)(m-2)(m-3)p_\eta p_\zeta p_\xi p_\iota \\
\mathrm{Exp}(z_\eta^2 z_\zeta z_\xi) &= m(m-1)(m-2)p_\eta p_\zeta p_\xi \\
&+ m(m-1)(m-2)(m-3)p_\eta^2 p_\zeta p_\xi \\
\mathrm{Exp}(z_\eta^2 z_\zeta^2) &= m(m-1)p_\eta p_\zeta \\
&+ m(m-1)(m-2)(p_\eta^2 p_\zeta + p_\eta p_\zeta^2) \\
&+ m(m-1)(m-2)(m-3)p_\eta^2 p_\zeta^2 \\
\mathrm{Exp}(z_\eta^3 z_\zeta) &= m(m-1)p_\eta p_\zeta \\
&+ 3m(m-1)(m-2)p_\eta^2 p_\zeta \\
&+ m(m-1)(m-2)(m-3)p_\eta^3 p_\zeta \\
\mathrm{Exp}(z_\eta^4) &= mp_\eta \\
&+ 7m(m-1)p_\eta^2 + 6m(m-1)(m-2)p_\eta^3 \\
&+ m(m-1)(m-2)(m-3)p_\eta^4
\end{aligned}$$

The proof of this lemma consists of straightforward calculations of the expected values according to the probability mass function of the multinomial distribution given by

$$\varphi(z_1, \dots, z_k) = \frac{m!}{z_1! \cdots z_k!} p_1^{z_1} \cdots p_k^{z_k}.$$

Next we give the proof of Theorem 1.

Proof: Let us start by writing the capacity C in the form

$$C = \frac{k}{m^2} \sum_{\eta=1}^{k} \left(z_\eta - \frac{m}{k} \right)^2 = \frac{k}{m^2} \sum_{\eta=1}^{k} z_\eta^2 - 1.$$

To compute the variance of the capacity it suffices to do it for the sum $\sum_{\eta=1}^{k} z_\eta^2$. We write

$$\mathrm{Var} \sum_\eta z_\eta^2 = \mathrm{Exp} \left(\sum_\eta z_\eta^2 \right)^2 - \left(\mathrm{Exp} \sum_\eta z_\eta^2 \right)^2 \tag{32}$$

$$= \mathrm{Exp} \left(\sum_\eta z_\eta^4 \right) \tag{33}$$

$$+ \; \mathrm{Exp} \left(\sum_\eta \sum_{\zeta \neq \eta} z_\eta^2 z_\zeta^2 \right) \tag{34}$$

$$- \left(\mathrm{Exp} \sum_\eta z_\eta^2 \right)^2 . \tag{35}$$

By Lemma 5 (33) can be expressed as

$$\begin{aligned}
&m + 7m(m-1)P_2 + 6m(m-1)(m-2)P_3 \\
&+ m(m-1)(m-2)(m-3)P_4,
\end{aligned}$$

where we have denoted

$$P_4 = \sum_{\eta=1}^{k} p_\eta^4.$$

Similarly, (34) can be expressed as

$$\begin{aligned}
&m(m-1) + (2m-5)m(m-1)P_2 \\
&+ m(m-1)(m-2)(m-3)P_2^2 - 2m(m-1)(m-2)P_3 \\
&- m(m-1)(m-2)(m-3)P_4,
\end{aligned}$$

and (35) as

$$m^2 + 2m^2(m-1)P_2 + m^2(m-1)^2 P_2^2.$$

By combining these expressions and multiplying by k^2/m^4, we get the claimed result for the variance of C. The derivation of the mean is similar, but simpler. ■

Appendix B
Proof of Theorem 2

Proof: The number of zeroes of $f(x)+a\cdot x$ can be written as

$$|\{x \in \mathbb{F}_2^n | f(x)=0, a\cdot x=0\}|+|\{x \in \mathbb{F}_2^n | f(x)=1, a\cdot x=1\}|,$$

where by Lemma 3, the number $\upsilon = |\{x \in \mathbb{F}_2^n \,|\, f(x) = 0,\, a\cdot x = 0\}|$ follows $\mathcal{H}G(2^n, 2^{n-1}, N_0)$. As f varies over all Boolean function, the number of zeroes N_0 follows the binomial distribution $\mathcal{B}(2^n, \frac{1}{2})$. That is

$$\Pr(N_0 = w) = \left(\frac{1}{2}\right)^{2^n} \binom{2^n}{w}.$$

Then

$$\Pr(\upsilon = k) = \sum_{w} \Pr(N_0 = w) \Pr(\upsilon = k \,|\, N_0 = w),$$

where the bounds for w and k are as follows

$$k \le w \le 2^{n-1} + k \text{ and } 0 \le k \le 2^{n-1}.$$

We get

$$\begin{aligned} \Pr(\upsilon = k) &= \left(\frac{1}{2}\right)^{2^n} \binom{2^{n-1}}{k} \sum_{w=k}^{2^{n-1}+k} \binom{2^{n-1}}{w-k} \\ &= \left(\frac{1}{2}\right)^{2^{n-1}} \binom{2^{n-1}}{k}, \end{aligned}$$

that is, $\upsilon \sim \mathcal{B}(2^{n-1}, \frac{1}{2})$. Similarly, it can be shown that

$$|\{x \in \mathbb{F}_2^n \,|\, f(x) = 1,\, a\cdot x = 1\}| \sim \mathcal{B}(2^{n-1}, \frac{1}{2}).$$

As the sum of two $\mathcal{B}(2^{n-1}, \frac{1}{2})$-variables, the number of zeroes of the function $f(x) + a \cdot x$ follows $\mathcal{B}(2^n, \frac{1}{2})$. ■

References

[1] A. Bogdanov, E. Tischhauser, and P. S. Vejre, “Multivariate profiling of hulls for linear cryptanalysis,” *IACR Trans. Symmetric Cryptol.*, vol. 2018, no. 1, pp. 101–125, Mar. 2018. [Online]. Available: https://tosc.iacr.org/

[2] M. Matsui, “The first experimental cryptanalysis of the data encryption standard,” in *Proc. 14th Annu. Int. Cryptol. Conf.*, in Lecture Notes in Computer Science, vol. 839, Santa Barbara, CA, USA, Y. Desmedt, Ed. Berlin, Germany: Springer, Aug. 1994, pp. 1–11, doi: 10.1007/3-540-48658-5_1.

[3] L. O’Connor, “Properties of linear approximation tables,” in *Proc. FSE*, in Lecture Notes in Computer Science, vol. 1008, B. Preneel, Ed., Heidelberg, Germany: Springer, 1995, pp. 131–136.

[4] J. Daemen and V. Rijmen, “Probability distributions of correlation and differentials in block ciphers,” *J. Math. Cryptol.*, vol. 1, no. 3, pp. 221–242, 2007.

[5] A. Bogdanov and V. Rijmen, “Linear hulls with correlation zero and linear cryptanalysis of block ciphers,” *Des., Codes Cryptogr.*, vol. 70, no. 3, pp. 369–383, Mar. 2014.

[6] A. Bogdanov and E. Tischhauser, “On the wrong key randomisation and key equivalence hypotheses in Matsui’s algorithm 2,” in *Proc. 20th Int. Workshop Fast Softw. Encryption*, in Lecture Notes in Computer Science, vol. 8424, Singapore, S. Moriai, Ed. Berlin, Germany: Springer, Mar. 2013, pp. 19–38, doi: 10.1007/978-3-662-43933-3_2.

[7] C. Blondeau and K. Nyberg, “Joint data and key distribution of simple, multiple, and multidimensional linear cryptanalysis test statistic and its impact to data complexity,” *Des., Codes Cryptogr.*, vol. 82, nos. 1–2, pp. 319–349, Jan. 2017, doi: 10.1007/s10623-016-0268-6.

[8] T. Ashur, T. Beyne, and V. Rijmen, “Revisiting the wrong-key-randomization hypothesis,” *J. Cryptol.*, vol. 33, no. 2, pp. 567–594, Apr. 2020.

[9] M. Hermelin, J. Y. Cho, and K. Nyberg, “Multidimensional extension of Matsui’s algorithm 2,” in *Proc. 16th Int. Workshop Fast Softw. Encryption*, in Lecture Notes in Computer Science, vol. 5665, Leuven, Belgium, O. Dunkelman, Ed. Berlin, Germany: Springer, Feb. 2009, pp. 209–227, doi: 10.1007/978-3-642-03317-9_13.

[10] M. Hermelin, J. Y. Cho, and K. Nyberg, “Multidimensional linear cryptanalysis,” *J. Cryptol.*, vol. 32, no. 1, pp. 1–34, Jan. 2019.

[11] A. Bogdanov, E. Tischhauser, and P. S. Vejre, “Multivariate linear cryptanalysis: The past and future of present,” *IACR, Cryptol. ePrint Arch.*, vol. 2016, p. 667, Jul. 2016. [Online]. Available: http://eprint.iacr.org/2016/667

[12] C. Blondeau and K. Nyberg, “Improved parameter estimates for correlation and capacity deviates in linear cryptanalysis,” *IACR, Trans. Symmetric Cryptol.*, vol. 2016, no. 2, pp. 162–191, 2016, doi: 10.13154/tosc.v2016.i2.162-191.

[13] K. Nyberg, “Affine linear cryptanalysis,” *Cryptogr. Commun.*, vol. 11, no. 3, pp. 367–377, May 2019.

[14] M. Khan and K. Nyberg, “Linear approximations of random functions and permutations,” *IACR Cryptol. ePrint Arch.*, vol. 2019, p. 934, Sep. 2019. [Online]. Available: https://eprint.iacr.org/2019/934

[15] K. Nyberg and M. Hermelin, “Multidimensional Walsh transform and a characterization of bent functions,” in *Proc. IEEE Inf. Theory Workshop Inf. Theory Wireless Netw.*, Solstrand, Norway, T. Helleseth, P. V. Kumar, and Ø. Ytrehus, Eds., Jul. 2007, pp. 1–4, doi: 10.1109/ITWITWN.2007.4318037.

[16] M. Hermelin, J. Y. Cho, and K. Nyberg, “Multidimensional linear cryptanalysis of reduced round Serpent,” in *Proc. 13th Australas. Conf. Inf. Secur. Privacy*, in Lecture Notes in Computer Science, vol. 5107, Wollongong, NSW, Australia, Y. Mu, W. Susilo, and J. Seberry, Eds. Berlin, Germany: Springer, Jul. 2008, pp. 203–215, doi: 10.1007/978-3-540-70500-0_15.

[17] J. N. K. Rao and D. R. Thomas, “Chi-squared tests for contingency tables,” in *Analysis of Complex Surveys*, C. J. Skinner, D. Holt, and T. M. F. Smith, Eds. Chichester, U.K.: Wiley, 1989, pp. 89–104.

[18] F. C. Drost, W. C. M. Kallenberg, D. S. Moore, and J. Oosterhoff, “Power approximations to multinomial tests of fit,” *J. Amer. Stat. Assoc.*, vol. 84, no. 405, pp. 130–141, Mar. 1989.

[19] A. Bogdanov, G. Leander, K. Nyberg, and M. Wang, “Integral and multidimensional linear distinguishers with correlation zero,” in *Proc. ASIACRYPT*, in Lecture Notes in Computer Science, vol. 7658, X. Wang and K. Sako, Eds. Berlin, Germany: Springer, 2012, pp. 244–261.

[20] G. Leander, “Small scale variants of the block cipher PRESENT,” *IACR Cryptol. ePrint Arch.*, vol. 2010, p. 143, Mar. 2010. [Online]. Available: http://eprint.iacr.org/2010/143

[21] S. Matyas, C. Meyer, and J. Oseas, “Generating strong one-way functions with cryptographic algorithms,” *IBM Tech. Discl. Bull.*, vol. 27, no. 10a, pp. 5658–5659, 1985.

[22] T. Shrimpton and M. Stam, “Building a collision-resistant compression function from non-compressing primitives,” in *Proc. 35th Int. Colloq. Automata, Lang., Program.*, in Lecture Notes in Computer Science, vol. 5126, Reykjavik, Iceland, L. Aceto, I. Damgård, L. A. Goldberg, M. M. Halldórsson, A. Ingólfsdóttir, and I. Walukiewicz, Eds. Berlin, Germany: Springer, Jul. 2008, pp. 643–654, doi: 10.1007/978-3-540-70583-3_52.

[23] J. Katz and Y. Lindell, *Introduction to Modern Cryptography*, 2nd ed. Boca Raton, FL, USA: CRC Press, 2014. [Online]. Available: https://www.crcpress.com/Introduction-to-Modern-Cryptography-Second-Edi% tion/Katz-Lindell/p/book/9781466570269

[24] M. Matsui, “Linear cryptanalysis method for DES cipher,” in *Proc. Workshop Theory Appl. Cryptograph. Techn.*, in Lecture Notes in Computer Science, vol. 765, Lofthus, Norway, T. Helleseth, Ed. Berlin, Germany: Springer, 1994, pp. 386–397.

[25] A. Biryukov, C. D. Cannière, and M. Quisquater, “On multiple linear approximations,” in *Proc. 24th Annu. Int. Cryptol. Conf.*, in Lecture Notes in Computer Science, vol. 3152, Santa Barbara, CA, USA, M. K. Franklin, Ed. Berlin, Germany: Springer, Aug. 2004, pp. 1–22, doi: 10.1007/978-3-540-28628-8_1.

[26] R. Beaulieu, D. Shors, J. Smith, S. Treatman-Clark, B. Weeks, and L. Wingers, “The SIMON and SPECK lightweight block ciphers,” in *Proc. 52nd Annu. Design Automat. Conf.*, San Francisco, CA, USA, Jun. 2015, pp. 175:1–175:6, doi: 10.1145/2744769.2747946.

[27] J. Alizadeh, H. AlKhzaimi, M. R. Aref, N. Bagheri, P. Gauravaram, and M. M. Lauridsen, "Improved linear cryptanalysis of round reduced SIMON," *IACR Cryptol. ePrint Arch.*, vol. 2014, no. 681, 2014. [Online]. Available: http://eprint.iacr.org/2014/681

[28] H. Chen and X. Wang, "Improved linear hull attack on round-reduced SIMON with dynamic key-guessing techniques," in *Proc. 23rd Int. Conf. Fast Softw. Encryption*, in Lecture Notes in Computer Science, vol. 9783, Bochum, Germany, T. Peyrin, Ed. Berlin, Germany: Springer, Mar. 2016, pp. 428–449, doi: 10.1007/978-3-662-52993-5_22.

Tomer Ashur received the Ph.D. degree from KU Leuven, Leuven, Belgium, in 2017, after completing the dissertation on cryptanalysis of symmetric-key primitives.

He was a Teaching Assistant with the University of Haifa, the CIO of Meditron Healthcare Services, the Project Manager with Katz Delivering Services, the Head of Support with Safend Inc., and a Communication Officer (OF-2) with the Israel Defense Forces. He is an FWO Post-Doctoral Fellow with KU Leuven and a Visiting Assistant Professor with TU Eindhoven, Eindhoven, The Netherlands.

Mohsin Khan received the B.Sc. degree in computer science and information technology from the Islamic University of Technology, Gazipur, Bangladesh, in 2006, the M.Sc. degree in foundations of advanced computing from Aalto University, Espoo, Finland, in 2015, and the Ph.D. degree in computer science from the University of Helsinki, Helsinki, Finland, in 2021.

From 2007 to 2012, he worked in several roles, such as a System Analyst and the Project Manager, at two subsidiaries of Telenor, Bangladesh. During 2013 to 2016, he worked as a Teaching Assistant in various graduate-level courses and as a Research Assistant in several short-term internship programs at Aalto University and the University of Helsinki. In the autumn of 2021, he joined Ericsson, Sweden, as a Researcher in security standardization. He is the author of nine articles. His research interests include cellular networks security, applied cryptography, and statistical cryptanalysis.

Kaisa Nyberg received the Ph.D. degree in mathematics from the University of Helsinki in 1980.

From 1987 to 1998, she worked as a Mathematician and a Cryptographer with the Finnish Defence Forces. In 1998, she joined Nokia and was responsible for cryptographic techniques in cellular security and related systems. Since 2005, she was a Full Professor at Aalto University, Espoo, Finland, until her retirement in 2016. She is an author of about 80 research publications and holds ten patents. She is also a coauthor of the book *UMTS Security* and was a member of the designer team of the Bluetooth secure simple pairing in 2002.

Prof. Nyberg became a member of the International Association for Cryptologic Research (IACR) in 1988 and was awarded the title of IACR Fellow in 2015. In 2006, she became a member of the Finnish Academy of Science and Letters. From 2010 to 2020, she was an Associate Editor of the *Journal of Cryptology*.