Publication V

Miia Hermelin, Joo Yeon Cho, and Kaisa Nyberg. 2009. Statistical tests for key recovery using multidimensional extension of Matsui's Algorithm 1. In: Postersession of the 28th Annual International Conference on the Theory and Applications of Cryptographic Techniques (Eurocrypt 2009). Cologne, Germany. 26-30 April 2009. Also appeared in Helena Handschuh, Stefan Lucks, Bart Preneel, and Phillip Rogaway (editors). Symmetric Cryptography. Dagstuhl Seminar 09031. Dagstuhl, Germany. 11-16 January 2009. Dagstuhl Seminar Proceedings, number 09031.

Statistical Tests for Key Recovery Using Multidimensional Extension of Matsui's Algorithm 1

Miia Hermelin[1], Joo Yeon Cho[1], and Kaisa Nyberg[1,2]

[1] Department of Information and Computer Science
Helsinki University of Technology
[2] Nokia, Finland
miia.hermelin@tkk.fi, joo.cho@tkk.fi, kaisa.nyberg@tkk.fi

Abstract. In one dimension, there is essentially just one binomially distributed statistic, bias or correlation, for testing correctness of a key bit in Matsui's Algorithm 1. In multiple dimensions, different statistical approaches for finding the correct key candidate are available. The purpose of this work is to investigate the efficiency of such test in theory and practice, and propose a new key class ranking statistic using distributions based on multidimensional linear approximation and generalisation of the ranking statistic presented by Selçuk.

1 Introduction

In 1993, Matsui introduced linear cryptanalysis in [1]. He presented two key recovery attacks, Algorithm 1 (Alg. 1) and Algorithm 2 (Alg. 2), against the block cipher DES. Using Alg. 1 one bit of information about the secret key of the cipher can be extracted. The goal of Alg. 2 is to recover a part of the last round key. Later Robshaw and Kaliski [2] and Biryukov, et al., [3] proposed the use of several linear approximations that were assumed to be statistically independent. While the goal of Robshaw and Kaliski was to reduce the data complexity using more than one approximation of one key bit, Biryukov, et al., presented a generalisation of Alg. 1 for finding several key bits simultaneously. They used a special purpose statistical test for finding the right key parity bits. In 2004, Baignères, et al., studied a truly multidimensional linear distinguisher [4] and its data complexity.

Algorithms for computing large probability distributions related to multidimensional linear distinguishers were studied for example in [5]. However, this approach is not feasible for block-size over 32 bits. In [6] it was shown how a multidimensional linear approximation can be constructed in practice from one-dimensional approximations, and how it

could be used in a multidimensional version of Alg. 1 for finding more than one bit of information of the key. The multidimensional attack was compared to the experimental results of [7] of the method of Biryukov, et al., and shown to perform better at least in the case of four-round Serpent.

In one dimension, there is only one statistic (the bias or equivalently, the correlation) that is used for testing the statistical hypotheses in Alg. 1. However, in multiple dimensions there are different ways of realising the attack. The purpose of this work is to investigate the efficiency of a few statistical key recovery methods and suggest a ranking of the key candidates by adopting the ranking statistic and the concept of advantage proposed by Selçuk [8] to the case of Alg. 1 and multidimensional linear approximation.

The structure of this paper is as follows: The basic statistical notions are given in Sect. 2. In Sect. 3 we present the multidimensional generalisation of Alg. 1 and the key recovery problem. In Sect. 4 we study the different ranking methods in theory. Practical experiments were done on reduced round Serpent and are presented in Sect. 5. Finally, Sect. 6 draws conclusions.

2 Mathematical Tools

We denote by V_n the linear space of n-bit vectors, for $n = 1, 2, \ldots$. Let $f : V_n \to V_1$ be a Boolean function. The correlation between f and zero (aka. the correlation of f) is

$$c(f) = c(f, 0) = 2^{-n} \left(\#\{\xi \in V_n \mid f(\xi) = 0\} - \#\{\xi \in V_n \mid f(\xi) = 1\}\right).$$

If random variable (r.v.) Y is distributed according to probability distribution (p.d.) p we denote $Y \sim p$. Let $X \sim \theta$, the uniform distribution and $f : V_n \to V_m$ be a vector Boolean function. We call $p = (p_0, \ldots, p_{2^m-1})$ the probability distribution (p.d.) of f if the r.v. $f(X) \sim p$. For $m = 1$, we have $p = (\frac{1}{2}(1 + c(f)), \frac{1}{2}(1 - c(f)))$.

2.1 Kullback-Leibler Distance and Log-Likelihood Ratio

We recall the following definitions from [4] and [6]:

Definition 1. *Let $p = (p_0, \ldots, p_M)$ and $q = (q_0, \ldots, q_M)$ be two p.d.'s. Their* relative entropy *or* Kullback-Leibler distance *is*

$$D(p||q) = \sum_{\eta=0}^{M} p_\eta \log \frac{p_\eta}{q_\eta}, \tag{1}$$

where we use the convention $0 \log 0/b = 0$, $b \neq 0$ *and* $b \log b/0 = \infty$.

Definition 2. *The capacity between two p.d.'s p and q is defined by*

$$C(p,q) = \sum_{\eta=0}^{M} \frac{(p_\eta - q_\eta)^2}{q_\eta}. \tag{2}$$

If $q = \theta$, *then* $C(p,\theta)$ *will be denoted by* $C(p)$ *and is called the capacity of p.*

Property 1. *We say that p.d p* is close to *p.d. q if* $|p_\eta - q_\eta| \ll q_\eta$, *for all* $\eta = 0, 1, \ldots, M$.

If p is close to q, we can approximate their Kullback-Leibler distance using the Taylor series [4] such that $D(p||q) = \frac{1}{2}C(p,q) + \mathcal{O}(\epsilon^3)$, where $\epsilon = \max_{\eta \in V_m} |p_\eta - q_\eta|$.

The normed normal distribution with mean 0 and variance 1 is denoted by $\mathcal{N}(0,1)$. Its probability density function (p.d.f.) is denoted by ϕ and the cumulative distribution function (c.d.f.) by Φ. The normal distribution with mean μ and variance σ^2 is denoted by $\mathcal{N}(\mu, \sigma^2)$.

Let us assume we are given N words of independently and identically distributed (i.i.d.) data, $\hat{z}_1, \ldots, \hat{z}_N$ drawn either from p.d. p or $q \neq p$. Let the word-size of the data be m. The empirical p.d. $\hat{q} = (\hat{q}_0, \ldots, \hat{q}_{2^m-1})$ corresponding to the data has components $\hat{q}_\eta = \#\{i = 1, \ldots, N \mid \hat{z}_i = \eta\}$. The hypothesis testing problem with null hypothesis H_0 stating the data is drawn from p and alternative hypothesis H_1 stating it is drawn from q can be solved using the log-likelihood ratio (LLR) calculated from data as follows:

$$\mathrm{LLR}(\hat{q}, p, q) = \sum_{\eta=0}^{2^m-1} N\hat{q}_\eta \log \frac{p_\eta}{q_\eta}. \tag{3}$$

We accept (reject) the null hypothesis if $\mathrm{LLR}(\hat{q}, p, q) \geq \gamma$ ($\leq \gamma$), where γ is the threshold of the test. The Neyman-Pearson lemma states that this statistic is the optimal distinguisher for these two hypotheses [9]. The proof for the following theorem can be found in [4].

Theorem 1. *The statistic* $\mathrm{LLR}(\hat{q}, p, q)$ *defined by* (3) *is asymptotically normal with mean and variance* $N\mu_0$ *(respectively* $N\mu_1$*) and* $N\sigma_0^2$ *(respectively* $N\sigma_1^2$*), if the data is drawn i.i.d. from p (respectively q). The means and variances are given by*

$$\begin{gathered} \mu_0 = D(p||q) \text{ and } \mu_1 = -D(q||p) \\ \sigma_0^2 = \sum_{\eta=0}^{M} p_\eta \log^2 \frac{p_\eta}{q_\eta} - \mu_0^2 \text{ and } \sigma_1^2 = \sum_{\eta=0}^{M} q_\eta \log^2 \frac{p_\eta}{q_\eta} - \mu_1^2. \end{gathered} \tag{4}$$

Moreover, if p is close to q, we have $\mu_0 \approx -\mu_1 \approx \frac{1}{2}C(p,q)$ and $\sigma_0^2 \approx \sigma_1^2 \approx C(p,q)$.

3 Multidimensional Linear Approximation

Let us study a block cipher with encryption function f and with block size n. Let x be the plaintext, K the expanded inner key, that is, a vector consisting of all (fixed) round key bits and $y = f(x, K)$ the ciphertext. Then an m-dimensional linear approximation of the block cipher can be considered as a vector Boolean function

$$V_n \times V_n \to V_m,\ (x, y) \mapsto Ux + Wy + VK, \tag{5}$$

where U and W are $m \times n$ binary matrices. The matrix V has also m rows and it divides the expanded keys, and therefore also the keys, to 2^m equivalence classes $g = VK$, $g \in V_m$. Our problem is now to find the right inner key class, denoted by g_0.

The most complex task in linear cryptanalysis is determining the probability distribution of the Boolean function (5). A generalised concept of correlation for vector Boolean functions is given in [10] and used in deriving some theoretical results about Boolean functions [11]. However, in practical applications we use another approach.

If correlations for each of the 2^m-1 one-dimensional nonzero linear approximations related to (5) are known then p can be calculated as shown in [6]. That is, p is determined based on one-dimensional projections, which is a well-known statistical method due to Cramér and Wold (1936) [12]. In practice, only estimates of the correlations are available. Hence only an approximation p of the true distribution can be achieved. Nevertheless, the problem of finding p reduces to that of finding several (say, m) strong one-dimensional, linearly independent approximations and then determining correlations for the remaining $2^m - m - 1$ approximations and selecting the ones with non-negligible correlations as described in [6].

4 Finding the Right Inner Key Class

We start by drawing N plaintext-ciphertext pairs $(\hat{x}_i, \hat{y}_i)$, $i = 1, \ldots, N$ from the cipher. The empirical data is then $\hat{z}_i = U\hat{x}_i + W\hat{y}_i, i = 1, \ldots, N$ with observed empirical p.d. $\hat{q}$. For $m = 1$, we denote by c ($\hat{c}$) the theoretical (empirical) correlation of $u \cdot x + w \cdot y$.

Let p be the p.d. of (5). For each $g \in V_m$, the data $\hat{z}_i, i = 1, \ldots, N$ is drawn from p.d. p^g, a fixed permutation of p determined by g. Then

all p.d. p^g, $g \in V_m$, are each other's permutations, and in particular, $p^g_{\eta\oplus k} = p^{g\oplus k}_\eta$, for all $g, \eta, k \in V_m$.

The decision in Alg. 1 in one dimension is based on the following test: The key class bit g is chosen to be 0 if $c\hat{c} > 0$. Otherwise, $g = 1$. In other words, the statistical decision problem is to determine which of the two distributions $(\frac{1}{2}(1 \pm c), \frac{1}{2}(1 \mp c))$ gives the best fit with the data.

We have studied three different ways to generalise the one-dimensional Alg. 1 to multiple dimensions. Since the data is drawn i.i.d. from the p.d. p^{g_0} and not from any other p.d $p^g, g \neq g_0$, we can interpret the problem of finding g_0 as a generalisation of the goodness-of-fit test where one determines whether given data is drawn from p.d. p^g or not. The inner key class $g \in V_m$ which is most strongly indicated by this test to fit the data is chosen to be the right key class. The classical goodness-of-fit tests are the χ^2-test and the G-test based on the Kullback-Leibler distance. In this paper we investigate generalisations of these tests into the case of multiple distributions, i.e., finding one distribution from a set of distributions. The χ^2-method based on the χ^2-test and the log-likelihood method based on the G-test are studied in Sections 4.1 and 4.2, respectively.

In [13] the problem of distinguishing one *known* p.d. from a set of other p.d.'s was studied. It was then possible to use the optimal distinguisher, the LLR-statistic, in solving the problem. However, since g_0 is unknown, we cannot apply the results of [13] in our work directly. Rather, we will use the following heuristic. Since the data $\hat{z}_i$, $i = 1, \ldots, N$ is drawn from the unknown p.d. $p^{g_0} \neq \theta$, it should be easiest to distinguish the right p.d. p^{g_0} rather than any other p.d. p^g, $g \neq g_0$ from the uniform distribution using the LLR-statistic. Hence, the inner key class $g \in V_m$ that gives the strongest distinguisher between the corresponding p.d p^g and θ is chosen to be the right key. The log-likelihood ratio or LLR-method is studied in Sect. 4.3.

In all our analysis, it is assumed that p^g and $p^{g'}$, for $g \neq g'$, are close to each other and all these distributions p^g are close to θ in the sense of Property 1. Then the following results apply:

$$D(p||\theta) = D(p^g||\theta) \text{ and } C(p) = C(p^g), \text{ for all } g \in V_m, \tag{6}$$

and

$$\min_{g,g\neq h} D(p^g||p^h) = \min_{g\neq 0} D(p^g||p) \text{ and } \min_{g,g\neq h} C(p^g, p^h) = \min_{g\neq 0} C(p^g, p), \text{ for all } h \in V_m, \tag{7}$$

where the minimum Kullback-Leibler distance and capacity will be denoted by $D_{\min}(p)$ and $C_{\min}(p)$, respectively. The assumption about closeness must be verified with practical experiments. Moreover, if $D_{\min}(p) =$

0, we need to join the corresponding key classes to one class such that we may assume $D_{\min}(p) \neq 0$ and $C_{\min}(p) \neq 0$. The number $2^m - 1$ will henceforth be denoted by M.

4.1 χ^2-method

The χ^2-statistic for each key $g \in V_m$ is defined as follows:

$$S(g) = N \sum_{\eta=0}^{M} \frac{(\hat{q}_\eta - p_\eta^g)^2}{p_\eta^g}, \tag{8}$$

with M degrees of freedom, where N is the amount of data used in the attack. The empirical distribution $\hat{q}$ should be near to the correct p.d. p^{g_0} while being further away from all the other p.d.'s p^g, $g \neq g_0$. Hence, the key corresponding to the smallest $S(g)$ is chosen to be the right key class.

By [14], one may approximate the distribution of $S(g)$ by $\chi^2_M\left(NC(p^g, p^{g_0})\right)$, the non-central χ^2-distribution with mean $\mu_g = M + NC(p^g, p^{g_0})$ and variance $\sigma_g^2 = 2(M + 2NC(p^g, p^{g_0}))$. We can approximate $S(g) \sim \mathcal{N}(\mu_g, \sigma_g^2)$, provided that $\mu_g > 30$ [15].

With similar calculations and approximations that were done in Sect. 4 in [4] or in the proof of Thm. 2 in [6] we get that the upper bound for the data complexity for finding g_0 is proportional to

$$N_{\chi^2} = \frac{4m - 4\gamma_S + 2\sqrt{2M(m - \gamma_S)}}{C_{\min}(p)}, \tag{9}$$

where $\gamma_S = \ln(\sqrt{2\pi} \ln P_S^{-1})$. Note the exponential dependence of N_{χ^2} on m as $M = 2^m - 1$.

4.2 The Log-Likelihood Method

Another popular goodness-of-fit test is the log-likelihood test, also known as G-test. The experiments on Alg. 1 done in [6] used this test. It is based on the Kullback-Leibler -distance $G(g) = D(\hat{q}||p^g)$ between the empirical p.d. $\hat{q}$ and the theoretical p.d. p^g. In [14] it is shown that for each key $g \in V_m$ the statistic can be approximated to be distributed as $G(g) \sim \chi^2_M(\delta_g) + \xi_g$, where $\delta_g = N \sum_{\eta=0}^{M} p_\eta^g \log^2 \frac{p_\eta^g}{p_\eta^{g_0}} - ND(p^g||p^{g_0})^2$ and $\xi_g = 2ND(p^g||p^{g_0}) - \delta_g$. Since p^g are near to p^{g_0}, the parameters $\delta_g \approx NC(p^g, p^{g_0})$ and $\xi_g \approx 0$ and the G-test is the same as the χ^2-test [14].

4.3 Log-Likelihood Ratio Method

The log-likelihood ratio is the optimal statistic for distinguishing two distributions [9]. It is also asymptotically normal as stated in Theorem 1. Hence, we would like to use it as a key ranking statistic. The method was shortly described in [16]. The idea is that the empirical distribution can be used for distinguishing the p.d. p^{g_0} related to the correct key class from the uniform p.d. with large LLR value, while any wrong p.d. $p^g, g \neq g_0$ is less distinguishable from θ. For each $g \in V_m$ we compute

$$l(g) = \mathrm{LLR}(\hat{q}, p^g, \theta). \tag{10}$$

We choose the key class g with largest $l(g)$ to be the right key class.

Key ranking has classically been used only in Alg. 2, as there are several, say, n-bit key candidates $k \in V_n$ among which the right key candidate has to be found, whereas in Alg. 1, the key parity bit $g \in V_1$ can only have two values. In multiple dimensions there are several key candidates $g \in V_m$ that are ranked, that is, sorted, according to a statistic $T(g)$ in decreasing (or increasing) order of magnitude. Writing the ordered r.v.'s as $T_1 \geq T_2 \geq \cdots \geq T_{2^m}$, we call T_r the rth order statistic. The higher the right key is on the list, the better the ranking statistic T is. Biryukov, et al., used a special purpose quantity called "gain" to measure the strength of their ranking statistic [3]. However, a more generally applicable measure is the advantage, defined by Selçuk in [8] as follows:

Definition 3. *We say that a key recovery attack for an m-bit key achieves an advantage of a bits over exhaustive search, if the correct key is ranked among the top $r = 2^{m-a}$ out of all 2^m key candidates.*

In this paper, we attempt to transfer the idea to Alg. 1. We wish to find a relationship between the data complexity N and the advantage a for the ranking statistic $l(g)$ defined in (10). Let us define the success probability P_S of ranking g_0 among the r highest ranking keys to be

$$P_S = \Pr(l(g_0) > l_r), \tag{11}$$

where l_r is the rth (wrong key) order statistic. The problem is now to solve N as a function of a and vice versa from (11). We cannot apply [13] here as the result would be too optimistic. The task is to distinguish an unknown p^{g_0} from a set of p.d.'s $\{p^g \mid g \in V_m\}$. In [13] one distinguishes only p^{g_0} from $p^{g_0'}$, $g_0' \neq g_0$, the p.d. closest to p^{g_0} in Kullback-Leibler distance. We need to consider all the other key candidates as well, which increases the data complexity. We will have the following result:

Theorem 2. *Assume that the r.v.'s $l(g)$ are s.i. If the p.d.'s p^g, $g \in V_m$ and θ are close to each other, the advantage of the LLR-method using statistic* (3) *can be approximated by*

$$a \approx \left(\frac{1}{2} \sqrt{NC(p)} - \Phi^{-1}(P_S) \right)^2, \tag{12}$$

where $P_S (\geq 0.5)$ is the probability of success, N is the amount of data used in the attack and $C(p)$ and m are the capacity and the dimension of the linear approximation (5)*, respectively.*

Before the proof let us investigate the assumption about s.i. of $l(g)$'s. Since the *same data* is used in calculating all of them, they are actually statistically dependent. However, the derivations become impossible in a general case with dependent r.v.'s. The experiments presented in Sect. 5 show that the statistical dependence of $l(g)$'s does not weaken the attack, hence, making the assumption does not give too optimistic results at least for Serpent. On the other hand, the assumption does not affect the actual method, it just makes it possible to calculate an upper bound for the data complexity and get the "worst-case" trade-off between the data complexity and advantage. Note that the assumption about statistical independence of $l(g)$, $g \in V_m$ does not mean that the one-dimensional linear approximations used in deriving p should be statistically independent.

Proof. Let us proceed first by finding the p.d.'s for the r.v.'s $l(g), g \in V_m$. By Theorem 1 and property (6) r.v. $l(g_0) \sim \mathcal{N}(N\mu_R, N\sigma_R^2)$, where $\mu_R \approx C(p)/2$ and $\sigma_R^2 \approx C(p)$. If $g \neq g_0$, we heuristically claim that $l(g) \sim \mathcal{N}(\mu_W, \sigma_W^2)$, where $\mu_W \leq 0$ and $\sigma_W^2 \approx C(p)$. The normal distribution is based on the law of large numbers [9], the approximation of variance is commonly used for example in [17] and the approximation of mean is based on the idea that the empirical data is not closer to any $p^g, g \neq g_0$ than θ. In the "worst-case", with largest data complexity, $\mu_W = 0$. We denote the p.d.f. and c.d.f. of $l(g), g \neq g_0$ by f_W and F_W, respectively.

Calculating the c.d.f. of the order statistic l_r of statistically dependent r.v.'s in a general case is not possible. The asymptotic c.d.f. of the maximum of normal, identically distributed but dependent r.v.'s for large M ($m \geq 7$) is derived in Sect. 9.3. in [18]. However, the r.v.'s $l(g_0)$ and $\max_{g \neq g_0} l(g)$ are still statistically dependent, and calculating the c.d.f. of their difference is not feasible in a general case. Hence, we assume them to be s.i. to carry out the calculations.

We may now use Theorem 1 in [8] to obtain that, for $g \neq g_0$, $l_r \sim \mathcal{N}(\mu_a, \sigma_a^2)$, where $\mu_a = F_W^{-1}(1 - 2^{-a}) = \sigma_W b$, $b = \Phi^{-1}(1 - 2^{-a})$ and

$\sigma_a^2 \approx \frac{2^{-(m+a)}}{f_W^2(\mu_a)}$. Basic approximations such as $b^2 \approx a$, can be then used in showing that $\sigma_a^2/\sigma_W^2 < 2^{-m}$ such that $\sigma_a^2 \ll \sigma_R^2$. Then

$$\begin{aligned} P_S &= \Pr(l(g_0) > l_r) \\ &= \Phi\left(\frac{\mu_R - \mu_a}{\sigma_R}\right) \\ &= \Phi\left(\frac{NC(p)/2 - \sqrt{NC(p)b}}{\sqrt{NC(p)}}\right), \end{aligned} \tag{13}$$

from which we can solve N as a function of a to be

$$N = \frac{4(\Phi^{-1}(P_S) + b)^2}{C(p)} \approx \frac{4(\Phi^{-1}(P_S) + \sqrt{a})^2}{C(p)}, \tag{14}$$

and by inversion, we get a as a function of N as desired. □

The experimental advantages for the different methods are studied in the next section. In the special case $a = m$ the data complexity is predicted to be

$$N_{\mathrm{LLR}} \approx \frac{16m}{C(p)}. \tag{15}$$

Hence, the dependence of N_{LLR} on m is linear. In practice $C(p) \approx C_{\min}(p)$, and hence, the LLR-method should be more efficient than the χ^2-method.

5 Experiments on 4-Round Serpent

We tested the different methods for multidimensional Alg. 1 described in this paper on 4-round Serpent by selecting linearly independent one-dimensional base approximations $u_i \cdot x \oplus w \cdot y$, $i = 1, \ldots, m$ to construct a linear approximation of the form (5) with $m = 7$ and $m = 10$. The output mask was $w = (0x00007000, 0x03000000, 0x00000000, 0x00000000)$ for all the approximations. The input masks u_i and the corresponding correlations are given in Table 1. We checked the assumption about closeness of the hypothetical distributions p^g and θ and saw that it holds as $|p_\eta^g - p_\eta^{g'}| < \frac{1}{150} p_\eta^g$, for all g, g' and $\eta \in V_m$. We also checked that $C_{\min}(p) \neq 0$ and actually, $C_{\min}(p) \approx C(p)$. The experiments showed that the empirical advantage when ranking the key classes was exactly the same for all methods. Hence, Figures 1 and 2 only depict the LLR-method. In particular, all methods were equally efficient in determining the correct key class. Equations (14) and (9) predict that the LLR-method should

Table 1. Input masks of base approximations and the corresponding correlations

	mask = (MSB, ..., LSB)	c_i
u_0	(0x70000000, 0x00000000, 0x00000000, 0x07000900)	-2^{-11}
u_1	(0x70000000, 0x00000000, 0x00000000, 0x07000B00)	2^{-11}
u_2	(0x70000000, 0x00000000, 0x00000000, 0x0B000900)	2^{-11}
u_3	(0xB0000000, 0x00000000, 0x00000000, 0x07000900)	2^{-11}
u_4	(0x70000000, 0x00000000, 0x00000000, 0x07000500)	2^{-12}
u_5	(0x70000000, 0x00000000, 0x00000000, 0x07000600)	2^{-12}
u_6	(0x70000000, 0x00000000, 0x00000000, 0x07000C00)	-2^{-12}
u_7	(0x70000000, 0x00000000, 0x00000000, 0x01000900)	-2^{-12}
u_8	(0x70000000, 0x00000000, 0x00000000, 0x0A000900)	2^{-12}
u_9	(0xB0000000, 0x00000000, 0x00000000, 0x03000B00)	-2^{-12}

be the most efficient: when m increases, the data requirement of χ^2-based tests increase exponentially with m whereas the increase is linear for the LLR-method. It is possible that the variance of the χ^2-method is not as large as the theory predicts, or the statistical dependence of r.v.'s $S(g)$ strengthens the χ^2-method more than expected. In [16] the χ^2-based method was proven to be weaker than the LLR-based method when used for multidimensional Alg. 2 attack. The statistical model of the relationship between the advantage a and data complexity N derived in this paper was tested in experiments. The results are given in Fig. 1 and 2 for $m = 7$ and $m = 10$, respectively. The empirical advantage using the LLR-method is depicted and it is compared with the theoretical advantage given by (14) with three different values of P_S. Possibly the statistical dependence of r.v.'s $l(g)$ explains why the experimental curves start higher than the theoretical curves. In both cases, we also show how much better the m-dimensional LLR-method is compared to the binomial method where the same set of m one-dimensional approximations and Matsui's Alg. 1 is used to determine each key class bit separately and independently. The m-bit key classes are then ranked according to the product of $|\hat{c}_i|$. This approach is similar than the method described in [3] where the key classes $g \in V_m$ are ranked using the sum of squares of the differences $(-1)^{g_i} c_i - \hat{c}_i$, and shown previously in [6] to be weaker than the multidimensional method.

We also observed that as m increases the data requirement decreases as long as the ratio $C_{\max}(p)/m$ increases. This gives an upper bound for m to be used in practice. In case of 4-round Serpent, the practical upper bound is around $m = 12$.

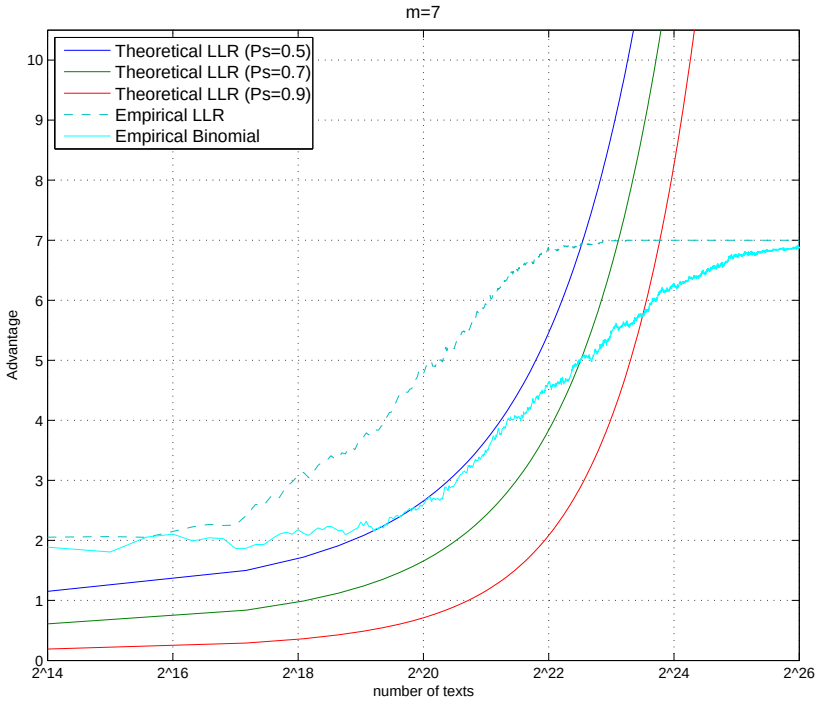

Fig. 1. The theoretical and empirical advantage as a function of data complexity using LLR-method with $m = 7$ for 4-round Serpent

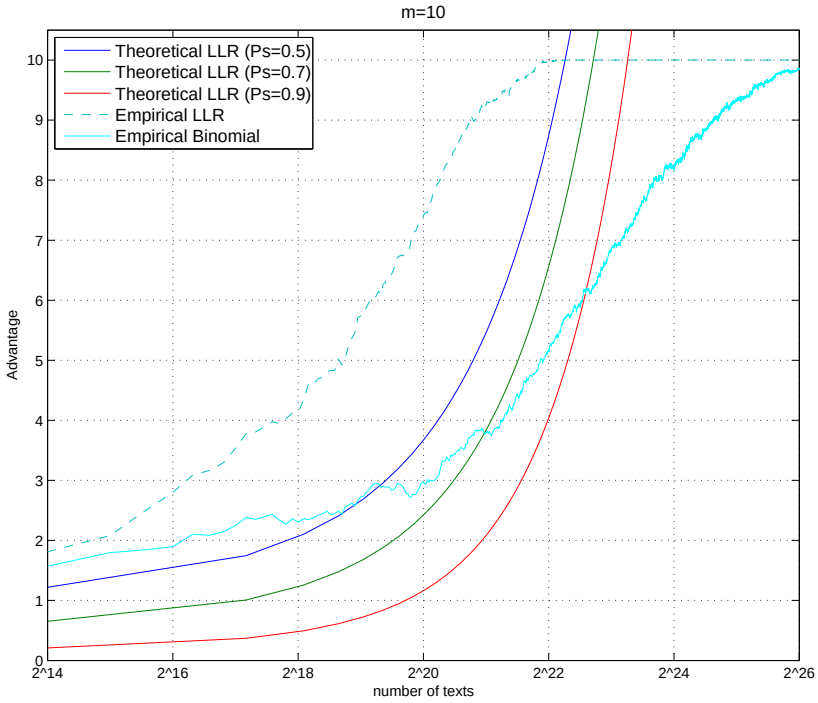

Fig. 2. The theoretical and empirical advantage as a function of data complexity using LLR-method with $m = 10$ for 4-round Serpent

6 Conclusions

In this paper three statistical methods for key recovery using multidimensional linear cryptanalysis were investigated. The problem of finding the right key among several key candidates can be interpreted as a goodness-of-fit problem or a multiple hypothesis testing problem. The goodness-of-fit based ranking statistics χ^2- and log-likelihood were shown to be equivalent and lead to the same χ^2-test. We solved the multiple hypothesis testing problem, i.e., distinguished one unknown p.d. from a given set of p.d.'s, using the LLR-statistic and observed the same correct key class as with the χ^2-method.

The methods were compared by using the advantage, modified to be used in multidimensional Alg. 1 from the original theory by Selçuk. The statistical model of the LLR and χ^2-method were tested in experiments on four-round Serpent. While the theory predicted a greater advantage for LLR than χ^2, they seemed to work equally well in the case of Serpent. For LLR, the empirical results were somewhat better than those predicted by the model, but still around the same range. It remains for future work to test the model on different ciphers, which also may show separation between the LLR-method and the χ^2-method. It is also an open question whether one can remove or better justify the assumption about statistical independence used in Theorem 2.

References

1. Matsui, M.: Linear Cryptanalysis Method for DES Cipher. In Helleseth, T., ed.: Advances in Cryptology – EUROCRYPT '93. Volume 765 of Lecture Notes in Computer Science., Berlin/Heidelberg, Springer (1994) 386–397
2. Burton S. Kaliski, J., Robshaw, M.J.B.: Linear Cryptanalysis Using Multiple Approximations. In Desmedt, Y.G., ed.: Advances in Cryptology – CRYPTO '94. Volume 839 of Lecture Notes in Computer Science., Berlin/Heidelberg, Springer (1994) 26–39
3. Biryukov, A., Cannière, C.D., Quisquater, M.: On Multiple Linear Approximations. In Franklin, M., ed.: Advances in Cryptology – CRYPTO '04. Volume 3152 of Lecture Notes in Computer Science., Berlin/Heidelberg, Springer (2004) 1–22
4. Baignères, T., Junod, P., Vaudenay, S.: How Far Can We Go Beyond Linear Cryptanalysis? In Lee, P.J., ed.: Advances in Cryptology – ASIACRYPT '04. Volume 3329 of Lecture Notes in Computer Science., Berlin/Heidelberg, Springer (2004) 432–450
5. Englund, H., Maximov, A.: Attack the Dragon. In Maitra, S., Madhavan, C.V., eds.: Progress in Cryptology – INDOCRYPT '05. Volume 3797 of Lecture Notes in Computer Science., Berlin/Heidelberg, Springer (2005) 130–142
6. Hermelin, M., Nyberg, K., Cho, J.Y.: Multidimensional Linear Cryptanalysis of Reduced Round Serpent. In Yi Mu, Willy Susilo, J.S., ed.: Informa-

tion Security and Privacy. Volume 5107 of Lecture Notes in Computer Science., Berlin/Heidelberg, Springer (2008) 203–215
7. Collard, B., Standaert, F.X., Quisquater, J.J.: Experiments on the Multiple Linear Cryptanalysis of Reduced Round Serpent. In Nyberg, K., ed.: Fast Software Encryption. Volume 5086 of Lecture Notes in Computer Science., Springer (2008) 382–397
8. Selçuk, A.A.: On probability of success in linear and differential cryptanalysis. Journal of Cryptology **21**(1) (January 2008) 131–147
9. Cover, T.M., Thomas, J.A.: Elements of Information Theory. 2nd edn. Wiley Series in Telecommunications and Signal Processing. Wiley-Interscience (2006)
10. Nyberg, K., Hermelin, M.: Multidimensional Walsh Transform and a Characterization of Bent Functions. In Tor Helleseth, P.V.K., Ytrehus, O., eds.: Proceedings of the 2007 IEEE Information Theory Workshop on Information Theory for Wireless Networks, IEEE (2007) 83–86
11. Hermelin, M., Nyberg, K.: Multidimensional Linear Distinguishing Attacks and Boolean Functions. In: Fourth International Workshop on Boolean Functions: Cryptography and Applications. (2008)
12. Cramèr, H., Wold, H.: Some theorems on distribution functions. J. London Math. Soc. **s1-11**(4) (Oct 1936) 290–295
13. Baignères, T., Vaudenay, S.: The Complexity of Distinguishing Distributions (Invited Talk). In Safavi-Naini, R., ed.: Information Theoretic Security. Volume 5155 of Lecture Notes in Computer Science., Berlin/Heidelberg, Springer (2008) 210–222
14. Drost, F., Kallenberg, W., D.S.Moore, J.Oosterhoff: Power Approximations to Multinomial Tests of Fit. Journal of the American Statistican Association **84**(405) (Mar 1989) 130–141
15. Cramér, H.: Mathematical Methods of Statistics. 7 edn. Princeton Mathematical Series. Princeton University Press (1957)
16. Hermelin, M., Cho, J.Y., Nyberg, K.: Multidimensional Extension of Matsui's Algorithm 2. In: Fast Software Encryption. Lecture Notes in Computer Science, Springer (2009) To appear.
17. Vaudenay, S.: An experiment on DES statistical cryptanalysis. In: CCS '96: Proceedings of the 3rd ACM conference on Computer and communications security, New York, NY, USA, ACM (1996) 139–147
18. David, H.A.: Order Statistics. 1 edn. A Wiley Publication in Applied Statistics. John Wiley & Sons, Inc. (1970)