Generalized Correlation Analysis of Vectorial Boolean Functions

Claude Carlet[1], Khoongming Khoo[2], Chu-Wee Lim[2], and Chuan-Wen Loe[2]

[1] University of Paris 8 (MAATICAH)
also with INRIA, Projet CODES, BP 105, 78153, Le Chesney Cedex, France
[2] DSO National Laboratories, 20 Science Park Drive, S118230, Singapore.
claude.carlet@inria.fr, kkhoongm@dso.org.sg, lchuwee@dso.org.sg,
lchuanwe@dso.org.sg

Abstract. We investigate the security of n-bit to m-bit vectorial Boolean functions in stream ciphers. Such stream ciphers have higher throughput than those using single-bit output Boolean functions. However, as shown by Zhang and Chan at Crypto 2000, linear approximations based on composing the vector output with any Boolean functions have higher bias than those based on the usual correlation attack. In this paper, we introduce a new approach for analyzing vector Boolean functions called generalized correlation analysis. It is based on approximate equations which are linear in the input x but of free degree in the output $z = F(x)$. Based on experimental results, we observe that the new generalized correlation attack gives linear approximation with much higher bias than the Zhang-Chan and usual correlation attacks. Thus it can be more effective than previous methods.

First, the complexity for computing the generalized nonlinearity for this new attack is reduced from $2^{2^m \times n+n}$ to 2^{2n}. Second, we prove a theoretical upper bound for generalized nonlinearity which is much lower than the unrestricted nonlinearity (for Zhang-Chan's attack) or usual nonlinearity. This again proves that generalized correlation attack performs better than previous correlation attacks. Third, we introduce a generalized divide-and-conquer correlation attack and prove that the usual notion of resiliency is enough to protect against it. Finally, we deduce the generalized nonlinearity of some known secondary constructions for secure vector Boolean functions.

Keywords. Vectorial Boolean Functions, Unrestricted Nonlinearity, Resiliency.

1 Introduction

In this paper, we consider n-bit to m-bit vectorial Boolean functions when they are used in stream ciphers. There are two basic designs for such stream ciphers based on linear feedback shift registers (LFSR). One is a combinatorial generator [12] which consists of n LFSR's and a vector function $F(x)$. At each clock, one bit is tapped from the secret state of each LFSR as an input bit of $F(x)$ to produce m bits of output keystream. This keystream is then XORed with the plaintext to form the ciphertext. The other model is the filter function generator [12] where n bits are tapped from one LFSR as input to $F(x)$ to produce the keystream output. The advantage of using vector Boolean functions is that the stream ciphers have higher throughput, i.e. the encryption and decryption speed is m times faster than single output Boolean functions. However, we need to study its security when compared to the single-bit output case.

Basic attacks on these stream ciphers are the correlation attack by Siegenthaler [14] and its improvements (see e.g. [2]). In the case of the filter function model, a linear approximation is formed between the LFSR state bits and output keystream. If the approximation has probability $p \neq 1/2$, then we can recover the secret LFSR bits when enough keystream bits are known. In the case of the combinatorial model, an approximation of the output is made by the combination of t out of the n input bits produced by the LFSRs and it is shown in [2] that the attack is optimal with the linear combination of these t bits. Siegenthaler's attack was described for single-output Boolean functions but it can be generalized naturally to the vector output case where we take any linear combination of output bits [3].

This attack can be improved as shown by Zhang and Chan at Crypto 2000 [15] where they consider linear approximation based on any Boolean function of the output vector (instead of just linear combination of the output vector). Since there are 2^{2^m+n} linear approximations to choose from in the Zhang-Chan approach compared to just 2^{n+m} linear approximations in the usual approach, it is easier to choose one with higher bias, i.e. where probability p is further away from $1/2$.

In Section 2, we introduce the generalized correlation attack by considering linear approximations which are linear in the input x and of free degree in the output $z = F(x)$. Now there are $2^{2^m \times (n+1)}$ linear approximations from which we can choose one with even higher bias than the Zhang-Chan and usual correlation attack. However, choosing the best linear approximation out of that many choices is infeasible. Therefore in Section 3, we reduce the complexity of choosing the best linear approximation for generalized correlation attack from $2^{2^m \times (n+1)+n}$ to 2^{2n}, which is much more manageable.

The generalized nonlinearity is an analogue of the usual nonlinearity, which measures the effectiveness of a function against generalized correlation attack. Based on efficient computation for finding the best generalized linear approximation, we computed the generalized nonlinearity of highly nonlinear vector functions and randomly generated vector functions in Section 3.2. We observe that the generalized nonlinearity is much lower than the usual nonlinearity and unrestricted nonlinearity (corresponding to Zhang-Chan's attack) for these functions. For example, when the inverse function on $GF(2^8)$ is restricted to $5, 6, 7$ output bits, the usual and unrestricted nonlinearities are non-zero while the generalized nonlinearity is already zero. That means the stream cipher can be attacked as a deterministic linear system while the Zhang-Chan and usual correlation attack are still probabilistic. Theoretical results on the generalized nonlinearity are also studied. In Section 4, we derive an upper bound for generalized nonlinearity which is much lower than the upper bound for usual correlation attack (covering radius bound [3]) and that for Zhang-Chan's attack (unrestricted nonlinearity bound [4]). Thus it gives further evidence that generalized correlation attack is more effective than the other correlation attacks on vector Boolean functions.

The Siegenthaler divide-and-conquer attack on combinatorial stream ciphers [14] resulted in the notion of t-th order correlation immune function (called t-resilient when the function is balanced, which is necessary). To protect against this attack, we require the combining function $F(x)$ to be t-resilient for large t. In Section 5.1, we introduce the concept of generalized divide-and-conquer correlation attack and generalized resiliency to protect against it. We observe that usual resiliency is equivalent to generalized resiliency and thus is sufficient to protect the cipher against the generalized divide-and-conquer attack.

In Section 6, we investigate the generalized nonlinearity of two secondary constructions for vector Boolean functions that are resilient and/or possess high nonlinearity. We conclude output composition (e.g. dropping output bits) of balanced vector functions may increase generalized nonlinearity. For a concatenated function to possess high generalized nonlinearity, we require all component functions to possess high generalized nonlinearity.

2 Generalized Correlation Analysis of Vector Output Stream Ciphers

We consider a stream cipher where the state bits of one or more linear feedback shift registers are filtered by a vector Boolean function $F : GF(2)^n \to GF(2)^m$ to form keystream bits. The keystream bits will be XORed with the plaintext to form the ciphertext. Traditionally, an adversary who wants to perform correlation attack on this stream cipher tries to find an approximation of a linear combination of output bits by a linear combination of input bits $u \cdot F(x) \approx w \cdot x$. For correlation attack to be successful, we require that the bias defined by:

$$Bias = |Pr(u \cdot F(x) = w \cdot x) - 1/2|,\ u \in GF(2)^m,\ w \in GF(2)^n,$$

is large. Conversely, if all linear approximations of $u \cdot F(x)$ have small bias, then it is secure against correlation attack. A concept related to the correlation attack is the *Hadamard transform* $\hat{f} : GF(2)^n \to \mathbb{R}$ of a Boolean function $f : GF(2)^n \to GF(2)$ which is defined as $\hat{f}(w) = \sum_{x \in GF(2)^n} (-1)^{f(x)+w\cdot x}$. Based on the Hadamard transform, we can define the *nonlinearity* [3, 10] of $F(x)$ as:

$$N_F = 2^{n-1} - 1/2 \max_{0 \neq u \in GF(2)^m, w \in GF(2)^n} |\widehat{u \cdot F}(w)|. \tag{1}$$

From the above equation, we deduce that a high nonlinearity ensures protection against correlation attack. It is well known that $0 \leq N_F \leq 2^{n-1} - 2^{n/2-1}$ [3].

At Crypto 2000, Zhang and Chan [15] observed that instead of taking linear combination of the output bit functions $u \cdot F(x)$, we can compose $F(x)$ with any

Boolean function $g : GF(2)^m \to GF(2)$ and consider the probability:

$$Pr(g(z) = w \cdot x) \text{ where } z = F(x). \tag{2}$$

Because $z = F(x)$ corresponds to the output keystream which is known, then $g(z)$ is also known. Therefore $g(z) \approx w \cdot x$ is a linear approximation which can be used in correlation attacks. Since we are choosing from a larger set of equations now, we can find linear approximations with larger bias $|Pr(g(z) = w \cdot x) - 1/2|$. Let us define the unrestricted nonlinearity [4] which measures the effectiveness of the Zhang-Chan attack. Denote by $wt(f)$ the number of ones among the output of $f : GF(2)^n \to GF(2)$.

Definition 1. *Let $F : GF(2)^n \to GF(2)^m$ and let $\mathcal{G}$ = Set of m-bit Boolean functions $g : GF(2)^m \to GF(2)$. We define the* unrestricted nonlinearity *as:*

$$UN_F = \min\{\min_{0 \neq u \in GF(2)^m}(wt(u \cdot F), 2^n - wt(u \cdot F)), nonlin_{UN}F\}$$

where

$$nonlin_{UN}F = 2^{n-1} - \frac{1}{2}\max_{w \neq 0, g \in \mathcal{G}} \widehat{g \circ F}(w). \tag{3}$$

Remark 1. If $w = 0$ in equation (2), then it does not involve the input x and it is not useful for correlation attack. Thus we let $w \neq 0$ when computing $nonlin_{UN}F$ which gauge the effectiveness of equation (2) for correlation attack. The other part $\min_{u \neq 0}(wt(u \cdot F), 2^n - wt(u \cdot F))$ ensures that $F(x)$ is close to balanced when UN_F is high.

From equation (3), we deduce that a high unrestricted nonlinearity is required for protection against correlation attack on $g \circ F(x)$.

In this paper, we introduce a linear approximation for performing correlation attack, which is more effective than the Zhang-Chan attack [15]. The idea is to consider implicit equations which are linear in the input variable x and of any degree in the output variable $z = F(x)$. In the pre-processing stage, we compute

$$Pr(g(z) + w_1(z)x_1 + w_2(z)x_2 + \cdots + w_n(z)x_n = 0), \tag{4}$$

where $z = F(x)$ and $w_i : GF(2)^m \to GF(2)$. In other words, we consider the probability $Pr(g(F(x)) + w_1(F(x))x_1 + w_2(F(x))x_2 + \cdots + w_n(F(x))x_n = 0)$, where x uniformly ranges over F_2^n. Then in the attack, because $z = F(x)$ corresponds to the output keystream which is known, $g(z)$ and $w_i(z)$ are known for all $i = 1, \ldots, n$. This means that we can substitute the known values $z = F(x)$ and treat equation (4) as a linear approximation.

We call the attack based on this linear approximation the *generalized correlation attack*. This attack can be considered as a generalization of Zhang-Chan's correlation attack because if we let $w_i(z)$ constant for $i = 1 \ldots n$, equation (4) becomes equation (2). Since we are choosing from a larger set than that of Zhang and Chan, it is easier to find a linear approximation with larger bias $|Pr(g(z) + w_1(z)x_1 + w_2(z)x_2 + \cdots w_n(z)x_n = 0) - 1/2|$.

In relation to the approximation of equation (4), we make the following definition:

Definition 2. *Let $F : GF(2)^n \rightarrow GF(2)^m$. The* generalized Hadamard transform *$\hat{F} : (GF(2)^{2^m})^{n+1} \rightarrow \mathbb{R}$ is defined as:*

$$\hat{F}(g(\cdot), w_1(\cdot), \dots, w_n(\cdot)) = \sum_{x \in GF(2)^n} (-1)^{g(F(x))+w_1(F(x))x_1+\cdots w_n(F(x))x_n},$$

where the input is an $(n+1)$-tuple of Boolean functions $g, w_i : GF(2)^m \rightarrow GF(2)$, $i = 1, \dots, n$.

Let $\mathcal{G}$ be defined as in Definition 1 and let

$$\begin{aligned} \mathcal{W} = \; & \textit{Set of all n-tuple functions } \{w(\cdot) = (w_1(\cdot), \dots, w_n(\cdot)) | w_i \in \mathcal{G}\} \\ & \textit{such that } w(z) = (w_1(z), \dots, w_n(z)) \neq (0, \dots, 0) \textit{ for all } z \in GF(2)^m. \end{aligned}$$

The generalized nonlinearity *is defined as:*

$$GN_F = \min\{ \min_{0 \neq u \in GF(2)^m} (wt(u \cdot F), 2^n - wt(u \cdot F)), nonlin_{gen}F\},$$

where

$$nonlin_{gen}F = 2^{n-1} - \frac{1}{2} \max_{g \in \mathcal{G}, w \in \mathcal{W}} \hat{F}(g(\cdot), w_1(\cdot), \dots, w_n(\cdot)). \qquad (5)$$

Remark 2. We introduce the set $\mathcal{W}$ to give a meaningful definition to the generalized nonlinearity. This is because if there exists $z \in GF(2)^m$ such that $(w_1(z), \dots, w_n(z)) = (0, \dots, 0)$, then the equation (4) resulting from this value of z will not involve the input x and will therefore not be useful in the attack stage. Thus we let $w \in \mathcal{W}$ when computing $nonlin_{gen}F$. The other part $\min_{u \neq 0}(wt(u \cdot F), 2^n - wt(u \cdot F))$ ensures that $F(x)$ is close to balanced when GN_F is high.

From equation (5), we deduce that a high generalized nonlinearity is required for protection against generalized correlation attack.

In Proposition 1, we show that the generalized nonlinearity is lower than the other nonlinearity measures and thus provides linear approximations with better bias for correlation attack. The proof follows naturally from the definitions of the various nonlinearities.

Proposition 1 *Let $F : GF(2)^n \rightarrow GF(2)^m$. Then the nonlinearity, unrestricted nonlinearity and generalized nonlinearity are related by the following inequality:*

$$GN_F \leq UN_F \leq N_F. \qquad (6)$$

I.e., the generalized correlation attack is more effective than the Zhang-Chan's correlation attack, which itself is more effective than the usual correlation attack.

Remark 3. A vector function $F : GF(2)^n \rightarrow GF(2)^m$ is said to be balanced if $|F^{-1}(z)| = 2^{n-m}$ for all $z \in GF(2)^m$. It can be deduced that $wt(u \cdot F) = 2^{n-1}$ for all $u \in GF(2)^m - \{0\}$ if and only if F is balanced [3]. Thus

$$\begin{aligned} GN_F &= \min\{ \min_{0 \neq u \in GF(2)^m} (wt(u \cdot F), 2^n - wt(u \cdot F)), nonlin_{gen}F\} \\ &= \min(2^{n-1}, nonlin_{gen}F) = nonlin_{gen}F, \end{aligned}$$

because $GN_F \leq N_F \leq 2^{n-1} - 2^{(n-1)/2}$ [3]. Therefore $GN_F = nonlin_{gen}F$ if F is balanced. In a similar way, $UN_F = nonlin_{UN}F$ if F is balanced.

3 Efficient Computation of the Generalized Nonlinearity

To compute the generalized nonlinearity GN_F, we first compute $\min_{u\neq 0}(wt(u \cdot F), 2^n - wt(u \cdot F))$ with complexity 2^{m+n}. Then we need to compute $nonlin_{gen}F$ which requires computation of the generalized Hadamard transform over all input. But the complexity of computing $\hat{F}(g(\cdot), w_1(\cdot), \ldots, w_n(\cdot))$ directly for all possible $(n+1)$-tuple of m-bit functions is $\approx 2^n \times 2^{2^m \times (n+1)}$. This is because for each fixed $(g(\cdot), w_1(\cdot), \ldots, w_n(\cdot))$, we sum over 2^n elements x to compute $\hat{F}$ and there are approximately[1] $2^{2^m \times (n+1)}$ tuples of functions $g, w_i : GF(2)^m \to GF(2)$, $i = 1, \ldots, n$. This computation quickly becomes unmanageable even for small values of n, m. Since the bulk of the computational time comes from $nonlin_{gen}F$, we need to make it more efficient to compute. Theorem 1 below gives an efficient way to compute $nonlin_{gen}F$.

Theorem 1 *Let $F : GF(2)^n \to GF(2)^m$ and let $w(\cdot)$ denote the n-tuple of m-bit Boolean functions $(w_1(\cdot), \ldots, w_n(\cdot))$. Then*

$$nonlin_{gen}F = 2^{n-1} - 1/2 \sum_{z \in GF(2)^m} \max_{w(z) \in GF(2)^n - \{0\}} \left| \sum_{x \in F^{-1}(z)} (-1)^{w(z) \cdot x} \right| .$$

Because of editorial constraints, the proof of Theorem 1 can be found in the Appendix, Section 7.2.

Remark 4. The proof of Theorem 1 also provides the functions $g(\cdot), w_i(\cdot)$, $i = 1, \ldots, n$, for the best generalized linear approximation. At each z, the optimal $g(z)$ is the one that makes the inner sum positive while the optimal tuple $(w_1(z), \ldots, w_n(z))$ is the n-bit vector that maximizes the inner sum.

3.1 Reduction in Complexity

To compute $nonlin_{gen}F$ based on Theorem 1, we first perform a pre-computation to identify the sets $\{x : x \in F^{-1}(z)\}$ with complexity 2^n and store them with memory of size $n \times 2^n$. This is needed in computing the sum $\sum_{x \in F^{-1}(z)} (-1)^{w(z) \cdot x}$.

To compute $nonlin_{gen}F$, we consider the 2^m elements $z \in GF(2^m)$. For each z, we find $w(z) \in GF(2)^n$ which maximizes the sum $\left|\sum_{x \in F^{-1}(z)} (-1)^{w(z) \cdot x}\right|$. Thus the computational complexity is:

$$\begin{aligned} \text{Complexity} &= \sum_{z \in GF(2)^m} 2^n \times |\{x : x \in F^{-1}(z)\}| = 2^n \sum_{z \in GF(2)^m} |\{x : x \in F^{-1}(z)\}| \\ &= 2^n \times |Domain(F)| = 2^n \times 2^n = 2^{2n}. \end{aligned}$$

[1] We say approximately $2^{2^m \times (n+1)}$ functions because we do not range over all tuples of functions $(w_1(\cdot), \ldots, w_n(\cdot))$ but only over those in the set $\mathcal{W}$ of Defintion 2

Together with a complexity of 2^{m+n} to compute $\min_{0\neq u\in GF(2)^m}(wt(u\cdot F), 2^n - wt(u\cdot F))$, the total complexity for computing GN_F is:

$$\text{Precomputation } = 2^n, \text{ Memory } = n\times 2^n, \text{ Time Complexity } = 2^{m+n} + 2^{2n}.$$

This is much less than a time complexity of $2^{m+n} + 2^{n+2^m\times(n+1)}$ by the direct approach.

3.2 Experimental Results

Based on Theorem 1, we can compute the generalized nonlinearity of some highly nonlinear functions. We also computed the unrestricted nonlinearity of these functions for comparison. We shall apply Proposition 6 (in Section 7.3 of the Appendix) to help us compute $nonlin_{UN}F$ efficiently. First, let us look at bent functions, which have the highest nonlinearity.

Example 1. Consider the bent function $F : GF(2)^4 \to GF(2)^2$ (i.e. the function whose component functions $u\cdot F$, $u \neq 0$, are all bent) defined by $F(x_1, x_2, x_3, x_4) = (z_1, z_2) = (x_1+x_1x_4+x_2x_3, x_1+x_1x_3+x_1x_4+x_2x_4)$. The truth table of F (which lists the output $F(0000), F(0001), \ldots, F(1111)$ where every number represents its binary representation) is as follows.

```
0  0  0  0  0  1  2  3  3  0  2  1  3  1  0  2
```

The various nonlinearity and bias take the following values:

$$\text{Usual nonlinearity } N_F = 6 \Rightarrow Bias = 0.125$$
$$\text{unrestricted nonlinearity } UN_F = 5 \Rightarrow Bias = 0.1875$$
$$\text{Generalized nonlinearity } GN_F = 2 \Rightarrow Bias = 0.375.$$

From Remark 4, we deduce that the following approximation holds with bias 0.375.

$$Pr(z_1 + z_2 = (z_1 + 1)(z_2 + 1)x_2 + z_1x_3 + z_2x_4) = \frac{14}{16},$$

where $x = 0100, 1110$ are the only two points not satisfying the relation.

Next we look at the inverse S-box on $GF(2^8)$ with truncated output.

Example 2. Let $GF(2^8)$ be the finite field defined by the relation $\alpha^8 + \alpha^4 + \alpha^3 + \alpha^2 + 1 = 0$. Consider the S-box $Inv : GF(2)^8 \to GF(2)^8$ defined by $Inv(0) = 0$ and $Inv(x) = x^{-1}$ if $x \neq 0$. We use the correspondence:

$$(x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8) \leftrightarrow x_1\alpha^7 + x_2\alpha^6 + \cdots + x_7\alpha + x_8$$

Consider $Inv(x)$ restricted to the least significant m bits. Then the nonlinearity, unrestricted nonlinearity and generalized nonlinearity are given by Table 1. We see that the generalized nonlinearity for the inverse function restricted to m output bits is lower than the usual and unrestricted nonlinearities. Therefore generalized correlation attack works better in this case. Moreover, for $m \geq 5$ output bits, the generalized nonlinearity is already 0 which means the system can be broken by linear algebra with very few keystream bits.

Table 1. Nonlinearities for x^{-1} on $GF(2^8)$ restricted to m least significant output bits.

m	1	2	3	4	5	6	7
N_F	112	112	112	112	112	112	112
UN_F	112	108	100	94	84	70	56
GN_F	112	80	66	40	0	0	0

Example 3. Lastly in Table 2, we tabulate the average nonlinearity measures for 100 randomly generated balanced functions $F : GF(2)^n \to GF(2)^m$, $n = 2m$, for various n. Again, we see that the average generalized nonlinearity is much lower than the unrestricted and usual nonlinearities. Therefore generalized correlation attack can be more effective.

Table 2. Average nonlinearity for randomly generated balanced functions, $n = 2m$

n	6	8	10	12	14
N_F	18	100	443	1897	7856
UN_F	16	88	407	1768	7454
GN_F	6	36	213	1101	5224

4 Upper Bound on Generalized Nonlinearity

In this Section, we prove an upper bound for the generalized nonlinearity. This allows us to gauge theoretically the effectiveness of the generalized correlation attack.

Theorem 2 *Let $F : GF(2)^n \to GF(2)^m$. Then the following inequality holds.*

$$nonlin_{gen}F \leq 2^{n-1} - \frac{1}{4} \sum_{z \in GF(2)^m} \sqrt{\frac{2^{n+2}|F^{-1}(z)| - 4|F^{-1}(z)|^2}{2^n - 1}}.$$

Furthermore if $F(x)$ is balanced, then we have:

$$GN_F \leq 2^{n-1} - 2^{n-1}\sqrt{\frac{2^m - 1}{2^n - 1}}$$

Proof. According to Theorem 1, we have:

$$nonlin_{gen}F = 2^{n-1} - 1/2 \sum_{z \in GF(2)^m} \max_{a \in GF(2)^n - \{0\}} \left| \sum_{x \in F^{-1}(z)} (-1)^{a \cdot x} \right|.$$

Let $\phi_z(x)$ be the indicator function of $F^{-1}(z)$. I.e., $\phi_z(x) = 1$ if $F(x) = z$ else $\phi_z(x) = 0$.

$$\sum_{x \in F^{-1}(z)} (-1)^{a \cdot x} = \sum_{x \in GF(2)^n} \phi_z(x)(-1)^{a \cdot x} = \sum_{x \in GF(2)^n} \frac{1 - (-1)^{\phi_z(x)}}{2}(-1)^{a \cdot x}$$

$$= -\frac{1}{2} \sum_{x \in GF(2)^n} (-1)^{\phi_z(x) + a \cdot x} = -\frac{1}{2}\widehat{\phi_z}(a), \text{ when } a \neq 0.$$

Thus

$$nonlin_{gen}F = 2^{n-1} - 1/4 \sum_{z \in GF(2)^m} \max_{a \in GF(2)^n - \{0\}} \left|\widehat{\phi_z}(a)\right|.$$

In a similar way to the computation of $\sum_{x \in F^{-1}(z)}(-1)^{a \cdot x}$, we can prove that $|F^{-1}(z)| = \sum_{x \in F^{-1}(z)}(-1)^{0 \cdot x} = 2^{n-1} - \frac{1}{2}\widehat{\phi_z}(0)$. This implies $\widehat{\phi_z}(0) = 2^n - 2|F^{-1}(z)|$.

By Parseval's relation,

$$\sum_{a \in GF(2)^n - \{0\}} \widehat{\phi_z}(a)^2 = 2^{2n} - \widehat{\phi_z}(0)^2$$

$$= 2^{2n} - (2^n - 2|F^{-1}(z)|)^2 = 2^{n+2}|F^{-1}(z)| - 4|F^{-1}(z)|^2.$$

By the pigeon hole principle, we deduce that

$$\max_{a \in GF(2)^n - \{0\}} \widehat{\phi_z}(a)^2 \geq \frac{2^{n+2}|F^{-1}(z)| - 4|F^{-1}(z)|^2}{2^n - 1}.$$

and therefore

$$nonlin_{gen}F \leq 2^{n-1} - \frac{1}{4} \sum_{z \in GF(2)^m} \sqrt{\frac{2^{n+2}|F^{-1}(z)| - 4|F^{-1}(z)|^2}{2^n - 1}}.$$

When $F(x)$ is balanced, $nonlin_{gen}F = GN_F$, $|F^{-1}(z)| = 2^{n-m}$ for all $z \in GF(2)^m$ and we deduce:

$$GN_F \leq 2^{n-1} - 2^{n-1}\sqrt{\frac{2^m - 1}{2^n - 1}}$$

□

This upper bound is much lower than the covering radius bound $2^{n-1} - 2^{n/2-1}$ and the upper bound for UN_F deduced in [4]:

$$UN_F \leq 2^{n-1} - \frac{1}{2}\left(\frac{2^{2m} - 2^m}{2^n - 1} + \sqrt{\frac{2^{2n} - 2^{2n-m}}{2^n - 1} + \left(\frac{2^{2m} - 2^m}{2^n - 1} - 1\right)^2} - 1\right).$$

when $F : GF(2)^n \to GF(2)^m$ is balanced. Thus Theorem 2 provides further evidence that generalized correlation attack can be more effective than the usual and Zhang-Chan correlation attacks on vector Boolean functions.

5 Spectral Characterization and Generalized Correlation Immunity

In Theorem 3, we express the generalized correlation in terms of the Hadamard transform (also called the spectrum) of $F(x)$. This allows us to deduce general correlation properties based on the spectral distribution.

Theorem 3 *Let $F : GF(2)^n \to GF(2)^m$ and $w_i : GF(2)^m \to GF(2)$. Let $w(\cdot)$ denote the n-tuple of m-bit Boolean functions $(w_1(\cdot), \dots, w_n(\cdot))$. Then the generalized Hadamard transform can be expressed as:*

$$\hat{F}(g(\cdot), w_1(\cdot), \dots, w_n(\cdot)) = \frac{1}{2^m} \sum_{z \in GF(2)^m} (-1)^{g(z)} \sum_{v \in GF(2)^m} (-1)^{v \cdot z} \widehat{v \cdot F}(w(z)).$$

The proof of Theorem 3 is easy and can be found in the Appendix, Section 7.4.

Remark 5. Based on Theorem 3 and equation (5), we get

$$nonlin_{gen}F = 2^{n-1} - \frac{1}{2^{m+1}} \sum_{z \in GF(2)^m} \max_{\substack{0 \neq w(z) \in \\ GF(2)^n}} \left| \sum_{v \in GF(2)^m} (-1)^{v \cdot z} \widehat{v \cdot F}(w(z)) \right| . \quad (7)$$

If the Hadamard transform distribution of $F(x)$ is known, then we can have a more efficient computation of GN_F. By equation (7), we compute $nonlin_{gen}F$ by an outer sum over 2^m elements z, each of which finds the maximum inner sum (over 2^m elements v) for 2^n choices of $w(z)$. Thus the complexity of computing $nonlin_{gen}F$ is 2^{n+2m}. Together with a complexity of 2^{m+n} for determining the balanceness of $F(x)$, the complexity for computing GN_F is $2^{m+n} + 2^{n+2m}$. This is more efficient than the computation of Theorem 1 because usually, m is much smaller than n in applications. Furthermore, we do not need pre-computation and memory to store the sets $\{x : x \in F^{-1}(z)\}$ as in Theorem 1. Some examples of vector functions with known spectral distribution is the Maiorana-McFarland class which can be used to construct bent functions and highly nonlinear resilient functions, e.g. see [3, 5].

5.1 Equivalence of Generalized Correlation Immunity and Usual Correlation Immunity

In this section, we extend the definition of correlation immunity (resiliency) for vectorial Boolean function to the generalized case with respect to the correlation attack based on equation (4). Then we show that the usual correlation immunity (resiliency) implies generalized correlation immunity (resiliency). First let us recall the definition of correlation immune vectorial Boolean functions. For a vector $w \in GF(2)^n$, denote by $wt(w)$ the number of ones in w.

Definition 3. *The vector function $F : GF(2)^n \to GF(2)^m$ is correlation immune of order t (denoted $CI(t)$) if*

$$u \cdot F(x) + w \cdot x \text{ is balanced, or equivalently } \widehat{u \cdot F}(w) = 0$$

whenever $1 \leq wt(w) \leq t$. Moreover if $F(x)$ is balanced, then $F(x)$ is t-resilient.

Resiliency is essential for protection against divide-and-conquer correlation attack on combinatorial generator (for more details, please see Siegenthaler [14]).

Next, let us describe a generalized divide-and-conquer attack against vector combinatorial stream ciphers. In a combinatorial generator involving n LFSR's, suppose there is a subset of outputs $z \in GF(2)^m$ for which the linear approximations $Pr(w_{i_1}(z)x_{i_1} + \cdots + w_{i_t}(z)x_{i_t} = g(z)) = p_z \neq 1/2$, involve the corresponding set of t linear feedback shift registers, $LFSR_{i_1}, \ldots, LFSRi_t$. The attacker guesses the initial state of $LFSR_{i_1}, \ldots, LFSR_{i_t}$. If the guess is correct, then this relation should hold between the t LFSR's states and the relevant output[2] z with probability $p_z \neq 1/2$. If the guess is wrong, then the LFSR states and the output are uncorrelated. Thus the complexity of guessing the secret initial state is reduced because we only need to guess the content of t instead of n LFSR's. To protect against such an attack, we define the concept of generalized correlation immunity and resiliency as follows.

Definition 4. *Let $F : GF(2)^n \to GF(2)^m$ and $g, w_i : GF(2)^m \to GF(2)$. We say $F(x)$ is* generalized correlation immune *of order t (generalized $CI(t)$) if*

$$g(F(x)) + w_1(F(x))x_1 + \cdots + w_n(F(x))x_n \textit{ is balanced,}$$

or equivalently,

$$\hat{F}(g(\cdot), w_1(\cdot), \ldots, w_n(\cdot)) = 0,$$

whenever $1 \leq wt(w_1(z), \ldots, w_n(z)) \leq t$ for all $z \in GF(2)^m$. Moreover if $F(x)$ is balanced, then we say $F(x)$ is generalized t-resilient.

Generalized t-resiliency ensures protection against generalized divide-and-conquer correlation attack on t or less LFSR's in a combinatorial stream cipher.

Theorem 4 *Let $F : GF(2)^n \to GF(2)^m$. Then $F(x)$ is $CI(t)$ (t-resilient) if and only if $F(x)$ is generalized $CI(t)$ (generalized t-resilient).*

Proof. If $F(x)$ is generalized $CI(t)$ (resp. generalized t-resilient), then it follows from Definitions 3 and 4 that $F(x)$ is $CI(t)$ (resp. t-resilient). Now assume $F(x)$ is $CI(t)$, we shall prove that $F(x)$ is generalized $CI(t)$. Suppose $1 \leq wt(w_1(z), \ldots, w_n(z)) \leq t$ for all $z \in GF(2)^m$. Then $\widehat{v \cdot F}(w(z)) = 0$ for all $v, z \in GF(2)^m$ because $F(x)$ is $CI(t)$. By Theorem 3, we see that

$$\hat{F}(g(\cdot), w_1(\cdot), \ldots, w_n(\cdot)) = \frac{1}{2^m} \sum_{z \in GF(2)^m} (-1)^{g(z)} \sum_{v \in GF(2)^m} (-1)^{v \cdot z} \widehat{v \cdot F}(w(z)) = 0.$$

This is because the inner summands is a sum of $\widehat{v \cdot F}(w(z))$ which are zeroes for all $v, z \in GF(2)^m$. Thus $F(x)$ is generalized $CI(t)$. The proof that t-resiliency implies generalized t-resiliency is identical to the $CI(t)$ case except that $F(x)$ is now balanced.

□

Thus we see that usual resiliency is sufficient to ensure generalized resiliency.

[2] By relevant output, we mean those $z \in GF(2)^m$ for which there exist a linear approximation with positive bias involving the same set of input $x_{i_1}, \ldots, x_{i_t}$

6 Generalized Nonlinearity of Secondary Constructions

Secondary constructions produce Boolean functions with high nonlinearity, resiliency and other good cryptographic properties from other Boolean functions as building blocks. With respect to the generalized correlation attack, it would be useful to check if these constructions yield functions with high generalized nonlinearity. Moreover by Theorem 4, vector functions that satisfy the usual correlation immunity are also generalized correlation immune. Thus we would also like to check that secondary construction for resilient functions have high generalized nonlinearity.

The first secondary construction we look at is output composition. One common candidate for output composition is the projection function, i.e. dropping output bits. For example, there are many known permutations with high nonlinearity [1] and by dropping output bits, we form vectorial Boolean functions with the same or higher nonlinearity.

Proposition 2 *Let* $F : GF(2)^n \to GF(2)^m$ *and* $G : GF(2)^m \to GF(2)^k$ *be balanced functions. Then* $GN_{G\circ F} \geq GN_F$*. If* $G(z)$ *is a permutation, then* $GN_{G\circ F} = GN_F$*.*

The proof of Proposition 2 can be found in the Appendix, Section 7.5. By Proposition 2, we see that output composition, e.g. dropping output bits, is good for enhancing security as it may increase the generalized nonlinearity.

Another common construction for vectorial resilient functions is concatenation. Let us look at the known results on this construction.

Proposition 3 *([16, Corollary 4]) Let* $F_1 : GF(2)^{n_1} \to GF(2)^{m_1}$ *be a* t_1*-resilient function and* $F_2 : GF(2)^{n_2} \to GF(2)^{m_2}$ *be a* t_2*-resilient function. Then* $F_1||F_2 : GF(2)^{n_1+n_2} \to GF(2)^{m_1+m_2}$ *defined by*

$$F_1||F_2(x, y) = (F_1(x), F_2(y))$$

is a t*-resilient function where* $t = \min(t_1, t_2)$*.*

By Proposition 3, given two smaller vector Boolean functions which are t-resilient, we can form a bigger Boolean function which is t-resilient. With respect to generalized correlation attack, we would like to know its generalized nonlinearity.

Proposition 4 *Let* $F_1 : GF(2)^{n_1} \to GF(2)^{m_1}$ *and* $F_2 : GF(2)^{n_2} \to GF(2)^{m_2}$ *be balanced functions. Then the generalized nonlinearity of their concatenation* $F(x, y) = F_1(x)||F_2(y)$ *satisfies:*

$$GN_F \leq 2^{n_1+n_2-1} - \frac{1}{2}(2^{n_1} - 2GN_{F_1})(2^{n_2} - 2GN_{F_2}).$$

The proof of Proposition 4 can be found in the Appendix, Section 7.6. By Proposition 4, we see that for a concatenated function to possess high generalized nonlinearity, both the component functions have to possess high generalized nonlinearity.

References

1. A. Canteaut, P. Charpin and H. Dobbertin, “Binary m-sequences with three-valued cross correlation: a proof of Welch’s conjecture”, *IEEE Trans. Inform. Theory*, vol. 46 no. 1, pp. 4-8, 2000.
2. A. Canteaut and M. Trabbia. Improved fast correlation attacks using parity-check equations of weight 4 and 5, *Advanced in Cryptology-EUROCRYPT 2000. Lecture notes in computer science* 1807, pp. 573-588, 2000.
3. C. Carlet, “Vectorial Boolean Functions for Cryptography”, to appear in *Boolean Methods and Models* published by Cambridge University Press, Eds Yves Crama and Peter Hammer. Can be found at http://www-rocq.inria.fr/codes/Claude.Carlet/chap-vectorial-fcts.pdf.
4. C. Carlet and E. Prouff, “On a New Notion of Nonlinearity Relevant to Multi-Output Pseudo-Random Generators”, LNCS 3006, *Selected Areas in Cryptography 2003*, pp. 291-305, Springer-Verlag, 2003.
5. C. Carlet and E. Prouff, “Vectorial Functions and Covering Sequences”, LNCS 2948, *International Conference on Finite Fields and Applications*, pp. 215-248, Springer-Verlag, 2003.
6. B. Chor, O. Goldreich, J. Hastad, J. Friedman, S. Rudich and R. Smolensky, “The Bit Extraction Problem or t-resilient Functions”, *IEEE Symposium on Foundations of Computer Science 26*, pp. 396-407, 1985.
7. J.F. Dillon, “Multiplicative Difference Sets via Additive Characters”, *Designs, Codes and Cryptography*, vol. 17, pp. 225-235, 1999.
8. R. Gold, “Maximal Recursive Sequences with 3-valued Cross Correlation Functions”, *IEEE Trans. Inform. Theory*, vol. 14, pp. 154-156, 1968.
9. K.C. Gupta and P. Sarkar, “Improved Construction of Nonlinear Resilient S-boxes”, LNCS 2501, *Asiacrypt 2002*, pp. 466-483, Springer-Verlag, 2002.
10. K. Nyberg, “On the Construction of Highly Nonlinear Permutations”, LNCS 658, *Eurocrypt’92*, pp. 92-98, Springer-Verlag, 1993.
11. E. Pasalic and S. Maitra, “Linear Codes in Constructing Resilient Functions with High Nonlinearity”, LNCS 2259, *Selected Areas in Cryptography 2001*, pp. 60-74, Springer-Verlag, 2001.
12. R. Rueppel, *Analysis and Design of Stream Ciphers*, Springer-Verlag, 1986.
13. P. Sarkar, “The Filter-Combiner Model for Memoryless Synchronous Stream Ciphers”, LNCS 2442, *Crypto 2002*, pp. 533-548, Springer-Verlag, 2002.
14. T. Siegenthaler, “Decrypting a Class of Stream Ciphers using Ciphertexts only”, *IEEE Transactions on Computers*, vol. C34, no. 1, pp. 81-85, 1985.
15. M. Zhang and A. Chan, “Maximum Correlation Analysis of Nonlinear S-boxes in Stream Ciphers”, LNCS 1880, *Crypto’2000*, pp. 501-514, Springer-Verlag, 2000.
16. X.M. Zhang and Y. Zheng, “On Cryptographically Resilient Functions”, *IEEE Transaction on Information Theory*, Vol. 43, no.5, pp. 1740-1747, 1997. (Also presented at *Eurocrypt’95*, LNCS 921, pp. 274-288, Springer-Verlag, 1995).

7 Appendix

7.1 The Single-Bit Output Case and Bilinear Cryptanalysis

It is easy to see that in the single output case ($m = 1$), the Zhang-Chan correlation attack is equivalent to the usual correlation attack, i.e. $UN_F = N_F$.

However, it is not so obvious whether the generalized correlation attack is better than the usual correlation attack. The four functions from $GF(2)$ to $GF(2)$ are of the form $w(z) = az + b$, where $a, b \in GF(2)$. Hence, the expression used for the generalized correlation attack is a bilinear approximation:

$$Pr(a_0z+b_0+(a_1z+b_1)x_1+(a_2z+b_2)x_2+\cdots+(a_nz+b_n)x_n = 0),\ a_i, b_i \in GF(2),$$

where for any $z \in GF(2)$, we have $(a_1z + b_1, \ldots, a_nz + b_n) \neq (0, \ldots, 0)$. The above equation can also be written as:

$$Pr(za'(x) = a(x)) \text{ where } a(x), a'(x) \text{ are affine functions,} \tag{8}$$

such that $za'(x)+a(x)$ is a non-constant function for every $z \in GF(2)$. In Proposition 5, we show that generalized nonlinearity is equal to the usual nonlinearity in the single output case.

Proposition 5 *Let $f : GF(2)^n \to GF(2)$. Then $GN_f = N_f$.*

Proof. In this proof, $a(x)$, $a'(x)$ and $a''(x) = a(x)+a'(x)+1$ are affine functions. We also require that $a(x)$ and $a''(x)$ be non-constant functions, so that the approximation in equation (8) is useful for correlation attack. From equation (5) and the discussion in Section 7.1, we deduce that:

$$\begin{aligned} nonlin_{gen}F &= \min_{a(x),a'(x)} |\{x : f(x)a'(x) = a(x)\}| \\ &= \min_{a(x),a'(x)} (|\{x : f(x) = a(x) = 0\}| + |\{x : f(x) = a(x) + a'(x) + 1 = 1\}|) \\ &= \min_{a(x)} |\{x : f(x) = a(x) = 0\}| + \min_{a''(x)} |\{x : f(x) = a''(x) = 1\}|. \end{aligned}$$

On the other hand, we see from equation (1) that:

$$\begin{aligned} N_f &= \min_{a(x)} |\{x : f(x) = a(x)\}| \\ &= \min_{a(x)}(|\{x : f(x) = a(x) = 0\}| + |\{x : f(x) = a(x) = 1\}|). \end{aligned}$$

But

$$\begin{aligned} |\{x : f(x) = a(x) = 1\}| &= |f^{-1}(1)| + |\{x : f(x) = a(x) = 0\}| - |a^{-1}(0)| \\ &= |f^{-1}(1)| + |\{x : f(x) = a(x) = 0\}| - 2^{n-1}. \end{aligned}$$

Thus

$$N_f = \min_{a(x)}(2 \times |\{x : f(x) = a(x) = 0\}| + c) \text{ where } c = |f^{-1}(1)| - 2^{n-1}.$$

From this, we deduce that:

$$\min_{a(x)} |\{x : f(x) = a(x) = 0\}| = \frac{N_f - c}{2},\ \ \min_{a''(x)} |\{x : f(x) = a''(x) = 1\}| = \frac{N_f + c}{2}.$$

By combining the above two expressions, we get:

$$N_f = \min_{a(x)} |\{x : f(x) = a(x) = 0\}| + \min_{a''(x)} |\{x : f(x) = a''(x) = 1\}| = nonlin_{gen}F.$$

Also $\min_{0 \neq u \in GF(2)^m}(wt(u \cdot F), 2^n - wt(u \cdot F)) \geq N_f$. Thus $GN_f = N_f$.

□

Although generalized correlation attack does not improve on the usual correlation attack when $m = 1$, we can see in Section 3.2 many examples where generalized correlation attack yields better results than the usual and Zhang-Chan correlation attack when the number of output bits is $m \geq 2$.

7.2 Proof of Theorem 1

Proof. We have:

$$\begin{aligned}
&\max_{g \in \mathcal{G}, w \in \mathcal{W}} \hat{F}(g(\cdot), w_1(\cdot), \ldots, w_n(\cdot)) \\
= &\max_{g \in \mathcal{G}, w \in \mathcal{W}} \sum_{z \in GF(2)^m} (-1)^{g(z)} \sum_{x \in F^{-1}(z)} (-1)^{w(z) \cdot x} \\
= &\sum_{z \in GF(2)^m} \max_{g(z) \in GF(2), w(z) \in GF(2)^n - \{0\}} (-1)^{g(z)} \sum_{x \in F^{-1}(z)} (-1)^{w(z) \cdot x}.
\end{aligned}$$

To maximize this expression, we choose $g(z) = 0$ if $\sum_{x \in F^{-1}(z)} (-1)^{w(z) \cdot x} > 0$, else we choose $g(z) = 1$. Thus we can equivalently write the expression as:

$$\max_{g \in \mathcal{G}, w \in \mathcal{W}} \hat{F}(g(\cdot), w_1(\cdot), \ldots, w_n(\cdot)) = \sum_{z \in GF(2)^m} \max_{w(z) \in GF(2)^n - \{0\}} \left| \sum_{x \in F^{-1}(z)} (-1)^{w(z) \cdot x} \right|.$$

By substituting this expression in equation (5), we get $nonlin_{gen}F$.

□

7.3 Efficient Computation of Unrestricted Nonlinearity

The bulk of the work in computing UN_F comes from the computation of $nonlin_{UN}F$. Proposition 6 gives an efficient way to compute $nonlin_{UN}F$.

Proposition 6 *Let $F : GF(2)^n \to GF(2)^m$. Then $nonlin_{UN}F$ can be computed as:*

$$nonlin_{UN}F = 2^{n-1} - \frac{1}{2} \max_{w \neq 0} \sum_{z \in GF(2)^m} \left| \sum_{x \in F^{-1}(z)} (-1)^{w \cdot x} \right|.$$

Proof.

$$\begin{aligned}\max_{w\neq 0, g\in\mathcal{G}} \widehat{g\circ F}(w) &= \max_{w\neq 0, g\in\mathcal{G}} \sum_{x\in GF(2)^n} (-1)^{g\circ F(x)+w\cdot x} \\ &= \max_{w\neq 0, g\in\mathcal{G}} \sum_{z\in GF(2)^m} (-1)^{g(z)} \sum_{x\in F^{-1}(z)} (-1)^{w\cdot x} \\ &= \max_{w\neq 0} \sum_{z\in GF(2)^m} \left| \sum_{x\in F^{-1}(z)} (-1)^{w\cdot x} \right|.\end{aligned}$$

where we choose $g(z) = 0$ if the inner sum is positive and $g(z) = 1$ is the inner sum is negative. By substituting this expression in equation (3), Proposition 6 is proved.

□

7.4 Proof of Theorem 3

Proof. Let $\phi_z(x)$ be as defined in the proof of Theorem 2. For a fixed $z \in GF(2)^m$,

$$\begin{aligned}\sum_{x\in F^{-1}(z)} (-1)^{w(z)\cdot x} &= \frac{1}{2^m} \sum_{x\in GF(2)^n} (-1)^{w(z)\cdot x} \times 2^m \phi_z(x) \\ &= \frac{1}{2^m} \sum_{x\in GF(2)^n} (-1)^{w(z)\cdot x} \times \sum_{v\in GF(2)^m} (-1)^{v\cdot(F(x)+z)} \\ &\quad (\text{because} \sum_{v\in GF(2)^m} (-1)^{v\cdot a} = 2^m \text{ if and only if } a = 0) \\ &= \frac{1}{2^m} \sum_{v\in GF(2)^m} (-1)^{v\cdot z} \times \sum_{x\in GF(2)^n} (-1)^{w(z)\cdot x + v\cdot F(x)} \\ &= \frac{1}{2^m} \sum_{v\in GF(2)^m} (-1)^{v\cdot z} \widehat{v\cdot F}(w(z)).\end{aligned}$$

By substituting this expression in Lemma 1, the proof is complete.

□

7.5 Proof of Proposition 2

Proof. Let $\mathcal{G}, \mathcal{W}$ and $\mathcal{G}', \mathcal{W}'$ be the set of m-bit and k-bit Boolean functions in Definitions 1 and 2 respectively.

$$\begin{aligned}\max_{g'\in\mathcal{G}', w'\in\mathcal{W}'} \widehat{G\circ F}(g', w'_1, \ldots, w'_n) &= \max_{g'\in\mathcal{G}', w'\in\mathcal{W}'} \hat{F}(g'\circ G, w'_1\circ G, \ldots, w'_n\circ G) \\ &\leq \max_{g\in\mathcal{G}, w\in\mathcal{W}} \hat{F}(g, w_1, \ldots, w_n).\end{aligned}$$

Therefore by equation (5), $nonlin_{gen} G \circ F \geq nonlin_{gen} F$. Note that $w' \in W'$ implies $w' \circ G \in W$ in the above inequality.

Since $F(x)$ is balanced, $nonlin_{gen} F = GN_F$ by remark 3. It is easy to deduce that $G \circ F$ is balanced if both F and G are balanced. Thus $nonlin_{gen} G \circ F = GN_{G \circ F}$ by remark 3 and we have $GN_{G \circ F} \geq GN_F$.

If $G(z)$ is a permutation, then $\{g \circ G | g \in \mathcal{G}\} = \mathcal{G}$ and $\{(w_1 \circ G, \ldots, w_n \circ G) | w \in \mathcal{W}\} = \mathcal{W}$. Thus we have $nonlin_{gen} G \circ F = nonlin_{gen} F$ which implies $GN_{G \circ F} = GN_F$.

□

7.6 Proof of Proposition 4

Proof. Consider any $g_i : GF(2)^{m_i} \to GF(2)$, $i = 1, 2$ and any $w_{i,1}, \ldots, w_{i,n_i} : GF(2)^{m_i} \to GF(2)$, $i = 1, 2$ where for all $z \in GF(2)^{m_i}$, $(w_{i,1}(z), \ldots, w_{i,n_i}(z)) \neq (0, \ldots, 0)$. We see that:

$$\begin{aligned}
&\widehat{F_1}(g_1(\cdot), w_{1,1}(\cdot), \ldots, w_{1,n_1}(\cdot)) \times \widehat{F_2}(g_2(\cdot), w_{2,1}(\cdot), \ldots, w_{2,n_2}(\cdot)) \\
&= \sum_x (-1)^{g_1(F_1(x)) + w_{1,1}(F_1(x))x_1 + \ldots + w_{1,n_1}(F_1(x))x_{n_1}} \\
&\times \sum_y (-1)^{g_2(F_2(y)) + w_{2,1}(F_2(y))y_1 + \ldots + w_{2,n_2}(F_2(y))y_{n_2}} \\
&= \sum_{x,y} (-1)^{g(F_1(x),F_2(y)) + w_1(F_1(x),F_2(y))x_1 + \ldots + w_{n_1+n_2}(F_1(x),F_2(y))y_{n_2}} \\
&= \widehat{(F_1, F_2)}(g(\cdot), w_1(\cdot), \ldots, w_{n_1+n_2}(\cdot)).
\end{aligned}$$

where we let $g : GF(2)^{m_1+m_2} \to GF(2)$ be defined by $g(z_1, z_2) = g_1(z_1) + g_2(z_2)$. Let

$$\begin{aligned}
&w_1(z_1, z_2) = w_{1,1}(z_1), \ldots, w_{n_1}(z_1, z_2) = w_{1,n_1}(z_1), \\
&w_{n_1+1}(z_1, z_2) = w_{2,1}(z_2), \ldots, w_{n_1+n_2}(z_1, z_2) = w_{2,n_2}(z_2).
\end{aligned}$$

For all $(z_1, z_2) \in GF(2)^{m_1+m_2}$, it is obvious that $(w_1(z_1, z_2), \ldots, w_{n_1+n_2}(z_1, z_2)) \neq (0, \ldots, 0)$.

Since on the left hand side of the above equations $g(\cdot)$ and $w_{i,j}(\cdot)$ can be any functions while the g, w_i defined on the right hand side are only functions on $(z_1, z_2) \in GF(2)^{m_1+m_2}$ of a special form, we have:

$$\begin{aligned}
&\max_{g_1, w_{1,i}} \widehat{F_1}(g_1(\cdot), w_{1,i}(\cdot)) \times \max_{g_2, w_{2,i}} \widehat{F_2}(g_2(\cdot), w_{2,i}(\cdot)) \\
&\leq \max_{g, w_i} \widehat{(F_1 || F_2)}(g(\cdot), w_1(\cdot), \ldots, w_{n_1+n_2}(\cdot).
\end{aligned}$$

By substituting this inequality in equation (5), we get

$$nonlin_{gen}(F_1 || F_2) \leq 2^{n_1+n_2-1} - \frac{1}{2}(2^{n_1} - 2nonlin_{gen} F_1)(2^{n_2} - 2nonlin_{gen} F_2). \tag{9}$$

Since $F_1(x)$ and $F_2(y)$ are balanced functions, we have $nonlin_{gen}F_i = GN_{F_i}$ by remark 3. Furthermore, it is easy to see that $(F_1(x), F_2(y))$ is a balanced function. Thus $nonlin_{gen}(F_1||F_2) = GN_{(F_1||F_2)}$ by remark 3. Thus we can substitute all the $nonlin_{gen}F$ in equation (9) by GN_F and we are done.

□