Systematic Generation of Cryptographically Robust S-boxes [*]

Jennifer Seberry
Xian-Mo Zhang
Yuliang Zheng

The Centre for Computer Security Research
Department of Computer Science
The University of Wollongong
Wollongong, NSW 2522, AUSTRALIA
E-mail: {jennie,xianmo,yuliang}@cs.uow.edu.au

[*] A preliminary version of the paper appeared in *The Proceedings of the First ACM Conference on Computer and Communications Security*, Fairfax, Virginia, USA, November 1993. The first author was supported in part by the Australian Research Council under the reference numbers A49130102, A9030136, A49131885 and A49232172, the second author by A49130102, and the third author by A49232172.

Systematic Generation of Cryptographically Robust S-boxes

Abstract

Substitution boxes (S-boxes) are a crucial component of DES-like block ciphers. This research addresses problems with previous approaches towards constructing S-boxes, and proposes a new definition for the robustness of S-boxes to differential cryptanalysis, which is the most powerful cryptanalytic attack known to date. A novel method based on group Hadamard matrices is developed to systematically generate S-boxes that satisfy a number of critical cryptographic properties. Among the properties are the high nonlinearity, the strict avalanche characteristics, the balancedness, the robustness against differential cryptanalysis, and the immunity to linear cryptanalysis. An example is provided to illustrate the S-box generating method.

1 Introduction

Differential cryptanalysis discovered by Biham and Shamir [?, ?] is currently the most powerful cryptanalytic attack to (secret-key) block ciphers, especially to DES-like substitution-permutation ciphers. The attack applies also to other cryptographic primitives such as one-way hash functions.

Since differential cryptanalysis was introduced, researchers have devoted a large number of efforts to designing substitution boxes (S-boxes) in order to strengthen the security of a block cipher against the attack [?, ?, ?, ?, ?, ?]. Although these S-boxes are interesting in terms of their security against differential cryptanalysis, they bear a number of shortcomings which render them unattractive in practice. These shortcomings will be fully addressed in Section **??**. Here we mention briefly two of them: (1) The S-boxes are based on permutation polynomials on finite fields, and hence have an equal number of input and output bits. Note that existing ciphers including DES, LOKI and FEAL employ S-boxes with less output bits than input bits. Though dropping an appropriate number of component functions from a permutation polynomial yields an S-box with less output bits, there is no guarantee that the resulting S-box is robust against differential cryptanalysis. (2) None of the component functions of the S-boxes satisfies the strict avalanche criterion (SAC). The SAC is considered as an indispensable requirement for S-boxes employed by a modern block cipher.

This research initiates the investigation of methods for systematically constructing S-boxes with a number of essential cryptographic properties. These properties include: security against differential cryptanalysis, immunity to the very recently discovered linear cryptanalysis [?], the SAC, balancedness, high nonlinearity, and uncorrelatedness. (Two or more Boolean functions are said to be uncorrelated if their sum gives a nonlinearly balanced function). A novel S-box construction method based on group Hadamard matrices is presented. An n-input s-output S-box (namely, an $n \times s$ S-box) constructed using this method, where $s > \lfloor n/2 \rfloor$, has the features now described.

1. It is at least $(1-2^{-t})$-robust against differential cryptanalysis, where t is a parameter subject to the condition that $(s - \lfloor n/2 \rfloor) > t \geqq 3$. For instance, when t =3, 5, or 7, the robustness is 0.875, 0.97 or 0.99 respectively. (See Section **??** for the definition of robustness.)

2. The sum of any subset of the component functions is a nonlinearly balanced function. Hence the component functions are all uncorrelated.

3. The nonlinearity of any component function is at least $2^{n-1} - 2^{s-t-1}$, which is a very high value, and its maximum algebraic degree is $n - s + t + 1$.

4. All component functions satisfy the SAC.

5. For each s-bit vector y, there are exactly 2^{n-s} n-bit vectors that are mapped to y. That is, the S-box is a regular many-to-one mapping.

These statements are very informal. The interested reader is directed to Section **??** for precise descriptions.

Section **??** introduces basic notations and definitions, and Section **??** addresses problems with previously proposed methods for constructing S-boxes. A new definition for robustness against differential cryptanalysis is introduced in the same section. Our first attempt to construct S-boxes is described in Section **??**, while improvements towards the robustness of the S-boxes are described in Section **??**. This is followed by a discussion of further refinement in Section **??**. An analysis of the number of different S-boxes that can be obtained by our method is conducted in Section **??**. Section **??** shows that the S-boxes constructed are also immune to linear cryptanalysis. An interesting relation between the SAC and the profile of the difference distribution table of an S-box is revealed in the same section. To illustrate the construction method, an example is shown in Section **??**. The paper is closed by some final remarks in Section **??**.

2 Basic Definitions

The vector space of n tuples of elements from $GF(2)$ is denoted by V_n. Vectors in V_n and integers in $[0, 2^n - 1]$ have a natural one-to-one correspondence. This allows us to switch from a vector in V_n to its corresponding integer in $[0, 2^n - 1]$, and vice versa.

Let f be a (Boolean) function from V_n to $GF(2)$ (or simply, a function on V_n). The *sequence* of f is defined as $((-1)^{f(\alpha_0)}$, $(-1)^{f(\alpha_1)}$, ..., $(-1)^{f(\alpha_{2^n-1})})$, while the *truth table* of f is defined as $(f(\alpha_0)$, $f(\alpha_1)$, ..., $f(\alpha_{2^n-1}))$, where α_i, $i = 0, 1, \ldots, 2^n - 1$, denote the vectors in V_n. f is said to be balanced if its truth table has an equal number of zeros and ones.

We call $h(x) = a_1x_1 \oplus \cdots \oplus a_nx_n \oplus c$ an affine function, where $x = (x_1, \ldots, x_n)$ and $a_j, c \in GF(2)$. In particular, h will be called a linear function if $c = 0$. The sequence of an affine (linear) function will be called an affine (linear) sequence.

The *Hamming weight* of a vector x, denoted by $W(x)$, is the number of ones in x. Let f and g be functions on V_n. Then $d(f, g) = \sum_{f(x)\neq g(x)} 1$, where the addition is over the reals, is called the *Hamming distance* between f and g. Let $\varphi_0, \ldots, \varphi_{2^{n+1}-1}$ be the affine functions on V_n. Then $N_f = \min_{i=0,\ldots,2^{n+1}-1} d(f, \varphi_i)$ is called the *nonlinearity* of f. It is well-known that the nonlinearity of f on V_n satisfies $N_f \leqq 2^{n-1} - 2^{\frac{1}{2}n-1}$. An extensive investigation of highly nonlinear balanced functions has been carried out in [?].

Let $\alpha = (a_1, \ldots, a_n) \in V_n$ and $\beta = (b_1, \ldots, b_n) \in V_n$. Then the scalar product of α and β, denoted by $\langle\alpha, \beta\rangle$, is defined by $\langle\alpha, \beta\rangle = \bigoplus_{j=1}^{n} a_jb_j$, where the addition and the multiplication are over $GF(2)$. A function f on V_n is said to be bent if

$$2^{-\frac{n}{2}} \sum_{x\in V_n} (-1)^{f(x)\oplus\langle\beta,x\rangle} = \pm 1$$

for every $\beta \in V_n$, where $x = (x_1, \ldots, x_n)$ [?]. Here $f(x) \oplus \langle\beta, x\rangle$ is considered as a real valued function. Bent functions exist only when n is even, and they achieve the maximum nonlinearity of $2^{n-1} - 2^{\frac{1}{2}n-1}$ [?, ?].

The concept of SAC was originally introduced in [?].

Definition 1 *Let f be a function on V_n and let $x = (x_1, \ldots, x_n)$. If $f(x) \oplus f(x \oplus \alpha)$ for every $\alpha \in V_n$ with $W(\alpha) = 1$, we say that f satisfies the strict avalanche criterion (SAC).*

Let f_0 and f_1 be functions on V_t. Then $f(x_0, x_1, \ldots, x_t) = (1 \oplus x_0)f_0(x_1, \ldots, x_t) \oplus x_0f_1(x_1, \ldots, x_t)$ is a function on V_{t+1}. The truth table of f is obtained by concatenating the truth tables of f_0 and f_1. For this reason we say that f is the concatenation of f_0 and f_1. Similarly we can define the concatenation of 2^s functions on V_t. To simplify the description of the concatenation of functions, we introduce a new notation. Let $s \geqq 1$ and $\delta = (i_1, \ldots, i_s)$ be a vector in V_s. Then D_δ is a function on V_s defined by

$$D_\delta(y) = (\bar{i_1} \oplus y_1) \cdots (\bar{i_s} \oplus y_s)$$

where $y = (y_1, \ldots, y_s)$ and $\bar{i} = 1 \oplus i$. For instance, when $s = 2$ we have $D_{0,0}(y_1, y_2) = (1 \oplus y_1)(1 \oplus y_2)$, and when $s = 3$ we have $D_{1,0,1}(y_1, y_2, y_3) = y_1(1 \oplus y_2)y_3$. Clearly $D_\delta(y) = 1$ if and only if $y = \delta$. To further simplify our description, D_δ will also be denoted by D_i where i is the integer in $[0, 2^s - 1]$ whose binary representation is δ.

Let f_0, f_1, ..., f_{2^s-1} be functions on V_t. Then the concatenation of these functions is

$$f(y, x) = \bigoplus_{i=0}^{2^s-1} [D_i(y) f_i(x)]$$

where $y = (y_1, \ldots, y_s)$ and $x = (x_1, \ldots, x_t)$. Note that f is a function on V_{s+t}. The following lemma is derived from Theorems 4 and 5 of [**?**].

Lemma 1 *Let $t \geqq s$, f_0, f_1, ..., f_{2^s-1} be distinct nonzero linear functions on V_t, and r be an arbitrary function on V_s. Also let*

$$g(y, x) = \bigoplus_{i=0}^{2^s-1} [D_i(y) f_i(x)] \oplus r(y).$$

Then

1. *g is balanced,*
2. *the nonlinearity of g satisfies $N_g \geqq 2^{s+t-1} - 2^{t-1}$,*
3. *$g(z) \oplus g(z \oplus \gamma)$ is balanced for all $\gamma = (\beta, \alpha)$ with $W(\beta) \neq 0$, where $\beta \in V_s$ and $\alpha \in V_t$.*

A mapping (tuple of functions) $(f_1, \ldots, f_s)$, where each f_i is a function on V_n and $n \geqq s$, is said to be *regular* if for each vector $y \in V_s$ there are exactly 2^{n-s} vectors in V_n that are mapped to y. In [**?**], the following result is proved:

Theorem 1 *A mapping $(f_1, \ldots, f_s)$, where each f_i is a function on V_n and $n \geqq s$, is regular if and only if all nonzero linear combinations of f_1, ..., f_s are balanced.*

A good S-box must be a regular mapping. Otherwise some output vectors appear more often than others when the input to the S-box is chosen uniformly at random, and a cryptosystem that employs such an S-box might be vulnerable to a cryptanalyst who exploits the bias.

3 Differential Cryptanalysis

The essence of differential cryptanalysis is that it exploits particular entries in the difference distribution tables of S-boxes employed by a block cipher. Entries with higher values are particularly useful to the attack. The difference distribution table of an $n \times s$ S-box is a $2^n \times 2^s$ matrix. The rows of the matrix, indexed by the vectors in V_n, represent the change in the input, while the columns, indexed by the vectors in V_s, represent the change in the output of the S-box. An entry in the table indexed by $(\Delta X, \Delta Y)$ indicates the number of input vectors which, when changed by ΔX (in the sense of bit-wise XOR), result in a change in the output by ΔY (also in the sense of bit-wise XOR). Note that an entry in the table can only take an even value, the sum of the values in a row is always 2^n, and the first row is always $(2^n, 0, \ldots, 0)$. Also note that the first column indicates the *smoothness* of the S-box, namely the characteristic that a change in the input does not result in a change in the output. As is discussed below, the smoothness is an extremely useful characteristic to differential cryptanalysis.

To thwart differential cryptanalysis, the difference distribution tables of the S-boxes employed by a DES-like block cipher must not contain entries with large values (not counting the first entry in the first row). Based on this observation, the initial reaction was to construct S-boxes with flat (i.e. uniform) difference distribution tables [?, ?]. However, as was pointed out in [?, ?], having no large values is not sufficient to prevent differential cryptanalysis, and in fact, a block cipher that employs S-boxes with flat difference distribution tables is easily breakable by differential cryptanalysis that exploits the *iterative characteristics* of the cipher (see Definition 12 of [?]). Among the various possible iterative characteristics, the one that uses the smoothness of an S-box, i.e., values in the first column of the difference distribution table, is particularly effective. In conjunction with other techniques, this characteristic can be used to break, at least in principle, a DES-like block cipher with an arbitrary number of rounds. Sections 6 and 7 of [?] provide a comprehensive description of this topic. Therefore, in addition to the requirement of having no large values, the difference distribution table of an S-box should also contain as less nonzero entries as possible in its first column. This prompts us to introduce the following definition:

Definition 2 *Let $F = (f_1, \ldots, f_s)$ be an $n \times s$ S-box, where f_i is a function on V_n, $i = 1, \ldots, s$, and $n \geqq s$. Denote by L the largest value in the difference distribution table of F, and by R the number of nonzero entries in the first column of the table. In either case the value 2^n in the first row is not counted. Then we say that F is ε-robust against differential cryptanalysis, where ε is defined by*

$$\varepsilon = (1 - \frac{R}{2^n})(1 - \frac{L}{2^n}).$$

Note that there is another issue with the profile of the difference distribution table of an S-box, namely the fraction of nonzero entries contained by the table. In general, if an S-box is *not* robust against differential cryptanalysis, then the smaller the fraction of nonzero entries in the table, the faster the differential cryptanalytic attack [?, ?]. That is, the performance of differential cryptanalysis is proportional to the fraction of zero entries. This problem, however, is not significantly relevant to *robust* S-boxes, including those constructed in this paper, and hence has not been taken into consideration in defining robustness.

The robustness of an $n \times s$ S-box is small if R or L is large. For instance, the robustness of an $n \times s$ S-box is merely $\frac{1}{2^n}(1 - \frac{L}{2^n}) < \frac{1}{2^n}$ if its difference distribution table contains only nonzero entries in its first column. Such an S-box is extremely prone to differential cryptanalysis. Examples of such weak S-boxes include those with flat difference distribution tables proposed in [?, ?].

Large robustness is obtained only when both R and L are small. An S-box attains the maximum robustness when R and L achieve their smallest possible values simultaneously. Clearly, the smallest possible value for L is 2^{n-s}. As an S-box which achieves this value has a flat difference distribution table, we have $R = 2^n - 1$ and hence the robustness is less than $\frac{1}{2^n}$. Therefore to make R small, L must be at least 2^{n-s+1}. In the following discussions we suppose that $L \geqq 2^{n-s+1}$. Two cases, $n > s$ and $n = s$, are considered in order to determine the set of possible small values for R.

When $n > s$, an S-box defines a many-to-one mapping. For such an S-box, we have $R \geqq 1$. Thus the robustness against differential cryptanalysis is bounded from above by $(1 - \frac{1}{2^n})(1 - 2^{-s+1})$. To decide S-boxes which achieve the upper bound for robustness, consider an $n \times s$ S-box whose difference distribution table has the following profile: each row, except the first, of the table contains an equal number of zero and nonzero entries, and the nonzero entries all contain a value 2^{n-s+1}. Thus we have $L = 2^{n-s+1}$. The upper bound would be achieved if $R = 1$. However, it has been proved in [?] that if each row, except the first, of the table contains an equal number of zero and nonzero entries, then R must be $2^{n-1} - 2^{s-1}$. Consequently the robustness of the S-box is less than $\frac{3}{4}$. This example indicates that finding a good combination of R and L is not easy. It is not clear to the authors whether or not the upper bound $(1 - \frac{1}{2^n})(1 - 2^{-s+1})$ is actually attainable. Nevertheless, it will be seen in Sections **??** and **??** that there exist S-boxes whose robustness is very close the upper bound.

Next we consider the case when $n = s$, namely when an S-box is a permutation V_n. As any change in the input to a permutation results in a change in the output, the first column of its difference distribution table contains only zeros except for the first entry. Therefore the maximum robustness against differential cryptanalysis is $(1 - 2^{-n+1})$. The maximum robustness is attained by a permutation with the following difference distribution table: except for the first row, half of the entries in a row contain the value 2 while the other half contain the value 0. Such S-boxes have been extensively investigated in [?, ?, ?, ?]. These S-boxes, however, suffer some or all of the drawbacks described below, which render them unattractive in practice.

1. Their component functions are quadratic. This is true for all the permutations in [?, ?], the first type of permutations in [?], and some of the permutations in [?]. A block cipher that employs functions with such a low algebraic degree as S-boxes would be vulnerable to more classic cryptanalytic attacks than the state-of-the-art differential cryptanalysis.

2. It has been suggested that an $n \times s$ S-box, where $s < n$, be constructed by omitting component functions from a permutation on V_n [?, ?, ?, ?]. However, in general, omitting component functions of a $(1-2^{-n+1})$-robust permutation does not yield a robust $n \times s$ S-box. In particular, we have proved in [?] that for any $n \times n$ S-box whose component functions are quadratic, dropping a component function results in an $n \times (n-1)$ S-box whose robustness against differential cryptanalysis is only $\frac{2^{n-1}}{2^n}(1 - 2^{-n+2}) < \frac{1}{2}$. The robustness decays drastically as more component functions are dropped. We conjecture that a similar phenomenon happens even in the more general case where component functions of an $n \times n$ S-box are not quadratic.

3. An S-box is said to satisfy the SAC if its component functions all satisfy the SAC. This property is considered to be at least as essential as the robustness against differential cryptanalysis. This issue has been completely neglected in [?, ?, ?, ?, ?], and none of the S-boxes constructed in those papers satisfies the SAC.

4. The S-boxes, with the following two exceptions, only accept an odd number of input bits. Applications of such S-boxes are limited.

   The first exception is some of the S-boxes constructed in [?] which accept an even number of input bits. Unfortunately the component functions of these S-boxes are all quadratic.

   The second exception is the inverse function on $GF(2^n)$ defined by

$$F(X) = \begin{cases} 0 & \text{if } X = 0 \\ 1/X & \text{otherwise} \end{cases}$$

   Results proved in [?] indicate that the robustness of $F(X)$ against differential cryptanalysis is $(1 - 2^{-n+1})$ when n is odd, and $(1 - 2^{-n+2})$ when n is even. As the input to the function has to be checked against the value zero, it would be very inconvenient to use the function in practical applications. Although this inconvenience can be removed by using look up tables, the amount of memory required in storing the tables becomes intolerable when n is large.

Table 1: Robustness of S-boxes Used by DES

S-Box	L	R	ε
S_1	16	37	0.316
S_2	16	33	0.363
S_3	16	37	0.316
S_4	16	24	0.469
S_5	16	31	0.387
S_6	16	33	0.363
S_7	16	35	0.340
S_8	16	36	0.328

L : The largest value in the difference distribution table, not counting the value 2^6 in the first row.

R : The number of nonzero entries in the first column of the difference distribution table, not counting the first entry containing a value 2^6.

ε : Robustness against differential cryptanalysis. It is calculated by $\varepsilon = (1 - \frac{R}{2^6})(1 - \frac{L}{2^6})$.

Interesting results on constructing S-boxes have been presented in [**?**]. These include a few 5×5 S-boxes which are $(1 - 2^{-4})$-robust against differential cryptanalysis. Although these S-boxes satisfy the SAC, they all bear the other three shortcomings. In addition, since the method relies on exhaustive search, it is beyond the currently available computing power to find a larger, say 7×7, S-box with similar properties.

A final remark is that the construction methods used in [**?**, **?**, **?**, **?**, **?**, **?**] are essentially the same from a technical point of view: they are all based on permutation polynomials on $GF(2^n)$. Although such permutations are easy to analyze, they have a very restricted form and consist of only a small portion among all the permutations on $GF(2^n)$.

In the following sections we take a completely different approach, which is based on group Hadamard matrices, towards constructing S-boxes. The S-boxes generated using the new approach will free of all the drawbacks addressed above. Before going into the description of the new approach, we note that DES employs eight 6×4 S-boxes. The difference distribution tables of the S-boxes can be found in [**?**]. ¿From the tables it can be seen that the fractions of nonzero entries in the tables are between 0.70 and 0.80. Table **??** shows that the robustness of the eight S-boxes against differential cryptanalysis is between 0.316 and 0.469. The values are far less than $(1 - \frac{1}{64})(1 - 2^{-3}) = 0.861$, the upper bound for the robustness of a 6×4 S-box. This might partially explain why differential cryptanalysis of DES was so successful.

4 Constructing S-boxes (Part I) — The First Attempt

We present our method for constructing robust S-boxes in three steps. The first step which is described in this section shows how to construct S-boxes whose component functions are highly nonlinear and also satisfy the SAC. A shortcoming of these S-boxes is that they are not robust against differential cryptanalysis. This shortcoming is removed in the second step which is described in the next section. This is followed by another section describing the third step which discusses further refinement on the results.

4.1 Bent Functions Which Form a Group

In [?], bent functions which form an additive group were constructed. These functions are the starting point of our method for generating S-boxes, and hence are reviewed in the following.

A $(1,-1)$-matrix of order n will be called a Hadamard matrix if $HH^T = nI_n$, where H^T is the transpose of H [?]. A Sylvester-Hadamard matrix (or Walsh-Hadamard matrix) is a matrix of order 2^n generated in the following way:

$$H_n = \begin{bmatrix} H_{n-1} & H_{n-1} \\ H_{n-1} & -H_{n-1} \end{bmatrix}, n = 1, 2, \ldots, H_0 = 1.$$

Let G be a group under operation $\cdot$ (dot), and let $p = (p_1, \ldots, p_n)$, $q = (q_1, \ldots, q_n)$ be two vectors of length n, whose entries p_j, q_j come from G. Define the operation $\odot$ such that $p \odot q = (p_1 \cdot q_1, \ldots, p_n \cdot q_n)$, and the inverse of q such that $q^{-1} = (q_1^{-1}, \ldots, q_n^{-1})$. We say that p and q are *s-orthogonal* if $p \odot q^{-1} = (p_1 \cdot q_1^{-1}, \ldots, p_n \cdot q_n^{-1})$ contains every element in G precisely s times.

A *generalized Hadamard matrix* [?, ?] of type s for the group G is a square matrix with entries from G whose rows and columns are both s-orthogonal. A *group Hadamard matrix* [?] is a generalized Hadamard matrix whose rows and columns both form a group under the operation $\odot$. Note that in a group Hadamard matrix of type s for G, there exist a row acting the role of identity, and each of the other rows contains each element of G precisely s times. A similar observation applies to the columns of the matrix.

Now let ε be a primitive element of $GF(2^k)$, and let C be a $(2^k-1) \times (2^k-1)$ matrix whose (i,j)th entry, $0 \leqq i, j \leqq 2^k - 2$, is defined as $c_{ij} = \varepsilon^{j+i \pmod{2^k-1}}$. Denote by D the extended $2^k \times 2^k$ matrix $\begin{bmatrix} 0 & \cdots & 0 \\ \vdots & C & \\ 0 & & \end{bmatrix}$.

Note that each entry of D is a polynomial in ε, whose algebraic degree is at most $k-1$. Therefore each entry can be expressed as $a_0 \oplus a_1\varepsilon \oplus \cdots \oplus a_{k-1}\varepsilon^{k-1}$, where $a_i \in GF(2)$. Replacing ε^i by x_{i+1}, where $0 \leqq i \leqq k-1$, we obtain a multi-variable polynomial $a_0x_1 \oplus a_1x_2 \oplus \cdots \oplus a_{k-1}x_k$, which can be viewed as a linear function on V_k. Denote by E be the matrix obtained from D by applying the replacement to all its entries. In [?], the following interesting result was proved

Lemma 2 *Denote by Γ_k the additive group consisting of all linear functions on V_k. Then E is a group Hadamard matrix of type 1 for Γ_k. That is, both the rows and the columns of the matrix E form a group under the component-wise polynomial addition with the zero row and the zero column as their identify elements respectively, and each linear function on V_k appears precisely once in each nonzero row and also in each nonzero column.*

Concatenating the linear functions in the ith row of E results in a function f_i on V_{2k}:

$$f_i(y, x) = \bigoplus_{j=0}^{2^k-1} [D_j(y)e_{ij}(x)] \tag{1}$$

where $y = (y_1, \ldots, y_k)$ and $x = (x_1, \ldots, x_k)$. From [?], we know that f_1, f_2, ..., f_{2^k-1} are all distinct bent functions on V_{2k}, and that f_0, f_1, ..., f_{2^k-1} form a additive group with $f_0 = 0$ as its identify element. In the same paper it was also shown that

Theorem 2 *The following statements hold:*

*(i) let f be a nonzero linear combination of the k functions f_1, f_2, ..., f_k that are defined by (**??**), namely $f(y,x) = \bigoplus_{j=1}^{k}[c_jf_j(y,x)]$, where $y = (y_1, \ldots, y_k)$, $x = (x_1, \ldots, x_k)$ and $c_j \in GF(2)$. Then $f = f_i$ for some $1 \leqq i \leqq 2^k - 1$. Conversely, any f_i, $1 \leqq i \leqq 2^k - 1$, can be expressed as a nonzero linear combination of f_1, f_2, ..., f_k;*

(ii) for any $1 \leqq j \leqq 2^k - 1$*, write*

$$\begin{aligned} e_{1j} &= a_{11}x_1 \oplus \cdots \oplus a_{1k}x_k, \\ e_{2j} &= a_{21}x_1 \oplus \cdots \oplus a_{2k}x_k, \\ &\vdots \\ e_{kj} &= a_{k1}x_1 \oplus \cdots \oplus a_{kk}x_k, \end{aligned}$$

then $A = (a_{ij})$*, whose entries come from* $GF(2)$*, is a nondegenerate matrix of order* k*.*

4.2 S-boxes Satisfying the SAC

We have shown that concatenating the functions in a row of E, except the first row, results in a bent function. Note that a bent function is not balanced. In the following we consider the concatenation of an incomplete or partial row in E.

Let n be an integer with $k < n < 2k$. We select 2^{n-k} distinct columns from the $2^k - 1$ nonzero columns of E. Denote by $H = (h_{ij})$ the $2^k \times 2^{n-k}$ matrix consisting of the 2^{n-k} selected columns, where $0 \leqq i \leqq 2^k - 1$ and $0 \leqq j \leqq 2^{n-k} - 1$.

Let g_i be the function obtained by concatenating the ith row of $H = (h_{ij})$, namely

$$g_i(y, x) = \bigoplus_{j=0}^{2^{n-k}-1} [D_j(y) h_{ij}(x)] \tag{2}$$

where $0 \leqq i \leqq 2^k - 1$, $y = (y_1, \ldots, y_{n-k})$ and $x = (x_1, \ldots, x_k)$.

Lemma 3 *Let* g *be a nonzero linear combination of* $g_1, \ldots, g_k$ *that are defined in (**??**), namely* $g(y, x) = \bigoplus_{i=1}^k [c_i g_i(y, x)]$*, where* $c_i \in GF(2)$*. Then*

(i) g *is balanced,*

(ii) the nonlinearity of g *satisfies* $N_g \geqq 2^{n-1} - 2^{k-1}$*,*

(iii) $g(z) \oplus g(z \oplus \gamma)$ *is balanced for any* $\gamma = (\beta, \alpha)$ *with* $W(\beta) \neq 0$*, where* $\beta \in V_{n-k}$ *and* $\alpha \in V_k$*,*

(iv) the maximum algebraic degree of g *is* $n - k + 1$*,*

(v) $G = (g_1, \ldots, g_k)$ *is a regular mapping.*

Proof. (i) of Theorem **??** implies that g, a nonzero linear combination of $g_1, \ldots, g_k$, matches g_i for some $1 \leqq i \leqq 2^k - 1$. Note that $g_1, \ldots, g_{2^k-1}$ are all concatenations of nonzero linear functions. By Lemma **??**, (i), (ii) and (iii) hold.

Now we show that (iv) is true. First we note that since the rows of the matrix E from which H is obtained form a group (see Lemma **??**), there is a $1 \leqq t \leqq 2^k - 1$ such that g can be expressed as the concatenation of the functions in a row of H indexed by t, namely, $g(y, x) = \bigoplus_{j=0}^{2^{n-k}-1} [D_j(y) h_{tj}(x)]$. Consider the function g_1 which is defined by $g_1(y, x) = \bigoplus_{j=0}^{2^{n-k}-1} [D_j(y) h_{1j}(x)]$. When the following condition is satisfied

$$\bigoplus_{j=0}^{2^{n-k}-1} h_{1j}(x) \neq 0 \tag{3}$$

the term $y_1 \cdots y_{n-k} \bigoplus_{j=0}^{2^{n-k}-1} h_{1j}(x)$ will not be canceled in the final expression of g_1, and hence g_1 achieves the maximum algebraic degree $n - k + 1$.

Now suppose that the the condition (**??**) is satisfied. Recall that the columns of E form a group as well (see Lemma **??**), and that each linear function in V_k appears precisely once in each nonzero column. These properties of E, together with the fact that $\bigoplus_{j=0}^{2^{n-k}-1} h_{0j}(x) = 0$, implies that when the condition (**??**) is satisfied, we have $\bigoplus_{j=0}^{2^{n-k}-1} h_{ij}(x) \neq 0$ for all $2 \leqq i \leqq 2^k - 1$. In other words, g_2, …, g_{2^k-1} all achieve the maximum algebraic degree $n-k+1$.

To ensure that the condition (**??**) is satisfied, first we select $2^{n-k}-1$ columns from the nonzero columns of E. Next we select a column from the nonzero columns of E that have not been touched so far, and check $\bigoplus_{j=0}^{2^{n-k}-1} h_{1j}(x)$. The selection and checking step continues until the condition (**??**) is satisfied. Since each linear function on V_k appears precisely once in a nonzero row of E, after the first $2^{n-k}-1$ columns are selected, there is at most one column in the untouched columns of E such that $\bigoplus_{j=0}^{2^{n-k}-1} h_{1j}(x) = 0$. Therefore the maximum algebraic degree is always achievable. This proves (iv).

(v) follows from (i) and Theorem **??**. □

A problem with $G = (g_1, \ldots, g_k)$ is that it does not satisfy the SAC. Using the following Lemma **??** which was first proved in [**?**], the problem can be circumvented by a suitable nondegenerate linear transformation on the coordinates of the mapping. Note that the balancedness, the nonlinearity and the algebraic degree of a function are not affected by a nondegenerate linear transformation on coordinates [**?**].

Lemma 4 *Let f_1, f_2, …, f_m be functions on V_n. Suppose that A is an $n \times n$ nondegenerate matrix on $GF(2)$ with the property that for each row α_i of A, $1 \leqq i \leqq n$, and for each function f_j, $1 \leqq j \leqq m$, $f_j(x) \oplus f_j(x \oplus \alpha_i)$ is balanced. Then $f_1(xA)$, $f_2(xA)$, …, $f_m(xA)$ all satisfy the SAC.*

Let A be a $n \times n$ nondegenerate matrix with nonzero values in the first $n-k$ entries of its rows. A simple example follows:

$$A = \begin{bmatrix} I_{n-k} & 0_{(n-k)\times k} \\ J_{k\times(n-k)} & I_k \end{bmatrix} \tag{4}$$

where I denotes the identity matrix, 0 the zero matrix, and J the matrix whose entries are all ones. Another example that introduces more inter-coordinate dependencies is as follows:

$$\begin{aligned} A &= \begin{bmatrix} I_{n-k} & 0_{(n-k)\times k} \\ B_{k\times(n-k)} & I_k \end{bmatrix} \begin{bmatrix} I_{n-k} & C_{(n-k)\times k} \\ 0_{k\times(n-k)} & I_k \end{bmatrix} \\ &= \begin{bmatrix} I_{n-k} & C_{(n-k)\times k} \\ B_{k\times(n-k)} & B_{k\times(n-k)}C_{(n-k)\times k} \oplus I_k \end{bmatrix} \end{aligned} \tag{5}$$

where B is a matrix not containing zero rows and C is an arbitrary matrix, both on $GF(2)$.

Denote by Π the mapping after applying the linear transformation A to the coordinates of $G = (g_1, \ldots, g_k)$, namely,

$$\begin{aligned} \Pi(x) &= (\pi_1(x), \ldots, \pi_k(x)) \\ &= (g_1(xA), \ldots, g_k(xA)). \end{aligned} \tag{6}$$

¿From (iii) of Lemma **??** and Lemma **??** it follows:

Theorem 3 *The nonzero linear combinations of the component functions of $\Pi = (\pi_1, \ldots, \pi_k)$ which is defined by (**??**) are all nonlinearly balanced and fulfill the SAC. Their nonlinearity is at least $2^{n-1} - 2^{k-1}$, and their maximum algebraic degree is $n-k+1$.*

Although $\Pi = (\pi_1, \ldots, \pi_k)$ satisfies some of the main requirements for an S-box with regard to nonlinearity, SAC and balancedness, the majority of the rows in its difference distribution table contain no zeros. By a similar argument to that for Lemma **??** in Subsection **??**, it can be shown that the difference distribution table has the following profile:

1. in $2^k - 1$ cases, 2^{n-k} out of the 2^k entries in a row contain a value 2^k, while the other $2^k - 2^{n-k}$ entries contain a value zero;
2. in the other $2^n - 2^k$ cases (not counting the first row), all the entries in a row contain a value 2^{n-k}.

Hence the robustness of Π against differential cryptanalysis is only $\frac{2^k}{2^n}(1 - \frac{1}{2^{n-k}}) < \frac{1}{2^{n-k}}$.

This shortcoming will be removed in the following section. Before going into the detailed description of how it is removed, we note that Lemma **??**, together with the discussions about the SAC fulfilling properties and the difference distribution tables of $G = (g_1, \ldots, g_k)$ and $\Pi = (\pi_1, \ldots, \pi_k)$, also holds in the case when g_i is defined in the following more general form:

$$g_i(y, x) = \bigoplus_{j=0}^{2^{n-k}-1} [D_j(y)h_{ij}(x)] \oplus r_i(y) \tag{7}$$

where r_i is an arbitrary function on V_{n-k}.

5 Constructing S-boxes (Part II) — Improvement

This section discusses how to strengthen S-boxes constructed in (**??**) so that they are much more robust against differential cryptanalysis. We start with a permutation on V_3 which has many desirable properties. Next we combine an $s \times k$ S-box $G = (g_1, \ldots, g_k)$ with the permutation on V_3 to obtain an $n \times (k + 3)$ S-box, where g_i is constructed by (**??**). Then we show that the new S-box is very robust against differential cryptanalysis.

5.1 A Permutation on V_3

Recall that each primitive polynomial defines an m-sequence (see [**?**]). Consider $(1, 0, 0, 1, 0, 1, 1)$, an m-sequence of length 7 generated by the primitive polynomial $1 \oplus x \oplus x^3$ with $(1,0,0)$ as its starting vector. Shifting cyclically the m-sequence to the left gives two new m-sequences $(0, 0, 1, 0, 1, 1, 1)$ and $(0, 1, 0, 1, 1, 1, 0)$. The three m-sequences can be viewed as the truth tables of functions on V_3 after appending a zero at the left end of each of the sequences. The functions corresponding to the three truth tables are

$$\left.\begin{array}{l} m_1(w) = y_1 \oplus y_3 \oplus y_2y_3 \\ m_2(w) = y_1 \oplus y_2 \oplus y_1y_2 \oplus y_2y_3 \\ m_3(w) = y_1y_2 \oplus y_2y_3 \oplus y_1y_3 \end{array}\right\} \tag{8}$$

where $w = (y_1, y_2, y_3)$. The three functions define a mapping on V_3:

$$M_3 = (m_1, m_2, m_3).$$

It is not hard to verify that M_3 is a permutation on V_3. In addition, by using properties of m-sequences or by straightforward verification, one can see that M_3 has the two properties described below.

1. Let $m(w) = c_1m_1(w) \oplus c_2m_2(w) \oplus c_3m_3(w)$ be a nonzero linear combination of m_1, m_2, m_3, where $c_1, c_2, c_3 \in GF(2)$. Then m is a nonlinearly balanced function. The nonlinearity of m is 2. Note that 2 is the maximum nonlinearity of a function on V_3.

2. Let α be a nonzero vector in V_3. When w runs through V_3, $M_3(w) \oplus M_3(w \oplus \alpha)$ runs through 4 vectors in V_3 twice each, and never through the other 4 vectors.

5.2 Robust S-boxes

Now we combine the permutation on V_3 with functions constructed by (**??**) to obtain an S-box much more robust against differential cryptanalysis. Let n and s be integers with $n \geqq s > (\lfloor n/2 \rfloor + 3)$, and let $k = s - 3$. Also let $r_1 = r_2 = \cdots = r_k = 0$, $r_{k+1} = m_1$, $r_{k+2} = m_2$ and $r_{k+3} = m_3$. Define $s = k + 3$ functions on V_n in the following way:

$$f_i(y_1, \ldots, y_{n-k}, x_1, \ldots, x_k) \;=\; g_i(y_1, \ldots, y_{n-k}, x_1, \ldots, x_k) \oplus r_i(y_1, y_2, y_3) \tag{9}$$

where g_i is defined by (**??**) and $i = 1, \ldots, k+3$.

The following lemma will be used in discussing properties of the functions constructed by (**??**).

Lemma 5 *Let $g(x_1, \ldots, x_s)$ be a function on V_s. Extend g into a function f on V_{s+t} by adding t dummy-coordinates, namely, $f(x_1, \ldots, x_s, y_1, \ldots, y_t) = g(x_1, \ldots, x_s)$. Then*

(i) if g is balanced then f is balanced,

(ii) $N_f \geqq 2^t N_g$, where N_f and N_g denote the nonlinearities of f and g respectively.

Proof. Note that

$$\begin{aligned} f(x_1, \ldots, x_s, y_1, \ldots, y_t) &= f(y_1, \ldots, y_t, x_1, \ldots, x_s) \\ &= \bigoplus_{i=0}^{2^t-1} [D_i(y_1, \ldots, y_t) g(x_1, \ldots, x_s)]. \end{aligned}$$

Thus f is obtained by concatenating g for 2^t times. This proves (i).

Let ξ be the sequence of g. Then $\eta = (\xi, \ldots, \xi)$ is the sequence of f. Let L be an arbitrary affine sequence of length 2^{t+s}. By Lemma 10 of [**?**], L is a row of $H_{t+s} = H_t \otimes H_s$, where H_n is the Sylvester-Hadamard matrix of order 2^n and $\otimes$ denotes the Kronecker product. Then L can be expressed as $L = \ell_t \otimes \ell_s$ where ℓ_t is an affine sequence of length 2^t and ℓ_s is an affine sequence of length 2^s. Let $\ell_t = (a_1, \ldots, a_{2^t})$. Then $L = (a_1 \ell_s, \ldots, a_{2^t} \ell_s)$ and

$$|\langle \eta, L \rangle| \leqq \sum_{j=1}^{2^t} |\langle \xi, \ell_s \rangle| = 2^t |\langle \xi, \ell_s \rangle|.$$

Since the nonlinearity of g is N_g, by Lemma 12 of [**?**], we have $|\langle \xi, \ell_s \rangle| \leqq 2^s - 2N_g$. Hence

$$|\langle \eta, L \rangle| \leqq 2^t (2^s - 2N_g)$$

As L is arbitrary, again by Lemma 12 of [**?**], we have $N_f \geqq 2^t N_g$. □

Now we have the following result:

Lemma 6 *Let $y = (y_1, \ldots, y_{n-k})$, $x = (x_1, \ldots, x_k)$, $w = (y_1, y_2, y_3)$ and $z = (y, x)$. Let $f(y, x) = \bigoplus_{j=1}^{k+3} [c_j f_j(y, x)]$ be a nonzero linear combination of $f_1, \ldots, f_{k+3}$ that are defined in (**??**). Then*

(i) f is balanced,

(ii) when $f(z) \neq \bigoplus_{j=k+1}^{k+3}[c_j r_j(w)]$, the nonlinearity of f is at least $2^{n-1} - 2^{k-1}$, and the maximum algebraic degree of f is $n - k + 1$. Otherwise, the nonlinearity of f is at least 2^{n-2}, and the algebraic degree of f is 2,

(iii) when $f(z) \neq \bigoplus_{j=k+1}^{k+3}[c_j r_j(w)]$, $f(z) \oplus f(z \oplus \gamma)$ is balanced for any $\gamma = (\beta, \alpha)$ with $W(\beta) \neq 0$, where $\beta \in V_{n-k}$ and $\alpha \in V_k$,

(iv) $(f_1, \dots, f_{k+3})$ is a regular mapping.

Proof. Note that f can be written as

$$f(z) = \bigoplus_{j=1}^{k+3}[c_j g_j(z)] \oplus \bigoplus_{j=k+1}^{k+3}[c_j r_j(w)].$$

It is easy to see that $f(z) \neq 0$, and there are only two cases to be considered

Case 1 — $f(z) = \bigoplus_{j=1}^{k+3}[c_j g_j(z)] \oplus \bigoplus_{j=k+1}^{k+3}[c_j r_j(w)]$ with $\bigoplus_{j=1}^{k+3}[c_j g_j(z)] \neq 0$.

Case 2 — $f(z) = \bigoplus_{j=k+1}^{k+3}[c_j r_j(w)] = c_{k+1}m_1(w) \oplus c_{k+2}m_2(w) \oplus c_{k+3}m_3(w)$.

¿From Lemma **??** and the discussion on the construction (**??**) at the end of Subsection **??**, it follows that f is balanced in Case 1. And due to the first property of the permutation on V_3 (see section **??**) and (i) of Lemma **??**, f is balanced in Case 2. This proves (i).

The first half of (ii), which corresponds to Case 1, follows from Lemma **??**, as well as the discussion on the construction (**??**). In Case 2, the algebraic degree of f is clearly 2. By (ii) of Lemma **??**, the nonlinearity of f is at least $2^{n-3} \cdot 2 = 2^{n-2}$.

Finally (iii) follows from Lemma **??**, while (iv) follows from (i) and Theorem **??**. □

Let A be a $n \times n$ nondegenerate matrix, whose ith row γ_i, $i = 1, \dots, k+3$, can be written as $\gamma_i = (\beta_i, \alpha_i)$, where $\beta_i \in V_{n-k}$, $W(\beta_i) \neq 0$ and $\alpha_i \in V_k$. Then by Lemma **??**, f_1, f_2, ..., f_{k+3} defined by (**??**) can all be transformed into SAC-fulfilling functions:

$$\begin{aligned} \Psi(z) &= (\psi_1(z), \dots, \psi_{k+3}(z)) \\ &= (f_1(zA), \dots, f_{k+3}(zA)). \end{aligned} \tag{10}$$

Thus we have the following theorem:

Theorem 4 *Let Ψ, ψ_1, ..., ψ_{k+3} and A be the same as in (**??**). Let $\psi(z) = \bigoplus_{j=1}^{k+3}[c_j \psi_j(z)]$ be a nonzero linear combination of $\psi_1, \dots, \psi_{k+3}$, where $z = (z_1, \dots, z_{k+3})$ and $c_j \in GF(2)$. Then*

(i) ψ is balanced,

(ii) in $2^{k+3} - 8$ cases, which include the cases when $\psi = \psi_j$, $j = 1, \dots, k+3$, the nonlinearity of ψ is at least $2^{n-1} - 2^{k-1}$, and the maximum algebraic degree of ψ is $n - k + 1$. In the other 7 cases, the nonlinearity of ψ is at least 2^{n-2}, and the algebraic degree of ψ is 2,

(iii) ψ satisfies the SAC if $\psi(z) \neq \bigoplus_{j=k+1}^{k+3}[c_j r_j(zA)]$,

(iv) $\Psi = (\psi_1, \dots, \psi_{k+3})$ is a regular mapping.

In the following we prove that the robustness of $\Psi = (\psi_1, \dots, \psi_{k+3})$ against differential cryptanalysis is $(\frac{7}{8} + 2^{-n+k-3} - 2^{-2n+2k})$. When $n = k+3$, Ψ is a permutation on V_n, and its robustness against differential cryptanalysis is $\frac{7}{8}$.

5.3 Profile of the Difference Distribution Table

Now we discuss the difference distribution table of $\Psi = (\psi_1, \ldots, \psi_{k+3})$ constructed by (**??**). The following results will simplify our discussions.

Let g_j be a function on V_n, $j = 1, \ldots, s$, and let $G = (g_1, \ldots, g_s)$. Also let A be a nondegenerate matrix of order s over $GF(2)$. Consider $F(x) = (g_1(x), \ldots, g_s(x))A$. Note that A is applied to the output of G. For any $\tau \in V_s$, $G(x) = (g_1(x), \ldots, g_s(x)) = \tau$ if and only if $F(x) = (g_1(x), \ldots, g_s(x))A = \tau A$. Therefore, while x runs through V_n, $G(x)$ runs through τ exactly the same number of times as that $F(x)$ runs through τA.

Now let B be a nondegenerate matrix of order n over $GF(2)$, and let $F(x) = (g_1(xB), \ldots, g_s(xB))$. Since $G(x) = F(xB^{-1})$, $G(x) = \tau$ if and only if $F(xB^{-1}) = \tau$, where $\tau \in V_s$. This implies that, while x runs through V_n, $G(x)$ and $F(x)$ run through τ the same number of times.

In summary, the profile of the difference distribution table of an S-box is not altered by a nondegenerate linear transformation on outputs or a nondegenerate linear transformation on inputs. The observation is used in analyzing the difference distribution table of $\Psi = (\psi_1, \ldots, \psi_{k+3})$.

Lemma 7 *Let $\Psi = (\psi_1, \ldots, \psi_{k+3})$ be an S-box constructed in (**??**). Also let $z = (z_1, \ldots, z_n)$ and $\gamma = (\beta, \alpha)$ be a nonzero vector in V_n. Then*

(i) for $2^k - 1$ cases of γ, $\Psi(z) \oplus \Psi(z \oplus \gamma)$ runs through 2^{n-k} vectors in V_{k+3} 2^k times each, but not through the other $2^{k+3} - 2^{n-k}$ vectors,

(ii) for other $2^{n-3} - 2^k$ cases of γ, $\Psi(z) \oplus \Psi(z \oplus \gamma)$ runs through 2^k vectors in V_{k+3} 2^{n-k} times each, but not through the other $2^{k+3} - 2^k$ vectors,

(iii) for the remaining $2^n - 2^{n-3}$ cases of γ, $\Psi(z) \oplus \Psi(z \oplus \gamma)$ runs through 2^{k+2} vectors in V_{k+3} 2^{n-k-2} times each, but not through the other 2^{k+2} vectors,

(iv) the first column of the difference distribution table of Ψ contains a value 2^{n-k} in $(2^{n-k-3} - 1)2^k$ entries, and a value zero in the other entries (not counting the first entry).

Proof. Let $F = (f_1, \ldots, f_{k+3})$, where f_i is constructed by (**??**). Then $\Psi(z) = F(zA)$, and hence $\Psi(z) \oplus \Psi(z \oplus \gamma) = F(zA) \oplus F(zA \oplus \gamma A)$. Thus the problem of discussing the difference distribution table of Ψ is reduced to that of F.

Let $z = (y, x)$, $y = (y_1, \ldots, y_{n-k})$, $x = (x_1, \ldots, x_k)$ and $w = (y_1, y_2, y_3)$. Write $\gamma = (\beta, \alpha)$, where $\beta \in V_{n-k}$ and $\alpha \in V_k$, and $\beta = (\mu, \nu)$ where $\mu \in V_3$ and $\nu \in V_{n-k-3}$. By (**??**) we have

$$\begin{aligned} F(z) &= (g_1(z), \ldots, g_k(z), g_{k+1}(z) \oplus m_1(w), \\ &\quad g_{k+2}(z) \oplus m_2(w), g_{k+3}(z) \oplus m_3(w)). \end{aligned}$$

Hence

$$\begin{aligned} F(z) \oplus F(z \oplus \gamma) &= (g_1(z) \oplus g_1(z \oplus \gamma), \ldots, g_k(z) \oplus g_k(z \oplus \gamma), \\ &\quad g_{k+1}(z) \oplus g_{k+1}(z \oplus \gamma) \oplus m_1(w) \oplus m_1(w \oplus \mu), \\ &\quad g_{k+2}(z) \oplus g_{k+2}(z \oplus \gamma) \oplus m_2(w) \oplus m_2(w \oplus \mu), \\ &\quad g_{k+3}(z) \oplus g_{k+3}(z \oplus \gamma) \oplus m_3(w) \oplus m_3(w \oplus \mu)). \end{aligned}$$

As g_{k+1}, g_{k+2} and g_{k+3} are nonzero linear combinations of g_1, $\ldots$, g_k, $F(z) \oplus F(z \oplus \gamma)$ can be written as $F(z) \oplus F(z \oplus \gamma) = (Q(z) \oplus Q(z \oplus \gamma))B$ for some nondegenerate matrix B, where

$$Q(z) = (g_1(z), \ldots, g_k(z), m_1(w), m_2(w), m_3(w)).$$

Thus the problem is further simplified, and we only have to discuss how $Q(z) \oplus Q(z \oplus \gamma)$ runs through the vectors in V_{k+3}.

¿From (**??**), we have

$$\begin{aligned} Q(z) \oplus Q(z \oplus \gamma) \;\; = \;\; & (\bigoplus_{\sigma \in V_{n-k}} [D_\sigma(y)(h_{1,\sigma}(x)] \oplus h_{1,\sigma\oplus\beta}(x \oplus \alpha)), \ldots, \\ & \bigoplus_{\sigma \in V_{n-k}} [D_\sigma(y)(h_{k,\sigma}(x)] \oplus h_{k,\sigma\oplus\beta}(x \oplus \alpha)), \\ & m_1(w) \oplus m_1(w \oplus \mu), m_2(w) \oplus m_2(w \oplus \mu), \\ & m_3(w) \oplus m_3(w \oplus \mu)). \end{aligned}$$

Note that we have switched from integers to vectors in describing indexes. We distinguish the following two cases: $W(\beta) = 0$ and $W(\beta) \neq 0$.

Case 1: $W(\beta) = 0$ and hence $W(\alpha) \neq 0$ and $W(\mu) = 0$. In this case we have

$$\begin{aligned} Q(z) \oplus Q(z \oplus \gamma) \;\; = \;\; & (\bigoplus_{\sigma \in V_{n-k}} [D_\sigma(y) h_{1,\sigma}(\alpha)], \ldots, \\ & \bigoplus_{\sigma \in V_{n-k}} [D_\sigma(y) h_{k,\sigma}(\alpha)], 0, 0, 0) \end{aligned}$$

where $h_{i,\sigma}(\alpha) = h_{i,\sigma}(x) \oplus h_{i,\sigma}(x \oplus \alpha)$ (Note that $h_{i,\sigma}(x)$ is a linear function).

As $D_\delta(y) = 1$ if and only if $y = \delta$, for any fixed $\delta \in V_{n-k}$, we have

$$(Q(z) \oplus Q(z \oplus \gamma))|_{y=\delta} = (h_{1,\delta}(\alpha), \ldots, h_{k,\delta}(\alpha), 0, 0, 0).$$

Now let $y = \delta$ run through V_{n-k}. Then $(Q(z) \oplus Q(z \oplus \gamma))|_{y=\delta}$ will run through 2^{m-k} vectors in V_{k+1}, 2^k times each. This follows from the fact that, if $\delta \neq \delta'$, then

$$(Q(z) \oplus Q(z \oplus \gamma))|_{y=\delta} \neq (Q(z) \oplus Q(z \oplus \gamma))|_{y=\delta'}.$$

To show that the fact is true we only have to show

$$(h_{1,\delta}(\alpha), \ldots, h_{k,\delta}(\alpha)) \neq (h_{1,\delta'}(\alpha), \ldots, h_{k,\delta'}(\alpha))$$

or equivalently

$$(h_{1,\delta}(\alpha) \oplus h_{1,\delta}(\alpha), \ldots, h_{k,\delta}(\alpha) \oplus h_{k,\delta}(\alpha)) \neq (0, \ldots, 0).$$

Since the rows of the matrix E introduced in Subsection **??** form a group, there exists a $\delta'' \neq (0, \ldots, 0)$ such that

$$(h_{1,\delta}(\alpha) \oplus h_{1,\delta}(\alpha), \ldots, h_{k,\delta}(\alpha) \oplus h_{k,\delta}(\alpha)) \;\; = \;\; (h_{1,\delta''}(\alpha), \ldots, h_{k,\delta''}(\alpha)).$$

As $W(\alpha) \neq 0$, it becomes clear that

$$(h_{1,\delta''}(\alpha), \ldots, h_{k,\delta''}(\alpha)) \neq (0, \ldots, 0).$$

This shows that the fact is indeed true.

To summarize Case 1, while z runs through V_n, $Q(z) \oplus Q(z \oplus \gamma)$ runs through 2^{n-k} vectors in V_{k+3}, 2^k times each, and not through the other $2^{k+1} - 2^{n-k}$ vectors.

Case 2: $W(\beta) \neq 0$. Then

$$\begin{aligned}(Q(z) \oplus Q(z \oplus \gamma))|_{y=\delta} &= (h_{1,\delta}(x) \oplus h_{1,\delta\oplus\beta}(x \oplus \alpha), \ldots, h_{k,\delta}(x) \oplus h_{k,\delta\oplus\beta}(x \oplus \alpha), \\ &\quad m_1(\rho) \oplus m_1(\rho \oplus \mu), m_2(\rho) \oplus m_2(\rho \oplus \mu), \\ &\quad m_3(\rho) \oplus m_3(\rho \oplus \mu))\end{aligned}$$

where $\delta = (\rho, \varrho)$, $\rho \in V_3$, $\varrho \in V_{n-k-3}$. Note that since h_{ij} is a linear function, we have $h_{1,\delta\oplus\beta}(x \oplus \alpha) = h_{1,\delta\oplus\beta}(x) \oplus h_{1,\delta\oplus\beta}(\alpha)$

Again as the columns of E defined in Subsection **??** form a group, there is a $\delta' \neq (0, \ldots, 0)$ such that

$$\begin{aligned}(Q(z) \oplus Q(z \oplus \gamma))|_{y=\delta} &= (h_{1,\delta'}(x) \oplus d_1, \ldots, h_{k,\delta'}(x) \oplus d_k, \\ &\quad m_1(\rho) \oplus m_1(\rho \oplus \mu), m_2(\rho) \oplus m_2(\rho \oplus \mu), \\ &\quad m_3(\rho) \oplus m_3(\rho \oplus \mu))\end{aligned}$$

where $d_i = h_{i,\delta\oplus\beta}(\alpha)$, $i = 1, \ldots, k$.

Recall that $\beta = (\mu, \nu)$ where $\mu \in V_3$ and $\nu \in V_{n-k-3}$. Two cases should be considered: $W(\mu) = 0$ and $W(\mu) \neq 0$.

Case 2.1: $W(\beta) \neq 0$ and $W(\mu) = 0$. We have

$$(Q(z) \oplus Q(z \oplus \gamma))|_{y=\delta} = (h_{1,\delta'}(x) \oplus d_1, \ldots, h_{k,\delta'}(x) \oplus d_k, 0, 0, 0).$$

By (ii) of Theorem **??**, $(h_{1,\delta'}(x) \oplus d_1, \ldots, h_{k,\delta'}(x) \oplus d_k)$ forms a permutation on V_k when δ, and hence δ', is fixed. Thus for any $\delta \in V_{n-k}$, $(h_{1,\delta'}(x) \oplus d_1, \ldots, h_{k,\delta'}(x) \oplus d_k)$ runs through each vector in V_k once while x runs through V_k. This is equivalent to say that $(Q(z) \oplus Q(z \oplus \gamma))|_{y=\delta}$ runs through each $(c_1, \ldots, c_k, 0, 0, 0) \in V_n$ precisely once. Consequently, when $y = \delta$ runs through all the 2^{n-k} vectors in V_{n-k}, $(Q(z) \oplus Q(z \oplus \gamma))|_{y=\delta}$ runs through each $(c_1, \ldots, c_k, 0, 0, 0)$ 2^{n-k} times, but never through the other vectors in V_n.

Case 2.2: $W(\beta) \neq 0$ and $W(\mu) \neq 0$. Recall that for any μ with $W(\mu) \neq 0$, while ρ runs through V_3, $(m_1(\rho) \oplus m_1(\rho \oplus \mu), m_2(\rho) \oplus m_2(\rho \oplus \mu), m_3(\rho) \oplus m_3(\rho \oplus \mu))$ runs through 4 vectors in V_3 twice each, but not through the other 4 vectors. Since $\delta = (\rho, \varrho)$, ρ runs through each vector in V_3 2^{n-k-3} times while $y = \delta$ runs through V_{n-k}. Taking into account the fact that $(h_{1,\delta'}(x) \oplus d_1, \ldots, h_{k,\delta'}(x) \oplus d_k)$ forms a permutation on V_k for any fixed $\delta \in V_{n-k}$, we can see that in the case when $W(\mu) \neq 0$, $Q(z) \oplus Q(z \oplus \gamma)$ runs through $4 \cdot 2^k = 2^{k+2}$ vectors in V_{k+3}, $2 \cdot 2^{n-k-3} = 2^{n-k-2}$ times each, but never through the other 2^{k+2} vectors in V_{k+3}.

Note that γ can take $2^k - 1$ different nonzero vectors in V_n for Case 1, $2^{n-3} - 2^k$ in Case 2.1, and $2^n - 2^{n-3}$ in Case 2.2, and that $Q(z) \oplus Q(z \oplus \gamma)$ and $F(z) \oplus F(z \oplus \gamma)$ are related by $F(z) \oplus F(z \oplus \gamma) = (Q(z) \oplus Q(z \oplus \gamma))B$, while $F(z)$ and $\Psi(z)$ are related by $\Psi(z) = F(zA)$. This proves the first three parts of the theorem.

Finally we consider the first column of the difference distribution table. Recall that the first column differs from the rest of the table in the sense that it indicates the smoothness of the S-box and that it is of particular importance to differential cryptanalysis. When $s = k + 3 = n$, the S-box is a permutation on V_n, and the first column in its difference distribution table is $(2^n, 0, \ldots, 0)^T$. To examine the case when $n > s$, we consider the solutions of the equation

$$\Psi(z) \oplus \Psi(z \oplus \gamma) = (0, \ldots 0, 0, 0, 0), \tag{11}$$

where $\gamma = (\beta, \alpha) \neq (0, \ldots, 0)$, $\beta \in V_{n-k}$ and $\alpha \in V_k$.

Similarly it can be discussed in the two cases: Case 1 where $W(\beta) = 0$ and Case 2 where $W(\beta) \neq 0$. The latter can be further divided into Case 2.1 where $W(\beta) \neq 0$ and $W(\mu) = 0$, and Case 2.2 where $W(\beta) \neq 0$ and $W(\mu) \neq 0$. It is not hard to verify that the equation (**??**) has 2^{n-k} solutions for z in Case 2.1, but no solutions in Case 1 and Case 2.2. The number of rows corresponding to Case 2.1 is $(2^{n-k-3}-1)2^k$. This completes the proof. □

The difference distribution table of the S-box has the following profile:

1. the largest number in the $2^k - 1$ rows corresponding to Case 1 is 2^k, while it is 2^{n-k} for the $2^n - 2^k$ rows corresponding to Case 2. When n is large, the number of rows for Case 2 is significantly larger than that for Case 1;

2. the first column contains a value 2^{n-k} in $(2^{n-k-3}-1)2^k$ entries, and a value zero in the other entries (not counting the first entry);

3. each row contains zero entries, and the fraction of nonzero entries in the table is between $0.44(= 0.5 - 2^{-4})$ and 0.5.

As a consequence, the robustness ε of $\Psi = (\psi_1, \ldots, \psi_{k+3})$ against differential cryptanalysis is

$$\begin{aligned} \varepsilon &= [1 - (2^{n-k-3}-1)2^k/2^n](1 - 2^{-n+k}) \\ &= \frac{7}{8} + 2^{-n+k-3} - 2^{-2n+2k} \\ &\geqq \frac{7}{8}. \end{aligned}$$

Thus we have proved:

Theorem 5 $\Psi = (\psi_1, \ldots, \psi_{k+3})$ *constructed in (??) is* $(\frac{7}{8} + 2^{-n+k-3} - 2^{-2n+2k})$*-robust against differential cryptanalysis.*

As their robustness against differential cryptanalysis is bounded from below by $\frac{7}{8}$, we expect S-boxes constructed by (??) are good enough in most practical applications. Nevertheless, we will show in the following section how to construct even more robust S-boxes. These S-boxes can meet even more stringent requirements imposed by certain applications.

6 Constructing S-boxes (Part III) — Refinement

We have shown that S-boxes constructed by (??) are at least $\frac{7}{8}$-robust against differential cryptanalysis, and that they are also very promising in terms of their nonlinearity, algebraic degrees and strict avalanche characteristics. Recall that (??) is obtained from (??) by applying a suitable nondegenerate linear transformation on coordinates, while (??) is the result of combining an S-box defined in (??) with a permutation M_3 on V_3 whose component functions are defined by (??). We have used the two properties of M_3 (see Subsection ??) in proving that combining (??) with (??) gives much better S-boxes. This approach can be generalized to further improve the robustness of an S-box.

Let $t \geqq 3$ and $M_t = (m_1, \ldots, m_t)$ a permutation on V_t that has the following properties:

1. any nonzero linear combination m of $m_1, \ldots, m_t$ is a nonlinearly balanced function;

2. for any nonzero vector $\alpha \in V_t$, when w runs through V_t, $M_t(w) \oplus M_t(w \oplus \alpha)$ runs through half of the vectors in V_t twice each, but never through the other half vectors.

For odd $t \geqq 3$, permutation polynomials based on the "cubing" technique [?, ?, ?, ?, ?, ?] satisfy the two requirements.

Let n, s and t be integers with $n \geqq s > (\lfloor n/2 \rfloor + t)$ and $t \geqq 3$, and let $k = s - t$. Now (**??**) can be generalized to

$$f_i(y_1, \ldots, y_{n-k}, x_1, \ldots, x_k) \;=\; g_i(y_1, \ldots, y_{n-k}, x_1, \ldots, x_k) \oplus r_i(y_1, \ldots, y_t) \qquad (12)$$

where $i = 1, \ldots, k+t$, g_i is defined by (**??**), and $r_1 = r_2 = \cdots = r_k = 0$, $r_{k+1} = m_1$, …, $r_{k+t} = m_t$.

Let f be a nonzero linear combination of the $k+t$ functions. Then when $f(z) \neq \bigoplus_{j=k+1}^{k+t}[c_j r_j(w)]$, $f(z) \oplus f(z \oplus \gamma)$ is balanced for any $\gamma = (\beta, \alpha)$, where $\beta \in V_{n-k}$, $W(\beta) \neq 0$ and $\alpha \in V_k$. Let A be a $(k+t) \times (k+t)$ nondegenerate matrix, whose ith row γ_i, $i = 1, \ldots, k+t$, can be written as $\gamma_i = (\beta_i, \alpha_i)$, where $\beta_i \in V_{n-k}$, $W(\beta_i) \geqq 1$ and $\alpha_i \in V_k$. Then (**??**) is generalized to:

$$\begin{aligned} \Psi(z) &= (\psi_1(z), \ldots, \psi_{k+t}(z)) \\ &= (f_1(zA), \ldots, f_{k+t}(zA)). \end{aligned} \qquad (13)$$

Note that all but $2^t - 1$ nonzero linear combinations of the component functions of Ψ satisfy the SAC.

Theorem **??** is generalized to:

Theorem 6 *Let n, s and t be integers with $n \geqq s > \lfloor n/2 \rfloor + t$. Let $k = s - t$. Also let Ψ, ψ_1, …, ψ_s and A be the same as in (**??**), and $\psi(z) = \bigoplus_{j=1}^{s}[c_j \psi_j(z)]$ be a nonzero linear combination of $\psi_1, \ldots, \psi_s$, where $z = (z_1, \ldots, z_n)$ and $c_j \in GF(2)$. Then*

(i) ψ is balanced,

(ii) in $2^{k+t} - 2^t$ cases, which include the cases when $\psi = \psi_j$, $j = 1, \ldots, k+t$, the nonlinearity of ψ is at least $2^{n-1} - 2^{k-1}$, and the maximum algebraic degree of ψ is $n - k + 1$. In the other $2^t - 1$ cases, the nonlinearity of ψ is at least $2^{n-t}N_{M_t}$, and the algebraic degree of ψ is at least 2, where N_{M_t} denotes the minimum among the nonlinearities of m_1, …, m_t,

(iii) ψ satisfies the SAC, except in $2^t - 1$ cases. In particular, ψ satisfies the SAC when $\psi = \psi_j$, $j = 1, \ldots, k+t$,

(iv) $\Psi = (\psi_1, \ldots, \psi_{k+t})$ is a regular mapping.

Lemma **??** can be generalized accordingly. In particular, it can be shown that the fraction of nonzero entries in the difference distribution table of $\Psi = (\psi_1, \ldots, \psi_s)$ constructed in (**??**) is between $(0.5 - 2^{-(t+1)})$ and 0.5, that the largest value in the table is 2^k, and that the first column of the table contains a value 2^{n-k} in $(2^{n-k-t} - 1)2^k$ entries, and a value zero in the other entries (not counting the first entry). Hence Theorem **??** is generalized to:

Theorem 7 *The robustness of $\Psi = (\psi_1, \ldots, \psi_s)$ constructed in (**??**) against differential cryptanalysis is $(1 - 2^{-t} + 2^{-n+s-2t} - 2^{-2(n+s-t)})$. The lower bound $1 - 2^{-t}$ is attained only when Ψ is a permutation.*

Consequently, when $t = 5$, the robustness of $\Psi = (\psi_1, \ldots, \psi_s)$ is at least 0.96875, and when $t = 7$ it is at least 0.9921875.

7 Counting Robust S-boxes

Two S-boxes $F = (f_1, \ldots, f_s)$ and $G = (g_1, \ldots, g_s)$ are said to be different if the two function sets $\{f_1, \ldots, f_s\}$ and $\{g_1, \ldots, g_s\}$ differ. We are interested in the number of different S-boxes that can be generated by our method.

Let n, s and t be integers with $n \geqq s > (\lfloor n/2 \rfloor + t)$ and $t \geqq 3$, and let $k = s - t$. The matrix H consists of 2^{n-k} columns selected from the matrix E (see Subsection **??**.) The total number of ways in which H is a selected is $\left(\begin{array}{c} 2^k - 1 \\ 2^{n-k} \end{array} \right)$. Each way gives a different matrix H. To achieve the maximum algebraic degree $n - k + 1$, we first select $2^{n-k} - 1$ columns from E and then select a column from the rest of the columns of E in such a way that the condition (**??**) is satisfied. This shows that the number of ways of achieving the maximum algebraic degree is $\left(\begin{array}{c} 2^k - 1 \\ 2^{n-k} - 1 \end{array} \right) (2^k - 2^{n-k} - 1)$.

It is easy to verify that permuting the 2^{n-k} columns of the matrix H results in a different matrix, and that discussions made above, in particular Lemma **??**, and Theorems **??** and **??**, also hold in this case. Note that there are $2^{n-k}!$ different ways to permute the columns of H.

It should be pointed out that S-boxes generated in the above two steps, selecting and permuting, contain all those which can be obtained by selecting a different primitive polynomial of algebraic degree $k - 1$. In other words, selecting a different primitive polynomial does not yield more S-boxes.

On the other hand, Theorems **??** and **??** also hold when $g_{k+1}, \ldots, g_{k+t}$, which are used to obtain $f_{k+1}, \ldots, f_{k+t}$ in the construction (**??**), are replaced by any distinct functions chosen from $g_1, \ldots, g_{2^k-1}$. There are $\left(\begin{array}{c} 2^k - 1 \\ t \end{array} \right)$ possible choices, each of which gives a different S-box.

Finally, we can obtain more S-boxes by selecting a different nondegenerate matrix in transforming $f_1, \ldots, f_{k+t}$ into SAC-fulfilling functions. These transformations, however, do not always produce different S-boxes.

In summary, the total number of different S-boxes is at least

$$2^{n-k}! \left(\begin{array}{c} 2^k - 1 \\ t \end{array} \right) \left(\begin{array}{c} 2^k - 1 \\ 2^{n-k} \end{array} \right)$$

and when the maximum algebraic degree $n - k + 1$ is required, it is at least

$$2^{n-k}! \left(\begin{array}{c} 2^k - 1 \\ t \end{array} \right) \left(\begin{array}{c} 2^k - 1 \\ 2^{n-k} - 1 \end{array} \right) (2^k - 2^{n-k} - 1).$$

8 Remarks

This section discusses the following two additional issues: immunity of the S-boxes against linear cryptanalysis and a relation between the SAC and the profile of a difference distribution table.

8.1 Immunity to Linear Cryptanalysis

Linear cryptanalysis is yet another powerful cryptanalytic attack discovered very recently by Matsui [**?**]. This cryptanalytic method exploits the low nonlinearity of S-boxes employed by a block cipher, and it has been successfully applied in attacking FEAL and DES.

Given an $n \times s$ S-box $(f_1, \ldots, f_s)$, where each f_i is a function on V_n, a linear cryptanalyst calculates the number of times that

$$f(x_1, \ldots, x_n) = \bigoplus_{i=1}^{n} (a_i x_i) \oplus \bigoplus_{j=1}^{s} [b_j f_j(x_1, \ldots, x_n)] \tag{14}$$

assumes the value zero, for all nonzero vectors $(a_1, \ldots, a_n) \in V_n$ and nonzero vectors $(b_1, \ldots, b_s) \in V_s$. The cryptanalyst then examines how far the numbers deviate from 2^{n-1}. Those which deviate the farthest are particularly useful for linear cryptanalysis.

In the original exposition of linear cryptanalysis [?], only counting the number of times that f assumes the value zero was described. This approach, however, captures only half of the information that is useful for linear cryptanalysis. The other half is obtained by counting the number of times that f assumes the value one. The two halves are complementary in the sense that one can be derived from the other. We can treat these two halves in a unified way by calculating the number of times that

$$g(x_1, \ldots, x_n) = [a_0 \oplus \bigoplus_{i=1}^{n}(a_i x_i)] \oplus \bigoplus_{j=1}^{s}[b_j f_j(x_1, \ldots, x_n)] \tag{15}$$

assumes the value one, where $a_0 \in GF(2)$. The first half of the information is obtained when $a_0 = 1$, while the second half is obtained when $a_0 = 0$.

Note that the number of times that the function g defined by (**??**) assumes the value one is the Hamming distance between $\bigoplus_{j=1}^{s}[b_j f_j(x_1, \ldots, x_n)]$, a nonzero linear combination of the component functions, and $a_0 \oplus \bigoplus_{i=1}^{n}(a_i x_i)$, an affine function on V_n. To immunize an S-box against linear cryptanalysis, it suffices for the Hamming distance between any nonzero linear combination of the component functions and any affine function not to deviate too far from 2^{n-1}. Alternatively we have,

> An S-box is immune to linear cryptanalysis if the nonlinearity of each nonzero linear combination of its component functions is high.

As is indicated by Theorem **??**, for the S-boxes constructed in this paper all nonzero linear combinations of the component functions are highly nonlinear. Hence we conclude that they are immune against linear cryptanalysis.

With S-boxes constructed in [?, ?, ?], any nonzero linear combination of the component functions is a bent function. Hence these S-boxes have the strongest possible immunity to linear cryptanalysis. Unfortunately, as was discussed before, their component functions are not balanced, and even worse, their difference distribution tables are flat and hence they are not immune to differential cryptanalysis.

8.2 SAC vs Difference Distribution Table

We have shown that the component functions of a robust S-box $\Psi = (\psi_1, \ldots, \psi_{k+t})$ constructed by (**??**) in Section **??** all satisfy the SAC. In fact we have shown a much stronger result, namely, all but $2^t - 1$ of their nonzero linear combinations satisfy the SAC. This should be compared to $\Pi = (\pi_1, \ldots, \pi_k)$ constructed by (**??**). Π is not robust against differential cryptanalysis. However, all nonzero linear combinations of its component functions satisfy the SAC. This raises a question as to whether all nonzero linear combinations of the component functions of a very robust S-box, whose difference distribution table contains zero entries in all its rows, can satisfy the SAC.

We prove that the answer to the question is negative. In other words, for any S-box whose difference distribution table contains zero entries in all its rows, at least one nonzero linear combinations of its component functions does not satisfy the SAC.

Theorem 8 *Let $F = (f_1, \ldots, f_s)$ be an $n \times s$ S-box, where f_i is a function on V_n and $n \geqq s$. If the difference distribution table of F contains zero entries in all its rows, then at least one nonzero linear combination of $f_1, \ldots, f_s$ does not satisfy the SAC.*

Proof. Let $x = (x_1, \ldots, x_n)$. Since all rows in the difference distribution table of F contain zero entries, we know that for any nonzero vector $\alpha \in V_s$, $F(x) \oplus F(x \oplus \alpha)$ does not run through some vectors in V_s, while x runs through V_n, or equivalently, $F(x) \oplus F(x \oplus \alpha)$ is not a regular mapping. Note that

$$F(x) \oplus F(x \oplus \alpha) = (f_1(x) \oplus f_1(x \oplus \alpha), \ldots, f_s(x) \oplus f_s(x \oplus \alpha)).$$

Theorem **??** implies that there is at least one nonzero vector $(a_1, \ldots, a_s) \in V_s$ such that

$$\begin{aligned} \bigoplus_{i=1}^{s} \{a_i[f_i(x) \oplus f_i(x \oplus \alpha)]\} &= \bigoplus_{i=1}^{s} [a_i f_i(x)] \oplus \bigoplus_{i=1}^{s} [a_i f_i(x \oplus \alpha)] \\ &= f_\alpha(x) \oplus f_\alpha(x \oplus \alpha) \end{aligned}$$

is not balanced, where $f_\alpha(x) = \bigoplus_{i=1}^{s}[a_i f_i(x)]$. In particular, the argument is true when $W(\alpha) = 1$. That is, f_α does not satisfies the SAC. □

9 An Example

The procedure for generating an $n \times s$ S-box, where $n \geqq s > \lfloor n/2 \rfloor + t$, can be described in the following steps.

1. Select a primitive polynomial of algebraic degree $k-1$, where $k = s-t$. Construct from the polynomial a matrix $D = \begin{bmatrix} 0 & \cdots & 0 \\ \vdots & C & \\ 0 & & \end{bmatrix}$, where $C = (c_{ij})$, $c_{ij} = \varepsilon^{j+i \pmod{2^k-1}}$, $0 \leqq i, j \leqq 2^k - 2$. Note that only $c_0 = (c_{00}, c_{01}, \ldots, c_{0,2^k-3}, c_{0,2^k-2})$ has to be calculated. The other rows of C can be obtained by rotating c_0 to the left. That is, $c_1 = (c_{01}, c_{02}, \ldots, c_{0,2^k-2}, c_{00})$, $c_2 = (c_{02}, c_{03}, \ldots, c_{00}, c_{01})$, and so on.
2. Obtain from D a matrix E of linear functions on V_k by substituting ε^i with x_{i+1}, where $0 \leqq i \leqq k-1$. Note that E is a $2^k \times 2^k$ matrix, and that the first row and the first column of E contain only zeros.
3. Obtain a $2^k \times 2^{n-k}$ matrix H by selecting 2^{n-k} distinct nonzero columns from E. When the maximum algebraic degree $n - k + 1$ is required, E should be chosen so that the condition (**??**) is satisfied.
4. Permute the columns of H.
5. Construct $k+t$ functions f_1, …, f_{k+t} by (**??**). Note that g_{k+1}, …, g_{k+t} can be any distinct functions chosen from g_1, …, g_{2^k-1}.
6. Select a $(k+t) \times (k+t)$ nondegenerate matrix A so that its ith row γ_i, $i = 1, \ldots, k+t$, can be written as $\gamma_i = (\beta_i, \alpha_i)$, where $\beta_i \in V_{n-k}$, $W(\beta_i) \geqq 1$ and $\alpha_i \in V_k$.
7. Output $(f_1(zA), \ldots, f_{k+t}(zA))$ as an S-box.

Now we construct a 12×10 S-box to illustrate the generating procedure. Let $n = 12$, $s = 10$, $t = 3$ and $k = 7$. Choose $x^7 \oplus x \oplus 1$ as the primitive polynomial. Let ε be a root of $x^7 \oplus x \oplus 1 = 0$.

The first row of the 127×127 matrix C (see Subsection **??**) is ε^0, ε^1, …, ε^{126}, that is

$$1, \varepsilon, \varepsilon^2, \varepsilon^3, \varepsilon^4, \varepsilon^5, \varepsilon^6, 1 \oplus \varepsilon, \varepsilon \oplus \varepsilon^2, \ldots, 1 \oplus \varepsilon^6.$$

The second row of C is obtained by rotating the first row to the left by one position, the third row by rotating the second row to the left by one position, and so on. Then we have an extended 128×128 matrix $D = \begin{bmatrix} 0 & \cdots & 0 \\ \vdots & C & \\ 0 & & \end{bmatrix}$. By substituting ε^i with x_{i+1}, $i = 0, 1, 2, 3, 4, 5, 6$, we obtain a matrix $E = (e_{ij})$, $0 \leqq i, j \leqq 127$. In particular, the first row of E contains only zeros, and the second row of E is

$$0, x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_1 \oplus x_2, x_2 \oplus x_3, \ldots, x_1 \oplus x_7$$

Next we select $2^{12-7} = 32$ different nonzero columns from E so that the condition (**??**) is satisfied. Then we permute randomly the selected rows. In this way we obtain a matrix $H = (h_{ij})$, where $0 \leqq i \leqq 127$ and $0 \leqq j \leqq 31$.

Now let $y = (y_1, y_2, y_3, y_4, y_5)$, $x = (x_1, x_2, x_3, x_4, x_5, x_6, x_7)$, $w = (y_1, y_2, y_3)$, $z = (y, x)$, and let

$$g_i(y,x) = \bigoplus_{j=0}^{31}[D_j(y)h_{ij}(x)], i = 1,2,3,4,5,6,7.$$

Let g_8, g_9 and g_{10} be three distinct nonzero linear combinations of $g_1, \ldots, g_7$. Set

$$\begin{aligned} f_j(z) &= g_j(z), j = 1,2,3,4,5,6,7, \\ f_{j+7}(z) &= g_{j+7}(x) \oplus m_j(w), j = 1,2,3 \end{aligned}$$

where $m_j(w) = m_j(y_1, y_2, y_3)$ is constructed in Subsection **??**. Let A be the following nondegenerate matrix

$$A = \begin{bmatrix} 1 & 0 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}.$$

The final S-Box is $\Psi = (\psi_1, \ldots, \psi_{10})$, where $\psi(z) = f_i(zA)$.

Let $\psi = \bigoplus_{j=1}^{10}[c_j\psi_j]$ be a nonzero linear combination of $\psi_1, \ldots, \psi_{10}$. By Theorem **??**, ψ has the properties described here.

1. ψ is balanced.
2. In $2^{10} - 8 = 1016$ cases including $\psi = f_i$, $i = 1, \ldots, 10$, the nonlinearity of ψ satisfies $N_\psi \geqq 2^{12-1} - 2^{7-1} = 1984$, and the algebraic degree of ψ is 6. In the other 7 cases, $N_\psi \geqq 2^{12-2} = 1024$, and the algebraic degree of ψ is 2.
3. ψ satisfies the SAC except when $\psi(z) = \bigoplus_{j=1}^{k+3}[c_j r_j(zA)]$.

The difference distribution table of the S-box has the profile described here:

1. In $2^7 - 1 = 127$ cases, $2^{12-7} = 32$ out of the $2^{10} = 1024$ entries in a row contain a value $2^7 = 128$, and the other $2^{10} - 2^5 = 992$ entries contain a value zero.
2. In other $2^9 - 2^7 = 384$ cases, $2^7 = 128$ out of the 1024 entries in a row contain a value $2^5 = 32$, and the other $2^{10} - 2^7 = 896$ entries contain a value zero.
3. In the remaining $2^{12} - 2^9 = 3584$ cases (not counting the first row), half of the 1024 entries in a row contain a value $2^3 = 8$, and the other half contain a value zero.
4. In the first column, the first entry contains a value $2^{12} = 4096$, $(2^{12-10} - 1)2^7 = 384$ other entries contain a value $2^{12-7} = 32$, and the remaining 3711 entries contain a value zero.

Consequently, the robustness of the S-box against differential cryptanalysis is $(\frac{7}{8} + 2^{-5})(1 - 2^{-5}) \approx 0.878$.

10 Conclusion

We have presented a method for systematically generating cryptographically strong S-boxes. The method is based on an interesting combinatorial structure called group Hadamard matrices. We have shown that the method is much superior to previous approaches, and that it generates promising S-boxes in terms of their robustness against differential cryptanalysis, immunity to linear cryptanalysis, SAC fulfilling properties, high nonlinearities and algebraic degrees. We have also illustrated the construction method by an example of 12×10 S-boxes. Future research directions include the investigation of possible further improvements on the algebraic degrees, the nonlinearities and the profiles of the difference distribution tables of the S-boxes.