Aalto University
School of Science

Improving Accuracy of Statistical Cryptanalysis

Kaisa Nyberg

Aalto University School of Science
kaisa.nyberg@aalto.fi

Second Lightweight Crypto Day
Haifa
March 12, 2015

Outline

Outline

Cryptanalysis at the design phase

- All known cryptanalysis as applicable to the new design
- Add reasonable security margin
  For example, add more rounds to iterated block ciphers

Lightweight ciphers

- Optimized with respect to platform specific performance requirements
- Secure: Resistant against all known cryptanalytic attacks

$\Rightarrow$ Minimize security margins

$\Rightarrow$ Must acquire better estimates for complexity of cryptanalytic attacks

Statistical cryptanalysis of iterated block ciphers

Differential cryptanalysis[1]

- chosen-plaintext attack
- data complexity upperbounded based on the best differential probability
- historical note: ciphertext-only differential cryptanalysis using index of coincidence [Friedman 1922]

Linear cryptanalysis[1]

- known-plaintext attack
- data complexity upperbounded based on the largest magnitude of linear correlation (or bias, "linear probability)

[1] Obvious references omitted

Differential probability of cipher E

- α difference in plaintext x
- β difference in ciphertext $E(x)$
- $\Pr[\alpha \to \beta]$ probability that given α we observe ciphertext difference β

$$E(x + \alpha) + E(x) = \beta$$

$$\Pr[\alpha \to \beta] = \sum_{\chi} p_{\chi}(\alpha \to \beta)$$

where $\chi = (\alpha = \chi_0, \chi_1, \ldots, \chi_r = \beta)$ is a differential characteristic χ from α to β and (assuming round independence)

$$p_{\chi}(\alpha \to \beta) = \prod_{i=1}^{r} \Pr[\chi_{i-1} \to \chi_i]$$

Linear correlation of cipher E

- u linear mask on plaintext x
- v linear mask on ciphertext $E(x)$
- $\mathrm{cor}[u \to v]$ correlation between $u \cdot x$ and $v \cdot E(x)$

$$\mathrm{cor}[u \to v] = \sum_{\theta} c_\theta(u, v)$$

where $\theta = (u = \theta_0, \theta_1, \ldots, \theta_r = v)$ is a linear characteristic from u to v and

$$c_\theta(u, v) = \prod_{i=1}^{r} \mathrm{cor}[\theta_{i-1} \to \theta_i]$$

Assuming round independence

$$\mathrm{cor}^2[u \to v] = \sum_{\theta} c_\theta^2(u, v) = \sum_{\theta} \prod_{i=1}^{r} \mathrm{cor}^2[\theta_{i-1} \to \theta_i]$$

Outline

Computing good estimates

Computation of exact values takes multiplications of huge $2^n \times 2^n$ matrices

- transition matrices of Markov processes [Lai-Murphy-Massey 1991]
- correlation matrices [Daemen 1994], or
- matrices of squared correlations [N 1994]

see also

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).

Computing good estimates

In some cases it is possible to exploit the structure of the cipher

- to make these computations feasible; or
- to obtain good estimates, or
- reasonable upperbounds (for security claims)
- reasonable lower bounds (to bound attack complexity)

In some cases this is really hard

AES: The best known upperbounds for 4 and more rounds obtained by Keliher (2005), Canteaut (2015)

Provable security theorem

N-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

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

CRADIC

Cipher Resistant Against Differential Cryptanalysis
($\mathcal{KN}$-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 $\mathbb{F}_2^{33}$

Jakobsen & Knudsen FSE1997 break it with

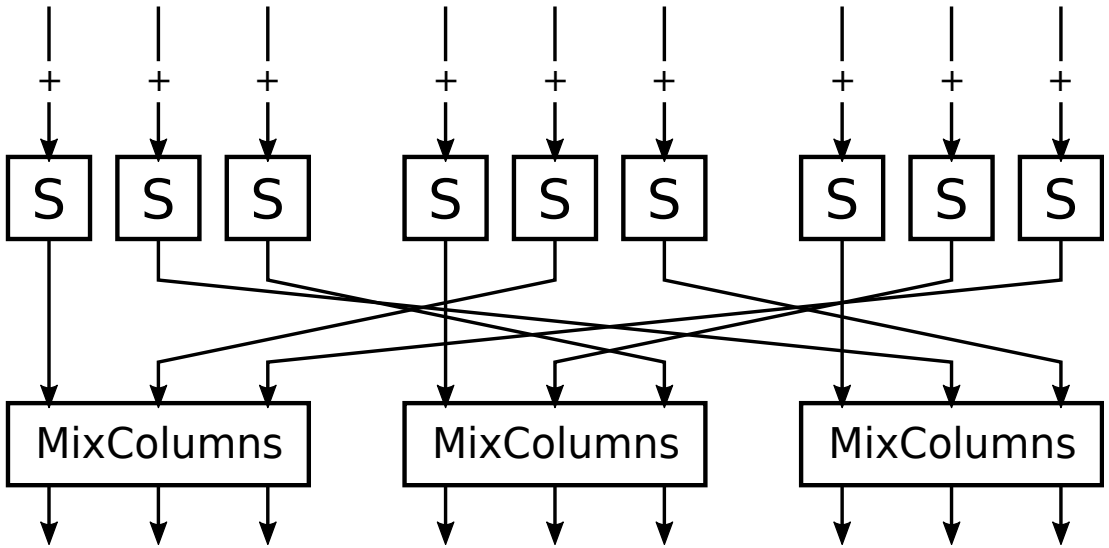
- with 512 chosen plaintexts and 2^{41} running time,
- or with 32 chosen plaintexts and 2^{70} running time
- using *higher order differential cryptanalysis*

Round-function based on the inverse mapping not any more resistant.
This approach was then abandonded
... higher algebraic degree does not help if the inverse has a low degree [Boura-Canteaut IEEE Trans. IT 2013].

Round functions based on the exponent function in $\mathbb{F}_p$ yet to be explored. Previously used for IDEA and SAFER family.

DEAN

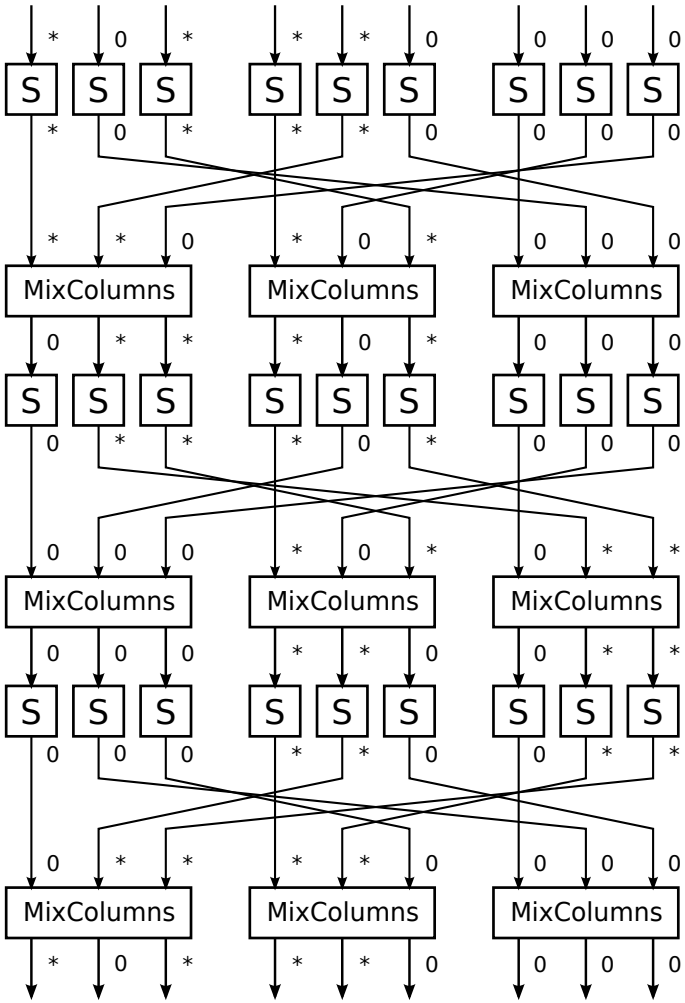
- Proposed by Baignères et al. 2007
- Encrypts blocks of nine elements from the additive group $\mathbb{Z}_{10} \times \mathbb{Z}_{10}$.
- Not a complete cipher, no key schedule, 8 rounds
- One round:



Linear setting of DEAN

- Select suitable linear approximation trails through DEAN
- Use MDS property of `MixColumns` and cover all trails for a minimum number of active S-boxes
- Restrict the (squared) correlation matrix to this subset
- obtain an upper bound to the data complexity and also an attack for this data complexity

Linear approximation of DEAN



Data complexity estimates

Full code book $10^{18} \approx 2^{61}$.
Estimated number of rounds needed to exceed full code book data complexity for a linear distinguisher:

designers	single trail estimate	four rounds
Hakala et al Inscrypt 2012	max two `MixColumns` wide trails and multiple approximations	seven rounds

PRESENT

- Proposed by Bogdanov et al. 2007
- designers: single linear and differential characteristics over more 15 rounds not effective
- specification has 31 rounds, that is, 16 round margin
- linear attack on 26 rounds [Cho 2010]
- the structure allows very accurate estimates of squared linear correlations over any number of rounds [Leander 2011].
- estimates of differential attacks hard to achieve

Correlation matrix of S-box of PRESENT

$\mathrm{cor}[u \xrightarrow{S} v]$

by focusing on single-bit masks

$u \backslash v$	1	2	3	4	5	6	7	8	9	a
1	0	0	0	0	-2^{-1}	0	-2^{-1}	0	0	0
2	0	2^{-2}	2^{-2}	-2^{-2}	-2^{-2}	0	0	2^{-2}	-2^{-2}	0
3	0	2^{-2}	2^{-2}	2^{-2}	-2^{-2}	-2^{-1}	0	-2^{-2}	2^{-2}	-2^{-1}
4	0	-2^{-2}	2^{-2}	-2^{-2}	-2^{-2}	0	2^{-1}	-2^{-2}	-2^{-2}	0
5	0	-2^{-2}	2^{-2}	-2^{-2}	2^{-2}	0	0	2^{-2}	2^{-2}	-2^{-1}
6	0	0	-2^{-1}	0	0	-2^{-1}	0	0	-2^{-1}	0
7	0	0	2^{-1}	2^{-1}	0	0	0	0	-2^{-1}	0
8	0	2^{-2}	-2^{-2}	0	0	-2^{-2}	2^{-2}	-2^{-2}	2^{-2}	0
9	2^{-1}	-2^{-2}	-2^{-2}	0	0	2^{-2}	-2^{-2}	-2^{-2}	-2^{-2}	-2^{-1}
a	0	2^{-1}	0	2^{-2}	2^{-2}	2^{-2}	-2^{-2}	0	0	0
b	-2^{-1}	0	0	-2^{-2}	-2^{-2}	2^{-2}	-2^{-2}	-2^{-1}	0	0
c	0	0	0	-2^{-2}	-2^{-2}	-2^{-2}	-2^{-2}	2^{-1}	0	0
d	2^{-1}	2^{-1}	0	-2^{-2}	-2^{-2}	2^{-2}	2^{-2}	0	0	0
e	0	2^{-2}	2^{-2}	-2^{-1}	2^{-1}	-2^{-2}	-2^{-2}	-2^{-2}	-2^{-2}	0
f	2^{-1}	-2^{-2}	2^{-2}	0	0	-2^{-2}	-2^{-2}	-2^{-2}	2^{-2}	2^{-1}

PRESENT structure

- single-bit masks give high correlations over the S-box
- linear layer is a bit permutation
- trails with masks of more than one bit can be ignored

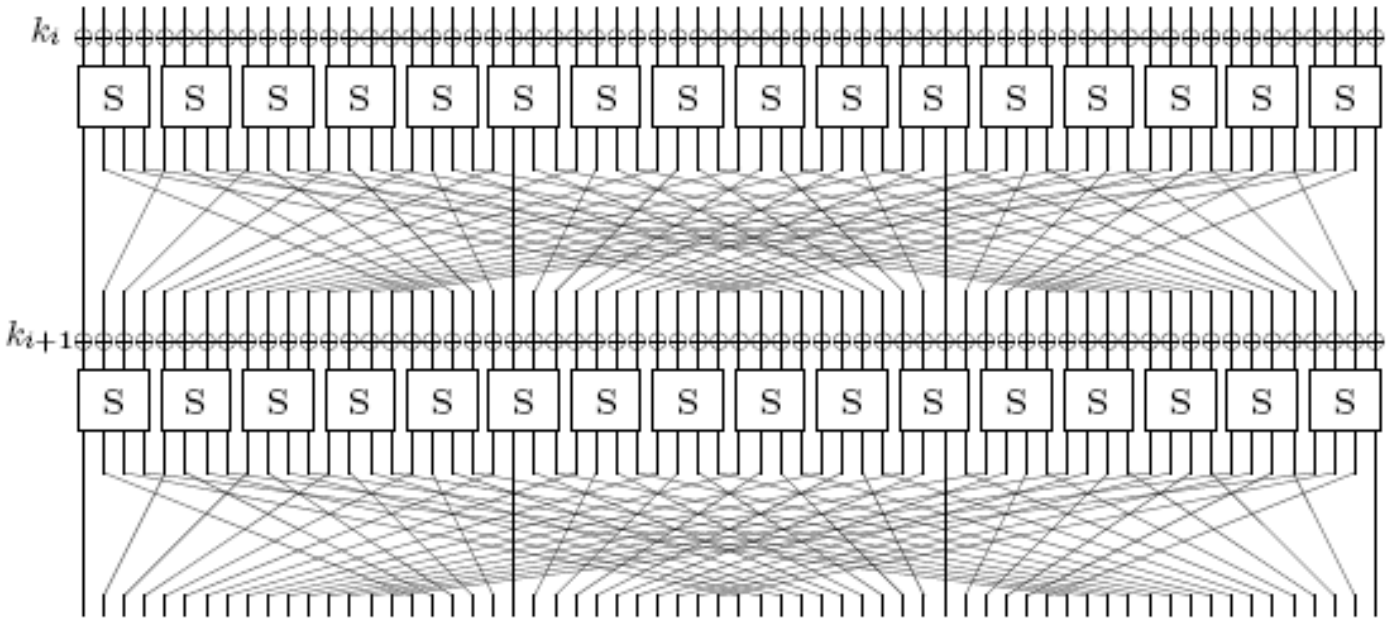


Fig. 2. . The S/P network for PRESENT.

Outline

Statistical attacks

LINEAR CONTEXT

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

Differential-Linear Cryptanalysis [Langford, Hellman 94]

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

Statistical Saturation [Collard, Standaert 09]

Zero Correlation [Bogdanov, Rijmen 11]

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

Multidimensional Linear Cryptanalysis [Cho, Hermelin, Nyberg 08]

. . . . . .

DIFFERENTIAL CONTEXT

Differential Cryptanalysis [Biham, Shamir 90]

Truncated Differential Cryptanalysis [Knudsen 94]

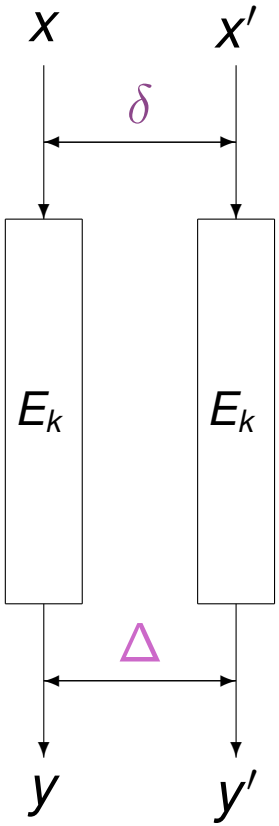
Higher Order Differential cryptanalysis [Lai 94] [Knudsen 94]

Impossible Differential Cryptanalysis [Knudsen 98]

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

. . . . . .

Truncated differential cryptanalysis



Input difference δ
Output difference Δ

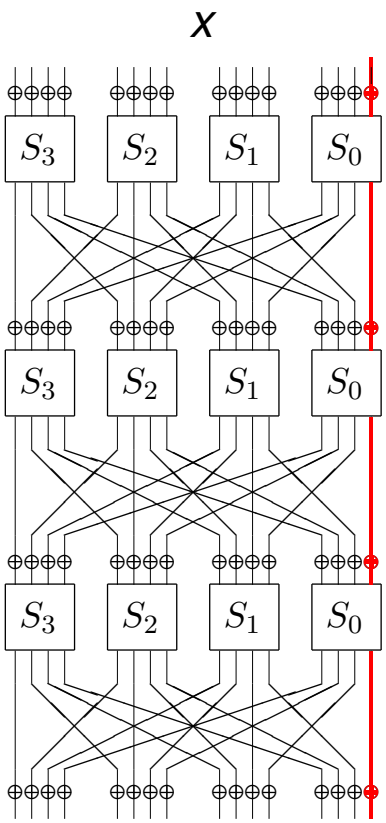
Set of input differences: $\delta \in C$

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

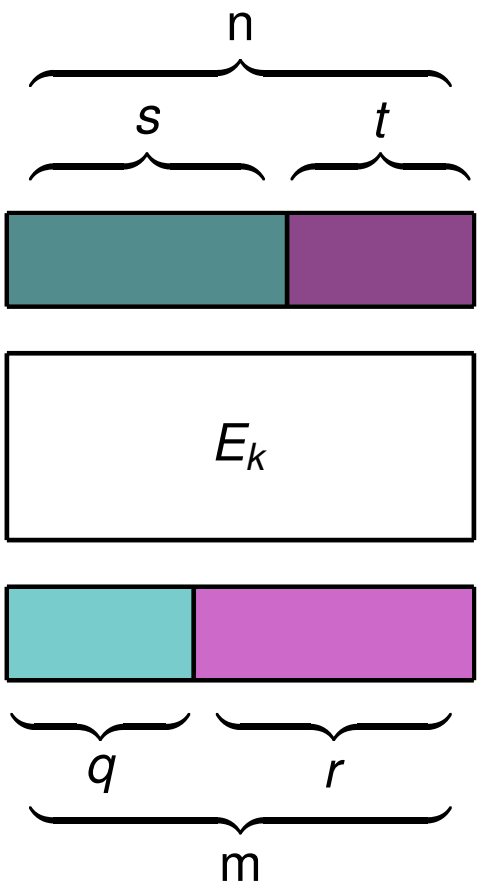
[Blondeau-N EC2013] :

Zero Correlation Linear $\Leftrightarrow$ Impossible Differential

[Blondeau-N EC2014] :

Multidimensional Linear $\Leftrightarrow$ Truncated Differential

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

Don't care about the right side:

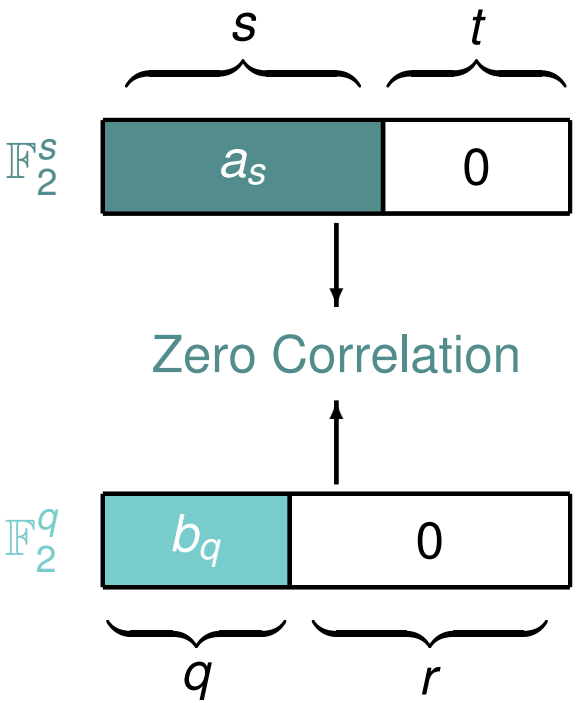
multidimensional linear context

- zero input and output masks

truncated differential context

- all input and output differences

Zero correlation linear

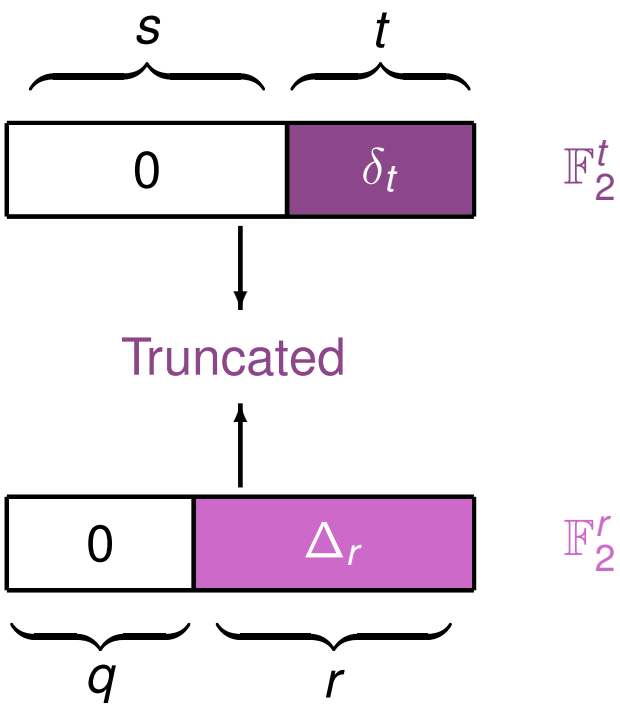


Zero Correlation Linear :

$$\mathrm{cor}_x((a_s, 0), (b_q, 0)) = 0$$

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

Impossible differential



Truncated Differential:

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

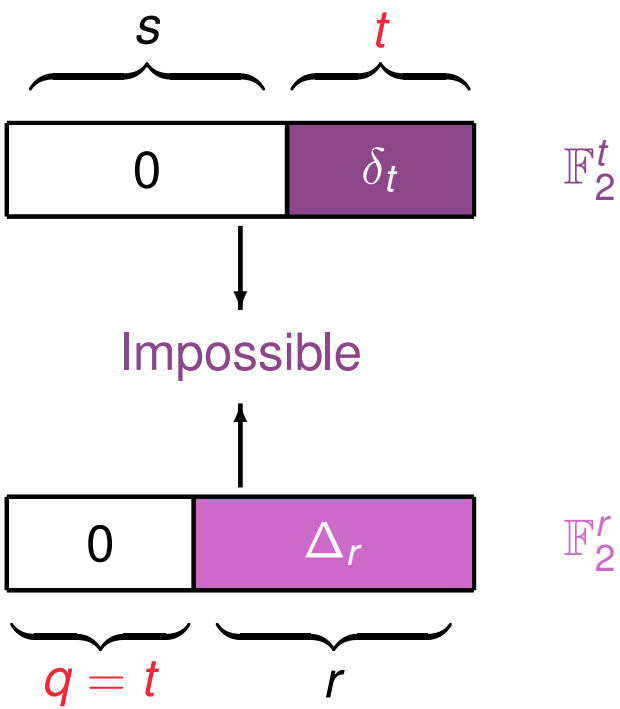
If t=q and $\delta_t \neq 0$

Impossible Differential:

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

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



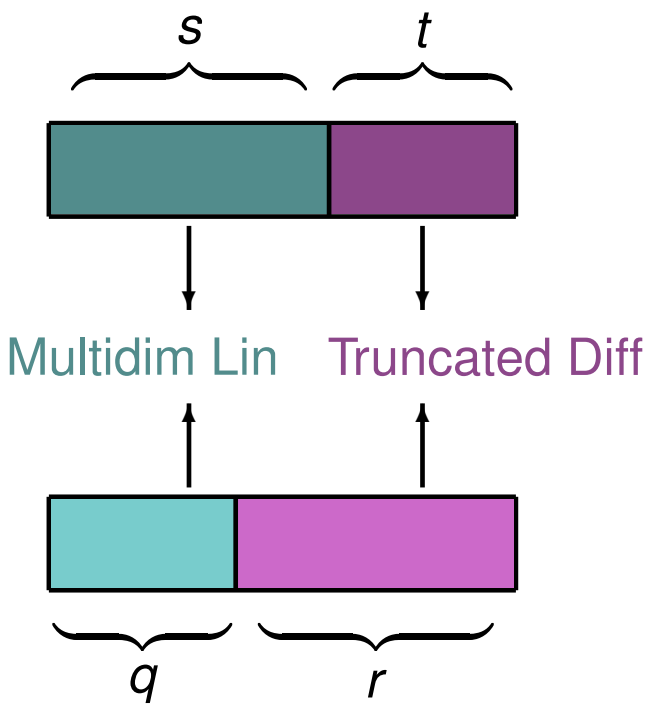
If $t = q$

Zero Correlation Linear Distinguisher

is equivalent to

Impossible Differential Distinguisher

Multidimensional linear and truncated differential



Multidimensional Linear Distinguisher

is equivalent to

Truncated Differential Distinguisher

The mathematical link

Capacity C is defined as

$$C = \sum_{(u_s, w_q) \neq 0} \mathrm{cor}[(u_s, 0) \to (w_q, 0)]^2.$$

Truncated differential probability P equals

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

Then it holds [Blondeau-N 2014]

Theorem

$$P = 2^{-q}(C + 1) = 2^s \sum_{x_s, y_q} \Pr(x_s, y_q)^2.$$

Focus on distributions

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

Multidimensional linear attack

- samples plaintexts x and corresponding ciphertexts y and examines the nonuniformity of the distribution of values (x_s, y_q)

Truncated differential attack

- samples in pairs of plaintexts
- counts collisions in values (x_s, y_q)

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

Corollary

Assume that the cipher is secure against classical linear and differential attacks.

Corollary

The cipher is secure against multidimensional linear attacks if and only if it is secure against truncated differential attacks.

Proof: Provable security requires accurate estimates of correlations of linear approximations (or probabilities of differentials).

Application to PRESENT

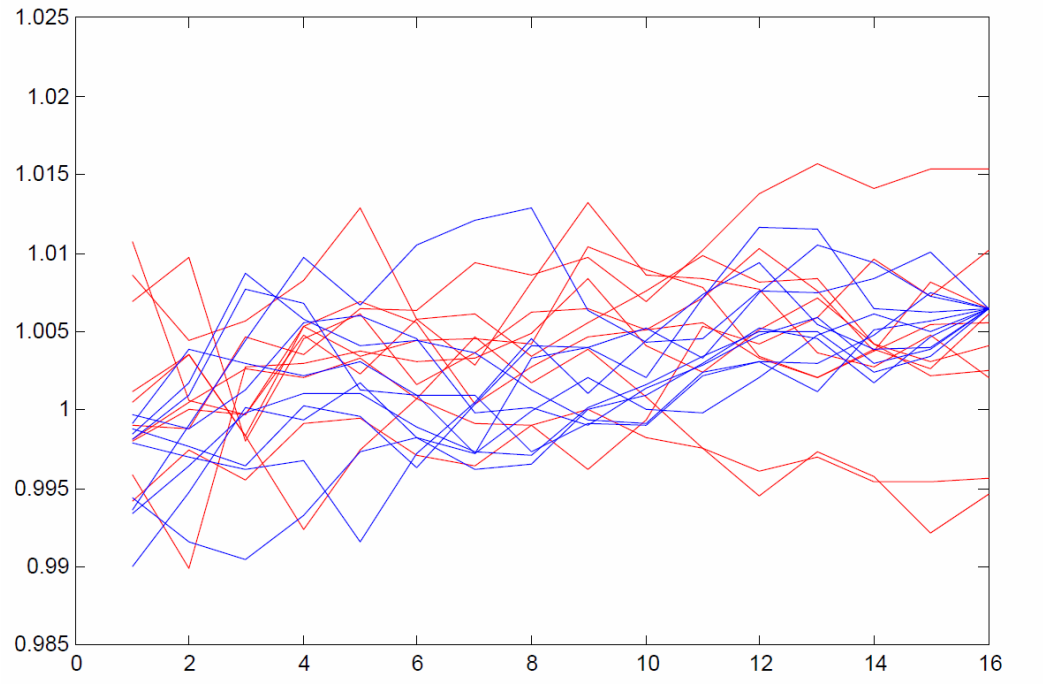
- PRESENT allows accurate estimation of correlations of its linear approximations over any number of rounds
- capacities of multidimensional linear approximations can be accurately evaluated
- The best known linear distinguisher is for 23 rounds and it seems that there is nothing better.

It follows that there is an efficient truncated differential distinguisher for 23 rounds, which can be used in a chosen plaintext key recovery attack.
If PRESENT is secure against linear attacks (as it seems) then it is also secure against differential attacks.

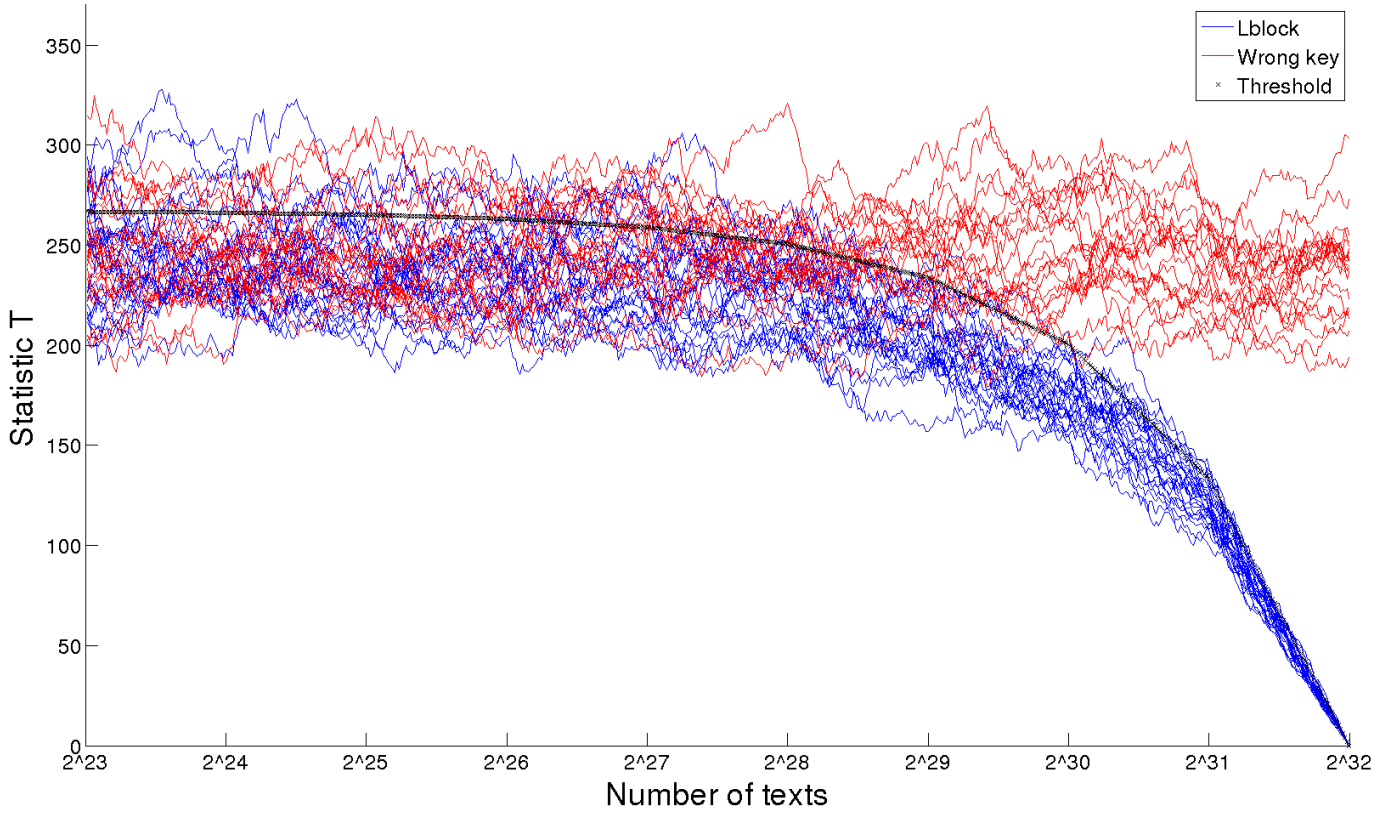
Outline

Sampling without replacement

- critical for zero-correlation cryptanalysis etc
- any statistical method gains in accuracy when close to full codebook [Blondeau-N WCC 2015]



Zero-correlation distinguisher on LBlock (Small Variant)



Outline

Conclusions

- Lightweight ciphers should allow accurate estimation of resistance against attacks to reduce unnecessary security margins
  - In this respect, PRESENT turns out to be very good
- This goal sets also requirements to attack models, which must be accurate. To improve accuracy, use
  - subsums of linear hulls and differentials instead of single linear and differential characteristics
  - multiple linear approximations or truncated differentials instead
  - estimate performance assuming non-repeating plaintext, in particular, when data complexity is close to full codebook.