Construction of High Degree Resilient S-Boxes With Improved Nonlinearity

Kishan Chand Gupta and Palash Sarkar

Cryptology Research Centre

Applied Statistics Unit

Indian Statistical Institute

203, B.T. Road, Kolkata 700 108, INDIA

E-mail: {kishan_t, palash}@isical.ac.in

Abstract

We present a simple method to use an $[n-d-1, m, t+1]$ code to construct an n-input, m-output, t-resilient function with degree $d > m$ and nonlinearity $2^{n-1} - 2^{n-\lceil\frac{d+1}{2}\rceil} - (m+1)2^{n-d-1}$. For any fixed values of parameters n, m, t and d, with $d > m$, the nonlinearity obtained by our construction is higher than the nonlinearity obtained by the *only* previously known construction which provides $d > m$ (due to Cheon in Crypto 2001).

Keywords : Cryptography, S-box, Resiliency, nonlinearity, algebraic degree, stream cipher.

1 Introduction

Resilient S-boxes were introduced by Chor et al [3] and Bennett et al [1]. The study of other important cryptographic properties of resilient S-boxes such as high nonlinearity and algebraic degree have been performed in [2, 5, 6, 8, 10]. In [2], Cheon used an $[n-d-1, m, t+1]$ code to construct an n-input, m-output, t-resilient S-box with degree $d > m$ and nonlinearity $(2^{n-1} - 2^{n-d-1}\lfloor\sqrt{2^n}\rfloor + 2^{n-d-2})$. *Till date this is the only known method which can be used to obtain degree $d > m$.* The construction of Cheon uses the algebraic structure of linearized polynomials and the nonlinearity calculation is based on the Hasse-Weil bound for higher genus curves.

In this paper we describe a *simple* construction of nonlinear resilient S-boxes. Given an $[n-d-1, m, t]$ code we construct an n-input, m-output, t-resilient S-box with degree $d > m$ and nonlinearity $2^{n-1} - 2^{n-\lceil\frac{d+1}{2}\rceil} - (m+1)2^{n-d-1}$. Further we *prove* that for any fixed values of the parameters n, m, t and d with $d > m$, the nonlinearity obtained by our method is in all cases higher than the nonlinearity obtained by Cheon's method.

2 Preliminaries

Let $F_2 = GF(2)$ be the finite field of two elements. We consider the domain of an n-variable Boolean function to be the vector space $(F_2^n, \oplus)$ over F_2, where $\oplus$ is used to denote the addition operator over both F_2 and the vector space F_2^n. The inner product of two vectors $u, v \in F_2^n$ will be denoted by $\langle u, v\rangle$.

The Walsh transform of an m-variable Boolean function g is an integer valued function $W_g : \{0,1\}^m \rightarrow [-2^m, 2^m]$ defined by (see [7, page 414]) $W_g(u) = \sum_{w\in F_2^m}(-1)^{g(w)\oplus\langle u,w\rangle}$. An m-variable function is called *t-resilient* if $W_g(u) = 0$ for all u with $0 \leq wt(u) \leq t$ [9]. The *nonlinearity* $\mathsf{nl}(f)$ of an n-variable Boolean function f, is defined as $\mathsf{nl}(f) = 2^{n-1} - \frac{1}{2}\max_{u\in F_2^n}|W_f(u)|$.

A Boolean function g can be uniquely represented by a multivariate polynomial over F_2. The degree of the polynomial is called the algebraic degree or simply the *degree* of g and is denoted by $\mathsf{deg}(g)$.

An (n,m) S-box (or vectorial function) is a map $f:\{0,1\}^n \to \{0,1\}^m$. Let $f:\{0,1\}^n \to \{0,1\}^m$ be an S-box and $g:\{0,1\}^m \to \{0,1\}$ be an m-variable Boolean function. The composition of g and f, denoted by $g \circ f$ is an n-variable Boolean function defined by $(g \circ f)(x) = g(f(x))$. An (n,m) S-box f is said to be t-resilient, if $g \circ f$ is t-resilient for every m-variable linear function g (see for example [5]). By an (n,m,t) S-box we mean t-resilient (n,m) S-box. Let f be an (n,m) S-box. Then nonlinearity of f, denoted by $\mathsf{nl}(f)$, is defined to be

$$\mathsf{nl}(f) = \min\{\mathsf{nl}(g \circ f) : g \text{ is a nonconstant } m\text{-variable linear function }\}.$$

Similarly the algebraic degree of f, denoted by $\mathsf{deg}(f)$, is defined to be

$$\mathsf{deg}(f) = \min\{\mathsf{deg}(g \circ f) : g \text{ is a nonconstant } m\text{-variable linear function }\}.$$

It is easy to see that if f is an (n,m) S-box, then $\mathsf{deg}(f) < n$. By an (n,m,t) S-box (or (n,m,t)-resilient function) we mean t-resilient (n,m) S-box. Similarly by an (n,m,t,d) S-box (or (n,m,t,d)-resilient function) we mean t-resilient (n,m) S-box with algebraic degree d.

3 Construction of (n,m,t)-Resilient S-box with Degree $> m$.

We will be interested in (n,m) S-boxes with maximum possible nonlinearity. If $n = m$, the S-boxes achieving the maximum possible nonlinearity are called maximally nonlinear [4]. If n is odd, then maximally nonlinear S-boxes have nonlinearity $2^{n-1} - 2^{\frac{n-1}{2}}$. For even n, it is possible to construct (n,m) S-boxes with nonlinearity $2^{n-1} - 2^{\frac{n}{2}}$, though it is an open question whether this value is the maximum possible [4]. The following result is well known(see for example [5]).

Theorem 3.1 *Let $\mathcal{C}$ be an $[n,m,t+1]$ binary linear code. Then we can construct an (n,m,t)-resilient function.*

The following result is a slight generalization of Theorem 3.1.

Lemma 3.1 *Let* $f : \{0,1\}^n \to \{0,1\}^m$ *be an* S*-box and* $f_1, f_2, \cdots, f_m$ *are the component functions. Let* $\mathcal{C}$ *be a* $[p, m, t+1]$ *binary linear code. We can construct a* t*-resilient* S*-box* $h : \{0,1\}^{n+p} \to \{0,1\}^m$ *with component functions* $h_1, h_2, \cdots, h_m$*, where* $\mathsf{nl}(h) = 2^p \mathsf{nl}(f)$ *and algebraic degree of* $h(x)$ *is same as algebraic degree of* $f(x)$*.*

Proof : A binary linear code $[p, m, t+1]$ is a vector space of dimension m over F_2. Let $\{C_1, C_2, \cdots, C_m\}$ be a basis where $C_i = (c_{i_1}, c_{i_2}, \cdots, c_{i_p}) \in F_2^p$. We construct the S-box $h : \{0,1\}^{n+p} \to \{0,1\}^m$ as follows. For $1 \leq i \leq m$, we define,

$$h_i(x_1, \cdots, x_n, x_{n+1}, \cdots, x_{n+p}) = f_i(x_1, \cdots, x_n) \oplus \langle C_i, (x_{n+1}, x_{n+2}, \cdots, x_{n+p}) \rangle.$$

Let h' be any nonzero linear combination of the component functions $h_1, \cdots, h_m$. So h' can be written as $h' = d_1 h_1 \oplus \cdots \oplus d_m h_m$ for some nonzero vector $(d_1, \cdots, d_m) \in F_2^m$. Hence

$$\begin{aligned} h' &= d_1 f_1 \oplus \cdots \oplus d_m f_m \oplus \langle d_1 C_1 \oplus \cdots \oplus d_m C_m, (x_{n+1}, \cdots, x_{n+p}) \rangle \\ &= d_1 f_1 \oplus \cdots \oplus d_m f_m \oplus \langle C', (x_{n+1}, \cdots, x_{n+p}) \rangle \end{aligned}$$

where $C' = d_1 C_1 \oplus \cdots \oplus d_m C_m$. We have weight of vector $C' \geq t+1$ since $\mathcal{C}$ is a $[p, m, t+1]$ linear code. Hence h' is t-resilient and so $h(x_1, \cdots, x_n, x_{n+1}, \cdots, x_{n+p})$ is t-resilient. As we are adding p new variables, $\mathsf{nl}(h) = 2^p \mathsf{nl}(f)$. Clearly $\mathsf{deg}(h) = \mathsf{deg}(f)$. ■

The next result provides a simple method to construct a $(d+1, m)$ S-box with degree d and very high nonlinearity.

Theorem 3.2 *It is possible to construct a* $(d+1, m)$ S*-box with degree* $d > m$ *and nonlinearity* $\mathsf{nl}(h) \geq 2^d - 2^{\lfloor \frac{d+1}{2} \rfloor} - (m+1)$*.*

Proof : Let f be a $(d+1, d+1)$ maximally nonlinear S-box whose component functions are $f_1, f_2, \cdots, f_{d+1}$. We construct a $(d+1, m)$ S-box h with component functions $h_1, h_2, \cdots, h_m$ in the following manner. For $1 \leq i \leq m$, define $\mu_i(x_1, \ldots, x_{d+1}) = x_1 \ldots x_{i-1}x_{i+1} \ldots x_{d+1}$, and

$$h_i(x_1, \cdots, x_{d+1}) = f_i(x_1, \cdots, x_{d+1}) \oplus \mu_i(x_1, \ldots, x_{d+1}).$$

By construction algebraic degree of S-box $h : \{0,1\}^{d+1} \to \{0,1\}^m$ is d. It is known that $\mathsf{nl}(f) \geq 2^d - 2^{\lfloor \frac{d+1}{2} \rfloor}$ [4]. We show that $\mathsf{nl}(h) \geq \mathsf{nl}(f) - (m+1)$. Let ϵ_i be the identity vector which has a one at the ith position and zero elsewhere. Let $\mathbf{1} = (1, \ldots, 1)$. From the definition of μ_i it is clear that $\mathsf{Sup}(\mu_\mathsf{i}) = \{(x_1, \ldots, x_{d+1}) : \mu_i(x_1, \ldots, x_{d+1}) = 1\} = \{\mathbf{1}, \bar{\epsilon_i}\}$. A nonzero linear combination h' of the component functions $h_1, \ldots, h_m$ can be written as

$$h' = f_{i_1} \oplus \cdots \oplus f_{i_r} \oplus \mu_{i_1} \oplus \ldots \oplus \mu_{i_r}, \text{ for some } \{i_1, i_2, \cdots, i_r\} \subseteq \{1, 2, \cdots, m\}.$$

We have $\cup_{j=1}^{r} \mathsf{Sup}(\mu_{i_j}) = \{\mathbf{1}, \epsilon_{i_1}, \ldots, \epsilon_{i_r}\}$ and so the weight of the function $\mu_{i_1} \oplus \ldots \oplus \mu_{i_r}$ is at most $r+1$. Hence $\mathsf{nl}(h') \geq \mathsf{nl}(f) - (r+1)$. Since $r \leq m$, it follows that $\mathsf{nl}(h) \geq \mathsf{nl}(f) - (m+1)$ which gives us the required result. ■

Now we are ready to describe our construction method.

Construction-I .

1. Input: Parameters n, m, t and d with $d > m$.
2. Output: An $(n, m, t, d)-$resilient function.

Procedure

1. Construct a $(d+1, m)$ S-box using Theorem 3.2.
2. Let $\mathcal{C}$ be a $[n-d-1, m, t+1]$ code. If no such code exists, then stop. The function cannot be constructed using this method.
3. Apply Lemma 3.1 on h and $\mathcal{C}$ to construct the required S-box g.

Theorem 3.3 *If an $[n-d-1, m, t+1]$ code exists, then Construction-I constructs an (n, m, t, d) S-box g with nonlinearity $2^{n-1} - 2^{n-\lceil \frac{d+1}{2} \rceil} - (m+1)2^{n-d-1}$.*

Proof : By Theorem 3.2, $\mathsf{nl}(h) = 2^d - 2^{\lfloor \frac{d+1}{2} \rfloor} - (m+1)$ and $\mathsf{deg}(h) = d$. By Lemma 3.1, g is t-resilient, $\mathsf{nl}(g) = 2^{n-d-1}\mathsf{nl}(h) = 2^{n-1} - 2^{n-\lceil \frac{d+1}{2} \rceil} - (m+1)2^{n-d-1}$ and $\mathsf{deg}(g) = \mathsf{deg}(h) = d$. ■

4 Comparison

In [2, Theorem 5], Cheon proved the following result.

Theorem 4.1 *For any non-negative integer d, if there exists $[n-d-1, m, t+1]$ linear code then there exists a (n, m, t)-resilient function with degree d and nonlinearity $(2^{n-1} - 2^{n-d-1}\lfloor\sqrt{2^n}\rfloor + 2^{n-d-2})$.*

The nonlinearity calculation in the above theorem is based on Hasse-Weil bound for higher genus curves. Till date, this is the only construction which provides (n, m, t) nonlinear resilient S-boxes with degree greater than m. In the next Theorem we prove that nonlinearity obtained by Construction-I is higher than nonlinearity obtained by Cheon's construction.

Theorem 4.2 *Let f be an (n, m, t, d)-resilient function f with $d > m$ and nonlinearity n_1 constructed by Cheon's method. Then it is possible to construct an (n, m, t, d)-resilient function g with nonlinearity n_2 using Construction-I. Further $n_2 > n_1$.*

Proof : As (n, m, t)-resilient function f is constructed by Cheon's method, there exists an $[n - d - 1, m, t + 1]$ code. Construction-I can be applied to obtain an (n, m, t)-resilient function g with degree d and nonlinearity $\mathsf{nl}(g) = n_2 = 2^{n-1} - 2^{n-\lceil \frac{d+1}{2} \rceil} - (m+1)2^{n-d-1}$. It remains to show that $n_2 > n_1$, which we show now. Recall $n_1 = 2^{n-1} - 2^{n-d-1}\lfloor\sqrt{2^n}\rfloor + 2^{n-d-2}$. Hence $n_2 - n_1 = -2^{n-\lceil \frac{d+1}{2} \rceil} + 2^{n-d-1}\lfloor\sqrt{2^n}\rfloor - 2^{n-d-2} - (m+1)2^{n-d-1}$. Thus we have $n_2 > n_1$ if $2^{n-d-1}\lfloor 2^{\frac{n}{2}} \rfloor > 2^{n-\lceil \frac{d+1}{2} \rceil} + 2^{n-d-2} + (m+1)2^{n-d-1}$. The last condition holds if and only if $\lfloor 2^{\frac{n}{2}} \rfloor > \frac{1}{2} + (m+1) + 2^{d+1-\lceil \frac{d+1}{2} \rceil}$. So $n_2 > n_1$ if $2^{\frac{n}{2}} - 1 > \frac{3}{2} + m + 2^{\frac{d+1}{2}}$, i.e., if $2^{\frac{n}{2}} > \frac{5}{2} + m + 2^{\frac{d+1}{2}}$. It is simple to check that the last condition holds for $n \geq d + 2$ for all positive integers n and $d > m$. The maximum possible degree of an S-box is $(n-1)$, so $d \leq n-1$. For $d = n-1$ the required code is $[0, m, t+1]$. Clearly such a code does not exist. Hence $n_2 > n_1$ holds for all d. ∎

Remark : We note that Cheon's method does not provide any nonlinearity for $d \leq \frac{n-1}{2}$, whereas Construction-I provides positive nonlinearity for $d > 2$.

5 Conclusion

In this paper we have presented a simple construction of nonlinear resilient S-boxes with algebraic degree greater than m. We *proved* that for any fixed values of the parameters n, m, t and d, the nonlinearity obtained by our construction is superior to the nonlinearity obtained by the only previously known construction [2] which can provide degree $d > m$.

References

[1] C. Bennett, G. Brassard and J. Robert. Privacy Amplification by Public Discussion. *SIAM Journal of Computing*, volume 17, pages 210–229, 1988.

[2] Jung Hee Cheon. Nonlinear Vector Resilient Functions. In *Advances in Cryptology - CRYPTO 2001*, pages 458–469, Lecture Notes in Computer Science, Springer-Verlag, 2001.

[3] 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*, volume 26, pages 396–407, 1985.

[4] Hans Dobbertin, Almost Perfect Nonlinear Power Functions on $GF(2^n)$: The Welch Case. *IEEE Transactions on Information Theory*, Vol 45 , No 4, pp. 1271-1275 , 1999.

[5] T. Johansson and E. Pasalic. A construction of resilient functions with high nonlinearity. *International Symposium on Information Theory*, 2000.

[6] K. Kurosawa, T. Satoh and K. Yamamoto. Highly nonlinear t-resilient functions . Journal of Universal Computer Science, vol.3, no. 6, pp. 721-729, Springer Publishing Company, 1997.

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

[8] E. Pasalic and S. Maitra. Linear Codes in Constructing Resilient Functions with High Nonlinearity. *Selected Areas in Cryptography 2001*, pages 60–74, Lecture Notes in Computer Science, Springer-Verlag, 2001.

[9] G. Xiao and J. L. Massey. A spectral characterization of correlation-immune combining functions. *IEEE Transactions on Information Theory*, pages 569–571, 1988.

[10] X.-M. Zhang and Y. Zheng, On Cryptographically Resilient Functions. *IEEE Transactions on Information Theory*, Vol 43 , No 5, pp. 1740-1747 , 1997.