Aalto University
School of Science

Links Between Differential and Linear Cryptanalysis and Boolean Functions

Kaisa Nyberg

Aalto University School of Science

kaisa.nyberg@aalto.fi

September 3, 2014

BFA, *Bergen*

Outline

Introduction

Historical Notes
- CRADIC
- Matsui's Algorithms
- Linear Hull

Links Between Statistical Attacks
- Newer Statistical Cryptanalysis
- Recent Links
- Multidimensional Linear and Truncated Differential Properties
- Index of Coincidence

Computing Differential Probabilities using Linear Correlations

Distinguishing Distributions

Conclusions

Acknowledgements

Thanks to Céline, Risto, Mohsin, Gregor and Andrey for their contributions to the developments of ideas described in this talk.

Thanks to Céline for letting me use some of her slides.

Outline

Introduction

Historical Notes
- CRADIC
- Matsui's Algorithms
- Linear Hull

Links Between Statistical Attacks
- Newer Statistical Cryptanalysis
- Recent Links
- Multidimensional Linear and Truncated Differential Properties
- Index of Coincidence

Computing Differential Probabilities using Linear Correlations

Distinguishing Distributions

Conclusions

Introduction

- Study of APN and PN functions is motivated by conventional differential cryptanalysis.
- Other types of attacks may (?) require stronger countermeasures.
- We will survey recent results on links between statistical attacks on block ciphers.
- The statistical models of distinguishers will be discussed.
- Some bent functions are more vulnerable than some others.

Outline

Introduction

Historical Notes
- CRADIC
- Matsui's Algorithms
- Linear Hull

Links Between Statistical Attacks
- Newer Statistical Cryptanalysis
- Recent Links
- Multidimensional Linear and Truncated Differential Properties
- Index of Coincidence

Computing Differential Probabilities using Linear Correlations

Distinguishing Distributions

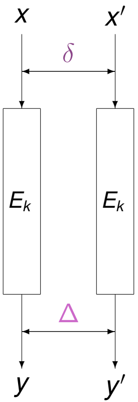
Conclusions

Brief History

- Biham-Shamir Crypto1990: Differential Cryptanalysis
- Lai, Massey, and Murphy EC1990: Markov Ciphers and Differential cryptanalysis
- K.N. EC1991: Perfect Nonlinear S-boxes

Differential Cryptanalysis

Difference between plaintext and ciphertext pairs



Input difference δ

Output Difference Δ

Differential Probability:

$$\Pr[\delta \xrightarrow{E_k} \Delta] \quad = \quad \Pr[\, E_k(x) \oplus E_k(x \;\oplus\; \delta) = \Delta]$$

Markov cipher $E_k = f_k \circ g_k$

$$\Pr[\delta \xrightarrow{E_k} \Delta] = \sum_{\gamma} \Pr[\delta \xrightarrow{g_k} \gamma]\Pr[\gamma \xrightarrow{f_k} \Delta]$$

Provable Security Theorem

with L. Knudsen, Crypto 1992 Rump Session, J Crypt 1995

Theorem ($\mathcal{KN}$-Theorem) *It is assumed that in a DES-like cipher with* $F : \mathbb{F}_2^m \to \mathbb{F}_2^n$ *the round keys are independent and uniformly random. Then the probability of an s-round differential,* $s \geq 4$, *is less than or equal to* $2p_{max}^2$.

Here

$$p_{\mathrm{max}} = \max_{\beta} \max_{\alpha \neq 0} \Pr[\alpha \xrightarrow{F} \beta]$$

If F bijective, then the claim of Theorem holds for $s \geq 3$.

Later Aoki showed that the constant 2 can be removed.

$$\text{Minimize } p_{\mathrm{max}} \quad \Leftrightarrow \quad F \text{ APN}$$

CRADIC

Cipher Resistant Against Differential Cryptanalysis

aka $\mathcal{KN}$-Cipher

6-round Feistel cipher with round function $f : \mathbb{F}_2^{32} \to \mathbb{F}_2^{32}$ based on the power three operation in $\mathbb{F}_2^{33}$

No key schedule, 198-bit key

Jakobsen & Knudsen FSE1997 break $\mathcal{KN}$-Cipher

- with 512 chosen plaintexts and 2^{41} running time,
- or with 32 chosen plaintexts and 2^{70} running time
- using *higher order differential cryptanalysis*

Round-function based on the inverse mapping not any more resistant.

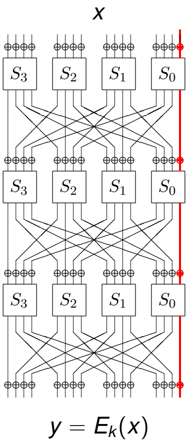
This approach was then abandonded

... but resumed again recently, see [Boura-Canteaut IEEE Trans. IT 2013].

Linear Cryptanalysis

- M. Matsui (EC1993 Bergen) Linear Cryptanalysis

Linear Cryptanalysis



Linear approximation with mask vector (u, τ, w) is a relation

$$u \cdot x + \tau \cdot k + w \cdot E_k(x)$$

Input mask u
Key mask τ
Output mask w

Bias:
$\varepsilon = 2^{-n} \#\{x \in \mathbb{F}_2^n | \; u \cdot x + \tau \cdot k + w \cdot y = 0\} - \frac{1}{2}$

Correlation: $\mathrm{cor}_x(u, w) = 2\varepsilon$

Matsui's Algorithms

Matsui's Algorithm 1 is a statistical cryptanalysis method for finding one bit of the key k based on the observed correlation of a linear approximation

$$u \cdot x + w \cdot E_k(x)$$

Matsui's Algorithm 2 is a statistical cryptanalysis method for finding a part of the last round key for a block cipher based on distinguishing cipher data from more random data using observed correlations of a linear approximation

$$u \cdot x + w \cdot E'_k(x)$$

Linear Hull

Or What is the Equivalent of Differential in Linear Cryptanalysis?

Correlation for Iterated Block Cipher

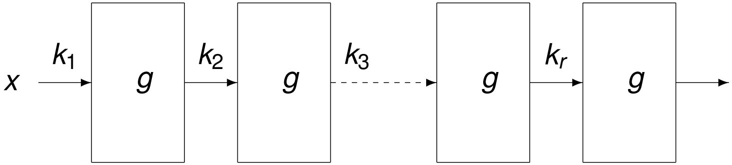
We focus on key alternating iterated block ciphers. Let $(k_1, k_2, \ldots, k_r)$ be the extended key with the round keys k_i derived from k and assume that E_k has the following structure

$$E_k(x) = g(\ldots g(g(g(x + k_1) + k_2) \ldots) + k_r).$$

Then [Daemen FSE1994]

$$\mathrm{cor}_x(u \cdot x + w \cdot E_k(x)) = \sum_{\tau} (-1)^{\tau \cdot k} \prod_{i=1}^{r} \mathrm{cor}_x(\tau_i \cdot x + \tau_{i+1} \cdot g(x)),$$

where $\tau = (\tau_1, \tau_2, \ldots, \tau_r)$, $\tau_1 = u$ and $\tau_{r+1} = w$.



Estimating Correlation

- Assumption for Matsui's algorithms: magnitudes of correlations about the same for all keys.
- In general, correlation magnitude varies with the key except when there is a single dominating trail with key mask τ and trail correlation

$$\begin{aligned}\tilde{c}(u, \tau, w) &= \prod_{i=1}^{r} \mathrm{cor}_x(\tau_i \cdot x + \tau_{i+1} \cdot g(x)) \\ &= \mathrm{Avg}_k \mathrm{cor}(u \cdot x + \tau \cdot k + w \cdot y)\end{aligned}$$

The Linear Hull Theorem

By Jensen's inequality

$$\mathrm{Avg}_k \mathrm{cor}_x(u \cdot x + \tau \cdot k + w \cdot E_k(x))^2 \geq \tilde{c}(u, \tau, w)^2,$$

for all τ, and in general a strict inequality holds. More accurately, the following theorem holds

The Linear Hull Theorem [K.N. EC1994, K.N. DAM 2001] If the round keys of a block cipher E_k are uniformly distributed, then

$$\mathrm{Avg}_k \mathrm{cor}_x(u \cdot x + w \cdot E_k(x))^2 = \sum_{\tau} \tilde{c}(u, \tau, w)^2$$

- Squared correlations of linear hull correspond to probabilities of differentials.
- An analogue of the $\mathcal{KN}$-Theorem for linear cryptanalysis is obtained.

More Generally: The Fundamental Theorem

$$f : \mathbb{F}_2^n \times \mathbb{F}_2^\ell \to \mathbb{F}_2, \quad \hat{f}(u,v) = \sum_{x \in \mathbb{F}_2^r,\, z \in \mathbb{F}_2^s} (-1)^{u \cdot x + v \cdot z + f(x,z)}$$

$$f_z(x) = f(x,z),\ f_z : \mathbb{F}_2^n \to \mathbb{F}_2,\ z \in \mathbb{F}_2^\ell$$

Theorem [K.N. EC1994] *For all* $u \in \mathbb{F}_2^n$

$$2^\ell \sum_{z \in \mathbb{F}_2^\ell} \widehat{f_z}(u)^2 \quad = \quad \sum_{v \in \mathbb{F}_2^\ell} \hat{f}(u,v)^2, \quad \textit{or equivalently,}$$

$$2^{-\ell} \sum_{z \in \mathbb{F}_2^\ell} \mathrm{cor}_x(u \cdot x + f_z(x))^2 \quad = \quad \sum_{v \in \mathbb{F}_2^\ell} \mathrm{cor}_{x,z}(u \cdot x + v \cdot z + f(x,z))^2.$$

A. Canteaut, C. Carlet, P. Charpin, C. Fontaine. On cryptographic properties of the cosets of r(1, m). IEEE Trans. IT 47(4), 14941513 (2001)

N. Linial, Y. Mansour and N. Nisan. Constant depth circuits, Fourier transform, and learnability. Journal of the ACM 40 (3), 607-620 (1993).

Estimation of Correlation

Methods to catch significant trails:

- Dominant trails: By hand
- Branch and Bound algorithm
- Transition matrices

Computing an Estimate of Correlation

$$\mathrm{Avg}_k \mathrm{cor}_x(u \cdot x + w \cdot E_k(x))^2 = \sum_{\tau_2,\ldots,\tau_r} \prod_{i=1}^{r} c_z(\tau_i \cdot z + \tau_{i+1} \cdot g(z))^2$$

$$= \sum_{\tau_r} c_z(\tau_r \cdot z + w \cdot g(z))^2 \sum_{\tau_{r-1}} c_z(\tau_{r-1} \cdot z + \tau_r \cdot g(z))^2$$

$$\cdots\cdots \quad \sum_{\tau_3} c_z(\tau_3 \cdot z + \tau_4 \cdot g(z))^2$$

$$\sum_{\tau_2} c_z(\tau_2 \cdot z + \tau_3 \cdot g(z))^2 c_z(u \cdot z + \tau_2 \cdot g(z))^2$$

- This expression gives an iterative algorithm: start from the bottom line to compute for each τ_3 the value on the last line.
- Can be made feasible by restricting to τ with low Hamming weight and keeping only the largest values from each iteration.
- Restrictions on τ will lead to a lower bound, which is still much larger than any single $\tilde{c}(u, \tau, w)^2$.

Outline

Introduction

Historical Notes

- CRADIC
- Matsui's Algorithms
- Linear Hull

Links Between Statistical Attacks

- Newer Statistical Cryptanalysis
- Recent Links
- Multidimensional Linear and Truncated Differential Properties
- Index of Coincidence

Computing Differential Probabilities using Linear Correlations

Distinguishing Distributions

Conclusions

Statistical Attacks

LINEAR CONTEXT

DIFFERENTIAL CONTEXT

Linear Cryptanalysis [Tardy, Gilbert 92] [Matsui 93]

Differential Cryptanalysis [Biham, Shamir 90]

Differential-Linear Cryptanalysis [Langford, Hellman 94]

Truncated Differential Cryptanalysis [Knudsen 94]

Higher Order Differential cryptanalysis [Lai 94] [Knudsen 94]

Square Attack, Integral $\cdots$ [Daemen, Rijmen, Knudsen 97]

Statistical Saturation [Collard, Standaert 09]

Zero Correlation [Bogdanov, Rijmen 11]

Impossible Differential Cryptanalysis [Knudsen 98]

Multiple Linear Cryptanalysis [Biryukov, de Cannière, Quisquater 04]

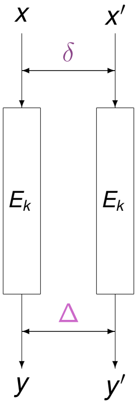
Multiple Differential Cryptanalysis [Albrecht, Leander 12] [Blondeau, Gérard, Nyberg 12]

Multidimensional Linear Cryptanalysis [Cho, Hermelin, Nyberg 08]

. . . . . .

. . . . . .

Truncated Differential Cryptanalysis



Input difference δ
Output difference Δ

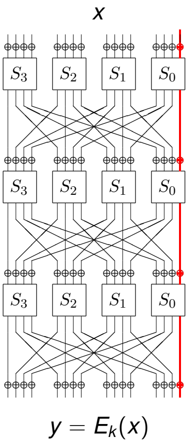
Set of input differences: $\delta \in C$

Set of output differences: $\Delta \in D$

Probability of truncated differential

$$\frac{1}{|C|} \sum_{\delta \in C} \sum_{\Delta \in D} P[\delta \xrightarrow{F} \Delta]$$

Multidimensional Linear Cryptanalysis



Multidimensional linear approximation:

Set of masks $(u, w) \in U \times W$

Capacity: $\sum_{u \in U} \sum_{w \in W} \mathrm{cor}_x(u \cdot x + w \cdot y)^2 - 1$

Recent Links

[Leander EC2011] :

Statistical Saturation $\Leftrightarrow$ Multidimensional Linear

[Bogdanov *et al* AsiaCrypt2012] :

Integral $\Leftrightarrow$ Zero Correlation Linear

Proofs follow from the Fundamental Theorem [N 1994]

[C.Blondeau-K.N. EC2013] :

Zero Correlation Linear $\Leftrightarrow$ Impossible Differential

[C.Blondeau-K.N. EC2014] :

Multidimensional Linear $\Leftrightarrow$ Truncated Differential

Proofs follow from the Chabaud-Vaudenay Link EC1994

Chabaud-Vaudenay Link

[Chabaud-Vaudenay EC1994]

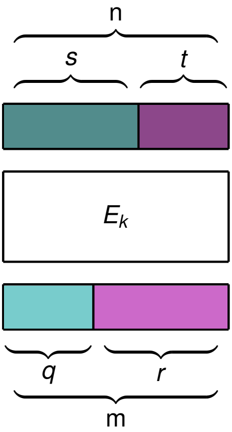
$$F : \mathbb{F}_2^n \to \mathbb{F}_2^m$$

Link between differential and linear cryptanalysis

$$\Pr[\delta \xrightarrow{F} \Delta] = 2^{-m} \sum_{u \in \mathbb{F}_2^n} \sum_{w \in \mathbb{F}_2^m} (-1)^{u \cdot \delta + w \cdot \Delta} \mathrm{cor}(u \cdot x + w \cdot F(x))^2$$

- Used for theory (almost bent $\Rightarrow$ APN)
- Not really used for cryptanalysis

Splitting the Spaces



Focus on the left side:

multidimensional linear context

- all non-zero input and output masks

truncated differential context

- zero input and output differences

Omit the right side:

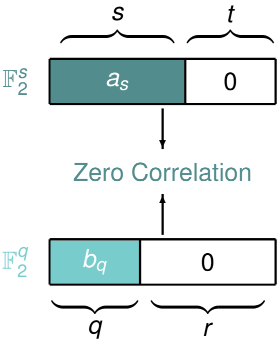
multidimensional linear context

- zero input and output masks

truncated differential context

- all input and output differences

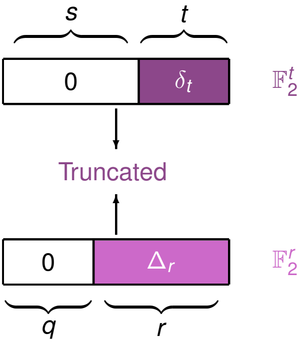
Zero Correlation Linear



Zero Correlation Linear :

$$\mathrm{cor}_x((a_s, 0), (b_q, 0)) = 0$$
$$\text{for all } (a_s, b_q) \in \mathbb{F}_2^s \times \mathbb{F}_2^q \setminus \{(0,0)\}$$

Impossible Differential



Truncated Differential:

$$\sum_{\delta_t \in \mathbb{F}_2^t} \sum_{\Delta_r \in \mathbb{F}_2^r} \Pr\left[(0, \delta_t) \to (0, \Delta_r)\right] = 2^{t-q}$$

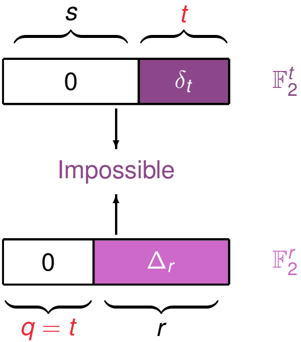
If t=q and $\delta_t \neq 0$

Impossible Differential:

$$\Pr\left[(0, \delta_t) \to (0, \Delta_r)\right] = 0$$

for all $(\delta_t, \Delta_r) \in \mathbb{F}_2^t \times \mathbb{F}_2^r \setminus \{(0, 0)\}$

Zero Correlation Linear and Impossible Differential



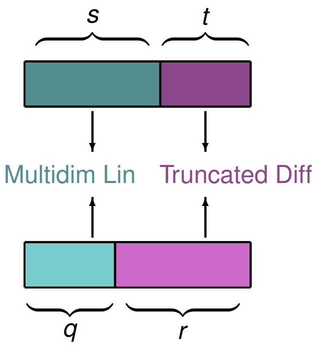
If $t = q$

Zero Correlation Linear Distinguisher

is equivalent to

Impossible Differential Distinguisher

Multidimensional Linear and Truncated Differential



Multidimensional Linear Distinguisher

is equivalent to

Truncated Differential Distinguisher

Statistical Saturation Attack

For fixed $x_s \in \mathbb{F}_2^s$ denote by $C(x_s)$ the capacity of the distribution of y_q.

Chosen plaintext sampling for evaluation of the uniformity of the distribution of y_q, for a fixed x_s.

Focus on Distributions

Distribution of values $(x_s, y_q) \in \mathbb{F}_2^s \times \mathbb{F}_2^q$

- Multidimensional Linear has

$$\Pr(x_s, y_q) = 2^{-(s+q)} \sum_{u_s, w_q} (-1)^{u_s \cdot x_s + w_q \cdot y_q} \mathrm{cor}((u_s, 0) \cdot x + (w_q, 0) \cdot y)$$

- Truncated Differential probability

$$P = 2^{-t} \sum_{\delta_t \in \mathbb{F}_2^t} \sum_{\Delta_r \in \mathbb{F}_2^r} \Pr\left[(0, \delta_t) \to (0, \Delta_r)\right]$$

These are just different approaches to sampling of the cipher data and measuring the nonuniformity of the same distribution of $(x_s, y_q) \in \mathbb{F}_2^s \times \mathbb{F}_2^q$.

The Mathematical Link

The capacity C of the multidimensional linear distribution is defined as

$$C = \sum_{(u_s, w_q) \neq 0} \mathrm{cor}((u_s, 0) \cdot x + (w_q, 0) \cdot y)^2.$$

We obtain the link [BN 2014]

$$P = 2^{-q}(C + 1),$$

or

$$P = 2^s \sum_{x_s, y_q} \Pr(x_s, y_q)^2.$$

Coincidences of (x_s, y_q)

Index of Coincidence

When solving the period of the key of a Vigènere cipher we count coincidences in letters to evaluate the nonuniformity of the distribution of the alphabet.[Friedman 1922]

Index of Coincidence is a method of ciphertext only differential cryptanalysis, but the idea generalizes to plaintext-ciphertext pairs:

$$\Pr\left[(0, \delta_t) \to (0, \Delta_r)\right] = \Pr(\, x_s \leftrightarrow x'_s,\, y_q = y'_q\,) \leftrightarrow \sum \Pr(x_s, y_q)^2$$

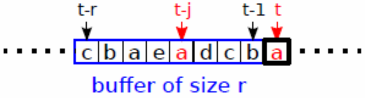
So we can evaluate the χ^2 statistic of the distribution of (x_s, y_q) using truncated differential frequences.

Differential (collision) approach is used in distribution context.

Efficient Online Entropy Estimator (Röck 2011)

- **Number of comparisons** before finding **last occurrence** of **current element**:

$$\ell_t^r = \begin{cases} r & \text{if } x_t \neq x_{t-j}, 1 \leq j \leq r, \\ \min\{0 \leq j \leq r-1 : x_{t-1-t} = x_t\} & \text{otherwise.} \end{cases}$$



- **Estimator:** $\hat{H}^r_{\text{pv}}(\mathbf{x}_{[t-r,t]}) = \frac{1}{\ln(2)} \sum_{j=1}^{\ell_t^r} \frac{1}{j}$

Outline

Introduction

Historical Notes
- CRADIC
- Matsui's Algorithms
- Linear Hull

Links Between Statistical Attacks
- Newer Statistical Cryptanalysis
- Recent Links
- Multidimensional Linear and Truncated Differential Properties
- Index of Coincidence

Computing Differential Probabilities using Linear Correlations

Distinguishing Distributions

Conclusions

Using Correlations to Compute Differential Probabilitites

- For some ciphers like PRESENT
  - it is easier to estimate linear correlations than differential probabilities
  - Single-bit linear trails are dominant
  - Computation of correlations using transition matrices as for instance in [Cho CT-RSA2010]

- Use the Chabaud-Vaudenay Link to compute differential probabilities using linear correlations [C.Blondeau-K.N. EC2013]
- Use the linear property of the cipher to mount a differential type of attack [C.Blondeau-K.N. EC2014]

Outline

Introduction

Historical Notes
- CRADIC
- Matsui's Algorithms
- Linear Hull

Links Between Statistical Attacks
- Newer Statistical Cryptanalysis
- Recent Links
- Multidimensional Linear and Truncated Differential Properties
- Index of Coincidence

Computing Differential Probabilities using Linear Correlations

Distinguishing Distributions

Conclusions

Distinguishing Test

- Distinguishing probability distributions over a large set of values of size M
  - Uniform distribution
  - Non-uniform distribution p with known capacity

$$C(p) = M \sum_{\eta=1}^{M} (p(\eta) - \frac{1}{M})^2.$$

- Problem. Determine the data complexity estimates of the χ^2 distinguisher.
- Solution. Use statistic

$$T = NM \sum_{\eta=1}^{M} (q(\eta) - \frac{1}{M})^2,$$

where q is the distribution obtained from the data of amount N.

- Need to determine the probability distribution of T in both cases.

χ^2 Distributions of T

- If q is drawn from uniform distribution, then

$$T = T_0 = \sum_{\eta=1}^{M} \frac{(Nq(\eta) - N/M)^2}{N/M} \sim \chi^2_{M-1}.$$

- If q is drawn from nonuniform distribution p, then

$$T = T_1 = \sum_{\eta=1}^{M} \frac{(Nq(\eta) - N/M)^2}{N/M} \sim \chi^2_{M-1}(\delta),$$

where

$$\delta = \sum_{\eta=1}^{M} \frac{(Np(\eta) - N/M)^2}{N/M} = NC(p).$$

- Denote $C(p) = C$.

Normal Approximations of Distributions of T

- If q is drawn from uniform distribution, then

$$T = T_0 \sim \mathcal{N}(M, 2M).$$

- If q is drawn from nonuniform distribution with capacity C, then

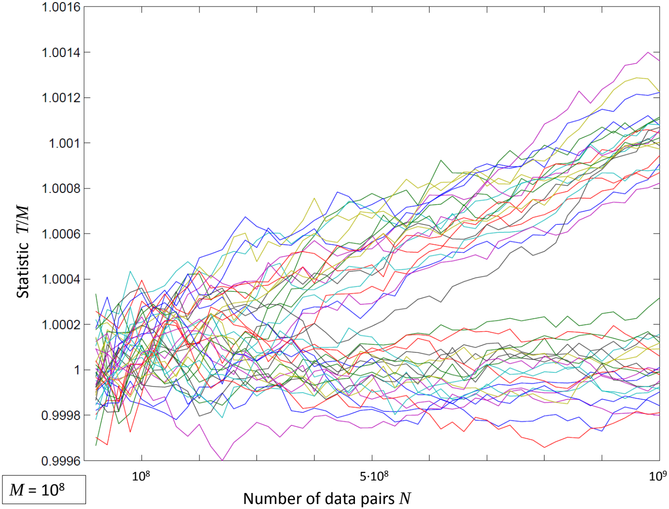
$$T = T_1 \sim \mathcal{N}(M + NC, 2(M + 2NC)).$$

- Data complexity

$$N \geq \frac{\sqrt{M}}{C} \phi.$$

For typical error probabilities, we take $\phi = 4$.

Experiment on a Large Distribution



Zero-Correlation Distribution

- With full code book of data the distribution of (x_s, y_q) should be exactly uniform
- We must do sampling without replacement
- Using hypergeometric distribution, with data size N and distribution size $M + 1 = 2^{s+q}$, we get

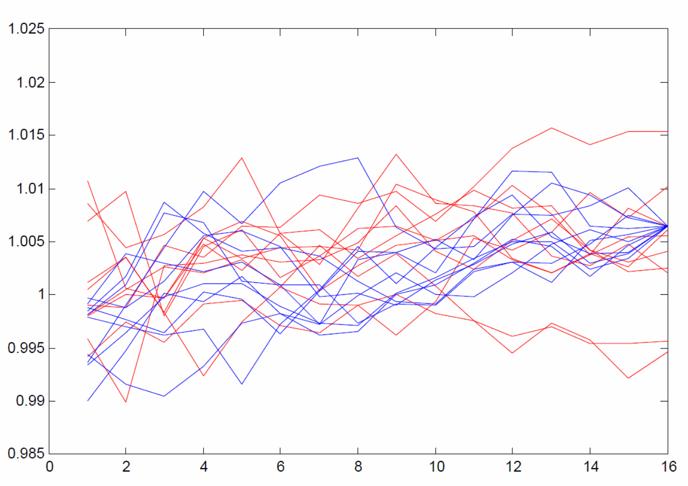
$$\mathrm{Exp}(T) = M\frac{2^n - N}{2^n - 1} \quad \text{and} \quad \mathrm{Var}(T) = 2M\left(\frac{2^n - N}{2^n - 1}\right)^2 .$$

- Using normal approximation, we get data-complexity

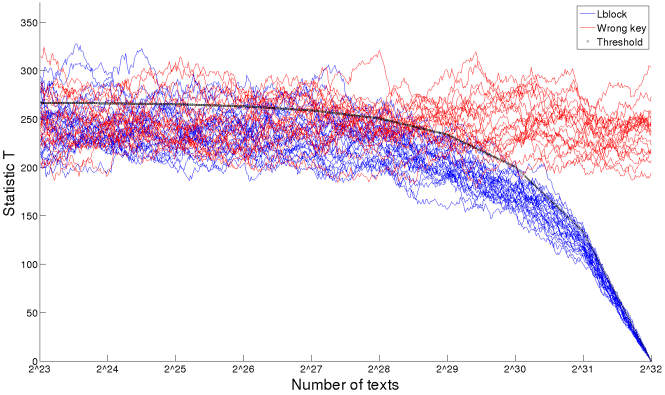
$$N \approx 2^{n - \frac{s+q}{2}}\phi$$

Data-sampling without replacement would be more correct also for ordinary linear cryptanalysis, in particular, when close to full code book.

Sampling Without Replacement



Zero-Correlations on LBlock (Small Variant)



Capacities of Bent Functions

- Capacities in some special multidimensional linear setting for certain vectorial Boolean functions were determined in [M.Hermelin-K.N. BFCA2008, M.Hermelin-K.N. CCDS2012].
- Capacity of multidimensional linear approximation of bent function $f : \mathbb{F}_2^n \to \mathbb{F}_2^m$

$$\begin{aligned} C &= \sum_{(a_s,b_q)\neq 0} \mathrm{cor}(a_s \cdot x, b_q \cdot f(x))^2 \\ &= 2^s(2^q - 1)2^{-n}, \end{aligned}$$

  where $0 \leq s \leq n$ and $0 < q \leq m$.
- A bent function can be distinguished from a random function using data size

$$N = 2^{n - \frac{s+q}{2}}\phi$$

Statistical Saturation Distinguisher of Maiorana-McFarland

- Consider Maiorana-McFarland function $f = (f_1, ...f_m)$
$$f_i(x_s, x_t) = A^i(x_s) \cdot x_t + g_i(x_s)$$
where $s = t = q = m = n/2$ [K.N. EC1991].
- For fixed $x_s \neq 0$, $f(x_s, x_t)$ is a linear function, and for $x_s = 0$ it is constant.
- The capacity of the multidimensional distribution of this bent function is equal to $2^{-s}(2^s - 1)$ and the multidimensional linear attack has data complexity $N = 2^s\phi$
- $C(x_s) = 0$, for $x_s \neq 0$, and $C(0) = 2^s - 1$.
- Pick random x_s. It takes a few data to verify if $f(x_s, x_t)$ is constant. If it is not constant, the distribution of $f(x_s, x_t)$ is uniform as the function is bijective. It takes about
$$N = 2^{s+2-\frac{s}{2}} = 2^{\frac{s}{2}}\phi$$
data to distinguish it from random.

Outline

Introduction

Historical Notes

CRADIC

Matsui's Algorithms

Linear Hull

Links Between Statistical Attacks

Newer Statistical Cryptanalysis

Recent Links

Multidimensional Linear and Truncated Differential Properties

Index of Coincidence

Computing Differential Probabilities using Linear Correlations

Distinguishing Distributions

Conclusions

Conclusions

- Since the invention of linear and differential cryptanalysis researchers have examined their relationships and discovered analogies between their properties.
- Linear hull vs. differential.
- We extended the Chabaud-Vaudenay link to truncated differentials and multidimensional linear approximations.
- Differential attacks can be seen as extensions of linear cryptanalysis.
- Distribution of cipher data values and χ^2 statistic offer a sufficiently general setting to handle both differential and linear statistical cryptanalysis.
- Chosen plaintext data sampling can be used for linear cryptanalysis and, *vice versa*, known plaintext data sampling for differential cryptanalysis.
- Chosen plaintext attack on the vectorial Maiorana-McFarland bent function.