New constructions of resilient Boolean functions with maximal nonlinearity

Yuriy Tarannikov

Mech. & Math. Department
Moscow State University
119899 Moscow, Russia
emails: yutaran@mech.math.msu.su, taran@vertex.inria.msu.ru

Abstract. In this paper we develop a technique that allows to obtain new effective constructions of highly resilient Boolean functions with high nonlinearity. In particular, we prove that the upper bound $2^{n-1} - 2^{m+1}$ on nonlinearity of m-resilient n-variable Boolean functions is achieved for $0.6n - 1 \leq m \leq n - 2$.

Keywords: *stream cipher, Boolean function, nonlinear combining function, correlation-immunity, resiliency, nonlinearity.*

1 Introduction

One of the most general types of stream cipher systems is several Linear Feedback Shift Registers (LFSRs) combined by nonlinear Boolean function. This function must satisfy certain criteria to resist different attacks (in particular, correlation attacks suggested by Siegenthaler [15] and different types of linear attacks). The following factors are considered as important properties of Boolean functions for using in stream cipher applications.

1. *Balancedness.* A Boolean function must output zeroes and ones with the same probabilities.

2. Good *correlation-immunity* (of order m). The output of Boolean function must be statistically independent of combination of any m its inputs. A balanced correlation-immune of order m Boolean function is called *m-resilient.*

3. Good *nonlinearity.* The Boolean function must be at the sufficiently high distance from any affine function.

Other important factors are large algebraic degree and simple implementation in hardware.

The variety of criteria and complicated trade-offs between them caused the next approach: to fix one or two parameters and try to optimize others. The most general model is when researchers fix the parameters n (number of variables) and m (order of correlation-immunity) and try to optimize some other cryptographically important parameters. Here we can call the works [13], [2], [5], [4] [6], [7], [8], [9], [16].

The present paper continues the investigations in this direction and gives new results. In Section 2 we give preliminary concepts and notions. In Section 3 we give a brief review of the investigations on the problem of maximal nonlinearity for n-variable m-resilient Boolean function. In Section 4 we discuss a concept of a linear and a pair of quasilinear variables which works in the following sections. In Section 5 we present our main construction method. This method is a generalization of a method described in [16]. This method allows to construct recursively the functions with good cryptographic properties using the functions with good cryptographic properties and smaller number of variables. The method is based on the existence of a *proper* matrix with prescribed properties. In Section 6 we give some examples of proper matrices and obtain new results on the maximal nonlinearity $nlmax(n,l)$ of m-resilient functions on V^n. Namely, we prove that $nlmax(n,m) = 2^{n-1} - 2^{m+1}$ for $\frac{5n-14}{8} \leq m \leq n-2$ and for $0.6n-1 \leq m \leq n-2$. In Section 7 we give some remarks on the combinatorial problem connected with proper matrices and give a geometrical interpretations of proper matrices.

2 Preliminary concepts and notions

We consider V^n, the vector space of n tuples of elements from $GF(2)$. A *Boolean function* is a function from V^n to $GF(2)$. The *weight* $wt(f)$ of a function f on V^n is the number of vectors $\widetilde{\sigma}$ on V^n such that $f(\widetilde{\sigma}) = 1$. A function f is said to be *balanced* if $wt(f) = wt(f \oplus 1)$. Obviously, if a function f on V^n is balanced then $wt(f) = 2^{n-1}$. A *subfunction* of the Boolean function f is a function f' obtained by substitution some constants for some variables in f. If we substitute in the function f the constants $\sigma_{i_1}, \ldots, \sigma_{i_s}$ for the variables $x_{i_1}, \ldots, x_{i_s}$ respectively then the obtained subfunction is denoted by $f^{\sigma_{i_1},\ldots,\sigma_{i_s}}_{x_{i_1},\ldots,x_{i_s}}$. If a variable x_i is not substituted by constant then x_i is called a *free* variable for f'.

It is well known that a function f on V^n can be uniquely represented by a polynomial on $GF(2)$ whose degree is at most n. Namely,

$$f(x_1, \ldots, x_n) = \bigoplus_{(a_1,\ldots,a_n) \in V^n} g(a_1, \ldots, a_n) x_1^{a_1} \ldots x_n^{a_n}$$

where g is also a function on V^n. This polynomial representation of f is called the *algebraic normal form* (briefly, ANF) of the function and each $x_1^{a_1} \ldots x_n^{a_n}$ is called a *term* in ANF of f. The *algebraic degree* of f, denoted by $\deg(f)$, is defined as the number of variables in the longest term of f. The *algebraic degree of variable* x_i in f, denoted by $\deg(f, x_i)$, is the number of variables in the longest term of f that contains x_i. If $\deg(f, x_i) = 1$, we say that x_i is a *linear* variable in f. The term of length 1 is called a *linear* term. If $\deg(f) \leq 1$ then f is called an *affine* function.

The *Hamming distance* $d(\widetilde{\sigma}_1, \widetilde{\sigma}_2)$ between two vectors $\widetilde{\sigma}_1$ and $\widetilde{\sigma}_2$ is the number of components where vectors $\widetilde{\sigma}_1$ and $\widetilde{\sigma}_2$ differ. For two Boolean functions f_1 and f_2 on V^n, we define the distance between f_1 and f_2 by $d(f_1, f_2) = \#\{\widetilde{\sigma} \in V^n | f_1(\widetilde{\sigma}) \neq f_2(\widetilde{\sigma})\}$. The minimum distance between f and the set of all affine functions is called the *nonlinearity* of f and denoted by $nl(f)$.

A Boolean function f on V^n is said to be *correlation-immune of order* m, with $1 \le m \le n$, if the output of f and any m input variables are statistically independent. This concept was introduced by Siegenthaler [14]. In equivalent non-probabilistic formulation the Boolean function f is called correlation-immune of order m if $wt(f') = wt(f)/2^m$ for any its subfunction f' of $n-m$ variables. A balanced mth order correlation immune function is called an *m-resilient* function. In other words the Boolean function f is called m-resilient if $wt(f') = 2^{n-m-1}$ for any its subfunction f' of $n-m$ variables. From this point of view we can consider formally any balanced Boolean function as 0-resilient (this convention is accepted in [1], [7], [9]) and an arbitrary Boolean function as (-1)-resilient. The concept of an m-resilient function was introduced in [3].

Siegenthaler's Inequality [14] states that if the function f is a correlation-immune function of order m then $\deg(f) \le n - m$. Moreover, if f is an m-resilient, $m \le n-2$, then $\deg(f) \le n - m - 1$.

The next lemma is well-known.

Lemma 1. *Let $f(x_1, \ldots, x_n)$ be a Boolean function represented in the form*

$$f(x_1, \ldots, x_n) = \bigoplus_{(\sigma_1, \ldots, \sigma_l)} (x_1 \oplus \sigma_1) \ldots (x_l \oplus \sigma_l) f(\sigma_1 \oplus 1, \ldots, \sigma_l \oplus 1, x_{l+1}, \ldots, x_n).$$

Suppose that all 2^l subfunctions $f(\sigma_1 \oplus 1, \ldots, \sigma_l \oplus 1, x_{l+1}, \ldots, x_n)$ are m-resilient. Then the function f is an m-resilient too.

The Lemma 1 was proved in a lot of papers including (for $l = 1$) the pioneering paper of Siegenthaler (Theorem 2 in [14]). General case follows immediately from the case $l = 1$.

3 The problem of maximal nonlinearity for resilient functions

Let m and m be integers, $-1 \le m \le n$. Denote by $nlmax(n, m)$ the maximal possible nonlinearity of m-resilient Boolean function on V^n. It is well-known that the nonlinearity of a Boolean function does not exceed $2^{n-1} - 2^{\frac{n}{2}-1}$ [12]. Thus, $nlmax(n, -1) \le 2^{n-1} - 2^{\frac{n}{2}-1}$, This value can be achieved only for even n. The functions with such nonlinearity are called *bent functions*. Thus, for even n we have $nlmax(n, -1) = 2^{n-1} - 2^{\frac{n}{2}-1}$. It is known [10, 11, 5] that for odd n, $n \le 7$, $nlmax(n, -1) = 2^{n-1} - 2^{(n-1)/2}$, and for odd n, $n \ge 15$, the inequality $nlmax(n, -1) > 2^{n-1} - 2^{(n-1)/2}$ holds. Bent functions are nonbalanced always, so, for balanced (0-resilient) n-variable function f we have $nl(f) < 2^{n-1} - 2^{\frac{n}{2}-1}$, and $nlmax(n, m) < 2^{n-1} - 2^{\frac{n}{2}-1}$ for $m \ge 0$. If f is n-variable m-resilient function, $m \ge n-2$, then by Siegenthaler's Inequality [14] $\deg(f) \le 1$, so $nlmax(n, m) = 0$. For some small values of parameters n and m exact values of maximal nonlinearity are known. The latest collection of such values is given in [8]. The upper bound $nlmax(n, m) \le 2^{n-1} - 2^{m+1}$ for $m \le n-1$ was proven independently in [8], [16] and [17], all three these manuscripts were submitted to Crypto 2000

although only the first was accepted). In [16] an effective construction of m-resilient function on V^n with nonlinearity $2^{n-1} - 2^{m+1}$ was given. Therefore $nlmax(n,m) = 2^{n-1} - 2^{m+1}$ for $\frac{2n-7}{3} \leq m \leq n-2$. In [8] it was proved that the nonlinearity of m-resilient function on V^n is divided by 2^{m+1}. Also in [8] it was proved that $nlmax(n,m) \leq 2^{n-1} - 2^{n/2-1} - 2^{m+1}$ for $m < (n/2) - 2$.

4 On linear and quasilinear variables

In this section we recall the concepts of linear and quasilinear variables. The last concept was introduced in [16].

Recall that a variable x_i is called *a linear* for a function $f = f(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_n)$ if $\deg(f, x_i) = 1$. Also we say that a function f depends on a variable x_i *linearly*. If a variable x_i is linear for a function f we can represent f in the form

$$f(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_n) = g(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n) \oplus x_i.$$

Other equivalent definition of a linear variable is that a variable x_i is linear for a function f if $f(\widetilde{\delta}_1) \neq f(\widetilde{\delta}_2)$ for any two vectors $\widetilde{\delta}_1$ and $\widetilde{\delta}_2$ that differ only in ith component. By analogy with the last definition we give a new definition for a pair of quasilinear variables.

Definition 1. *We say that a Boolean function $f = f(x_1, \ldots, x_n)$ depends on a pair of its variables (x_i, x_j)* quasilinearly *if $f(\widetilde{\delta}_1) \neq f(\widetilde{\delta}_2)$ for any two vectors $\widetilde{\delta}_1$ and $\widetilde{\delta}_2$ of length n that differ only in ith and jth components. A pair (x_i, x_j) in this case is called a* pair of quasilinear variables *in f.*

The proof of the next lemma is given in [16].

Lemma 2. *Let $f(x_1, \ldots, x_n)$ be a Boolean function. Then (x_i, x_j), $i < j$, is a pair of quasilinear variables in f iff f can be represented in the form*

$$f(x_1, \ldots, x_n) = g(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_{j-1}, x_{j+1}, \ldots, x_n, x_i \oplus x_j) \oplus x_i. \quad (1)$$

The next lemmas are obvious.

Lemma 3. *Let $f(x_1, \ldots, x_n)$ be a Boolean function. If f depends on some variable x_i linearly then f is balanced.*

Lemma 4. *$f(x_1, \ldots, x_n)$ be a Boolean function. If f depends on some variables x_{i_1}, x_{i_2}, $\ldots$, x_{i_s} linearly then f is $(s-1)$-resilient.*

Note that Lemma 4 agrees with our assumption that a balanced function is 0-resilient, and an arbitrary Boolean function is (-1)-resilient. (In the last case $s = 0$.)

Lemma 5. *Let $f(x_1, \ldots, x_n)$ be a Boolean function. If f depends on some pair of variables (x_i, x_j) quasilinearly then f is balanced.*

Lemma 6. *Let* $f(x_1,\ldots,x_n,x_{n+1}) = f(x_1,\ldots,x_n) \oplus cx_{n+1}$ *where* $c \in \{0,1\}$. *Then* $nl(f) = 2nl(g)$.

Lemma 7. *Let* $f(x_1,\ldots,x_n)$ *be a Boolean function on* V^n *and* f *depends on some pair of variables* (x_i,x_j) *quasilinearly. Then* $nl(f) = 2nl(g)$ *where* g *is a function used in the representation of* f *in Lemma 2.*

Lemma 8. *Let* f_1 *and* f_2 *be two Boolean functions on* V^n. *Moreover, there exist variables* x_i *and* x_j *such that* f_1 *depends on a pair of variables* (x_i,x_j) *quasilinearly whereas* f_2 *depends on the variables* x_i *and* x_j *linearly. Let* l *be an arbitrary affine function on* V^n. *Then at least one of two functions* $f_1 \oplus l$ *and* $f_2 \oplus l$ *is balanced.*

Proof. Let $l = \bigoplus_{r=1}^{n} u_r x_r \oplus u_0$, $u_r \in \{0,1\}$, $r = 0,1,\ldots,n$. If $u_i = 0$ (correspondently, $u_j = 0$) then $f_2 \oplus l$ depends on the variable x_i (x_j) linearly, therefore the function $f_2 \oplus l$ is balanced. The remained case is $u_i = u_j = 1$. But here it is easy to see that the function $f_1 \oplus l$ depends on a pair of variables (x_i,x_j) quasilinearly, therefore $f_1 \oplus l$ is balanced. □

5 A method of constructing

Suppose that $f_0, f_1, \ldots, f_{2^k-1}$ are Boolean functions on V^n. We denote f_r also as $f_{\sigma_1\ldots\sigma_k}$ where $\sigma_1\ldots\sigma_k$ is a binary representation of the number r. Suppose that $c = (c_1,\ldots,c_k)$ is an arbitrary binary vector. Put $s = \sum_{i=1}^{k} c_i$. We denote $X = \{x_i \,|\, i = 1,\ldots,n\}$, $Y = \{y_i \,|\, i = 1,\ldots,k\}$, $Z = \{z_i \,|\, c_i = 1,\ i = 1,\ldots,k\}$. We define

$$f(X,Y,Z) = \left(\bigoplus_{(\sigma_1,\ldots,\sigma_k)\in V^k} \left(\prod_{i=1}^{k} (y_i \oplus c_i z_i \oplus \sigma_i) \right) f_{\sigma_1\ldots\sigma_k}(X) \right) \oplus \bigoplus_{i=1}^{k} c_i z_i. \quad (2)$$

By construction the function f in (2) depends on $n+k+s$ variables. Below we formulate some properties of construction (2).

Remark. Some details of construction (2) can be understood more easily if we put $c_1 = \ldots = c_s = 1$, $c_{s+1} = \ldots = c_k = 0$. But for an effective implementation it is important in some cases to vary the vector c.

Lemma 9. *Suppose that all* 2^k *Boolean functions* f_0, f_1, …, f_{2^k-1} *in (2) are* m*-resilient. Then the function* $f(X,Y,Z)$ *is* $(m+s)$*-resilient.*

Proof. Substitute in (2) arbitrary $m+s$ constants for arbitrary $m+s$ variables. We obtain some $(n+k-m)$-variable subfunction f'. If $c_i = 1$ for some i and if both variables y_i and z_i are free in f' then the pair of variables $(y_i, z_i$ is a quasilinear pair in f' therefore by Lemma 5 the subfunction f' is balanced). Thus, we can assume that for each i such that $c_i = 1$ at least one of two variables y_i and

z_i is substituted by constant. Then at most m variables from X are substituted by constants in (2). All functions $f_0, f_1, \ldots, f_{2^k-1}$ are balanced, therefore the function $f(X,Y,Z)$ is balanced too. We have proved that an arbitrary $(n+k-m)$-variable subfunction of $f(X,Y,Z)$ is balanced. □

Lemma 10. *Suppose that the nonlinearity of all 2^k Boolean functions $f_0, f_1, \ldots, f_{2^k-1}$ in (2) is at least N_0. Moreover, for any two functions f_{r_1} and f_{r_2}, $0 \le r_1 \ne r_2 \le 2^k-1$, there exists a pair of variables (x_i, x_j) such that one of these two functions, say f_{r_1} depends linearly on the variables x_i and x_j whereas another function f_{r_2} depends quasilinearly on the pair (x_i, x_j). Then $nl(f) \ge 2^s(2^{n-1}(2^k-1)+N_0)$.*

Proof. It is obvious that if we replace $c_i = 0$ by $c_i = 1$ then we multiplate the nonlinearity by 2 (adding new variable). Thus we can assume that $s = 0$. Consider an arbitrary affine function l. Denote $l_r = l_{y_1, \ldots, y_k}^{\sigma_1 \oplus 1, \ldots, \sigma_k \oplus 1}$ where $\sigma_1 \ldots \sigma_k$ is a binary representation of the number r. Note that for any $r = 0, \ldots, 2^k-1$, we have $l_r = l_0$ or $l_r = l_0 \oplus 1$. Then $d(f,l) = \sum_{r=0}^{2^k-1} d(f_r, l_r)$. By Lemma 8 and the hypothesis of this lemma we have that $d(f_r, l_r) \ne 2^{n-1}$ for at most one value of r. Thus $d(f,l) \ge 2^{n-1}(2^k-1)+N_0$. An affine function l was chosen arbitrary. Therefore $nl(f) \ge 2^{n-1}(2^k-1)+N_0$. □

The construction (2) is a generalization of the construction in [16] where only the case $k = 1$ is considered.

The problem is to find the functions $f_0, f_1, \ldots, f_{2^k-1}$ with desirable distribution of linear and quasilinear variables. Below we give some approach that allows to construct such systems of functions.

Definition 2. *Let $B = (b_{ij})$ be $(2^k \times p)$ matrix of 2^k rows and p columns with entries from the set $\{1, 2, *\}$. Let k_0 and t be positive integers. We assume that*

(i) for every two rows i_1 and i_2 there exist a column j such that $b_{i_1 j} = 1$, $b_{i_2 j} = 2$ or $b_{i_1 j} = 2$, $b_{i_2 j} = 1$.

(ii) for every row i the inequality $\sum_{j=1}^{p} b_{ij} \le t$ holds (a sign $*$ does not give an influence to these sums).

(iii) in every row the number of ones does not exceed k_0.

If the matrix B satisfies all properties (i), (ii), (iii) we say that B is a proper *(k_0, k, p, t)-matrix.*

Definition 3. *Let F be a set of Boolean functions such that for every s, $0 \le s \le k$, the set F contains an $(m+s)$-resilient function on V^{n+s} with nonlinearity at least $2^s(2^{n-1} - 2^{m+\lambda})$ (λ is not necessary integer) that contains s disjoint pairs of quasilinear variables. Then we say that F is a $S_{n,m,k_0,\lambda}$-*system *of Boolean functions.*

Remark. To provide an existence of a $S_{n,m,k,\lambda}$-system of Boolean functions it is sufficiently to have only $(m+k)$-resilient function f on V^{n+k} with nonlinearity at least $2^k(2^{n-1} - 2^{m+\lambda})$ that contains k disjoint pairs of quasilinear variables. All other necessary functions of $S_{n,m,k,\lambda}$-system can be obtained from

f by substitutions of constants for the variables from different disjoint pairs of quasilinear variables. But note that the last way is not effective from the implementation point of view.

Lemma 11. *There exists an $S_{2,-1,2,1}$-system of Boolean functions.*

Proof. Put $f'_0 = x_1x_2$, $f'_1 = (x_1 \oplus x_2)x_3 \oplus x_1$, $f'_2 = (x_1 \oplus x_2)(x_3 \oplus x_4) \oplus x_1 \oplus x_3$. It is easy to verify that f'_s, $s = 0, 1, 2$, is a $(-1+s)$-resilient function on V^{2+s} with nonlinearity $2^s(2^{2-1} - 2^{-1+1})$, moreover, f'_s contains s disjoint pairs of quasilinear variables. □

Theorem 1. *Suppose that there exists an $S_{n,m,k_0,\lambda}$-system of Boolean functions F and there exists a proper (k_0, k, p, t)-matrix B, $n \geq 2p - t$. Then there exists an $S_{n+k+t,m+t,k,\lambda}$-system of Boolean functions.*

Proof. Consider the ith row of the matrix B, $i = 0, 1, \ldots, 2^k - 1$. Suppose that this row contains $s = s(i)$ ones. The matrix B is a proper, therefore $s \leq k_0$, $s \leq t$. By assumption there exists an $(m+s)$-resilient function f'_i on V^{n+s} that contains s disjoint pairs of quasilinear variables with nonlinearity at least $2^s(2^{n-1} - 2^{m+\lambda})$. Add $t - s$ new linear variables to the function f'_i. As a result we obtain the function f''_i on V^{n+t}. It is easy to see that the function f''_i is an $(m+t)$-resilient function with nonlinearity at least $2^t(2^{n-1} - 2^{m+\lambda})$, moreover f''_i contains s disjoint pairs of quasilinear variables and besides $t-s$ linear variables. Note that by the property (ii) of a proper matrix the value $t-s$ is not less than the number of 2's in ith row of B multiplied by 2. By this way we construct the functions f''_i on V^{n+t} for every i, $i = 0, 1, \ldots, 2^k - 1$. By assumption $n+t \geq 2p$. Next, for every i, $i = 0, 1, \ldots, 2^k - 1$, we permute the variables in $f''_i(x_1, \ldots, x_{n+t})$ obtaining the function f_i such that the function f_i depends on a pair of variables (x_{2j-1}, x_{2j}) quasilinearly if $b_{ij} = 1$, and the function f_i depends on the variables x_{2j-1} and x_{2j} linearly if $b_{ij} = 2$. By the arguments given above we have sufficient numbers of quasilinear and linear variables for this procedure. Now we are ready to apply the construction (2). By means of this construction varying the number of ones in the vector $(c_1, \ldots, c_k)$ we obtain the functions $f(X, Y, Z_s)$, $s = 0, 1, \ldots, k$. The function $f(X, Y, Z_s)$ by lemmas 9 and 10 is an $(m+t+s)$-resilient function on $V^{n+k+t+s}$ with the nonlinearity at least $2^s(2^{n+k+t-1} - 2^{m+t+\lambda})$. Moreover, the function $f(X, Y, Z_s)$ contains s disjoint pairs of quasilinear variables. Thus, we have constructed an $S_{n+k+t,m+t,k,\lambda}$-system of Boolean functions. □

An application of the construction given in Theorem 1 we denote by

$$S_{n,m,k_0,\lambda}T_{k_0,k,p,t} = S_{n+k+t,m+t,k,\lambda}.$$

If we add new linear variable to an m-resilient function f on V^n then we obtain $(m+1)$-resilient function f on V^{n+1} with nonlinearity $2nl(f)$. We denote this procedure by

$$S_{n,m,0,\lambda}T_{0,0,0,1} = S_{n+1,m+1,0,\lambda}.$$

6 Examples of proper matrices effective for our construction and new resilient Boolean functions with maximal nonlinearity

At first, we give some examples of proper matrices effective for the construction of Boolean functions with good combination of parameters. We denote a proper (k_0, k, p, t)-matrix by $B_{k_0,k,p,t}$.

$$B_{1,1,1,2} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \; B_{2,2,2,4} = \begin{pmatrix} 2 & 2 \\ 2 & 1 \\ 1 & 2 \\ 1 & 1 \end{pmatrix}, \; B_{3,2,3,3} = \begin{pmatrix} 2 & 1 & * \\ * & 2 & 1 \\ 1 & * & 2 \\ 1 & 1 & 1 \end{pmatrix},$$

$$B_{2,3,5,6} = \begin{pmatrix} 2 & 2 & 1 & 1 & * \\ 2 & 1 & 1 & 2 & * \\ 2 & 1 & * & 1 & 2 \\ 2 & 1 & 2 & * & 1 \\ 1 & 1 & * & 2 & 2 \\ 1 & 2 & 1 & * & 2 \\ 1 & * & 2 & 1 & 2 \\ 1 & * & 2 & 2 & 1 \end{pmatrix}, \; B_{3,3,4,5} = \begin{pmatrix} * & 1 & 2 & 2 \\ 2 & * & 1 & 2 \\ 2 & 2 & * & 1 \\ 1 & 2 & 2 & * \\ 2 & 1 & 1 & 1 \\ 1 & 2 & 1 & 1 \\ 1 & 1 & 2 & 1 \\ 1 & 1 & 1 & 2 \end{pmatrix},$$

$$B_{2,4,7,8} = \begin{pmatrix} 1 & 1 & * & * & 2 & 2 & 2 \\ * & 2 & 1 & 1 & * & 2 & 2 \\ * & * & 2 & 2 & 1 & 1 & 2 \\ 1 & * & * & 2 & 2 & 2 & 1 \\ 2 & 1 & 1 & * & * & 2 & 2 \\ * & * & 2 & 1 & 1 & 2 & 2 \\ * & * & 2 & 2 & 2 & 1 & 1 \\ 1 & 2 & 2 & 1 & 2 & * & * \\ * & 1 & 2 & 2 & 1 & 2 & * \\ * & * & 1 & 2 & 2 & 1 & 2 \\ 2 & * & * & 1 & 2 & 2 & 1 \\ 1 & 2 & * & 2 & 1 & 2 & * \\ * & 1 & 2 & * & 2 & 1 & 2 \\ 2 & * & 1 & 2 & * & 2 & 1 \\ 2 & 2 & * & 1 & * & 1 & 2 \\ 2 & 2 & 2 & * & 1 & * & 1 \end{pmatrix}, B_{4,4,6,6} = \begin{pmatrix} 2 & 2 & 2 & * & * & * \\ 1 & 2 & * & 1 & 2 & * \\ 1 & 2 & * & * & 1 & 2 \\ 1 & 2 & * & 2 & * & 1 \\ * & 1 & 2 & 1 & 2 & * \\ * & 1 & 2 & * & 1 & 2 \\ * & 1 & 2 & 2 & * & 1 \\ 2 & * & 1 & 1 & 2 & * \\ 2 & * & 1 & * & 1 & 2 \\ 2 & * & 1 & 2 & * & 1 \\ 2 & * & 1 & 1 & 1 & 1 \\ 1 & 2 & * & 1 & 1 & 1 \\ * & 1 & 2 & 1 & 1 & 1 \\ 1 & 1 & 1 & 2 & * & 1 \\ 1 & 1 & 1 & 1 & 2 & * \\ 1 & 1 & 1 & * & 1 & 2 \end{pmatrix}.$$

It is easy to verify that all matrices given above are proper matrices with correspondent parameters.

The simplest example of a proper matrix is the matrix $B_{1,1,1,2}$. If $\frac{2n-7}{3} \leq m \leq n-3$ then the numbers n and m can be represented in the form $n = 3r + s + 2$, $m = 2r + s - 1$, where r and s are nonnegative integers (an existence of this representation as well as an existence of the representations in Theorems 2 and 3 can be proved by the arguments from the elementary arithmetic). By Lemma 11 there exists a system $S_{2,-1,2,1}$. We apply

$$S_{2,-1,2,1}(T_{1,1,1,2})^r (T_{0,0,0,1})^s = S_{n,m,0,1}.$$

Therefore $nlmax(n,m) \geq 2^{n-1} - 2^{m+1}$ for $m \geq \frac{2n-7}{3}$. In Section 3 it was pointed out that $nlmax(n,m) \leq 2^{n-1} - 2^{m+1}$ for $m \leq n-2$. Therefore $nlmax(n,m) = 2^{n-1} - 2^{m+1}$ for $\frac{2n-7}{3} \leq m \leq n-2$. The above construction was given in [16].

Theorem 2. *$nlmax(n,m) = 2^{n-1} - 2^{m+1}$ for $\frac{5n-14}{8} \leq m \leq n-2$.*

Proof. Let n, m be integers. Note that $\lceil \frac{5n-14}{8} \rceil \geq \lceil \frac{2n-7}{3} \rceil$ for $n < 17$. If $n \geq 17$, $m > n-8$, then $m \geq \frac{2n-7}{3}$. If $n \geq 17$, $\frac{5n-13}{8} \leq m \leq n-8$, then the numbers n and m can be represented in the form $n = 8r_1 + 3r_2 + s + 17$, $m = 5r_1 + 2r_2 + s + 9$, where r_1, r_2 and s are nonnegative integers. We apply

$$S_{2,-1,2,1}T_{2,2,2,4}T_{2,3,5,6}(T_{3,3,4,5})^{r_1}(T_{1,1,1,2})^{r_2}(T_{0,0,0,1})^{s} = S_{n,m,0,1}.$$

If $n \geq 17$, $\frac{5n-14}{8} = m$, then the numbers n and m can be represented in the form $n = 8r + 22$, $m = 5r + 12$, where r is nonnegative integer. In this case we apply

$$S_{2,-1,2,1}T_{2,2,2,4}T_{2,3,5,6}(T_{3,3,4,5})^{r}T_{3,2,3,3} = S_{n,m,2,1}.$$

□

Theorem 3. *$nlmax(n,m) = 2^{n-1} - 2^{m+1}$ for $0.6n - 1 \leq m \leq n-2$.*

Proof. Let n, m be integers. Note that $0.6n - 1 \geq \frac{2n-7}{3}$ for $n \leq 20$. If $n \geq 20$, $m > n-9$, then $m \geq \frac{2n-7}{3}$. If $n \geq 20$, $0.6n - 1 \leq m \leq n-9$, excepting the case $m = 0.6n - 1$, $n \equiv 5 \pmod{10}$, then the numbers n and m can be represented in the form $n = 10r_1 + 8r_2 + 3r_3 + s + 20$, $m = 6r_1 + 5r_2 + 2r_3 + s + 11$, where r_1, r_2,r_3 and s are nonnegative integers. We apply

$$S_{2,-1,2,1}T_{2,2,2,4}T_{2,4,7,8}(T_{4,4,6,6})^{r_1}(T_{3,3,4,5})^{r_2}(T_{1,1,1,2})^{r_3}(T_{0,0,0,1})^{s} = S_{n,m,0,1}.$$

In the case $n = 10r + 25$, $m = 6r + 14$, where r is a nonnegative integer we apply

$$S_{2,-1,2,1}T_{2,2,2,4}T_{2,4,7,8}(T_{4,4,6,6})^{r}T_{3,2,3,3} = S_{n,m,2,1}.$$

□

7 Some remarks on combinatorial problem and geometrical interpretations

If there exists a proper (k,k,p,t)-matrix then using the technique described in the previous section we can prove that $nlmax(n,m) = 2^{n-1} - 2^{m+1}$ for $m > \frac{t}{t+k}n - c'$ where c' is a some constant. Note that the construction in [2] allows to achieve such nonlinearity only for $m \leq c''\frac{n}{4}(1+o(1))$. Therefore we are interesting to find a proper (k,k,p,t)-matrix where the ratio $\frac{t}{k}$ is as small as possible.

For given positive integer k we denote by $t(k)$ the minimal positive integer t such that for some p there exists a proper (k,k,p,t)-matrix. It is clear that we can consider only matrices without all-$*$ columns. Then obviously $p \leq t \cdot 2^k$. There exists a proper $(k,k,k,2k)$-matrix (all rows are different and without $*$). Thus, to find $t(k)$ it is sufficiently to consider only a finite set of matrices.

Proposition 1. *Let k_1 and k_2 be positive integers. Then $t(k_1 + k_2) \leq t(k_1) + t(k_2)$.*

Proof. By definition for some p_1 and p_2 there exist a proper $(k_1, k_1, p_1, t(k_1))$-matrix B' and a proper $(k_2, k_2, p_2, t(k_2))$-matrix B''. Compose a $(2^{k_1+k_2} \times (p_1 + p_2))$ matrix B where the rows of B are all possible concatenations of rows of matrices B' and B''. It is easy to see that B is a proper $(k_1 + k_2, k_1 + k_2, p_1 + p_2, t(k_1) + t(k_2))$-matrix. Therefore $t(k_1 + k_2) \leq t(k_1) + t(k_2)$. □

It is quite obvious that

Proposition 2. $t(k) \geq k$.

Propositions 1 and 2 follow that there exists the limit $\lim\limits_{k\to\infty} \frac{t(k)}{k}$.

A proper (k_0, k, p, t)-matrix B can be interpreted as a collection of 2^k disjoint subcubes in Boolean cube $\{1, 2\}^p$. Indeed, a row of B can be interpreted as a subcube where the components with $*$ are free whereas the components with 1 or 2 are substituted by correspondent constant. We illustrate this at the example of the matrix $B_{3,3,4,5}$:

row of B_{3345}	points of a subcube
$*122$	$\{(1,1,2,2),(2,1,2,2)\}$
$2*12$	$\{(2,1,1,2),(2,2,1,2)\}$
$22*1$	$\{(2,2,1,1),(2,2,2,1)\}$
$122*$	$\{(1,2,2,1),(1,2,2,2)\}$
2111	$\{(2,1,1,1)\}$
1211	$\{(1,2,1,1)\}$
1121	$\{(1,1,2,1)\}$
1112	$\{(1,1,1,2)\}$

The property (i) of a proper matrix provides that subcubes are disjoint. The properties (ii) and (iii) characterize the location of subcubes in a cube and the size of subcubes.

Estimating the numbers of points at different levels of Boolean cube that belong to some disjoint subcubes we are able to prove that

Proposition 3. *$t(1) = 2$, $t(2) = 4$, $t(3) = 5$, $t(4) = 6$, $t(5) = 8$, $t(6) = 9$, $t(7) = 11$, $t(8) = 12$, $t(10) = 15$.*

The author is grateful to Claude Carlet, Oktay Kasim-Zadeh and Maria Fedorova for helpful discussions.

References

1. P. Camion, C. Carlet, P. Charpin, N. Sendrier, On correlation-immune functions, Advances in Cryptology: Crypto '91, Proceedings, Lecture Notes in Computer Science, V. 576, 1991, pp. 86–100.

2. Seongtaek Chee, Sangjin Lee, Daiki Lee and Soo Hak Sung, On the Correlation Immune Functions and their Nonlinearity, Advances in Cryptology - Asiacrypt '96, Lecture Notes in Computer Science, V. 1163, 1996, pp. 232–243.
3. B. Chor, O. Goldreich, J. Hastad, J. Friedman, S. Rudich, R. Smolensky, The bit extraction problem or t-resilient functions, IEEE Symposium on Foundations of Computer Science, V. 26, 1985, pp. 396–407.
4. T. W. Cusick, On constructing balanced correlation immune functions, in Sequences and Their Applications, Proceedings of SETA '98, Springer Discrete Mathematics and Theoretical Computer Science, 1999, pp. 184-190.
5. E. Filiol, C. Fontaine, Highly Nonlinear Balanced Boolean Functions with a Good Correlation Immunity, Advanced in Cryptology, Eurocrypt '98, Helsinki, Finland, Lecture Notes in Computer Sciences, Vol. 1403, 1998, pp. 475–488.
6. S. Maitra, P. Sarkar, Highly nonlinear resilient functions optimizing Siegenthaler's Inequality, Crypto '99, Lecture Notes in Computer Science, Vol. 1666, 1999, pp. 198–215.
7. P. Sarkar, S. Maitra, Construction of nonlinear Boolean functions with important cryptographic properties, In Advanced in Cryptology: Eurocrypt 2000, Lecture Notes in Computer Science, V. 1807, 2000, pp. 485–506.
8. P. Sarkar, S. Maitra, Nonlinearity bounds and constructions of resilient Boolean functions, In Advanced in Cryptology: Crypto 2000, Proceedings, Lecture Notes in Computer Science, V. 1880, 2000, pp. 515–532.
9. E. Pasalic, T. Johansson, Further results on the relation between nonlinearity and resiliency for Boolean functions, IMA Conference on Cryptography and Coding, Lecture Notes in Computer Science, Vol. 1746, 1999, pp. 35–44.
10. N. J. Patterson, D. H. Wiedemann, The covering radius of the $\left[2^{15}, 16\right]$ Reed–Muller code is at least 16276, IEEE Transactions on Information Theory, V. 29, No. 3, pp. 354–356, May 1983.
11. N. J. Patterson, D. H. Wiedemann, Correction to [10], IEEE Transactions on Information Theory, V. 36, No. 2, p. 443, March 1990.
12. O. S. Rothaus, On bent functions, Journal of Combinatorial Theory, Series A20, pp. 300–305.
13. J. Seberry, X. Zhang, Y. Zheng, On Constructions and Nonlinearity of Correlation Immune Functions, Advances in Cryptology, Eurocrypt '93, Proceedings, Lecture Notes in Computer Science, V. 765, 1993, pp. 181–199.
14. T. Siegenthaler, Correlation-immunity of nonlinear combining functions for cryptographic applications, IEEE Transactions on Information theory, V. IT-30, No 5, 1984, p. 776–780.
15. T. Siegenthaler, Decrypting a Class of Stream Ciphers Using Ciphertext Only, IEEE Transactions on Computer, V. C-34, No 1, Jan. 1985, pp. 81–85.
16. Yu. Tarannikov. On resilient Boolean functions with maximal possible nonlinearity, Cryptology ePrint archive (http://eprint.iacr.org/), Report 2000/005, March 2000, 18 pp.
17. Y. Zheng, X.-M. Zhang, Improving upper bound on nonlinearity of high order correlation immune functions, to appear in SAC 2000, Lecture Notes in Computer Science, Springer Verlag, 2000.