Information Processing Letters 95 (2005) 413–417

Construction of high degree resilient S-boxes with improved nonlinearity

Kishan Chand Gupta *, Palash Sarkar

Cryptology Research Centre, Applied Statistics Unit, Indian Statistical Institute, 203, B.T. Road, Kolkata 700 108, India

Received 20 June 2002; received in revised form 16 December 2004; accepted 7 February 2005

Communicated by Y. Desmedt

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 $2^{n-1} - 2^{n-\lceil (d+1)/2 \rceil} - (m+1)2^{n-d-1}$. 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 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 [2,6,7,9,12]. 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 $(2^{n-1} - 2^{n-d-1}\lfloor\sqrt{2^n}\rfloor + 2^{n-d-2})$. 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 $2^{n-1} - 2^{n-\lceil (d+1)/2 \rceil} - (m+1)2^{n-d-1}$. 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.

1.1. Work since the submission of this paper

After submitting this work to IPL, we continued our study and applied more sophisticated techniques to ob-

* Corresponding author.
E-mail addresses: kishan_t@isical.ac.in (K.C. Gupta), palash@isical.ac.in (P. Sarkar).

tain higher nonlinearity. These results were recently published in [5] and to the best of our knowledge provides the currently best known nonlinearity.

The motivations for the current paper and [5] are different. The point of the current paper is that simple techniques can give good results. The point of [5] is to obtain the best possible nonlinearity through the use of sophisticated techniques. Further, the gain in nonlinearity obtained in [5] over the current paper is not by a large amount, so that there is really a trade-off between simplicity and gain in nonlinearity.

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 $W_g : \{0, 1\}^m \to [-2^m, 2^m]$ defined by (see [8, p. 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 $0 \leqslant wt(u) \leqslant t$ [11]. The *nonlinearity* $\mathsf{nl}(f)$ of an n-variable Boolean function f, is defined as

$$\mathsf{nl}(f) = 2^{n-1} - \frac{1}{2}\max_{u \in F_2^n}\left|W_f(u)\right|.$$

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, [6]). 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

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

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

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

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.

We are interested in S-boxes (as opposed to Boolean functions) and hence we will assume that $m > 1$. Also we are interested in S-boxes for which $d > m$ and for resilient S-boxes it is known [10] that $d < n$. So the following condition holds: $1 < m < d < n$. This implies that for the S-boxes in which we are interested, the minimum value of n is 4.

3. Construction of (n, m, t)-resilient S-box with degree greater than 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 $2^{n-1} - 2^{(n-1)/2}$. For even n, it is possible to construct (n, m) S-boxes with nonlinearity $2^{n-1} - 2^{n/2}$, though it is an open question whether this value is the maximum possible [4]. The following result is well known (see, for example, [6]).

Theorem 3.1. *Let 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 $f_1, f_2, \ldots, f_m$ are the component functions. Let 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 $h_1, h_2, \ldots, h_m$, 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 $\{C_1, C_2, \ldots, C_m\}$ be a basis where $C_i = (c_{i_1}, c_{i_2}, \ldots, c_{i_p}) \in F_2^p$. We construct the S-box $h : \{0,1\}^{n+p} \to \{0,1\}^m$ as follows. For $1 \leqslant i \leqslant m$, we define,

$$\begin{aligned} &h_i(x_1, \ldots, x_n, x_{n+1}, \ldots, x_{n+p}) \\ &\quad = f_i(x_1, \ldots, x_n) \oplus \langle C_i, (x_{n+1}, x_{n+2}, \ldots, x_{n+p}) \rangle. \end{aligned}$$

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

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

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

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* $\mathsf{nl}(h) \geqslant 2^d - 2^{\lfloor (d+1)/2 \rfloor} - (m+1)$.

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

$$\mu_i(x_1, \ldots, x_{d+1}) = x_1 \ldots x_{i-1} x_{i+1} \ldots x_{d+1},$$

and

$$\begin{aligned} h_i(x_1, \ldots, x_{d+1}) &= f_i(x_1, \ldots, x_{d+1}) \\ &\quad \oplus \mu_i(x_1, \ldots, x_{d+1}). \end{aligned}$$

By construction algebraic degree of S-box $h : \{0,1\}^{d+1} \to \{0,1\}^m$ is d. It is known that $\mathsf{nl}(f) \geqslant 2^d - 2^{\lfloor (d+1)/2 \rfloor}$ [4]. We show that $\mathsf{nl}(h) \geqslant \mathsf{nl}(f) - (m+1)$. 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 $\mathsf{Sup}(\mu_i) = \{(x_1, \ldots, x_{d+1}) \colon \mu_i(x_1, \ldots, x_{d+1}) = 1\} = \{\mathbf{1}, \bar{\varepsilon_i}\}$. A nonzero linear combination h' of the component functions $h_1, \ldots, h_m$ can be written as

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

We have $\bigcup_{j=1}^{r} \mathsf{Sup}(\mu_{i_j}) = \{\mathbf{1}, \varepsilon_{i_1}, \ldots, \varepsilon_{i_r}\}$ and so the weight of the function $\mu_{i_1} \oplus \cdots \oplus \mu_{i_r}$ is at most $r+1$. Hence $\mathsf{nl}(h') \geqslant \mathsf{nl}(f) - (r+1)$. Since $r \leqslant m$, it follows that $\mathsf{nl}(h) \geqslant \mathsf{nl}(f) - (m+1)$ 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* $2^{n-1} - 2^{n - \lceil (d+1)/2 \rceil} - (m+1) \cdot 2^{n-d-1}$.

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

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. First we need the following result.

Lemma 4.1. *For $n \geqslant d + 4$, and $2 \leqslant m < d < n$ we have $\lfloor 2^{n/2}\rfloor > \frac{1}{2} + (m + 1) + 2^{\lfloor (d+1)/2\rfloor}$.*

Proof. We have to show

$$\lfloor 2^{n/2}\rfloor > \tfrac{1}{2} + (m + 1) + 2^{\lfloor (d+1)/2\rfloor}. \qquad (1)$$

Since $\lfloor 2^{n/2}\rfloor \geqslant 2^{n/2} - 1$ and $2^{(d+1)/2} \geqslant 2^{\lfloor (d+1)/2\rfloor}$, Eq. (1) holds if

$$2^{n/2} - 1 > \tfrac{1}{2} + (m + 1) + 2^{(d+1)/2}. \qquad (2)$$

Since $m < d$, we have $m \leqslant d - 1$ and Eq. (2) holds if

$$2^{n/2} - 1 > \tfrac{1}{2} + (d - 1 + 1) + 2^{(d+1)/2}. \qquad (3)$$

Again since $n \geqslant d + 4$, we have that Eq. (3) holds if

$$2^{(d+4)/2} > \tfrac{3}{2} + d + 2^{(d+1)/2}. \qquad (4)$$

Thus, Eq. (1) holds if

$$2^{d/2}(4 - \sqrt{2}) > \tfrac{3}{2} + d. \qquad (5)$$

Clearly, Eq. (5) holds for all $d \geqslant 2$. Hence the proof. □

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

$$\mathrm{nl}(g) = n_2 = 2^{n-1} - 2^{n-\lceil (d+1)/2\rceil} - (m + 1)2^{n-d-1}.$$

It remains to show that $n_2 > n_1$, which we show now. Recall $n_1 = 2^{n-1} - 2^{n-d-1}\lfloor\sqrt{2^n}\rfloor + 2^{n-d-2}$. The maximum possible degree of an S-box is $(n - 1)$, so $d \leqslant n - 1$. For $d = n - 1, n - 2$ and $n - 3$ the required codes are $[0, m, t + 1]$, $[1, m, t + 1]$ and $[2, m, t + 1]$. Since $2 \leqslant m$ (see Section 2, last paragraph) the first two codes do not exist and so Cheon's method cannot be applied for $d = n - 1$ and $n - 2$. The third code exists only for $m = 2$ and $t = 0$, in which case the code is a $[2, 2, 1]$ code. In this case, $n - d - 1 = 2$, $m = 2$, $d \geqslant 3$ and so $n \geqslant 6$. Also, for this case, $n_2 > n_1$ if $4 \times (\lfloor 2^{n/2}\rfloor - 2^{\lfloor (n-2)/2\rfloor}) > 14$. This condition holds for $n \geqslant 6$. Hence, $n_2 > n_1$ for $d = n - 3$.

Now we consider the case $d \leqslant n - 4$. We have $n_2 - n_1 = -2^{n-\lceil (d+1)/2\rceil} + 2^{n-d-1}\lfloor\sqrt{2^n}\rfloor - 2^{n-d-2} - (m + 1)2^{n-d-1}$. Thus we have $n_2 > n_1$ if $2^{n-d-1}\lfloor 2^{n/2}\rfloor > 2^{n-\lceil (d+1)/2\rceil} + 2^{n-d-2} + (m + 1)2^{n-d-1}$. The last condition holds if and only if $\lfloor 2^{n/2}\rfloor > \frac{1}{2} + (m + 1) + 2^{\lfloor (d+1)/2\rfloor}$. Since $2 \leqslant m < d < n$ (see Section 2, last paragraph) and $n \geqslant d + 4$ we apply Lemma 4.1 and get $n_2 > n_1$. This completes the proof of the result. □

Remark. We note that Cheon's method does not provide any nonlinearity for $d \leqslant \frac{n-1}{2}$, 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, with $d > m$, the nonlinearity obtained by our simple construction is higher than the nonlinearity obtained by the more complicated algebraic construction of Cheon [2] in Crypto 2001.

References

[1] C. Bennett, G. Brassard, J. Robert, Privacy amplification by public discussion, SIAM J. Comput. 17 (1988) 210–229.

[2] J.H. Cheon, Nonlinear vector resilient functions, in: Advances in Cryptology—CRYPTO 2001, in: Lecture Notes in Comput. Sci., Springer-Verlag, Berlin, 2001, pp. 458–469.

[3] B. Chor, O. Goldreich, J. Hastad, J. Friedman, S. Rudich, R. Smolensky, The bit extraction problem or t-resilient functions, in: IEEE Symp. on Foundations of Computer Science, vol. 26, 1985, p. 396–407.

[4] H. Dobbertin, Almost perfect nonlinear power functions on $GF(2^n)$: the Welch case, IEEE Trans. Inform. Theory 45 (4) (1999) 1271–1275.

[5] K.C. Gupta, P. Sarkar, Improved construction of nonlinear resilient S-boxes, IEEE Trans. Inform. Theory 51 (1) (2005) 339–348.

[6] T. Johansson, E. Pasalic, A construction of resilient functions with high nonlinearity, in: Internat. Symp. on Information Theory, 2000.

[7] K. Kurosawa, T. Satoh, K. Yamamoto, Highly nonlinear t-resilient functions, J. Universal Comput. Sci. 3 (6) (1997) 721–729.

[8] F.J. MacWillams, N.J.A. Sloane, The Theory of Error Correcting Codes, North-Holland, Amsterdam, 1977.

[9] E. Pasalic, S. Maitra, Linear codes in constructing resilient functions with high nonlinearity, in: Selected Areas in Cryptography 2001, in: Lecture Notes in Comput. Sci., Springer-Verlag, Berlin, 2001, pp. 60–74.

[10] T. Siegenthaler, Correlation-immunity of nonlinear combining functions for cryptographic applications, IEEE Trans. Inform. Theory IT-30 (5) (1984) 776–780.

[11] G. Xiao, J.L. Massey, A spectral characterization of correlation-immune combining functions, IEEE Trans. Inform. Theory 34 (3) (1988) 569–571.

[12] X.-M. Zhang, Y. Zheng, On cryptographically resilient functions, IEEE Trans. Inform. Theory 43 (5) (1997) 1740–1747.