Constructions of vector output Boolean functions with high generalized nonlinearity

Abstract Carlet et al. recently introduced generalized nonlinearity to measure the ability to resist the improved correlation attack of a vector output Boolean function. This article presents a construction of vector output Boolean functions with high generalized nonlinearity using the $\epsilon$-biased sample space. The relation between the resilient order and generalized nonlinearity is also discussed.

Keywords Boolean functions, correlation attack, generalized nonlinearity, almost resilient functions

1 Introduction

The $n$-bit to $m$-bit vectorial Boolean functions are used in the pseudo-random generators of stream ciphers. The generation of the keystream consists of a linear part, producing a sequence with a large period. This part is usually composed of several linear feedback shift registers (LFSR) and a nonlinear combining function $F$, which produces the output, given the state of the linear part. Such stream ciphers have higher throughput than those using single-bit output Boolean functions.

Siegenthaler presented a basic attack named correlation attack on stream ciphers. He observed that if the linear approximation between the output of some LFSRs and keystream has probability biased form $1/2$, then, the secret bits (initial state bits) of some LFSRs may be recovered when enough keystream bits are known [1]. Zhang and Chan improved this attack at Crypto 2000. They composed vector output function with any Boolean function (not only linear function) [2]. Therefore, the number of possible linear approximations is $2^{n-m}$ compared to just $2^m$ in the Siegenthaler’s attack. To measure the effectiveness of a function against Zhang-Chan attack, unrestricted nonlinearity is proposed and the functions with high unrestricted nonlinearity are also constructed in Ref. [3].

Recently, Carlet et al. generalized the Zhang-Chan attack by considering the linear approximations of input, which are free degree in the output, instead of linear approximations of input in the Zhang-Chan case [4]. Thus, there are $2^{n-m}$ linear approximations from which we can choose one with higher bias than the Zhang-Chan attack and the usual correlation attack. Similar to the usual nonlinearity and unrestricted nonlinearity, a new notation of nonlinearity, generalized nonlinearity, denoted by $G_N(f)$ of a vector output Boolean function $f$, is then introduced to measure the ability to resist the generalized correlation attack. The upper and lower bounds of $G_N(f)$ are also derived in Ref. [4].

For single output Boolean functions, the generalized nonlinearity has been proven to be equivalent to the usual nonlinearity [4]. It is also easy to see that the unrestricted nonlinearity is equivalent to the usual one. Therefore, in the single output case, these three notions of nonlinearity are equivalent. However, for vector output functions, these are always different. In fact, there are functions with high nonlinearity but zero unrestricted and generalized nonlinearity. Furthermore, functions with high unrestricted nonlinearity but zero generalized nonlinearity are also illustrated in Refs. [3, 4]. Although the lower bounds of the generalized nonlinearity of some special functions with high nonlinearity such as vector almost bent functions have been obtained [4], little is known about the construction of a vector output function with high generalized nonlinearity. The article will address this problem.

The notations of $\epsilon$-biased sample space were introduced by Naor and Naor [5], which has been proven to have several cryptographic applications, such as multiple authentication codes [6], almost security cryptographic Boolean functions [7], and so on. For more information about $\epsilon$-biased sample space, see [8].

The article is organized as follows. Some definitions and preliminaries that will be used later in the article are described in Sect. 2. In Sect. 3, a general construction of vector output functions with high generalized nonlinearity is presented using the $\epsilon$-biased sample space and a concrete example is given.
As pointed out in Ref. [4], one open problem about generalized nonlinearity is to find better lower and upper bounds. In the last section of this article, the upper bound is improved for vector resilient functions. A trade-off relation between resilient order and generalized nonlinearity is also revealed.

2 Preliminaries

The vector spaces of binary \{0, 1\} \(n\)-tuples are denoted by \(F_2^n\). Let \(f\) be a function from \(F_2^n\) to \(F_2^n\), and \(m=1\) and \(m>1\) correspond to single output and vector output, respectively. They can be used as filter functions in the stream cipher based on LFSRs. The keystream bits will be exclusive or (XORed) with the plaintext to derive the ciphertext.

Let \(f: F_2^n \rightarrow F_2\) be a Boolean function. The Hadamard transform \(\hat{f}\) is defined as:

\[
\hat{f}(w) = \sum_{x \in F_2^n} (-1)^{f(x) + \langle x, w \rangle},
\]

To perform correlation attack, the adversary will try to get an approximation of the output bits (linear combination of output bits, in the vector output case, respectively) by a linear combination of input bits. In the vector output case, the adversary wishes to maximize the following bias

\[
\Pr[u \cdot f(x) = w \cdot x] - \frac{1}{2}, \quad u, w \in F_2^n.
\]

Vector output Boolean function is said to be balanced if the pre-image of each output vector is identical [9]. The nonlinearity of \(f\), denoted by \(N_f\), is an important criterion to measure the ability of \(f\) to resist the correlation attack, which can be characterized by the Hadamard transform

\[
N_f = 2^{n-1} - \frac{1}{2} \max_{u \in F_2^n} \| u \hat{f}(w) \|.
\]

Carlet C et al. introduced a new correlation attack [4], which can be considered as the generalization of Zhang-Chan’s correlation attack [2]. The ideal is to consider the probability of the expression

\[
\Pr[(g(f(x)) + w_i(f(x)) \cdot x_i + \cdots + w_n(f(x)) \cdot x_n) = 0]
\]

where, \(w_i: F_2^n \rightarrow F_2\), \(i = 1, 2, \ldots, n\). He called this attack the generalized correlation attack. In relation to the approximation of Eq. (4), the following definition is introduced:

**Definition 1** Let \(f: F_2^n \rightarrow F_2^m\) be a vector output Boolean function. The generalized Hadamard transform \(\hat{f}\) is defined as:

\[
\hat{f}(g, w_i, \ldots, w_n) = \sum_{x \in F_2^n} (-1)^{g(f(x)) + w_i(f(x)) \cdot x_i + \cdots + w_n(f(x)) \cdot x_n},
\]

where, \(g, w_i: F_2^n \rightarrow F_2\), \(i = 1, 2, \ldots, n\). Let \(B_n\) be the set of Boolean functions \(g: F_2^n \rightarrow F_2\) and \(W\) be the set of all \(n\)-tuple functions \(w() = (w_1, w_2, \ldots, w_n)\) such that \(w(z) = (w_1(z), w_2(z), \ldots, w_n(z)) \neq 0\) for all \(z \in F_2^n\).

The generalized nonlinearity is defined as:

\[
G_f(f) = \min_{0 \neq \epsilon \subseteq F_2^n} \{ \min_{0 \neq w \subseteq F_2^n} (H(u \cdot f), 2^x - H(u \cdot f), N_{\epsilon}(f)) \}
\]

where, \(H(.)\) denotes the Hamming weight and

\[
N_{\epsilon}(f) = 2^{n-1} - \frac{1}{2} \max_{g \in B_n, w \in W} \| \hat{f}(g, w_1, \ldots, w_n) \|.
\]

Here,

\[
\min_{0 \neq \epsilon \subseteq F_2^n} (H(u \cdot f), 2^x - H(u \cdot f))
\]

is used to ensure that \(f\) is nearly balanced when \(G_f(f)\) is high. Since \(G_f(f) \leq N_{\epsilon}(f) \leq 2^{n-1} - 2^{(n-1)/2}\), \(G_f(f) = N_{\epsilon}(f)\) if \(f\) is balanced [4, 9].

Low biased sample spaces have manifold applications in computer science and cryptography. This has been studied for years. Some bounds and construction methods have been obtained, see for example [5, 8]. Let \(\epsilon\) be a constant such that \(0 \leq \epsilon \leq 1\). An \(\epsilon\)-biased sample space is a subset \(S_\epsilon \subseteq F_2^n\) defined as follows:

**Definition 2** \(S_\epsilon\) is \(\epsilon\)-biased if

\[
\| \Pr_{x \in F_2^n}[X \cdot a = 0] - \Pr_{x \in F_2^n}[X \cdot a = 1] \| \leq \epsilon
\]

for any \(a \in F_2^n \setminus \{0\}\), and \(S_\epsilon\) is also called an \(\epsilon\)-biased sample space.

3 Construction of vector output Boolean functions with high generalized nonlinearity

In Ref. [4], the authors presented a lower and upper bound of generalized nonlinearity and considered the generalized nonlinearities of some special almost bent functions and secondary construction. However, general constructions of vector output Boolean functions with high generalized nonlinearity are not known until now. In this section, a general construction method is proposed using the \(\epsilon\)-sample space. These functions are also shown to have good almost resilient order. Balanceness is an essential property since an unbalanced keystream will reveal statistical information on the plaintext. Therefore, in this section, we only consider balanced vector output Boolean functions.

To reduce the complexity of computing the generalized nonlinearity, the generalized nonlinearity is rewritten as follows:

**Theorem 1** [4] Let \(f: F_2^n \rightarrow F_2^m\) and \(w()\) denote the \(n\)-tuple of \(m\)-bit Boolean functions \(w(1), w(2), \ldots, w(n)()\). Then,

\[
N_{\epsilon}(f) = 2^{n-1} - \frac{1}{2} \max_{w : 0 \neq w \subseteq F_2^n} \| \sum_{x \in F_2^n} (-1)^{w(x) \cdot x} \|
\]

For balanced vector output Boolean functions, \(G_f(f) = N_{\epsilon}(f)\). Therefore, by Theorem 1, to construct vector output
Boolean functions with high generalized nonlinearity, we must minimize
\[
\max_{x \in \mathbb{F}_2^n} \left| \sum_{x \in \mathbb{F}_2^n} (-1)^{x \cdot z} \right| \quad \text{for each } z \in \mathbb{F}_2^n
\]  
(12)

By the definition of an \( \varepsilon \)-biased sample space \( S_\varepsilon \), we have
\[
\Pr_{x \in \mathbb{F}_2^n}[X \cdot a = 0] - \Pr_{x \in \mathbb{F}_2^n}[X \cdot a = 1] = -2 \Pr_{x \in \mathbb{F}_2^n}[X \cdot a = 1] \leq \varepsilon \quad \text{for any } a \in \mathbb{F}_2^n \setminus \{0\}
\]
(13)

That is,
\[
\left\| S_\varepsilon \right\| - \left| \left\{ x \in S_\varepsilon : X \cdot a = 1 \right\} \right| \leq \sum_{x \in \mathbb{F}_2^n} (-1)^{x \cdot y} \leq \left\| S_\varepsilon \right\| \varepsilon
\]
(14)

**Definition 3** [7] An \( \varepsilon \)-biased sample space \( S_\varepsilon \) is called \( t \)-systematic if \( \left| S_\varepsilon \right| = 2^t \), and there exist \( t \) positions \( i_1 < i_2 < \cdots < i_t \) such that each \( t \)-bit string occurs exactly once in these positions.

Now, we construct a balanced vector output Boolean function such that the pre-image of each output vector is an \( \varepsilon \)-biased sample space, which is similar to Proposition 3.2 in Ref. [7].

**Lemma 1** If there exists a \( t \)-systematic \( \varepsilon \)-biased sample space \( S_\varepsilon \), then, there exists an \( n \)-bit input and \( n \)-bit output Boolean function such that the pre-image of each output vector is an \( \varepsilon \)-biased sample space.

**Proof** Without loss of generality, we may assume that the first \( t \) positions in \( S_\varepsilon \) run through all possible elements of \( \mathbb{F}_2^t \). We construct \( 2^{n-t} \) sample space \( S_\varepsilon^n \) indexed by \( a = (a_1, a_2, ..., a_n) \).

\[
S_\varepsilon^n = S_\varepsilon + (0, 0, ..., 0, a_1, a_2, ..., a_n)
\]
(15)

It is easy to see that \( \bigcup_{a \in \mathbb{F}_2^t} S_\varepsilon^n \) is a partition of \( \mathbb{F}_2^n \).

Define a function \( f : \mathbb{F}_2^n \to \mathbb{F}_2^{n-t} \) by the rule \( f(x) = a \) if and only if \( x \in S_\varepsilon^n \).

It is easy to verify that the definition is accurate, and for any \( a \in \mathbb{F}_2^t \) and \( w \in \mathbb{F}_2^n \),
\[
\left| \Pr_{x \in \mathbb{F}_2^n}[X \cdot w = 0] - \Pr_{x \in \mathbb{F}_2^n}[X \cdot w = 1] \right| = \left| \Pr_{x \in \mathbb{F}_2^n}[(X + \beta) \cdot w = 0] - \Pr_{x \in \mathbb{F}_2^n}[(X + \beta) \cdot w = 1] \right| = \left| \Pr_{x \in \mathbb{F}_2^n}[X \cdot w = \beta \cdot w + 1] \right|
\]
(16)

where, \( \beta = (0, 0, ..., 0, \alpha) \in \mathbb{F}_2^n \). By the definition of \( \varepsilon \)-sample space, the above formula is less than or equal to \( \varepsilon \). Thus, the proof is complete.

**Theorem 2** Given a \( t \)-systematic \( \varepsilon \)-sample space \( S_\varepsilon \), the function constructed in Lemma 1 has generalized nonlinearity
\[
G_\varepsilon(f) \geq 2^{n-t} - 2^{-1} \varepsilon
\]
(17)

**Proof** By Lemma 1, for any \( z \in \mathbb{F}_2^{n-t} \), the pre-image of \( f^{-1}(z) \) is a \( t \)-systematic \( \varepsilon \)-sample space. Therefore, by Eq. (14), for any \( w \in \mathbb{F}_2^n \setminus \{0\} \)
\[
\left| \sum_{x \in \mathbb{F}_2^n} (-1)^{x \cdot w} \right| \leq 2 \varepsilon
\]
(18)

By Theorem 1
\[
N_\varepsilon(f) \geq 2^{n-t} - \frac{1}{2} 2^{n-t} 2^{-1} \varepsilon = 2^{n-t} - 2^{-1} \varepsilon
\]
(19)

Note that \( f \) is balanced, \( G_\varepsilon(f) = N_\varepsilon(f) \), and the proof is complete.

In Ref. [4], an upper bound of generalized nonlinearity is obtained.

**Theorem 3** [4] If \( f : \mathbb{F}_2^n \to \mathbb{F}_2^m \) is balanced, then, we have
\[
G_\varepsilon(f) \leq 2^{n-t} - 2^{-1} \sqrt{\frac{2^{m} - 1}{2^{t} - 1}}
\]
(20)

Thus, if the sample space \( S_\varepsilon \) used in our construction has parameter \( \varepsilon \) approximate to \( \sqrt{(2^n - 1)/(2^t - 1)} \), then the vector output Boolean functions have the generalized nonlinearity very close to the upper bound given in Ref. [4]. The observation also gives a lower bound of the size of an \( \varepsilon \)-sample space.

**Corollary 1** If a \( t \)-systematic \((2^t, n)\) \( \varepsilon \)-sample space exists, then, \( t \geq n - \log_2(1 + 2^{-1}(2^n - 1)) \).

**Proof** If a \( t \)-systematic \((2^t, n)\) \( \varepsilon \)-sample space exists, we can construct \((n, n - t)\) vector output Boolean function with generalized nonlinearity lower bound by Theorem 2. According to Theorem 3,
\[
\varepsilon \geq \frac{2^{n-t} - 1}{\sqrt{2^t - 1}}
\]
(21)

Then, it is easy to obtain the corollary from the above inequality.

**Example** The exponential sum method based on Weil-Carlitz-Uchiyama bound is useful in the construction of low biased sample space [7, 8, 10]. By use of the Weil-Carlitz-Uchiyama bound, Kurosawa K constructed \( t \)-systematic \((2^t, n)\) \( \varepsilon \)-sample space, where, \( n = tD' \) and
\[
\varepsilon = \frac{2(D' - 1)}{\sqrt{2^t}}
\]
(22)

Then, according to Theorem 2, we can construct a \((n, m)\) Boolean function with generalized nonlinearity
\[
G_\varepsilon(f) \geq 2^{n-t} - 2^{n-m} \frac{2^{n - m - 1}}{2^{n - m}}
\]
(23)

where, \( m = n - 1 \). Note that if \( n = m/2 \), the lower bound is very close to the upper bound in Theorem 3.

Almost dependent sample space, which is a generalization of orthogonal array, can be used to construct almost resilient
function. For exact definition of almost resilient function, refer Ref. [7].

**Theorem 4** [8] An array in \( \varepsilon \)-sample space is also an \( \varepsilon \)-dependent sample space for some \( \varepsilon' \leq \varepsilon \).

According to Theorem 4, it is easy to prove that the functions we constructed are also almost resilient.

**4 The generalized nonlinearity of resilient functions**

As pointed out in Ref. [4], one open problem about generalized nonlinearity is to find better lower and upper bounds. In this section, the upper bound is improved for vector resilient functions.

An analogue result of the relation between immune order and nonlinearity of a Boolean function \( f : F_2^n \rightarrow F_2 \) is

\[
N_f \geq \frac{1}{2^{n-1}}\sqrt{2^n - \sum_{i=1}^{n} C_i} \leq 1 \tag{24}
\]

where, \( m \) is the immune order and \( N_f \) is the nonlinearity. See for example [11].

The function \( f \) is called an \((n, m, k)\)-resilient function if

\[
\Pr[f(x_1, x_2, \ldots, x_n) = (y_1, y_2, \ldots, y_n) | x_1 x_2 \cdots x_n = \alpha] = 2^{-n} \tag{25}
\]

for any \( k \) positions \( i_1 < i_2 < \cdots < i_k \), for any \( k \) bits string \( \alpha \in F_2^k \), and for any \((y_1, y_2, \ldots, y_n) \in F_2^n \), where, the values \( x_j (j \notin \{i_1, i_2, \ldots, i_k\}) \) are chosen independently at random.

**Theorem 5** Let \( f : F_2^n \rightarrow F_2^m \) be a \( t \)-resilient function, then

\[
\frac{G_f(f)}{2^{n-1}} + 2^{2^n - 1} \sqrt{\frac{2^n - 1}{2^n - \sum_{i=1}^{n} C_i}} \leq 1 \tag{26}
\]

**Proof** By Theorem 1 and that resilient function must be balanced, we have

\[
G_f(f) = 2^{n-1} - \frac{1}{2} \sum_{x \in F_2^n} \max_{w \in F_2^m} \left| \sum_{t \in \{0, 1\}} (-1)^{w(t)x} \right| \tag{27}
\]

Similar to the proof of Theorem 3, let \( \phi_t(x) \) be the indicator function of \( F^{-1}(z) \), i.e., \( \phi_t(x) = 1 \) if \( f(x) = z \) else \( \phi_t(x) = 0 \). Then, for \( w \neq 0 \)

\[
\sum_{x \in F_2^n} (-1)^{w(t)x} = \sum_{x \in F_2^n} \phi_t(x) (-1)^{w(t)x} = -\frac{1}{2}\phi_t(w) \tag{28}
\]

It is easy to see that

\[
\phi_t(0) = \sum_{x \in F_2^n} (-1)^{w(t)x} = 2^n - 2 | f^{-1}(z) | = 2^n - 2^{n+1-2t+1} \tag{29}
\]

and for any \( z \in F_2^n \), \( \phi_t(x) \) is the \( t \)-immune Boolean function. According to the Xiao-Massey Theorem [12], \( \phi_t(w) = 0 \) for any \( w \in F_2^n \), \( 0 < w_1(w) < t \). By Parseval Theorem,

\[
\sum_{w \in F_2^n} \phi_t^2(w) = 2^n \tag{30}
\]

Therefore,

\[
\max_{w \in F_2^n} \phi_t^2(w) = \frac{2^{2^n - 1} - 2^{n+1-2t+1}}{2^n - \sum_{i=1}^{n} C_i} \tag{31}
\]

Therefore,

\[
G_f(f) \leq 2^{n-1} - 2^{n-1} \frac{2^{2^n - 1} - 2^{n+1-2t+1}}{2^n - \sum_{i=1}^{n} C_i} = 2^{n-1} - 2 \frac{2^n - 1}{2^n - \sum_{i=1}^{n} C_i} \tag{32}
\]

Thus, the result is easily deduced.

Given an \((n, m, t)\)-resilient function, by comparing Inequalities (24) and (26), the generalized nonlinearity of the function is very likely to be lower than its nonlinearity.

**5 Conclusions**

In this article, a general construction of vector output Boolean function with high generalized nonlinearity is given using the \( \varepsilon \)-sample space. If the sample space is chosen properly, the generalized nonlinearity of the constructed function may be very close to the upper bound presented in Ref. [4]. At last, a tradeoff relation between resilient order and generalized nonlinearity is obtained for a vector resilient function.

**Acknowledgements** This work is supported by the National Natural Science Foundation of China (90604023), Fujian Province Young Talent Program (2006F3044), Natural Science Foundation of Fujian Province (2006J0189), the Open Funds of Key Laboratory of Fujian Province University Network Security and Cryptology (07B002) and Fujian Education Department Technology Program (JA07050).

**References**


Biographies:
KE Pin-hui, lecturer in Fujian Normal University. His interests include cryptography and information security.

ZHANG Sheng-yuan, professor at Fujian Normal University. He is interested in combinational mathematics and information security.

From p. 59

Biographies: ZHANG Lei, Ph. D. Candidate in Beijing University of Posts and Telecommunications, interested in MIMO detector and space-time coding.

YANG Da-cheng, professor in Beijing University of Posts and Telecommunications, interested in theory and application of wireless communication.