Asymptotic nonlinearity of Boolean functions

François Rodier

Institut de Mathématiques de Luminy
Marseille

Outline

1. Boolean functions
2. Higher order nonlinearity of Boolean functions
3. Nonlinearity of Vectorial Boolean functions
4. Resistance against linear cryptanalysis
5. Conclusion

Boolean functions

- Let m be a positive integer and $q = 2^m$.
- A **Boolean function** with m variables is a map from the space $V_m = \mathbb{F}_2^m$ into $\mathbb{F}_2$.
- A Boolean function is **linear** if it is a linear form on the vector space $\mathbb{F}_2^m$.
- It is **affine** if it is equal to a linear function up to a constant.

Cryptanalysis

- Booleaan functions are used to build cryptosystems, block ciphers or stream ciphers.

- The existence of affine approximations of the Boolean functions involved in a cryptosystem allows in various situations to build attacks on this system.

- It consists in

  simplifying the enciphering algorithm by a **linear approximation.**

- Therefore a function f is **the more resistant** to this attack that f is **distinct from a linear mapping.**

Non-linearity

We call **non-linearity** of a Boolean function $f : V_m \longrightarrow \mathbb{F}_2$
the distance from f to the set of affine functions with m variables:

$$nl(f) = \min_{h \text{ affine}} d(f, h)$$

where d is the Hamming distance.

The non-linearity is equal to $$nl(f) = 2^{m-1} - \frac{1}{2} S(f)$$

where

$$S(f) = \sup_{v \in V_m} \left| \sum_{x \in V_m} (-1)^{(f(x)+v \cdot x)} \right|$$

and $v \cdot x$ denote the usual scalar product in V_m.

$S(f)$ is **the spectral amplitude** of the Boolean function f.

Inequalities on the nonlinearity

$$2^{m/2} \quad \underset{\substack{\uparrow \\ \text{Parseval}}}{\leq} \quad 2^m - 2nl(f) = S(f) = \sup_{v \in V_m} \left| \sum_{x \in V_m} (-1)^{(f(x)+v \cdot x)} \right| \quad \underset{\substack{\uparrow \\ \text{clear}}}{\leq} \quad 2^m$$

For an even dimension m: bent functions reach the lower bound $2^{m/2}$.

For odd m: $2^{m/2}\sqrt{2}$ has been a long time the only known lower bound of the spectral amplitude $S(f)$.

Improvements of the bound by Patterson and Wiedemann and more recently, by Kavut, Maitra and Yücel have led to a conjecture:

$$\inf_f S(f) \sim 2^{m/2}$$

Boolean functions in cryptography

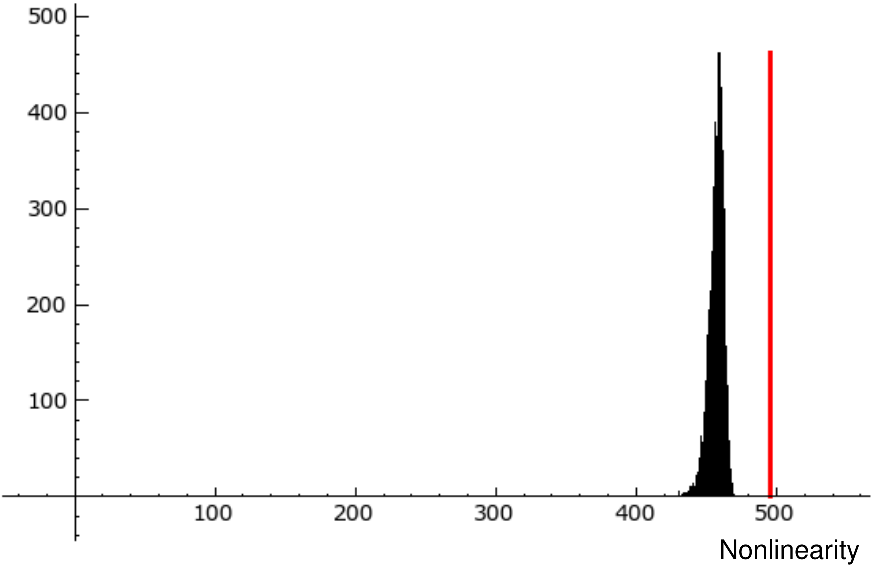
For security reasons functions need to have properties like

- high nonlinearity
- balancedness
- high algebraic degree,
- High algebraic immunity,
- ...

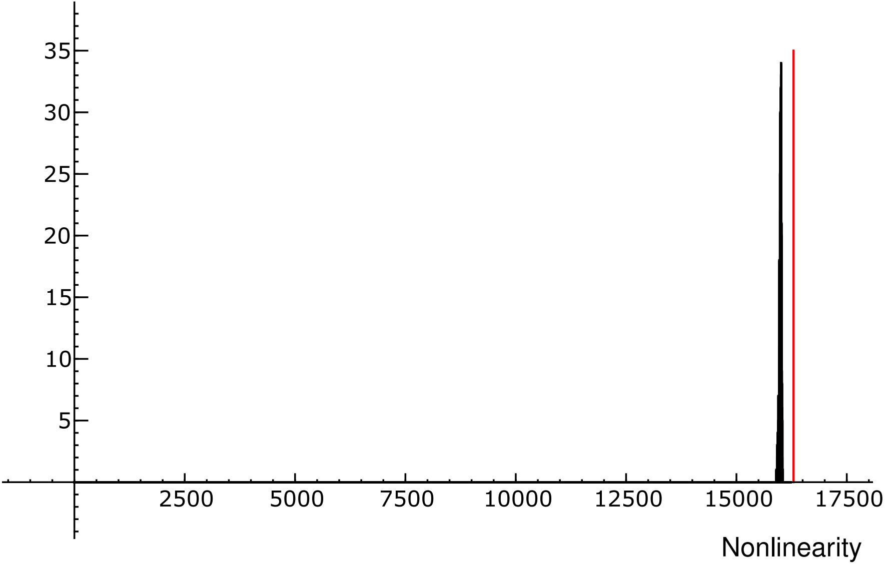
It is necessary to have the possibility of choosing among many Boolean functions,

- not only bent functions,
- but also functions which are **close to be bent,**

Distribution of the nonlinearity for $m = 10$



Distribution of the nonlinearity for $m = 15$



Distribution of the nonlinearity of the Boolean functions

Theorem (Olejar, Stanek, Carlet, FR)

The probability that

$$a\sqrt{2q\log q} < q - 2nl(f) = S(f) < b\sqrt{2q\log q}$$

tends to 1 as m goes to infinity for $0 < a < 1 < b$.

If f is a Boolean function, then, almost surely:

$$\lim_{m\to\infty} \frac{S(f)}{\sqrt{2q\log q}} = 1$$

Cryptanalysis of order r

- The cryptanalysis of order r consists in **simplify the enciphering algorithm** by making an **approximation** by the set of all functions whose algebraic degrees do not exceed r.

- Therefore a function f is **the more resistant** to that this attack that f is **distinct from a mapping of order r.**
- The nonlinearity of order r generalizes the usual nonlinearity.

  For a given function f, it is its Hamming distance to the set of all r-order functions
- Let $NL_r(f)$ denote the r-th order nonlinearity of f, we have

$$NL_r(f) = \min_{g \in \mathrm{RM(r,n)}} d_H(f, g).$$

Higher order nonlinearity of Boolean functions

Very little is known on $nl_r(f)$ for $r > 1$.

To be able to compare with the preceding theorem, we define the spectral amplitude of a Boolean function f the integer $S_r(f)$ such that

$$nl_r(f) = 2^{m-1} - \frac{1}{2} S_r(f).$$

Theorem (C. Carlet and S. Mesnager)

The minimum possible spectral amplitude of order r of Boolean functions, is bounded from below by $\sqrt{15}(1+\sqrt{2})^{r-2} \times 2^{m/2+1} + O(m^{r-2})$.

Higher order nonlinearity of Boolean functions

Asymptotically, C. Carlet proved that almost all Boolean functions have high r-th order nonlinearities, or low r-th order spectral amplitude.

S. Dib, K-U. Schmidt proved that this was the exact bound.

Theorem (C. Carlet, S. Dib, K-U. Schmidt)

The density of the set of functions satisfying

$$a2^{\frac{m+1}{2}}\sqrt{\binom{m}{r}\log 2} < 2^m - 2nl_r(f) = S_r(f) < b2^{\frac{m+1}{2}}\sqrt{\binom{m}{r}\log 2}$$

tends to 1 when m tends to infinity, if $0 < a < 1 < b$.

If f is a Boolean function, then, almost surely:

$$\lim_{m\to\infty}\frac{S(f)}{2^{\frac{m+1}{2}}\sqrt{\binom{m}{r}\log 2}} = 1$$

Vectorial Boolean functions - S-boxes

The linear cryptanalysis exploits **nonuniform statistical behaviors** in the process of encryption.

It consists in **simplifying the encryption algorithm** by making a **linear approximation**.

Therefore a function F is **the more resistant** to that this attack that F is **distinct from a linear mapping.**

Vectorial Boolean functions - S-boxes

A (m, n) vectorial Boolean function with m variables is a map from the space $V_m = \mathbb{F}_2^m$ into $V_n = \mathbb{F}_2^n$.

I define the component functions $u \cdot f$ as the functions $x \longmapsto (u \cdot f)(x) = u \cdot f(x)$ where $\cdot$ denote the usual scalar product of two elements of V_n.

Non-linearity

We call **non-linearity** of a vectorial Boolean function $f : V_m \longrightarrow V_n$ the minimum Hamming distance between all the component functions of f and all affine functions on m variables:

$$nl(f) = \min_{u \in V_n^*} \min_{h \text{ affine}} d(u \cdot f, h)$$

where d is the Hamming distance.

The non-linearity is equal to $nl(f) = 2^{m-1} - \frac{1}{2} S(f)$

where $S(f) = \sup_{u \in V_n^*} \sup_{v \in V_m} \left| \sum_{x \in V_m} (-1)^{(u \cdot f(x) + v \cdot x)} \right|$

Vectorial Boolean functions - S-boxes

Results by Nyberg

Theorem (Nyberg)

The minimal spectral amplitude of a vectorial function $f : \mathbb{F}_{2^m} \longrightarrow \mathbb{F}_{2^n}$ *such that* $n \leq m - 1$ *is such that*

$$S(f) \geq 2^{m/2}$$

This bound can be achieved with equality only if m *is even and* $n \leq m/2$ *by the so-called bent functions.*

Results by Chabaud-Vaudenay

Theorem (Chabaud-Vaudenay)

The minimal spectral amplitude of a vectorial function $\mathbb{F}_{2^m} \longrightarrow \mathbb{F}_{2^m}$ *is* $2^{\frac{m+1}{2}}$.

The functions reaching this bound are called **almost bent.** *They exists when* m *is odd.*

Vectorial Boolean functions - S-boxes

Theorem (S. Dib)

If f is a vectorial function $\mathbb{F}_q \longrightarrow \mathbb{F}_q$, then, almost surely: the probability that

$$2a\sqrt{q\log q} < q - 2nl(f) = S(f) < 2b\sqrt{q\log q}$$

tends to 1 as m goes to infinity for $0 < a < 1 < b$.

If f is a vectorial function $\mathbb{F}_q \longrightarrow \mathbb{F}_q$, then, a.s.:

$$\lim_{m\to\infty} \frac{S(f)}{2\sqrt{q\log q}} = 1.$$

Vectorial Boolean functions - S-boxes

Theorem (S. Dib)

If f is a vectorial function $\mathbb{F}_2^m \longrightarrow \mathbb{F}_2^n$, then, almost surely: the probability that

$$S(f) < b\sqrt{2^{m+1}(m+n)\log 2}$$

tends to 1 as m goes to infinity for $1 < b$.

Theorem (S. Dib)

If f is a vectorial function $\mathbb{F}_2^m \longrightarrow \mathbb{F}_2^n$, and $m \geq n$ then, almost surely: the probability that

$$a\sqrt{2^{m+1}(m+n)\log 2} < S(f)$$

tends to 1 as m goes to infinity for $0 < a < 1$.

Resistance against linear cryptanalysis

Let an r round cipher with

- $X \in \mathbb{F}_2^m$, the plain text,
- $K \in \mathbb{F}_2^\ell$ the key
- $Y(X, K) \in \mathbb{F}_2^m$ a function of X and K.

where all are random variables.

$$x_1 \xrightarrow{F_{K_1}} x_2 \xrightarrow{F_{K_2}} \dots \xrightarrow{F_{K_{r-2}}} x_{r-1} \xrightarrow{F_{K_{r-1}}} Y(X, K)$$

Let $a \in \mathbb{F}_2^m$, $b \in \mathbb{F}_2^m$ be linear masks.

$$a \cdot X = b \cdot Y(X, K) \ ?$$

Theorem (Nyberg)

In a DES-like cipher with more than 4 rounds, independent round keys and uniformly random plaintext and f be the S-box.

$$2^{-\ell} \sum_{K \in \mathbb{F}_2^\ell} \left(P_X\left(a \cdot X = b \cdot Y(X, K)\right) - \frac{1}{2}\right)^2 \leq 2^{-4m-1} S(f)^4$$

Example

$$2^{-\ell} \sum_{K \in \mathbb{F}_2^{\ell}} \left(P_X \left(a \cdot X = b \cdot Y(X, K) \right) - \frac{1}{2} \right)^2 \leq 2^{-4m-1} S(f)^4$$

- Let we consider 2 ciphers
  - Let a cipher on $\mathbb{F}_2^m$ with f be an almost bent function.
  - Let a cipher on $\mathbb{F}_2^{m'}$ with f' a function $\mathbb{F}_2^{m'} \longrightarrow \mathbb{F}_2^{m'}$ such that $S(f') \simeq 2\sqrt{2^{m'} m' \log 2}$.
- Then they give the same bound on probability of breaking the cipher if

$$m = m' - log_2(m') - 0.47$$

- Let $m' = 136$ for random function, then $m = 128$ for almost bent function.

Conclusion

- We have been interested in classifying the Boolean functions according to the nonlinearity.
- We found a concentration point for the nonlinearity of random functions in the case of
  - Boolean functions with r^{th} order nonlinearity
  - Vectorial Boolean functions
- We found that it is close to the maximum nonlinearity in these cases
- You don't lose very much by replacing an almost bent function by a random function

- To be done

  to find bounds for $\sqrt{q} \leq S(f) \leq \sqrt{2q \log q}$.

Thank you