

# Statistical Tests for the Multidimensional Extension of Matsui's Algorithm 1

Kaisa Nyberg

Department of Information and Computer Science  
Helsinki University of Technology

Dagstuhl, January 14, 2009



## Abstract

We consider extensions of Matsui's Algorithm 1 to using multiple linear approximations as statistical *goodness-of-fit* problems. The correct key should give the best fit with the observed data. We examine common statistical goodness-of-fit tests and show how to use them for finding among a number of estimated probability distributions the one which gives the best fit with the observed data. We test the statistical models using experimental results on a reduced-round Serpent.

The presentation is based on a paper accepted for presentation at Eurocrypt 2009 poster session and is joint work with Miia Hermelin and Joo Yeon Cho.



# Outline

Introduction

Basic Concepts

Goodness-of-Fit Tests for Key Recovery

Experiments

Conclusions



# History - Multiple linear approximations

- Matsui EUROCRYPT'93: Uses one biased approximate linear equation to recover one bit information of the key
- Robshaw and Kaliski CRYPTO'94: Recovers one bit of information of the key by using a number of statistical independent approximate linear equations
- Biryukov, et al., CRYPTO'04: Uses multiple approximate (statistically independent) linear equations, recovers multiple bits of information of the key, success measured using *gain*
- Collard, et al., FSE 2008: Experiments of Biryukov's algorithms on Serpent



# History - Probability distributions of multidimensional linear approximations

- Baignères, et al., ASIACRYPT'04: distinguishing probability distributions based on log-likelihood ratio LLR
- Maximov, 2006: algorithms for computing large probability distributions of multidimensional approximate linear equations
- Baignères, et al., ICITS'08, distinguishing one *known* probability distribution from a set of probability distributions using LLR
- Hermelin, et al., ACISP 2008, multidimensional Algorithm 1, using G-test and comparison with the algorithm of Biryukov, et al.
- Cho, et al., ICISC 2008, improved linear attack on 10-round Serpent



# Problems in previous solutions

- Assumption about *Statistical independence* of simultaneous linear approximations. Piling-up lemma: given two random variables  $X$  and  $Y$ , the three random variables  $X$ ,  $Y$  and  $X \oplus Y$  cannot be mutually statistically independent
- Algorithm 2 dependent on Algorithm 1
- Computing *large probability distributions*
- *Special-purpose* statistical tests and key ranking
- How to distinguish an unknown probability distribution?



# Our contributions

- Statistical goodness-of-fit approach
- Computing the probability distribution related to linear approximation of arbitrary dimension
- Using LLR (optimal distinguisher) for goodness-of-fit
- Application of standard order statistics and extension of the concept of *advantage* introduced by Selçuk, JoC'08
- Implementation of tests and running experiments using 4-round Serpent
- Complete separation of Algorithm 1 and Algorithm 2



# Matsui's Algorithm 1 as goodness-of-fit

Given a linear approximation  $\alpha \cdot X + \beta \cdot Y + \gamma \cdot K$  with correlation  $c$  which of the two distributions of values  $(0, 1)$

$$\left(\frac{1}{2}(1 \pm c), \frac{1}{2}(1 \mp c)\right)$$

fits better with the data  $\alpha \cdot \hat{X} + \beta \cdot \hat{Y}$ ?



# Outline

Introduction

**Basic Concepts**

Goodness-of-Fit Tests for Key Recovery

Experiments

Conclusions



# Correlation

- $f : V_n \rightarrow V$  Boolean function
- Correlation between  $f : V_n \rightarrow V$  and zero

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

- $f = (f_1, \dots, f_m) : V_n \rightarrow V_m$  an  $m$ -dimensional vector Boolean function
- For a generalised concept of correlation for vector Boolean functions, see our paper in ITW 2007 (may not be useful in practice)



# Probability distribution

- p.d. of  $Y = f(X)$ , where  $X$  is uniformly distributed, is called the p.d. of  $f : V_n \rightarrow V_m$
- probability distribution (p.d.)  $p = (p_0, \dots, p_M)$ ,  $M = 2^m - 1$

$$p_\eta = 2^{-m} \sum_{\alpha \in V_m} (-1)^{\alpha \cdot \eta} c_\alpha$$

where  $c_\alpha$  is the correlation of  $\alpha \cdot f(x)$

- Cramer-Wold type sampling technique: it suffices to use linear approximations  $\alpha$  with non-negligible correlations
- $\theta$  uniform distribution
- i.i.d. independent and identically distributed



# Kullback-Leibler distance

## Definition

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

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

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



# Capacity

## Definition

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

## Definition

The capacity between two p.d.'s  $p$  and  $q$  is defined by

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

If  $q$  is the uniform distribution, then  $C(p, q)$  will be denoted by  $C(p)$ . We call  $C(p)$  capacity (cf. Biryukov, et al.) and is identical to the notion of squared Euclidean imbalance of  $p$  used by Baignères, et al.



# Log-Likelihood Ratio (LLR)

- The i.i.d. data  $\hat{z}_1, \dots, \hat{z}_N, \hat{z}_i \in V_m$  is drawn from  $p$  or  $q$ ,  $p \neq q$
- LLR is the optimal distinguisher between the two p.d.'s (hypotheses)
- Empirical p.d.  $\hat{q} = (\hat{q}_0, \dots, \hat{q}_M)$
- We decide  $p$  if

$$\text{LLR}(\hat{q}, p, q) = \sum_{\eta=0}^M N\hat{q}_\eta \log \frac{p_\eta}{q_\eta} \geq \gamma$$

and otherwise we decide  $q$ , where  $\gamma$  is a threshold, usually taken equal to zero



# Asymptotic Behaviour of LLR

## Theorem

The r.v. LLR( $\hat{q}, p, q$ ) 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

$$\mu_0 = D(p\|q) \text{ and } \mu_1 = -D(q\|p)$$

$$\sigma_0^2 = \sum_{\eta=0}^M p_{\eta} \log^2 \frac{p_{\eta}}{q_{\eta}} - \mu_0^2 \text{ and } \sigma_1^2 = \sum_{\eta=0}^M q_{\eta} \log^2 \frac{p_{\eta}}{q_{\eta}} - \mu_1^2.$$



# Outline

Introduction

Basic Concepts

**Goodness-of-Fit Tests for Key Recovery**

Experiments

Conclusions



# Multidimensional Linear Approximation

- Multidimensional linear relation

$$Ux + Wy + g$$

with p.d.  $p$ , where  $g$  is  $m$  bits of unknown key information and  $U$  and  $W$   $m \times n$  binary matrices,  $x, y \in V_n$ .

- Search for the one-dimensional linear approximations with non-negligible correlation (using all known theory and heuristics and taking the linear hull effect into consideration if necessary) and compute an approximation of the expected probability distribution  $p$ .
- $2^m$  key classes  $g \in V_m$ , the problem is to determine the correct  $g_0$
- The data  $U\hat{x} + W\hat{y}$  is drawn from  $p^{g_0}$  (an unknown permutation of  $p$ )



# Statistical Setting

- Draw  $N$  plaintext-ciphertext pairs  $(\hat{x}_i, \hat{y}_i), i = 1, \dots, N$
- Data  $U\hat{x}_i + W\hat{y}_i, i = 1, \dots, N$ , with observed p.d.  $\hat{q}$
- Rank the keys  $g$  by measuring the goodness of fit of  $\hat{q}$  with  $p^g$
- The correct key  $g_0$  should give the best fit
- Estimate the efficiency of the ranking statistic



# The $\chi^2$ -method

- Ranking statistic

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



# The Log-Likelihood Method

- Ranking statistic based on the G-test:

$$G(g) = D(\hat{q} \| p^g)$$

- Log-likelihood (or G-test) is the same as  $\chi^2$ -test when  $p^g$ 's are close to each other



# The LLR-method

- Turn the LLR-test to a goodness-of-fit test
- Idea: For all  $g$ , test whether the data is drawn from  $p^g$  or  $\theta$
- Ranking statistic

$$\ell(g) = \text{LLR}(\hat{q}, p^g, \theta)$$

- Assumption:  $p^g$  is distinguishable from  $\theta$  with, and only with, the correct key  $g = g_0$



# The max-LLR-method

- A slight variation of the LLR-method
- Distinguish between  $g$  and  $\bar{g}$ , where  $\bar{g}$  is the key class farthest away from  $g$  in Kullback-Leibler distance

- Ranking statistic

$$\bar{\ell}(g) = \text{LLR}(\hat{q}, p^g, p^{\bar{g}})$$

- Basically the same as the LLR-method



# Data Complexity and Advantage

- Advantage  $a$  if the right key among the  $r = 2^{m-a}$  highest ranking keys
- $\ell(g_0) \sim \mathcal{N}(N\mu_R, N\sigma_R^2)$ , where  $\mu_R \approx C(p)/2$  and  $\sigma_R^2 \approx C(p)$ .
- $g \neq g_0 : \ell(g) \sim \mathcal{N}(\mu_W, \sigma_W^2)$ , where  $\mu_W = 0$  and  $\sigma_W^2 \approx C(p)$  (heuristic)
- Assume  $\ell(g), g \in V_m$  are statistically independent
- Order  $\ell(g), g \in V_m \Rightarrow r$ th order statistic  $\ell_r \sim \mathcal{N}(\mu_a, \sigma_a^2)$ , where  $\mu_a = \sigma_W b$ ,  $b = \Phi^{-1}(1 - 2^{-a})$  and  $\sigma_a^2 \ll \sigma_R^2$  (Thm. 1 in Selçuk, JoC'08)
- Connection between  $a$  and  $N$  given by 
$$P_S = \Pr(\ell(g_0) > \ell_r) = \Phi((\mu_R - \mu_a)/\sigma_R)$$



# Data Complexities

- Data complexity of ranking  $g_0$  on the top of the list:

$$N_{\chi^2} \approx \frac{\sqrt{2Mm}}{C_{\min}(p)} \quad N_{\text{LLR}} \approx \frac{4m}{C(p)}$$

- Data complexity as a function of advantage, success probability  $P_S$

$$N_{\chi^2}(a) = \frac{\sqrt{2M}(\Phi^{-1}(P_S) + \sqrt{a})}{C_{\min}(p)}$$

$$N_{\text{LLR}}(a) = \frac{4(\Phi^{-1}(P_S) + \sqrt{a})^2}{C(p)}$$

- $C_{\min}(p) = \min_{g, g \neq 0} C(p, p^g) = \min_{g, g \neq h} C(p^h, p^g)$  (by symmetry) and  $C_{\min}(p) \approx C(p)$  (cipher property)



# Outline

Introduction

Basic Concepts

Goodness-of-Fit Tests for Key Recovery

**Experiments**

Conclusions

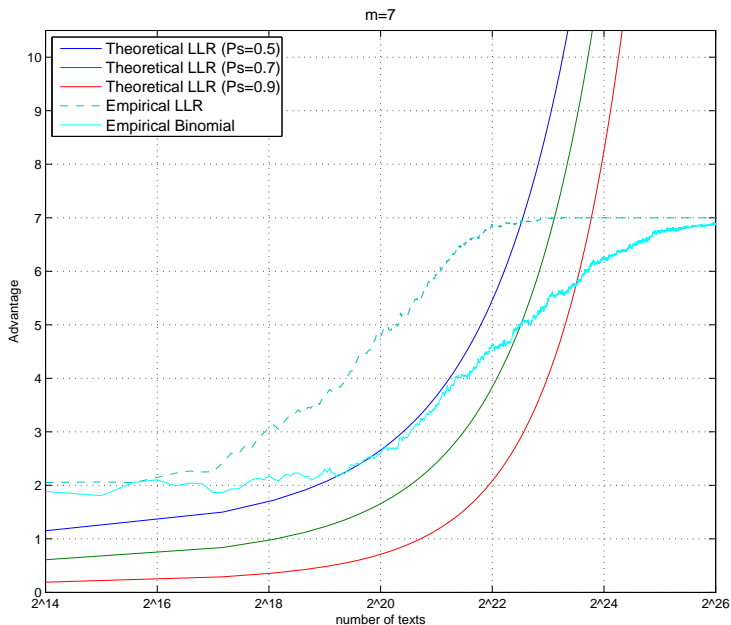


# Experimental Results

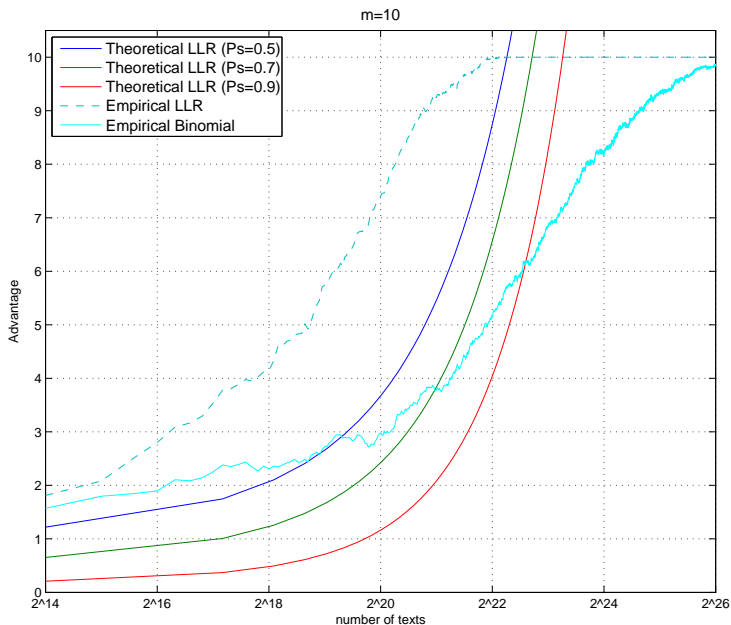
- Advantage increases with  $m$  until  $m = 12$  (for 4-round Serpent)
- In practice,  $\chi^2$  seems to have the same advantage as LLR
- Both  $\chi^2$ -method and Log-likelihood method (G-test) and LLR and max-LLR equally effective



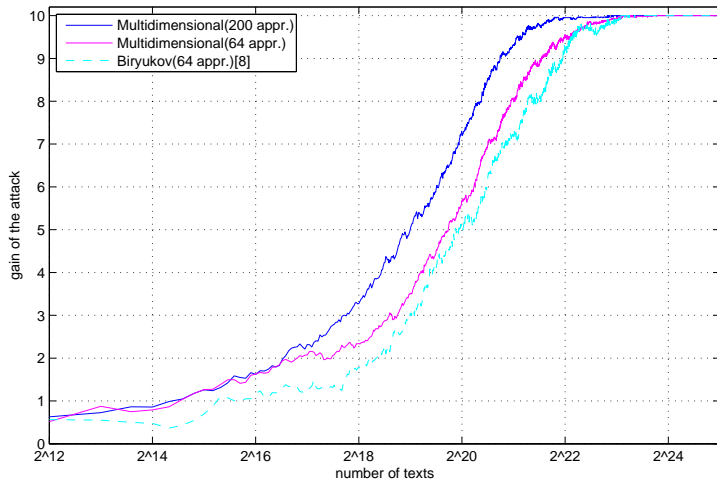
# Theoretical and Empirical Advantage for LLR with $m = 7$



# Theoretical and Empirical Advantage for LLR, $m = 10$



# Observed Gain for G-method and Birykov's method\*



\* see Hermelin, et al., ACISP 2008

# Outline

Introduction

Basic Concepts

Goodness-of-Fit Tests for Key Recovery

Experiments

**Conclusions**



# Results

- Multidimensional extensions of Matsui's Algorithm 1
- No assumption about statistical independence of the approximations
- No need to compute huge distributions, just one-dimensional correlations using the best state-of-the-art techniques
- Multidimensional Algorithm 1 gives higher advantage than one-dimensional or Biryukov's multiple linear cryptanalysis
- Order statistics (Selçuk) for measuring success of key ranking
- Estimates for data complexities calculated
- Different methods and dimensions can be easily compared



# Related Work and Open Questions

- In theory, the LLR-method should perform better than  $\chi^2$ , which was not observed in the experiments. Experiments on other ciphers needed.
- Possibly some ciphers may show difference between LLR and max-LLR?
- Can assumption about statistical independence of  $\ell(g)$ ,  $g \in V_m$ , be removed?
- The same tools can be used to construct a statistical model using  $\chi^2$  and LLR for a multidimensional Algorithm 2 (last round key recovery)
- Extensions to nonlinear cryptanalysis (cf. Baignères, et al.2004)?



# Thank you!

