CCZ-equivalence of single and multi output Boolean functions

Claude Carlet

University of Paris 8 (LAGA), France

Lilya Budaghyan,

University of Bergen

Outline

- Boolean and vectorial functions ; nonlinearity notions
- Affine, EA and CCZ (graph) equivalences
- CCZ / EA equivalence of single-output functions
- CCZ / EA equivalence of multi-output functions
- An equivalence notion on vectorial functions apparently more general than CCZ-equivalence
- Conclusion

Boolean and vectorial functions
Nonlinearity notions

The functions from $\mathbb{F}_2^n$ to $\mathbb{F}_2^m$ are called (n,m)-*functions* or *vectorial functions* or *S-boxes*.
$m=1 \rightarrow$ Boolean, single-output ; $m>1 \rightarrow$ multi-output.

The linear nonzero combinations of the coordinate functions of F, i.e. the functions $v\cdot F$; $v\neq 0$, are the *component functions* of F.

The *Walsh transform* of F :

$$(u,v)\in \mathbb{F}_2^n\times \mathbb{F}_2^m \rightarrow \sum_{x\in \mathbb{F}_2^n}(-1)^{v\cdot F(x)+u\cdot x}\in \mathbb{Z}.$$

The *algebraic normal form* (ANF) exists and is unique :

$$\sum_{I\subseteq\{1,\cdots,n\}} a_I \left(\prod_{i\in I} x_i\right)\ ;\ a_I \in \mathbb{F}_2^m.$$

The *algebraic degree* $d^\circ F$ of F is the global degree of its ANF.
F affine : $d^\circ F = 1$; F quadratic : $d^\circ F = 2$.

A second representation exists and is unique ($m = n$ or $m \mid n$) :

$$\mathbb{F}_2^n \sim \mathbb{F}_{2^n}\ ; \quad F(x) = \sum_{j=0}^{2^n-1} \delta_j x^j\ , \quad \delta_j \in \mathbb{F}_{2^n}\ .$$

Then $d^\circ F = \max_{j /\, \delta_j \neq 0} w_2(j),$ where $w_2(j)$ is the 2-weight of j (i.e. the Hamming weight of the binary expansion of j).

Note that, when $\mathbb{F}_2^n \sim \mathbb{F}_{2^n}$, we can take : $x \cdot y = tr_n(x\,y)$, where

$$tr_n(x) = x + x^2 + x^{2^2} + \cdots + x^{2^{n-1}}.$$

The *nonlinearity* $nl(F)$ of F is the minimum Hamming distance between all the component functions $v \cdot F$, $v \neq 0$, of F and all affine functions $u \cdot x + cst$ on n variables.

$$nl(F) = 2^{n-1} - \frac{1}{2} \max_{v \in \mathbb{F}_2^{m*};\ u \in \mathbb{F}_2^n} \left| \sum_{x \in \mathbb{F}_2^n} (-1)^{v \cdot F(x) + u \cdot x} \right|.$$

Main upper bounds on $nl(F)$:

- *Covering radius bound* :

$$nl(F) \leq 2^{n-1} - 2^{n/2-1}$$

is tight iff n is even and $m \leq n/2$ (Nyberg).

The (n,m)-functions achieving it with equality satisfy $\sum_{x\in\mathbb{F}_2^n}(-1)^{v\cdot F(x)+u\cdot x} \in \{\pm 2^{\frac{n}{2}}\}$ for every $v \neq 0$ and u. They are called *bent* or *perfect nonlinear* (PN).

F is bent iff all its *derivatives* $D_aF(x) = F(x) + F(x+a)$, $a \in \mathbb{F}_2^{n*}$, are balanced (i.e. have uniform output distribution).

- *Sidelnikov-Chabaud-Vaudenay (SCV)*, valid for $m \geq n-1$:

$$nl(F) \leq 2^{n-1} - \frac{1}{2}\sqrt{3 \times 2^n - 2 - 2\frac{(2^n-1)(2^{n-1}-1)}{2^m-1}};$$

It equals the covering radius bound when $m = n-1$.

The SCV bound is tight only for $m = n$ with n odd and states then $nl(F) \leq 2^{n-1} - 2^{\frac{n-1}{2}}$.

The (n,n)-functions achieving it with equality satisfy $\sum_{x \in \mathbb{F}_2^n} (-1)^{v \cdot F(x) + u \cdot x} \in \{0, \pm 2^{\frac{n+1}{2}}\}$ for every $v \neq 0$ and u, and are called *almost bent* (AB).

According to Chabaud-Vaudenay's proof of the SCV bound, any AB function is *almost perfect nonlinear* (APN), that is :
all its derivatives D_aF , $a \in \mathbb{F}_2^{n*}$, are 2-to-1.

An APN (n, n)-function contributes to an optimal resistance to the differential attack (Biham-Shamir).

A PN (n, m)-function contributes to an optimal resistance to the differential attack and to the linear attack (Matsui) ; an AB (n, n)-function as well.

AB $\Rightarrow$ APN ; APN $\not\Rightarrow$ AB (except when F is quadratic, n odd).

Affine, EA and CCZ (graph) equivalences

All these notions are invariant under affine, extended affine and CCZ equivalences. Two functions F, G are called :

- *affine equivalent* if $F = L \circ G \circ L'$; L, L' affine permutations ;

- *extended affine equivalent* (EA-equivalent) if $F = L \circ G \circ L' + L''$; L, L' affine permutations ; L'' affine function ;

- *CCZ-equivalent* (graph-equivalent) if the graphs

$$\{(x, y) \in \mathbb{F}_2^n \times \mathbb{F}_2^n \,|\, y = F(x)\} \;; \quad \{(x, y) \in \mathbb{F}_2^n \times \mathbb{F}_2^n \,|\, y = G(x)\}$$

are affine equivalent.

Hence, F and G are CCZ-equivalent if :

$$y = F(x) \Leftrightarrow L_2(x,y) = G(L_1(x,y)),$$

where $L_1 : \mathbb{F}_2^n \times \mathbb{F}_2^m \to \mathbb{F}_2^n$, $L_2 : \mathbb{F}_2^n \times \mathbb{F}_2^m \to \mathbb{F}_2^m$ and (L_1, L_2) is an affine automorphism of $\mathbb{F}_2^n \times \mathbb{F}_2^m$.

Equivalently : the indicators of the graphs of F and G are affine equivalent.

- *Question :*
  is the EA-equivalence of these indicators more general ?
  what about the CCZ-equivalence of these indicators ?

The algebraic degree is EA-invariant (if greater than 1) but not CCZ-invariant.

PN, APN, and AB-ness being notions naturally defined on the graphs of the functions, the proper equivalence notion in cryptographic framework is the CCZ equivalence.

But given some PN, APN or AB function F, it is difficult to construct a CCZ-equivalent function F' which is not EA-equivalent to F (while it is straightforward to construct EA-equivalent functions).

The only known examples (by L. B., C. C., A. Pott) are with F a Gold function $F(x) = x^{2^i+1}$ on $\mathbb{F}_{2^n} \sim \mathbb{F}_2^n$ (doing this with Kasami, Welch, Niho, Dobbertin or inverse function is an open problem).

Even checking that two given functions are CCZ-inequivalent may be quite hard if they share the same CCZ-invariant parameters.

Hence, identifying cases where CCZ-equivalence reduces to EA-equivalence is useful.

A first example (L. B., C. C.) of such case has already been pointed out : *two bent functions (Boolean or vectorial) are CCZ equivalent if and only if they are EA equivalent*.

CCZ / EA equivalence of single-output functions

Let $f' \sim_{CCZ} f$ and $f' \not\sim_{EA} f$.

Up to translation, there exist $L : \mathbb{F}_2^n \longrightarrow \mathbb{F}_2^n$, and $l : \mathbb{F}_2^n \longrightarrow \mathbb{F}_2$ both linear, and $a \in \mathbb{F}_2^n \setminus \{0\}$, $\eta \in \mathbb{F}_2$, such that

$$\mathcal{L}(x, y) = \big(L(x) + ay, l(x) + \eta y\big)$$

is a linear permutation of $\mathbb{F}_2^n \times \mathbb{F}_2$, and denoting :

$$\begin{aligned} F_1(x) &= L(x) + af(x), \\ f_2(x) &= l(x) + \eta f(x), \end{aligned}$$

F_1 is a permutation of $\mathbb{F}_2^n$ and

$$f'(x) = f_2 \circ F_1^{-1}(x).$$

We need characterizing the permutations of the form $L(x) + af(x)$.

We can wlog restrict us to two cases : $L(x) = x$ and $L(x) = x + x^2$ (with $\mathbb{F}_2^n$ identified with $\mathbb{F}_{2^n}$).

Lemma 1 *For every n,*
- $x + af(x)$ is a permutation if and only if it is an involution.
- $x + x^2 + af(x)$ is a permutation if and only if $tr_n(a) = 1$ and $f(x + 1) = f(x) + 1$ for every x.

Theorem 2 *Two Boolean functions on* $\mathbb{F}_{2^n}$ *(or equivalently on* $\mathbb{F}_2^n$*) are CCZ-equivalent if and only if they are EA-equivalent.*

A little more generally :

Theorem 3 *Let* f *be a Boolean function on* $\mathbb{F}_{2^n}$ *(on* $\mathbb{F}_2^n$*) and* f' *an* (n, m)*-function. Then* f *and* f' *are CCZ-equivalent as* (n, m)*-functions if and only if they are EA-equivalent.*

CCZ / EA equivalence of multi-output functions

Proposition 4 *Let* $n \geq 5$ *and* $m > 1$ *be any divisor of* n*, or* $n = m = 4$*. Then for* (n, m)*-functions, CCZ-equivalence is strictly more general than EA-equivalence.*

Sketch of proof : Let $tr_n^m(x) = x + x^{2^m} + x^{2^{2m}} + ... + x^{2^{(n/m-1)m}}$ and $F(x) = tr_n^m(x^3)$.

- if n is odd, $\mathcal{L}(x, y) = \Big(x + tr_n(x) + tr_m(y), y + tr_n(x) + tr_m(y)\Big)$ is an involution, and $F_1(x) = x + tr_n(x) + tr_n(x^3)$ is an involution too. This leads to the function :

$$tr_n^m(x^3) + tr_n^m(x^2 + x)tr_n(x) + tr_n^m(x^2 + x)tr_n(x^3)$$

which is CCZ-equivalent to F and nonquadratic if $n \geq 5$ and $m > 1$.
- if n is even, $\mathcal{L}(x, y) = (x + tr_m(y), y)$.

Proposition 5 *If F and F' are CCZ-equivalent and EA-inequivalent then $H(x) = (F(x), 0)$ and $H'(x) = (F'(x), 0)$ are also CCZ-equivalent and EA-inequivalent.*

This leads to :

Theorem 6 *Let $n \geq 5$ and $k > 1$ be the smallest divisor of n. Then for any $m \geq k$, the CCZ-equivalence of (n, m)-functions is strictly more general than EA-equivalence.*

In particular, when $n \geq 6$ is even, this is true for every $m \geq 2$.

An equivalence notion on vectorial functions apparently more general than CCZ-equivalence

Proposition 7 *Two (n,m)-functions F and F' are CCZ-equivalent if and only if the indicators of their graphs 1_{G_F} and $1_{G'_F}$ are EA-equivalent.*

Corollary 8 *Two (n,m)-functions F and F' are CCZ-equivalent if and only if the indicators of their graphs 1_{G_F} and $1_{G'_F}$ are CCZ-equivalent.*

Conclusion

• From CCZ-equivalence viewpoint, multi-output Boolean functions behave quite differently from single-output Boolean functions.

• However, some classes of vectorial functions behave similarly as single-output functions (i.e. in these classes, CCZ=EA) :

– An example of such subclass is that of bent (perfect nonlinear) functions.

– Are there other examples ?
Note that APN (and AB) functions are not such examples since the functions CCZ-equivalent and EA-inequivalent given in this talk are APN/AB.

- Even if CCZ-equivalence and EA-equivalence are identical for Boolean functions and for bent functions, it is possible to use CCZ-equivalence to obtain, from known bent functions, bent Boolean functions which are new up to EA-equivalence (L. B., C.C., WCC 2009)

Announcement : Next SETA conference

Sequences and Their Applications

will be held in ***Paris*** in ***September 13-17, 2010***.

General chair : Patrick Solé

PC co-chairs : A. Pott ; C. Carlet