



Aalto University
School of Science

“Provable” Security Against Differential and Linear Cryptanalysis

Kaisa Nyberg

Department of Information and Computer Science
Aalto University

FSE 2012

March 19, 2012

Introduction

CRADIC

Linear Hull

SPN and Two Strategies

Highly Nonlinear Functions

Generalized Linearity

Linear Approximations Are Universal

Distinguishing Distributions

Conclusions

Disclaimer

- ▶ Many more authors should have been mentioned
- ▶ ... and contributions should have been quoted
- ▶ In particular, I will not cover
 - decorrelation theory,
 - impossible differentials,
 - zero-correlation linear cryptanalysis,
 - weak keys,
 - etc.

Introduction

State of the Art

- ▶ HIGHT(CHES 2006)
128-bit keys - Block length 64 bits - 32 rounds - 3048 GE
31 round attack
- ▶ DESL (FSE 2007)
Is based on the general structure of DES, while using a specially selected S-box. (1848 GE)
- ▶ PRESENT (CHES 2007)
80-bit keys - Block length 64 bits - 31 rounds - 1570GE
26 rounds attack
- ▶ KATAN and KTANTAN (CHES 2009)
80-bit keys - Block length (32-48-64) bits - 254 rounds -
(462-1054)GE
Full round attack for KTANTAN

Do we know how to design block ciphers?

Brief History

- ▶ Biham-Shamir 1989: Differential Cryptanalysis
- ▶ Massey, Lai and Murphy 1990: Differentials and Markov ciphers
- ▶ 1991: Perfect nonlinear S-boxes

CRADIC

Cipher Resistant Against Differential Cryptanalysis

Provable Security Theorem

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

Theorem (KN Theorem) *It is assumed that in a DES-like cipher with $f : \mathbb{F}_2^m \rightarrow \mathbb{F}_2^n$ 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

$$\begin{aligned} p_{\max} &= \max_{\beta} \max_{\alpha_R \neq 0} \Pr[\alpha_L + f(E(X + \alpha_R)) + K + f(E(X) + K) = \beta_R] \\ &\leq p_f = \max_b \max_{a \neq 0} \Pr[f(Y + a) + f(Y) = b] \end{aligned}$$

If f bijective, then the claim of Theorem holds for $s \geq 3$.
Later Aoki showed that the constant 2 can be removed.

4-Round Feistel Differentials

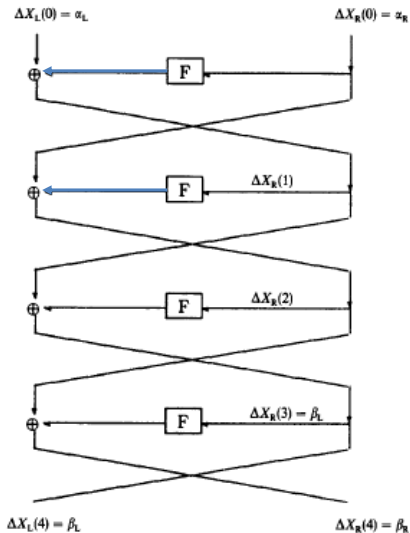


Fig. 1. The four-round differential.

CRADIC

aka KN-Cipher

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

No key schedule, 198-bit key

Jakobsen & Knudsen (1997) break 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 inversion operation not any more resistant.

This approach was then abandoned.

Applications and Further Developments

Feistel

- ▶ Schneier-Kelsey (FSE 1996) Unbalanced Feistel networks
- ▶ Nyberg (Asiacrypt 1996) Generalized Feistel networks
- ▶ Matsui (FSE 1997) Nested structure: MISTY I and II and (3GPP 1999) KASUMI
- ▶ Matsui, Moriai et al.(2000) CAMELLIA
- ▶ etc.

... and more generally and importantly

- ▶ the role of differentials:
single characteristic approach not sufficient

Linear Hull

Or What is the Equivalent of Differential in
Linear Cryptanalysis?

Linear Hull

Eurocrypt 1994 Rump Session

Theorem Let X , K and Y be random variables in \mathbb{F}_2^m , \mathbb{F}_2^ℓ , and \mathbb{F}_2^n , resp. where $Y = F(X, K)$ and X and K are independent. If K is uniformly distributed, then for all $a \in \mathbb{F}_2^m$ and $b \in \mathbb{F}_2^n$,

$$\text{Exp}_K \text{corr}(a \cdot X + b \cdot Y)^2 = \sum_{c \in \mathbb{F}_2^\ell} \text{corr}(a \cdot X + b \cdot Y + c \cdot K)^2.$$

Here, for random variable Z in \mathcal{Z} (binary strings)

$$\text{corr}(u \cdot Z) = \frac{1}{|\mathcal{Z}|} \sum_{z \in \mathcal{Z}} \Pr[z] (-1)^{u \cdot z}.$$

Approximate linear hull given a and b :

$$ALH(a, b) = \{a \cdot X + b \cdot Y + c \cdot K \mid c \in \mathbb{F}_2^\ell\}$$

Application to DES given.

An analogue of the KN Theorem for linear cryptanalysis achieved.

Fixed Key Approach

Correlation of Boolean Function

$f : \mathbb{F}_2^n \mapsto \mathbb{F}_2$ Boolean function

Given two vectors

$$a = (a_1, \dots, a_n), x = (x_1, \dots, x_n) \in \mathbb{F}_2^n$$

the inner product (dot product) is defined as

$$a \cdot x = a_1x_1 + \dots + a_nx_n.$$

Linear Boolean function: $f(x) = a \cdot x$, where $a \in \mathbb{F}_2^n$ is called a linear mask

Vector Boolean function: $f : \mathbb{F}_2^n \mapsto \mathbb{F}_2^m$ with $f = (f_1, \dots, f_m)$, where $b \cdot f_i$ are Boolean functions, for all $b \in \mathbb{F}_2^m$

Correlation between $b \cdot f(x)$ and $a \cdot x$

$$c_f(a, b) = \frac{1}{2^n} (\#\{x \in \mathbb{F}_2^n \mid b \cdot f(x) = a \cdot x\} - \#\{x \in \mathbb{F}_2^n \mid b \cdot f(x) \neq a \cdot x\})$$

Fixed Key Approach

Daemen (1994)

Correlation of a composed function computed as matrix product

$$c_{f \circ g}(a, b) = \sum_u c_g(a, u) c_f(u, b)$$

For key-alternating block cipher E , round functions $x \mapsto f_i(x + K_i)$, and fixed set of round keys K_0, \dots, K_r :

$$c_E(u_0, u_r) = \sum_{u_1, \dots, u_{r-1}} (-1)^{u_0 \cdot K_0 + \dots + u_r \cdot K_r} \prod_{i=1}^r c_{f_i}(u_{i-1}, u_i)$$

Assuming that the round keys are uniformly distributed and independent:

$$\text{Average}_{K_0, \dots, K_r} c_E(u_0, u_r)^2 = \sum_{u_1, \dots, u_{r-1}} \prod_{i=1}^r c_{f_i}(u_{i-1}, u_i)^2.$$

Trail Correlations

It is straightforward to check that for key-alternating block cipher with round functions $x \mapsto f_i(x + K_i)$, and independent and uniformly distributed key $K = K_0 || \dots || K_{r-1}$ we have

$$\text{corr}(a \cdot X + b \cdot Y + c \cdot K) = \prod_{i=1}^r c_{f_i}(u_{i-1}, u_i),$$

where $a = u_0$, $b = u_r$, and c is in unique correspondence with the trail masks u_1, \dots, u_{r-1} .

A Note on Key Scheduling

Design goal: the magnitudes of the correlations

$$c_E(u_0, u_r) = \sum_{u_1, \dots, u_{r-1}} (-1)^{u_0 \cdot K_0 + \dots + u_r \cdot K_r} \prod_{i=1}^r c_{f_i}(u_{i-1}, u_i)$$

should not vary too much with the key.

If all dominating trail correlations are of about equal magnitude and the map:

$$(u_1, \dots, u_{r-1}) \mapsto \text{sign} \left(\prod_{i=1}^r c_{f_i}(u_{i-1}, u_i) \right)$$

is highly nonlinear, the correlations $|c_E(u_0, u_r)|$ are bounded by a small linearity bound.

- ▶ The bent and cube mappings have highly nonlinear correlation sign functions.
- ▶ Correlation sign function of the cube mapping restricted to a half space is bent.

SPN and Two Strategies

Chopping Algebraic S-boxes

- ▶ Lesson learnt from CRADIC: To avoid algebraic attacks, no large algebraic building blocks can be used.
- ▶ Small S-boxes can be searched exhaustively
- ▶ Saarinen (SAC 2011): Complete classification of 4×4 S-boxes with respect to large number of cryptographic and implementation criteria.

Design of AES

- ▶ Get guarantees for the minimum number of active S-boxes
- ▶ MDS matrices for creating larger S-boxes with controlled diffusion
- ▶ Wide-Trail Strategy ensures that
 - ▶ collecting all dominant differential or linear trails becomes impossible
 - ▶ the full linear hull effect cannot be exploited
- ▶ Provable security in the sense of the KN Theorem
- ▶ The best known upperbounds for 4 and more rounds by Keliher (2005)

Design of PRESENT

- ▶ Bit permutation between rounds for optimal diffusion
- ▶ Hardware optimized S-box exhibits strong linear correlations with single-bit masks.
- ▶ Fairly accurate estimates of correlations achievable using single-bit linear approximation trails.
- ▶ **Bad news:** Linear attacks more powerful than expected by the designers (Cho, CT-RSA 2010)
- ▶ **Good news:** Better estimates of strength against linear attacks, including multidimensional linear attacks
- ▶ Leander, Eurocrypt 2011: Statistical Saturation Attack and Multidimensional Linear Cryptanalysis are the same attack
- ▶ Provable security under the assumption that the effect of the single-bit trails is almost complete.

Highly Nonlinear Functions

Bent Function

Correlation between $f : \mathbb{F}_2^n \rightarrow \mathbb{F}_2$ and linear function $x \mapsto u \cdot x$ is defined as

$$c_f(u) = \frac{1}{2^n} (\#\{x \in \mathbb{F}_2^n \mid f(x) = u \cdot x\} - \#\{x \in \mathbb{F}_2^n \mid f(x) \neq u \cdot x\})$$

Parseval's Theorem $\sum_{u \in \mathbb{F}_2^n} c_f(u)^2 = 1.$

A Boolean function is called *bent* if

$$|c_f(u)| = 2^{-\frac{n}{2}}, \text{ for all } u \in \mathbb{F}_2^n.$$

[Rothaus1976][Dillon1978]

If $f : \mathbb{F}_2^n \rightarrow \mathbb{F}_2$ is bent then n is even.

Meier and Staffelbach [1988] introduced the notion of perfect nonlinearity of Boolean functions as an important cryptographic criterion, and later observed that it is equivalent to bentness.

Vector Bent Functions

or Perfect Nonlinear S-Boxes (Eurocrypt 1991)

Vector function $f : \mathbb{F}_2^n \rightarrow \mathbb{F}_2^m$ is said to be *bent* if

- ▶ $w \cdot f$ is bent, for all $w \neq 0$; or what is equivalent,
- ▶ f is *perfect nonlinear* (PN), that is,

$$f(x + \alpha) + f(x)$$

is uniformly distributed as x varies, for all fixed $\alpha \in \mathbb{F}_2^n \setminus \{0\}$.

Theorem. If $f : \mathbb{F}_2^n \rightarrow \mathbb{F}_2^m$ is bent then $n \geq 2m$.

Efficient Constructions of Bent S-Boxes

Classical Maiorana-MacFarland (MM) construction

$$f(x, y) = \pi(x) \cdot y + g(x), \quad x, y \in \mathbb{F}_2^{n/2}$$

where π is a permutation and g any Boolean, is bent Boolean.

Carlet (Eurocrypt 1993): new classes \mathcal{C} and \mathcal{D} of bent Boolean functions.

FSE 1993: To construct a vector bent function from MM, \mathcal{C} and \mathcal{D} , take permutations $\pi_j, j = 1, \dots, \frac{n}{2}$ such that all their sums are permutations, as follows:

$$\pi_j = A^j$$

where A is a state transition matrix of an LFSR with irreducible polynomial of degree $n/2$.

Balanced output if input is restricted to $x \neq 0$. Happens naturally, if π_j implemented with LFSR without non-zero state.

APN S-Boxes

Eurocrypt 1993

S-box $f : \mathbb{F}_2^n \rightarrow \mathbb{F}_2^n$ is said to be *almost perfect nonlinear* (APN) if

$$\#\{x \mid f(x + \alpha) + f(x) = \beta\} \leq 2,$$

for all fixed $\alpha \in \mathbb{F}_2^n \setminus \{0\}$.

- ▶ Function

$$f : \mathbb{F}_2^n \rightarrow \mathbb{F}_2^n, f(x) = x^3,$$

and more generally,

$$f : \mathbb{F}_2^n \rightarrow \mathbb{F}_2^n, f(x) = x^{2^k+1},$$

with multiplication in \mathbb{F}_{2^n} is APN.

- ▶ Bijective only for odd n .

Differentially δ -Uniform S-Boxes

S-box $f : \mathbb{F}_2^n \rightarrow \mathbb{F}_2^n$ is said to be *differentially δ -uniform* if

$$\#\{x \mid f(x + \alpha) + f(x) = \beta\} \leq \delta,$$

for all fixed $\alpha \in \mathbb{F}_2^n \setminus \{0\}$.

Function $f : \mathbb{F}_2^n \rightarrow \mathbb{F}_2^n$, $f(x) = x^{-1}$,

- ▶ is bijective, differentially δ -uniform, $\delta = 2$, n odd, $\delta = 4$, n even
- ▶ all correlations $|\text{corr}(w \cdot f(x) + u \cdot x)|$ are upperbounded by $2^{-\frac{n}{2}+1}$
- ▶ complete Walsh spectrum determined using hyperelliptic curves by Lachaud and Wolfmann (1990).
- ▶ adapted as the core of the S-box for the Rijndael block cipher in 1998 to become the AES in 2001.

Small differential uniformity desirable, also distribution of the differences matters (Blondeau, Canteaut and Charpin, 2010)

Generalized Linearity

Generalized Bent Functions

Let $q \geq 2$ be integer and denote

$$e_q(x) = e^{\frac{2\pi x}{q} i}.$$

$f : \mathbb{F}_2^n \rightarrow \mathbb{F}_2$ is bent if and only if

$$\left| \sum_{x \in \mathbb{F}_2^n} e_2(f(x) + u \cdot x) \right| = 2^{\frac{n}{2}}, \text{ for all } u \in \mathbb{F}_2^n.$$

Kumar-Scholtz-Welch [1985]: $f : \mathbb{Z}_q^n \rightarrow \mathbb{Z}_q$ is *generalized bent* if

$$\left| \sum_{x \in \mathbb{Z}_q^n} e_q(f(x) - ux) \right| = q^{\frac{n}{2}}, \text{ for all } u \in \mathbb{Z}_q^n.$$

Example $f : \mathbb{Z}_p \rightarrow \mathbb{Z}_p$, $f(x) = x^2$, p odd prime.

Generalized Correlation

- ▶ Baignères, Vaudenay, Stern [2007]: Additive groups
- ▶ Drakakis, Requena, McGuire [2010]: \mathbb{Z}_p and \mathbb{Z}_{p-1}
- ▶ Feng, Zhou, Wu, Feng [2011]: Subsets of \mathbb{Z}_{2^n}
- ▶ For any positive integers q and p and $f : A \rightarrow \mathbb{Z}_p$, where A is a subset of \mathbb{Z}_q , we define

$$c_f(u, w) = \frac{1}{|A|} \sum_{x \in A} e_p(wf(x)) \overline{e_q(ux)}$$

8 × 8-bit S-boxes of SAFER

$$f(x) = (45^x \bmod 257) - 1, x \in \mathbb{Z}_{256}$$

and its inverse

$$f^{-1}(y) = \log_{45}(y + 1), y \in \mathbb{Z}_{256}$$

Nonlinearity?

Welch-Costas Functions

p odd prime

g generator of the multiplicative group \mathbb{F}_p^*

Exponential Welch-Costas function

$$f(x) = (g^x \bmod p) - 1, x \in \mathbb{Z}_{p-1}$$

and its inverse, *logarithmic Welch-Costas function*

$$f^{-1}(y) = \log_g(y + 1), y \in \mathbb{Z}_{p-1}$$

are bijections in \mathbb{Z}_{p-1} .

Hakala [2011] proved upperbound $\mathcal{O}(p^{-\frac{1}{2}} \log p)$ of magnitudes of p -ary correlations.

Binary nonlinearity unknown. Interesting cases $p - 1 = 2^n$.

Discrete Logarithm

α generator of the multiplicative group $\mathbb{F}_{2^n}^*$

$$f(x) = \begin{cases} \log_{\alpha}(x), & \text{for } x \neq 0 \\ (1, 1, \dots, 1,) & \text{for } x = 0. \end{cases}$$

gives an n -bit S-box.

Brandstätter, Lange, Winterhof (2005): For any single bit of f , its correlation with any linear function is upperbounded by

$$\mathcal{O}(n2^{-n/2}).$$

For multiple-bit masks, no useful general upperbound known.

Carlet, Feng (2009): Optimum algebraic immunity

Round function for CRADIC?

Linear Approximations are Universal

Linear Projections of Distributions

For random variable¹ Z in \mathcal{Z}

$$\text{corr}(u \cdot Z) = \frac{1}{|\mathcal{Z}|} \sum_{z \in \mathcal{Z}} \Pr[z] (-1)^{u \cdot z}.$$

Applying the inverse Walsh-Hadamard transform we get

$$p_z = \Pr[z] = \sum_{u \in \mathcal{Z}} \text{corr}(u \cdot Z) (-1)^{u \cdot z}$$

Z is a random variable which can be sampled from cipher data:

- ▶ multidimensional linear approximation
- ▶ difference
- ▶ ciphertext from chosen biased plaintext, etc.

anything expected to have non-random behaviour

¹ binary strings for notation only

Capacity of Distribution

Let $M = |\mathcal{Z}|$. We call the quantity

$$\frac{1}{M} \sum_{z \in \mathcal{Z}} \left(p_z - \frac{1}{M} \right)^2$$

the *capacity* of probability distribution of Z and denote it by $C(Z)$. Then

$$C(Z) = \sum_{u \neq 0} |\text{corr}(u \cdot Z)|^2$$

All this generalizes to Z that takes on values in any finite group. Linear (homomorphic) approximations are presented by group characters e_q .

Capacity of probability distribution is sufficient to determine data complexity of distinguishing samples of Z from random.

Distinguishing Distributions

The Best Distinguisher

- ▶ Two probability distributions $p = (p_z)$ and $p' = (p'_z)$
- ▶ Decide whether a given sample distribution $q(N) = (q_z(N))$ obtained from a sample of size N , is drawn from p or p' .
- ▶ Baignères and Vaudenay: Optimal distinguisher

$$\text{LLR}(q(N)) = \sum_{z \in \text{Supp}(q)} q_z(N) \log \frac{p_z}{p'_z}$$

- ▶ Distinguisher decides for p if $\text{LLR}(q)$ is above threshold, otherwise p' .
- ▶ Threshold determines error probability as a function of the sample size N .
- ▶ Error probability depends on Chernoff information between p and p'

Distinguishing from Random

- ▶ For close-to-uniform distributions, the Chernoff information between p and the uniform distribution of support size M can be approximated using the squared Euclidean distance

$$\frac{M}{8 \ln 2} \sum_z (p_z - \frac{1}{M})^2$$

- ▶ Here

$$M \sum_z (p_z - \frac{1}{M})^2 = \sum_{u \neq 0} |\text{corr}(u \cdot Z)|^2 = C(Z) = C(p)$$

the capacity of Z with distribution p .

Data Requirement of the LLR Distinguisher

- ▶ Baignères and Vaudenay (ICITS 2008): for close-to-uniform distributions, the data requirement for the LLR distinguisher is

$$N_{\text{LLR}} \approx \frac{\lambda}{C(p)},$$

where the constant λ depends only on the success probability.

- ▶ In practice,
 - ▶ alternative non-random p may vary with key
 - ▶ accurate estimate of p may not be available
 - ▶ while $C(p)$ may be about the same for almost all keys, or $C(p)$ takes on only a small number of values as key varies.
- ▶ Junod 2003: χ^2 test is asymptotically optimal distinguisher for distributions of binary variables.

Data Requirements

- ▶ For close-to-uniform distribution p (with support of any finite size), an upperbound to the data requirement of the LLR distinguisher can be given as:

$$N_{\text{LLR}} = \frac{\lambda}{C(p)},$$

where the constant λ depends only on the success probability.

- ▶ Vaudenay (ACM CCS 1995): For close-to-uniform distribution p with support of cardinality M , the data requirement of the χ^2 distinguisher can be given as:

$$N_{\chi^2} = \frac{\lambda' \sqrt{M}}{C(p)}, \quad \text{where}$$

$$\lambda' = (\sqrt{2} + 2)\Phi^{-1}(P_S) \approx \lambda$$

For details of this bound, see my presentation in Dagstuhl 2012.

What to use?

For attacks, minimize data requirement

- ▶ either, use LLR if it works
- ▶ else, use χ^2

For provable security, maximize data requirement

- ▶ either, prove that LLR does not work and that the data complexity can be derived using χ^2 estimates
- ▶ else, use complexity estimates for LLR

Conclusions

Conclusions

I discussed

- ▶ provable security against certain statistical attacks in average key setting
- ▶ block cipher design strategies
- ▶ role of linear cryptanalysis among statistical method
- ▶ nonlinearity of S-boxes

Acknowledgements

Thanks to Kimmo, Céline, Risto and Hadi for their help in preparing this presentation, and to the attendees of FSE 2012 for pointing out errors in the previous version of this presentation.

