Almost Security of Cryptographic Boolean Functions

Kaoru Kurosawa

Department of Computer and Information Sciences,
Ibaraki University
4-12-1 Nakanarusawa, Hitachi, Ibaraki, 316-8511, Japan
kurosawa@cis.ibaraki.ac.jp

Abstract

$PC(l)$ of order k is one of the most general cryptographic criteria of secure Boolean functions. In this paper, we introduce its ϵ-almost version. The new definition requires *only* that $f(X)+f(X+\Delta)$ is *almost* uniformly distiributed (while the original definition of $PC(l)$ of order k requires that it is *strictly* uniformly distiributed). We next show its construciton. Better parameters are then obtained than normal $PC(l)$ of order k functions.

(Keywords) Boolean functions, $PC(l)$ of order k, ϵ-almost version

1 Introduction

Several cryptographic criteria of Boolean functions have been studied by many researchers in order to design secure block ciphers. Among them, $PC(l)$ of order k [15, 16] is one of the most general criteria.

A Boolean function $f(X)$ satisfies $PC(l)$ if the output difference $f(X)+f(X+\Delta)$ is uniformly distributed for any input difference Δ such that the Hamming weight Δ is equal to l or less. (That is, $1 \leq wt(\Delta) \leq l$, where $wt(\Delta)$ denotes the Hamming weight of Δ.) Further suppose that $f(X)$ satisfies $PC(l)$ even if any k bits of $X = (x_1, \cdots, x_n)$ are fixed into any constants. Then we say that $f(X)$ satisfies $PC(l)$ of order k.

The famous strict avalanche criterion (SAC), which was introduced as a criterion of the security of S-boxes [17], is equivalent to $PC(1)$. SAC(k) is equivalent to $PC(1)$ of order k. Also, $f(X)$ is a bent function [10] if and only if $f(X)$ satisfies $PC(n)$ [15], where a bent function has the largest distance from the set of affine (linear) functions. (Hence it is directly related to the linear attack.) $PC(l)$ of order k in general is directly related to the security against differential attacks.

Kurosawa et al. gave a general method to design such functions by using linear codes [9]. Carlet extended it to nonlinear codes [4].

Boolean functions, however, do not need to satisfy the strict definitions of cryptographic criteria in general. These definitions are sometimes stronger than what we want, *i.e.,* attacks on block ciphers should be impossible. (It is well-known that bent functions cannot be balanced.) Therefore, as far as attacks are impossible, it will be better if better parameters are obtained by relaxing the definitions.

From this point of view, this paper intodUces a notion of ϵ-almost $PC(l)$ of order k. It requires *ony* that $f(X) + f(X + \Delta)$ is *almost* uniformly distiributed in the original definition of $PC(l)$ of order k. We then show that indeed better parameters are obtained than normal $PC(l)$ of order k functions.

We present a design method of ϵ-almost $PC(l)$ of order k functions using linear codes and a ϵ-biased sample spaces [11] which satisfy some property. Our construction offers smaller input length n than normal $PC(l)$ of order k functions for the same (l, k). (The input size n of Sboxes can be smaller for the security level (l, k).) In other words, we can obtain larger (l, k) for the same input length n. (Higher security level (l, k) can be obtained for the same input size n of Sboxes.) We further generalize our result to multiple output bits Boolean functions.

Related works: Suppose that $F(x_1, \cdots, x_n) = (y_1, \cdots, y_m)$ is uniformly distributed even if any k bits of $(x_1, \cdots, x_n)$ are fixed into any constants. We then say that F is an (n, m, k)-resilient function. This notion has been studied by several researchers from a view point of key renewal [5, 2, 7, 13, 3, 14]. Especially, [8] introduced a notion of ϵ-almost k-resilient functions. The authors presented its construction and showed that better parameters are obtained than normal k-resilient functions. Dodis et al. improved it by showing a probabilistic construction [6].

Our work can be considered as an extension of [8]. Indeed, we show that an ϵ-almost $PC(l)$ of order k function is obtained from a linear code and an ϵ-almost k-resilient function which satisfies some special property. However, we cannot use the previous constructions of ϵ-almost k-resilient functions

[8, 6] because it is not known if they satisfy our special property.

2 Preliminaries

$wt(\Delta)$ denotes the Hamming weight of a binary vector Δ. Let $\cdot$ denote the inner product of two binary vectors over $GF(2)$. For a set A, $|A|$ denotes the cardinality of A.

Let a linear $[N, m, d]$-code denote a binary linear code C of length N, dimension m and the minimum Hamming distance at least d. The dual code $C^{\perp}$ of a linear code C is defined as $C^{\perp} \stackrel{\triangle}{=} \{u \mid u \cdot v = 0 \text{ for all } v \in C\}$. The dual minimum Hamming distance $d^{\perp}$ of C is defined as the minimum Hamming distance of $C^{\perp}$.

2.1 Resilient Functions

Definition 2.1 *We say that $F(X) = (y_1, \cdots, y_m)$ is a (n, m, k)-resilient function if $F(X)$ is uniformly distributed even if any k variables $x_{i_1}, \cdots, x_{i_k}$ are fixed into any constants. That is,*

$$\Pr[f(x_1, \ldots, x_n) = (y_1, \ldots, y_m) \mid x_{i_1} x_{i_2} \cdots x_{i_k} = \alpha] = 2^{-m}$$

for any k positions $i_1 < \cdots < i_k$, for any k-bit string $\alpha \in \{0, 1\}^k$ and for any $(y_1, \cdots, y_m) \in \{0, 1\}^l$, where the values x_j ($j \notin \{i_1, \ldots, i_k\}$) are chosen independently at random.

Chor et al. showed that a (n, m, k)-resilient function can be obtained from a linear $[n, m, k + 1]$-code [5].

Proposition 2.1 *Let G be a generator matrix of a linear $[n, m, k + 1]$-code C. Then $F(X) = G \cdot X$ is a (n, m, k)-resilient function.*

(Proof) It is known that $F(X) = (y_1, \cdots, y_m)$ is a (n, m, k)-resilient function if and only if

$$a_1 y_1 + \cdots + a_m y_m \qquad (1)$$

is a $(n, 1, k)$-resilient function for any $(a_1, \cdots, a_m) \neq (0, \cdots, 0)$. (See [5].) In our case, eq.(1) becomes as follows.

$$(a_1, \cdots, a_m) \cdot GX = (b_1, \cdots, b_n) \cdot X,$$

where $(b_1, \cdots, b_n) = (a_1, \cdots, a_m) \cdot G$. Note that $wt(b_1, \cdots, b_n) \geq k+1$ because $(b_1, \cdots, b_n)$ is a nonzero codeword of C. Then it is easy to see that $(b_1, \cdots, b_n) \cdot X$ is k-resilient.

Q.E.D.

2.2 $PC(l)$ of order k

Definition 2.2 *[15, 16] We say that $f(X)$ satisfies $PC(l)$ of order k if $f(X) + f(X + \Delta)$ is k-resilient for any Δ such that $1 \leq wt(\Delta) \leq l$. (We also say that $f(X)$ is a $PC(l)$ of order k function.)*

Kurosawa et al. gave a general method to design $PC(l)$ of order k functions by using two linear codes [9].

Proposition 2.2 *Suppose that there exist*

1. *a linear $[n_1, m, k+1]$-code C_1 with the dual minimum Hamming distance at least $l+1$ and*
2. *a linear $[n_2, m, k+1]$-code C_2 with the dual minimum Hamming distance at least $l+1$.*

Then there exists a a $PC(l)$ of order k function $f(x_1, \cdots, x_n)$ such that

$$n = n_1 + n_2.$$

3 Almost Resilient Functions

In [8], the authors introduced a notion of *almost* resilient functions and showed its construction. Then better parameters are obtained than normal resilient functions.

3.1 Almost k-Wise Independent Sample Space

Let $S_n \subseteq \{0,1\}^n$. We consider that S_n is a $|S_n| \times n$ binary matrix and each row is randomly chosen.

Definition 3.1 *S_n is ϵ-biased if*

$$\left| \Pr_{X \in S_n}(X \cdot \alpha = 0) - \Pr_{X \in S_n}(X \cdot \alpha = 1) \right| \leq \epsilon$$

for any $\alpha \in \{0,1\}^n \setminus \{0^n\}$.

Definition 3.2 (almost k-wise independence). *Suppose that $X = x_1 \cdots x_N$ is chosen randomly from S_N. Then we say that S_N is (ϵ, k)-independent if for any k positions $i_1 < i_2 < \cdots < i_k$ and any k-bit string α, we have*

$$|\Pr[x_{i_1} x_{i_2} \cdots x_{i_k} = \alpha] - 2^{-k}| \leq \epsilon.$$

Proposition 3.1 *[11] Suppose that S_n is ϵ-biased. Let H be a parity check matrix of a $[N, N-n, k+1]$-linear code C. Define*

$$S_N \triangleq S_n \cdot H$$

Then S_N is $(\widetilde{\epsilon}, k)$-independent, where

$$\widetilde{\epsilon} = \left(1 - \frac{1}{2^k}\right) \cdot \epsilon$$

3.2 Almost Resilient Functions

Definition 3.3 *[8]* The function $f(X)$ is called an ϵ-*almost* (n, m, k)-*resilient function* if

$$|\Pr[f(x_1, \ldots, x_n) = (y_1, \ldots, y_m) \mid x_{i_1} x_{i_2} \cdots x_{i_k} = \alpha] - 2^{-m}| \leq \epsilon$$

for any k positions $i_1 < \cdots < i_k$, for any k-bit string $\alpha \in \{0,1\}^k$ and for any $(y_1, \cdots, y_m) \in \{0,1\}^m$, where the values x_j $(j \notin \{i_1, \ldots, i_k\})$ are chosen independently at random.

Definition 3.4 *[8]* An (ϵ, k)-independent sample space S_N is called t-*systematic* if $|S_N| = 2^t$, and there exist t positions $i_1 < \cdots < i_t$ such that each t-bit string occurs in these positions for exactly one N-tuple in S_N.

We define t-*systematic* ϵ-biased sample spaces similarly.

Proposition 3.2 *[8, Theorem 4.4] If there exists a t-systematic (ϵ, k)-independent sample space S_N, then there exists a balanced δ-almost $(N, N-t, k)$-resilient function, where $\delta = \epsilon/2^{N-t-k}$.*

The δ-almost $(N, N-t, k)$-resilient function given in Proposition 3.2 is constructed as follows [8]. Without loss of generality, assume that the first t positions in S_N run through all possible t-bit strings. We then obtain 2^{N-t} sample spaces E_α indexed by $\alpha = (\alpha_1, \ldots, \alpha_{N-t}) \in \{0,1\}^{N-t}$ by

$$E_\alpha = S_N + (\underbrace{0, 0, \ldots, 0}_{t}, \alpha_1, \ldots, \alpha_{N-t}).$$

Finally define a function $\phi : \{0,1\}^N \rightarrow \{0,1\}^{N-t}$ by the rule

$$\phi(x_1, \ldots, x_m) = \alpha \text{ if and only if } (x_1, \ldots, x_N) \in E_\alpha.$$

Then ϕ is a δ-almost $(N, N-t, k)$-resilient function, where δ is given in Proposition 3.2.

4 Almost $PC(l)$ of order k

In this section, we introduce a notion of almost $PC(l)$ of order k functions. We then show that an almost $PC(l)$ of order k function is obtained from a linear code and an ϵ-almost (n, m, k)-resilient function which satisfies some property.

Definition 4.1 We say that $f(X)$ satisfies *ϵ-almost $PC(l)$ of order k* if $f(X) + f(X + \Delta)$ is an *ϵ-almost $(n, 1, k)$-resilient function* for any Δ such that $1 \leq wt(\Delta) \leq l$. (We also say that $f(X)$ is an *ϵ-almost $PC(l)$ of order k function.*)

4.1 Basic Theorem

Definition 4.2 *For $\Phi : \{0,1\}^n \rightarrow \{0,1\}^m$, we define the domain distance d_Φ as follows. Suppose that*

$$\Phi(\beta) \neq \Phi(\beta + \Omega)$$

for any β and any Ω such that $1 \leq wt(\Omega) \leq l$. Then $d_\Phi \geq l + 1$.

Lemma 4.1 *Let $\phi(X) = (y_1, \cdots, y_m)$ be an ϵ-almost (n, m, k)-resilient function. Then $\phi(X) \cdot \Delta$ is a $(2^{m-1}\epsilon)$-almost $(n, 1, k)$-resilient function for any $\Delta \neq (0, \cdots, 0)$.*

(Proof) For any $\Delta \neq (0, \cdots, 0)$, let

$$A_0 = \{Y \mid Y \cdot \Delta = 0\}, \ A_1 = \{Y \mid Y \cdot \Delta = 1\}.$$

Then $|A_0| = |A_1| = 2^{m-1}$. Therefore,

$$\Pr(\phi(X) \cdot \Delta = 0) = \sum_{\alpha \in A_0} \Pr(\phi(X) = \alpha) \geq \sum_{\alpha \in A_0} (2^{-m} - \epsilon) = 1/2 - 2^{m-1}\epsilon.$$

Similarly we have

$$\Pr(\phi(X) \cdot \Delta = 0) \leq 1/2 + 2^{m-1}\epsilon.$$

Hence

$$|\Pr(\phi(X) \cdot \Delta = 0) - 1/2| \leq 2^{m-1}\epsilon.$$

Similarly,

$$|\Pr(\phi(X) \cdot \Delta = 1) - 1/2| \leq 2^{m-1}\epsilon.$$

Q.E.D.

Then our basic theorem is stated as follows.

Theorem 4.1 *Suppose that there exist*

1. *a linear $[n_1, m, k+1]$-code C_1 with the dual minimum Hamming distance at least $l+1$ and*
2. *an ϵ-almost (n_2, m, k)-resilient function Φ with the domain distance $d_\Phi \geq l+1$.*

Then there exists a $(2^{m-1}\epsilon)$-almost $PC(l)$ of order k function $f(x_1, \cdots, x_n)$ such that

$$n = n_1 + n_2.$$

(Proof) Let G_1 be a generator matrix of C_1. For $X = (x_1, \cdots, x_{n_1})$ and $Y = (y_1, \cdots, y_{n_2})$, define

$$f(X, Y) = \Phi(Y) \cdot G_1 X + g(Y),$$

where $g(Y)$ is any Boolean function. We show that $f(X, Y)$ satisfies $(2^{m-1}\epsilon)$-almost $PC(l)$ of order k. Let

$$f'(X, Y) = f(X, Y) + f(X + \Delta, Y + \Omega),$$

then

$$\begin{aligned} f'(X, Y) &= (\Phi(Y) + \Phi(Y + \Omega)) \cdot G_1 X + \Phi(Y + \Omega) \cdot G_1 \Delta \\ &\quad + g(Y) + g(Y + \Omega) \end{aligned}$$

Case 1. Suppose that $\Omega = 0$ and $1 \leq wt(\Delta) \leq l$. In this case,

$$f'(X, Y) = \Phi(Y) \cdot G_1 \Delta.$$

Then $G_1 \Delta \neq \mathcal{O}$ because Δ is not a codeword of $C_1^\perp$. Hence $f'(X, Y)$ is $(2^{m-1}\epsilon)$-almost k-resilient from Lemma 4.1.

Case 2. Suppose that $\Omega \neq 0$ and $1 \leq wt(\Delta) + wt(\Omega) \leq l$. Then for any β,

$$f'(X, \beta) = (\Phi(\beta) + \Phi(\beta + \Omega)) \cdot G_1 X + \gamma,$$

where $\gamma = \Phi(\beta + \Omega) \cdot G_1 \Delta + g(\beta) + g(\beta + \Omega)$ is a constant. Now $\Phi(\beta) \neq \Phi(\beta + \Omega)$ because $d_\Phi \geq l + 1$. Therefore, $f'(X, \beta)$ is k-resilient from the proof of Proposition 2.1.

This means that $f'(X, Y)$ is k-resilient.

Consequently, $f(X, Y)$ satisfies $2^m \epsilon$-almost $PC(l)$ of order k. Q.E.D.

Remark 4.1 Kurosawa et al. [8] and Dodis et al. [6] showed how to construct ϵ-almost (n, m, k)-resilient functions. However, it is not known if their constructions satisfy our condition on the *domain distance*.

4.2 Discussion

Proposition 2.2 is obtained as a corollary of Theorem 4.1. Indeed, it is easy to show that if there exists a linear $[n_2, m, k + 1]$-code with the dual minimum Hamming distance at least $l + 1$, then there exists a (n_2, m, k)-resilient function with the domain distance $d_\Phi \geq l + 1$.

Now suppose that there exists a linear $[n_2, m, k+1]$-code. In what follows, we show that there exists an ϵ-almost (n_2', m, k)-resilient function with the domain distance $d_\Phi \geq l + 1$ such that $n_2' < n_2$.

This means that we can obtain smaller input length n for the same (l, k). In other words, we can obtain larger (l, k) for the same n.

5 Construction

5.1 Overview

In this section, we show how to achieve the second condition of Theorem 4.1, i.e. how to construct an ϵ-almost (n, m, k)-resilient function with a domain distance at least $l + 1$.

For a (ϵ, k)-independent sample space S_N, we can consider a nonlinear code $C(S_N)$ such that each row of S_N is a codeword.

Definition 5.1 *For S_N, let d be the minimum Hamming distance of $C(S_N)$. Then we say that S_N has the domain distance d.*

1. We first show that the second condition of Theorem 4.1 is satisfied if there exists a t-systematic (ϵ, k)-independent sample space S_N whose domain distance is at least $l + 1$.

2. We next show that such S_N is obtained from a t-systematic ϵ-biased sample space S_n and a linear $[N, N-n, k+1]$-code with the dual minimu Hamming distance at least $l + 1$.

3. We finally show how to construct such S_n by using Weil-Carlitz-Uchiyama bound. (The same technique was used in [8] to construct a t-systematic (ϵ, k)-independent sample space S_N.)

5.2 General Construction

Theorem 5.1 *Suppose that there exists a t-systematic (ϵ, k)-independent sample space S_N with the domain distance at least $l+1$. Then there exists a balanced δ-almost $(N, N-t, k)$-resilient function ϕ with the domain distance d_ϕ at least $l+1$, where $\delta = \epsilon/2^{N-t-k}$.*

(Proof) Construct ϕ from S_N by using the method shown just after Proposition 3.2. Suppose that $d_\phi \leq l$. That is, $\phi(\beta) = \phi(\beta + \Omega) = \alpha$ for some α, β and Ω such that $1 \leq wt(\Omega) \leq l$. Then we see that

$$\beta + (0, \cdots, 0, \alpha) \in S_N \text{ and } \beta + \Omega + (0, \cdots, 0, \alpha) \in S_N.$$

This means that there are two codewords with the distance l or less in S_N. However, this is a contradiction because S_N has the domain distance at least $l + 1$.

Q.E.D.

Theorem 5.2 *Suppose that there exists a t-systematic ϵ-biased sample space S_n and a linear $[N, N-n, k+1]$-code C with the dual minimum Hamming distacen at least $l+1$. Then there exists a t-systematic $(\widetilde{\epsilon}, k)$-independent sample space S_N with the domain distance at least $l + 1$, where*

$$\widetilde{\epsilon} = \left(1 - \frac{1}{2^k}\right) \cdot \epsilon.$$

(Proof) Let $H = (I_n, \tilde{H})$ be a parity check matrix of C, where I_n is the $n \times n$ identity matrix. Let

$$S_N = S_n \cdot H. \tag{2}$$

Then S_N is $(\widetilde{\epsilon}, k)$-independent from Propposition 3.1, where

$$\widetilde{\epsilon} = \left(1 - \frac{1}{2^k}\right) \cdot \epsilon$$

Next it is easy to see that S_N is t-systematic if S_n is t-systematic. Finally, we show that S_N has the domain distacen at least $l+1$. From eq.(2), we see that S_N is a subset of all codewords of $C^{\perp}$. Therefore S_N has the domain distacen at least $l+1$ becasue C has the dual distance at least $l+1$.

Q.E.D.

5.3 Construction of Systematic ϵ-Biased Sample Space

We next show how to construct a t-systematic ϵ-biased sample space S_n by using Weil-Carlitz-Uchiyama boumd. For $x \in GF(2^t)$, let

$$\mathrm{Tr}(x) \stackrel{\triangle}{=} x + x^2 + x^{2^2} + \cdots + x^{2^{t-1}}.$$

It is well-known that $\mathrm{Tr}(x) = 0$ or 1 and $\mathrm{Tr}(x_1 + x_2) = \mathrm{Tr}(x_1) + \mathrm{Tr}(x_2)$.

Proposition 5.1 (Weil-Carlitz-Uchiyama Bound) *[18, 10] Let $f(x) = \sum_{i=1}^{D} f_i x^i \in GF(2^t)[x]$ be a polynomial such that $f(x) \neq g(x)^2 - g(x) + \theta$ for any polynomial $g(x) \in GF(2^t)[x]$ and for any constant $\theta \in F_{2^t}$. Then*

$$\left| \sum_{\alpha \in GF(2^t)} (-1)^{\mathrm{Tr}(f(\alpha))} \right| \leq (D-1)\sqrt{2^t}.$$

Remark 5.1 *It is easy to see that if $f(x)$ is an odd degree polynomial, then $f(x) \neq g(x)^2 - g(x) + \theta$ for any $g(x)$ and any θ.*

Now for two positive integers t and D', let $n = tD'$ and $D = 2D' - 1$, g be a primitive element of $GF(2^t)$ and $x_1, x_2, \cdots, x_{2^t}$ be the elements of $GF(2^t)$. For each $x_i \in GF(2^t)$, let X_i be a string of length $n = tD'$ such that

$$X_i \stackrel{\triangle}{=} (Z_{i,1}, Z_{i,2}, \cdots, Z_{i,D'}),$$

where

$$Z_{i,j} \stackrel{\triangle}{=} (\mathrm{Tr}(x_i^{2j-1}), \mathrm{Tr}(g x_i^{2j-1}), \cdots, \mathrm{Tr}(g^{t-1} x_i^{2j-1})).$$

The proposed ϵ-biased sample space is defined as

$$S_n \stackrel{\triangle}{=} \{X_1, X_2, \cdots, X_{2^t}\}.$$

Theorem 5.3 *The above $S_n \subseteq \{0,1\}^n$ is a t-systematic ϵ-biased sample space such that $n = tD', |S_n| = 2^t$ and*

$$\epsilon = \frac{2(D'-1)}{\sqrt{2^t}}.$$

(Proof) First it is a well known fact [8, page 245] that

$$Y_x = (Tr(x), Tr(gx), \ldots, Tr(g^{t-1}x))$$

runs through $\{0,1\}^t$ when x runs through $GF(2^t)$. Hence S_n is t-systematic.

Next consider $\alpha \in \{0,1\}^n \backslash \{0^n\}$. Let $\alpha = (\Lambda_1, \Lambda_2, \cdots, \Lambda_{D'})$ with

$$\Lambda_j = (\alpha_{0,2j-1}, \alpha_{1,2j-1}, \cdots, \alpha_{t-1,2j-1}).$$

Then since $\alpha_{i,j}$ is binary, we have that

$$\begin{aligned} X_i \cdot \alpha &= \sum_{j=1}^{D'} (\alpha_{0,2j-1}\mathrm{Tr}(x_i^{2j-1}) + \cdots + \alpha_{t-1,2j-1}\mathrm{Tr}(g^{t-1}x_i^{2j-1})) \\ &= \sum_{j=1}^{D'} \mathrm{Tr}(\alpha_{0,2j-1} + \alpha_{1,2j-1}g + \cdots + \alpha_{t-1,2j-1}g^{t-1})x_i^{2j-1})) \\ &= \mathrm{Tr}(a_1 x_j + a_3 x_i^3 + \cdots + a_D x_i^D), \end{aligned} \tag{3}$$

where

$$a_j \stackrel{\triangle}{=} \alpha_{0,j} + \alpha_{1,j}g + \cdots + \alpha_{t-1,j}g^{t-1}$$

Since g is a primitive element, $a_j = 0$ if an only if $(\alpha_{0,j}, \alpha_{1,j}, \cdots, \alpha_{t-1,j}) = (0, \cdots, 0)$. This implies that $(a_1, \cdots, a_D) \neq (0, \cdots, 0)$ because $\alpha \neq 0$.

Now define

$$f_i(x) \stackrel{\triangle}{=} a_1 x_i + a_3 x_i^3 + \cdots + a_D x_i^D$$

Let

$$A_0 \stackrel{\triangle}{=} \{x_i | \mathrm{Tr}(f(x_i)) = 0\}, \ A_1 \stackrel{\triangle}{=} \{x_i | \mathrm{Tr}(f(x_i)) = 1\}.$$

Then we see that

$$\begin{aligned} |\Pr(X \cdot \alpha = 0) - \Pr(X \cdot \alpha = 1)| &= \left| \frac{|A_0|}{2^t} - \frac{|A_1|}{2^t} \right| \\ &= \frac{1}{2^t} \left| \sum_{x_i \in GF(2^t)} (-1)^{\mathrm{Tr}(f(x_i))} \right|. \end{aligned}$$

Finally from Weil-Carlitz-Uchiyama bound (see Remark 5.1, too), we have

$$|\Pr(X \cdot \alpha = 0) - \Pr(X \cdot \alpha = 1)| \le \frac{(D-1)\sqrt{2^t}}{2^t} = \frac{D-1}{\sqrt{2^t}}.$$

Hence

$$\epsilon = \frac{D-1}{\sqrt{2^t}} = \frac{2(D'-1)}{2^{t/2}}.$$

Q.E.D.

5.4 Final Result

Corollary 5.1 *Suppose there exists a $[N, N-tD', k+1]$-code C with the dual distance at least $l+1$. Then there exists a balanced δ-almost $(N, N-t, k)$-resilient function with the domain distance at least $l+1$ such that*

$$\delta = \left(1 - \frac{1}{2^k}\right) \frac{2(D'-1)\sqrt{2^t}}{2^{N-k}}.$$

(Proof) From Theorem 5.1, 5.2 and 5.3.

Q.E.D.

From Theorem 4.1 and Corollary 5.1, we finally obtain the following corollary.

Corollary 5.2 *Suppose that there exist*

1. *a linear $[n'_1, m, k+1]$-code C'_1 with the dual minimum Hamming distance at least $l+1$ and*
2. *a linear $[n'_2, m-(D'-1)t, k+1]$-code C'_2 with the dual minimum Hamming distance at least $l+1$*

Then there exists a $(2^{m-1}\delta)$-almost $PC(l)$ of order k function $f(x_1, \cdots, x_n)$ such that

$$n' = n'_1 + n'_2,$$

where

$$\delta = \left(1 - \frac{1}{2^k}\right) \frac{2(D'-1)\sqrt{2^t}}{2^{N-k}}.$$

6 Comparison

Let's compare the parameters of our construction (Corollary 5.2) with normal $PC(l)$ of order k functions (Proposition 2.2).

We first show that our construction has a smaller input length than normal $PC(l)$ of order k functions for the same (l, k). Let n' denote the input length of our construction and n denote the input length of Proposition 2.2. We use C_1', C_2' to refer the linear codes of Corollary 5.2 and C_1, C_2 to refer the linear codes of Proposition 2.2, respectively. Suppose that

1. $C_1 = C_1'$ (hence $n_1 = n_1'$).
2. Each of C_2 and C_2' has the minimum Hamming distances at least $k+1$ and the dual minimum Hamming distances at least $l+1$.

Then

$$\begin{aligned}\text{the dimension of } C_2' &= m-(D'-1)t,\\ \text{the dimension of } C_2 &= m\end{aligned}$$

Therefore, $n_2' < n_2$ because $m-(D'-1)t < m$. Hence

$$n' < n.$$

This shows that our construction has a smaller input length for the same (l, k).

In other words, we can say that our construction has larger (l, k) for the same input length n.

7 Generalization to Multiple Output Bits

In this section, we generalize our result to multiple output Boolean functions.

Definition 7.1 *We say that $F(X) = (f_1, \cdots, f_p)$ satisfies ϵ-almost $PC(l)$ of order k if $a_1 f_1 + \cdots + a_p f_p$ satisfies ϵ-almost $PC(l)$ of order k for any $(a_1, \cdots, a_p) \neq (0, \cdots, 0)$.*

Theorem 7.1 *Suppose that there exist*

1. *a linear $[n_1, m, k+1]$-code C_1 with the dual minimum Hamming distance at least $l+1$ and*

2. *a linear* $[n_2, m-(D'-1)t, k+1]$*-code* C_2 *with the dual minimum Hamming distance at least* $l+1$

Then there exists a $(2^{m-1}\delta)$*-almost* $PC(l)$ *of order* k *function* $F(x_1,\cdots,x_n)=(y_1,\cdots,y_m)$ *such that* $n=n_1+n_2$*, where*

$$\delta=\left(1-\frac{1}{2^k}\right)\frac{2(D'-1)\sqrt{2^t}}{2^{N-k}}.$$

(Proof) Let G_1 be a generator matrix of a linear $[n_1, m, k+1]$-code C_1 with the dual minimum Hamming distacne at least $l+1$. Let $\Phi(Y)$ be an ϵ-almost (n_2, m, k)-resilient function with the domain distance $d_\Phi \geq l+1$.

Consider a linear feedback shift register of length m and with a primitive feedback polynomial. Let S be the state transition matrix of such a shift register. Let $X=(x_1,\cdots,x_{n_1})$ and $Y=(y_1,\cdots,y_{n_2})$. For $i=1,\cdots,m$, define

$$f_i(X,Y)\stackrel{\triangle}{=}\Phi(Y)\cdot S^{i-1}G_1X+g_i(Y)$$

where $g_i(Y)$ is any Boolean function. Then we show that $F(X,Y)=(f_1,\cdots,f_m)$ satisfies $(2^{m-1}\epsilon)$-almost $PC(l)$ of order k.

For $(a_1,\cdots,a_m)\neq(0,\cdots,0)$, we have

$$\begin{aligned} a_1f_1+\cdots a_mf_m &= \Phi(Y)\cdot(a_1I+a_2S+\cdots a_mS^{m-1})G_1X \\ &\quad +a_1g_1(Y)+\cdots a_mg_m(Y). \end{aligned}$$

It is easy to see that $a_1I+a_2S+\cdots a_mS^{m-1}$ is a permutation of the space $\{0,1\}^m$, as pointed out by Nyberg [12]. Therefore, this matrix is nonsingular. It implies that $(a_1I+a_2D+\cdots a_mD^{m-1})G_1$ is a generator matrix of the linear code C_1. Then from the proof of Theorem 4.1, we see that $a_1f_1+\cdots a_mf_m$ satisfies $(2^{m-1}\epsilon)$-almost $PC(l)$ of order k.

The rest of the proof is straitforward from Sec.5. Q.E.D.

References

[1] N. Alon, O. Goldreich, J. Hastad, and R. Peralta. Simple constructions of almost k-wise independent random variables. *Random Structures and Algorithms* **3** (1992), 289–304.

[2] C.H. Bennett, G. Brassard, and J.-M. Robert. Privacy amplification by public discussion. *SIAM Journal on Computing* **17** (1988), 210–229.

[3] J. Bierbrauer, K. Gopalakrishnan and D.R. Stinson. Bounds for resilient functions and orthogonal arrays. *Lecture Notes in Computer Science* **839** (1994), 247–257 (CRYPTO '94).

[4] C.Carlet. On the propagation criterion of degree l and order k. In *Advances in Cryptology — EUROCRYPT '98 Proceedings*, *Lecture Notes in Computer Science* 1403, pages 462–474. Springer-Verlag, 1998.

[5] B. Chor, O. Goldreich, J. Hastad, J. Friedman, S Rudich and R. Smolensky. The bit extraction problem or t-resilient functions. *26th IEEE symposium on Foundations of Computer Science*, pages 396–407, 1985.

[6] Y.Dodis, A.Sahai and A.Smith, "On Perfect and Adaptive Security in Exposure-Resilient Cryptography", *Lecture Notes in Computer Science* **2045**, EUROCRYPT '01, pp.301–324 (2001)

[7] J. Friedman. On the bit extraction problem. *33rd IEEE symposium on Foundations of Computer Science*, pages 314–319, 1992.

[8] K.Kurosawa, T.Johansson, D.Stinson: "Almost k-wise Independent Sample Spaces and Their Cryptologic Applications", Journal of Cryptology, Vol.14, No.4, pp.231–253 (2001)

[9] K.Kurosawa and T.Satoh. Design of SAC/PC(l) of order k Boolean functions and three other cryptographic criteria. In *Advances in Cryptology — EUROCRYPT '97 Proceedings*, *Lecture Notes in Computer Science* 1233, pages 434–449. Springer-Verlag, 1997.

[10] F.J. MacWilliams and N.J.A. Sloane. *The Theory of Error-Correcting Codes*. North-Holland, 1977.

[11] J. Naor and M. Naor. Small bias probability spaces: efficient constructions and applications. *SIAM Journal on Computing* **22** (1993), 838–856.

[12] K. Nyberg. Perfect nonlinear S-boxes. In *Advances in Cryptology — EUROCRYPT '91 Proceedings*, *Lecture Notes in Computer Science* 547, pages 378–386. Springer-Verlag, 1991.

[13] D. R. Stinson. Resilient functions and large set of orthogonal arrays. *Congressus Numerantium* **92** (1993), 105–110.

[14] D.R. Stinson and J.L. Massey. An infinite class of counterexamples to a conjecture concerning nonlinear resilient functions. *Journal of Cryptology* **8** (1995), 167–173.

[15] B.Preneel, W.Van Leekwijck, L.Van Linden, R.Govaerts, and J.Vandewalle. Propagation characteristics of Boolean functions. In *Advances in Cryptology — EUROCRYPT '90 Proceedings*, *Lecture Notes in Computer Science* 473, pages 161–173. Springer-Verlag, 1991.

[16] B. Preneel, R. Govaerts, and J. Vandewalle. Boolean functions satisfying higher order propagation criteria. In *Advances in Cryptology — EUROCRYPT '91 Proceedings*, *Lecture Notes in Computer Science* 547, pages 141–152. Springer-Verlag, 1991.

[17] A.F.Webster and S.E.Tavares. On the design of S-boxes. In *Advances in Cryptology — CRYPTO '85 Proceedings*, *Lecture Notes in Computer Science* 218, pages 523–534. Springer-Verlag, 1986.

[18] Andre Weil. Basic Number Theory. Springer-Verlag, New York, 1995