New Constructions of Resilient and Correlation Immune Boolean Functions achieving Upper Bounds on Nonlinearity

Enes Pasalic & Thomas Johansson
Lund University
P. O. Box 118, 221 00 Lund, SWEDEN
e-mail: {enes, thomas}@it.lth.se

Subhamoy Maitra & Palash Sarkar
Indian Statistical Institute
203, B.T. Road, Calcutta 700 035, INDIA
e-mail: {subho, palash}@isical.ac.in

Abstract

Recently weight divisibility results on resilient and correlation immune Boolean functions have received a lot of attention. These results have direct consequences towards the upper bound on nonlinearity of resilient and correlation immune Boolean functions of certain order. Now the clear benchmark in the design of resilient Boolean functions (which optimizes Siegenthaler's inequality) is to provide results which attain the upper bound on nonlinearity. Here we construct a 7-variable, 2-resilient Boolean function with nonlinearity 56. This solves the maximum nonlinearity issue for 7-variable functions with any order of resiliency. Using this 7-variable function, we also construct a 10-variable, 4-resilient Boolean function with nonlinearity 480. Construction of these two functions were justified as important open questions in Crypto 2000. Also we provide methods to generate an infinite sequence of Boolean functions on $n = 7 + 3i$ variables ($i \geq 0$) with order of resiliency $m = 2 + 2i$, algebraic degree $4 + i$ and nonlinearity $2^{n-1} - 2^{m+1}$, which were not known earlier. We conclude with a few interesting construction results on unbalanced correlation immune functions of 5 and 6 variables.

Keywords: Boolean functions, Nonlinearity, Correlation Immunity, Resiliency, Stream Ciphers.

1 Introduction

Very recently Sarkar and Maitra [16] have provided weight divisibility results on resilient Boolean functions which in turn present a nontrivial upper bound on the nonlinearity of such functions. Similar kinds of results related to weight divisibility and upper bound on nonlinearity of resilient and correlation immune Boolean functions have also been presented independently by Tarannikov [20] and Zheng and Zhang [21]. Currently Carlet [3] and Sarkar [14] have (independently and using different kinds of techniques) settled the weight divisibility results for resilient and correlation immune Boolean functions involving the algebraic degree too. Note that balanced correlation immune Boolean functions are also known as resilient Boolean functions.

These weight divisibility results have direct consequences to the upper bound on nonlinearity of these functions and a benchmark in design of such resilient Boolean functions has thus been settled. In other direction, construction of these functions achieving the upper bound on nonlinearity strengthens the tightness of the upper bound results.

In a more practical direction, these functions have immediate applications in stream cipher cryptosystems. A standard model of stream cipher [18, 19, 5] combines the outputs of several independent Linear Feedback Shift Register (LFSR) sequences using a nonlinear Boolean function to produce the keystream. This keystream is bitwise XORed with the message bitstream to produce the cipher. The decryption machinery is identical to the encryption machinery. Getting the kind of Boolean functions which we propose here provide the best possible trade-off among the parameters important to resist the known cryptanalytic techniques [19, 12, 9, 8, 10].

It is now well accepted that for a Boolean function to be used in stream cipher systems, it must satisfy the properties balancedness, high nonlinearity, high algebraic degree and high order of correlation immunity (see Section 2 for definitions). All of the above mentioned parameters are important for resisting different kinds of attacks. Also it is not possible to get the best possible values for each of these parameters separately and there are certain trade-offs involved among the above parameters. Siegenthaler showed [18] that for an n-variable function, of degree d and order of correlation immunity m, the following holds: $m + d \leq n$. Further, if the function is balanced then $m + d \leq n - 1$. Currently, the exact nature of trade-off among order of correlation immunity, nonlinearity and algebraic degree has also been investigated [15, 20, 21, 3, 14]. Earlier, a series of papers [2, 17, 4, 6, 11, 13, 15] have approached the construction problem by fixing the number of variables and the order of correlation immunity (and possibly the algebraic degree) and then trying to design balanced Boolean functions with as high nonlinearity as possible. However, the existence of the current papers [15, 20, 21, 3, 14] completely changed the motivation. Now either we have to design a function which provides the best possible trade-off among the parameters we are discussing, or we have to show that such a function cannot exist.

In this paper, for the first time we construct a 7-variable, 2-resilient Boolean function with nonlinearity 56. Earlier all the 7-variable resilient functions of different orders (except order 2) with maximum possible algebraic degree and maximum possible nonlinearity (equal to the upper bound) were known. We here close the issue by proving the case for order 2 also. Our method is basically a search technique, where we decrease the search space using different involved necessary conditions on the functions (see Section 3). We start with the table of 5-variable functions (48 different equivalence classes) provided by Berlekamp and Welch [1]. We use our necessary conditions to select very few classes out of those 48 and concatenate four 5-variable functions from those classes to get 7-variable functions.

It is known that for an n-variable m-resilient function ($m > \frac{n}{2} - 2$), the maximum possible nonlinearity is $2^{n-1} - 2^{m+1}$ and such a function must have the maximum possible algebraic degree $n - m - 1$ [16, 20, 3, 14]. In [16], the concept of saturated sequence SS for resilient Boolean functions achieving the best possible trade-off has been proposed. All the functions of SS(0) and SS(1) are already known. However, the initial functions for an SS(i) were not known earlier for $i > 1$. In fact the 7-variable, 2-resilient, nonlinearity 56 function is the initial function of the sequence SS(2). Using this 7 variable function we can construct a 10-variable, 4-resilient, nonlinearity 480 function, which was also presented as an open question in [16]. This function is the second function of SS(3). The initial function of SS(3) is a 9-variable, 3-resilient, nonlinearity 240 function, which is still an open question and we are working on it now. However, we explain how we can extend our techniques to find such a 9-variable function.

Tarannikov [20] has provided a construction technique of resilient Boolean functions with maximum possible nonlinearity. The method presents n-variable, m-resilient functions with nonlinearity $2^{n-1} - 2^{m+1}$ for $\frac{2n-7}{3} \leq m \leq n - 2$. Basically Tarannikov's construction is a recursive one and using this technique and taking n-variable, m-resilient, degree d, nonlinearity x functions one can generate $(n + 3)$-variable, $(m + 2)$-resilient, degree $(d + 1)$ and nonlinearity $2^{n+1} + 4x$ functions. These $(n+3)$-variable functions can again be used to generate $(n+6)$-variable resilient functions and so on. We here provide a much simplified modification of Tarannikov's construction, which gives the same quality results. We interpret Tarannikov's construction [20] as concatenation of Boolean functions. In Tarannikov's construction two functions are required as inputs and the functions must satisfy certain properties. Here we modify the construction so that it requires only one function in a *desired* form as input and also the resulting function becomes a *desired* one (See Subsection 3.1 for *desired form*). This construction is much easier to understand.

However, the most important thing is to get a *desired* resilient function which can be used as the initial function for this recursive construction. The 7-variable, 2-resilient function we construct here is in *desired* form. Starting with this 7-variable function as a single input and using the recursive construction proposed here, we show that for $n = 7 + 3i$, $m = 2 + 2i$, we can construct resilient functions with nonlinearity $2^{n-1} - 2^{m+1}$ and algebraic degree $n - m - 1$ (see recursive use of Construction 2 in Subsection 3.1). This improves Tarannikov's bound on m, as for $n = 7 + 3i$, it is possible to construct n-variable, m-resilient functions with nonlinearity $2^{n-1} - 2^{m+1}$ for $\frac{2n-7}{3} - 1 \leq m \leq n - 2$.

At the end we provide some important results on 5-variable and 6-variable unbalanced correlation immune functions. We present a search technique using inverse Walsh transform to get 5-variable, 2-ci (correlation immune of order 2), nonlinearity 12 functions. These functions can be used immediately to get 6-variable, 3-ci, nonlinearity 24 functions. Moreover, we run a search technique to show that 6-variable, 1-ci, nonlinearity 26 function exists. This question was open for quite sometime. Also this shows that the upper bound on nonlinearity provided by the weight divisibility results [16, 3, 14] is tight in this case.

2 Definitions and Notations

Definition 2.1 *The addition operator over $GF(2)$ is denoted by $\oplus$. For binary strings S_1, S_2 of same length λ, we denote by $\#(S_1 = S_2)$ (respectively $\#(S_1 \neq S_2)$), the number of places where S_1 and S_2 are equal (respectively unequal). The Hamming distance between S_1, S_2 is denoted by $d(S_1, S_2)$, i.e. $d(S_1, S_2) = \#(S_1 \neq S_2)$. The Walsh distance $wd(S_1, S_2)$, between S_1 and S_2, is defined as, $wd(S_1, S_2) = \#(S_1 = S_2) - \#(S_1 \neq S_2)$. Note that, $wd(S_1, S_2) = \lambda - 2\, d(S_1, S_2)$. Also the Hamming weight or simply the weight of a binary string S is the number of ones in S. This is denoted by $wt(S)$. By Ω_n we mean the set of n-variable Boolean functions. An n-variable function f is said to be balanced if its output column in the truth table contains equal number of 0's and 1's (i.e. $wt(f) = 2^{n-1}$).*

Definition 2.2 *An n-variable Boolean function $f(X_n, \ldots, X_1)$ can be considered to be a multivariate polynomial over $GF(2)$. This polynomial can be expressed as a sum of products representation of all distinct k-th order products $(0 \leq k \leq n)$ of the variables. More precisely, $f(X_n, \ldots, X_1)$ can be written as $a_0 \oplus (\bigoplus_{i=1}^{i=n} a_i X_i) \oplus (\bigoplus_{1 \leq i \neq j \leq n} a_{ij} X_i X_j) \oplus \ldots \oplus a_{12\ldots n} X_1 X_2 \ldots X_n$ where the coefficients $a_0, a_{ij}, \ldots, a_{12\ldots n} \in \{0, 1\}$. This representation of f is called the algebraic normal form (ANF) of f. The number of variables in the highest order product term with nonzero coefficient is called the algebraic degree, or simply degree of f.*

Definition 2.3 *Functions of degree at most one are called affine functions. An affine function with constant term equal to zero is called a linear function. The set of all n-variable affine (respectively linear) functions is denoted by $A(n)$ (respectively $L(n)$). The nonlinearity of an n variable function f is $nl(f) = min_{g \in A(n)}(d(f, g))$, i.e. the distance from the set of all n-variable affine functions.*

Definition 2.4 *Let $\overline{X} = (X_n, \ldots, X_1)$ and $\overline{\omega} = (\omega_n, \ldots, \omega_1)$ both belong to $\{0, 1\}^n$ and $\overline{X}.\overline{\omega} = X_n \omega_n \oplus \ldots \oplus X_1 \omega_1$. Let $f(\overline{X})$ be a Boolean function on n variables. Then the Walsh transform of $f(\overline{X})$ is a real valued function over $\{0, 1\}^n$ that can be defined as $W_f(\overline{\omega}) = \sum_{\overline{X} \in \{0,1\}^n} (-1)^{f(\overline{X}) \oplus \overline{X}.\overline{\omega}}$.*

Definition 2.5 [7] *A function* $f(X_n, \ldots, X_1)$ *is* m*-th order correlation immune (CI) iff its Walsh transform* W_f *satisfies* $W_f(\overline{\omega}) = 0$*, for* $1 \leq wt(\overline{\omega}) \leq m$*. If* f *is balanced then* $W_f(\overline{0}) = 0$*. Balanced* m*-th order correlation immune functions are called* m*-resilient functions. Thus, a function* $f(X_n, \ldots, X_1)$ *is* m*-resilient iff its Walsh transform* W_f *satisfies* $W_f(\overline{\omega}) = 0$*, for* $0 \leq wt(\overline{\omega}) \leq m$*.*

The relationship between Walsh transform and Walsh distance is [11] $W_f(\overline{\omega}) = wd(f, \bigoplus_{i=1}^{i=n} \omega_i X_i)$.

Before proceeding, we would like to introduce a few notations for future convenience. By an (n, m, d, x) function we mean an n-variable, m-resilient function with degree d and nonlinearity x. By $(n, 0, d, x)$ function we mean a balanced n-variable function with degree d and nonlinearity x. By $[n, m, d, x]$ we denote an n-variable unbalanced correlation immune function of order m, nonlinearity x and degree d. In the above notation a component is replaced by a '$-$' if we do not specify it, e.g., $(n, m, -, x)$, if we do not want to specify the degree. Further, given an affine function $l \in A(n)$, by $ndg(l)$ we denote the number of variables on which l is nondegenerate.

3 Construction of $(7, 2, 4, 56)$ Functions

From [20, 3, 14], it is clear that if an n-variable, m-resilient ($m > \frac{n}{2} - 2$) function achieves the maximum possible nonlinearity $2^{n-1} - 2^{m+1}$ then the function must have the maximum possible algebraic degree $n - m - 1$. Putting $n = 7, m = 2$, we find that maximum possible nonlinearity 56 can be obtained only when the algebraic degree is $n - m - 1 = 4$. Thus, a $(7, 2, -, 56)$ function must be a $(7, 2, 4, 56)$ function, i.e. of algebraic degree 4. Next we have the following result relating a $(7, 2, 4, 56)$ function with two $(6, 1, 4, 24)$ functions.

Proposition 3.1 *Let* f *be a* $(7, 2, 4, 56)$ *function. Then* f *can be represented as* $f = (1 \oplus X_7) f_1(X_6, \ldots, X_1) \oplus X_7 f_2(X_6, \ldots, X_1)$ *where* $f_1, f_2 \in \Omega_6$*, and both* f_1, f_2 *are* $(6, 1, 4, 24)$ *functions. That is* f *can be expressed as concatenation of* f_1 *and* f_2*.*

Proof : The function f can be represented as $f = (1 \oplus X_i) f_1(X_7, \ldots, X_{i+1}, X_{i-1}, \ldots, X_1) \oplus X_i f_2(X_7, \ldots, X_{i+1}, X_{i-1}, \ldots, X_1)$ where $f_1, f_2 \in \Omega_6$. Since degree of f is 4, we can select at least one term of degree 4 which does not contain some X_i, $1 \leq i \leq 7$. If we condition on that variable X_i, then both f_1, f_2 must be of degree 4. Thus, from Siegenthaler's inequality, f_1, f_2 can have resiliency of order at most 1. Also, since f is 2-resilient, f_1, f_2 must have resiliency of order at least 1. Thus, f_1, f_2 are 1-resilient. Since, $nl(f) = 56$, $nl(f_1), nl(f_2) \geq 56 - 32 = 24$. However, it is known [13, 16] that the maximum possible nonlinearity of 1-resilient functions on 6 variables is 24. Now, it is easy to see that if we permute the input variables, then X_i and X_7 can be interchanged. Then the output column of f in the truth table can be written as concatenation of two $(6, 1, 4, 24)$ functions. ■

The importance of this result is that if we can construct all the $(6, 1, 4, 24)$ functions, then concatenating every pair of them either we will find $(7, 2, 4, 56)$ functions or if we do not get such a function then we can guarantee that $(7, 2, 4, 56)$ function does not exist. Next we consider the results related to Walsh spectra of $(6, 1, 4, 24)$ functions f_1, f_2.

Proposition 3.2 *Let* f *be a* $(7, 2, 4, 56)$ *function made by concatenation of two* $(6, 1, 4, 24)$ *functions* f_1, f_2*. Then we have the following results. (1)* $wd(f_i, l) = 0$*, for* $ndg(l) \leq 1$*,* $1 \leq i \leq 2$*. (2)* $wd(f_i, l) = 0$ *or* ± 8*, for* $ndg(l) = 2$*,* $1 \leq i \leq 2$*. (3)* $wd(f_i, l) = \pm 16$ *iff* $wd(f_j, l) = 0$ *for* $1 \leq i \neq j \leq 2$*. (4)* $wd(f_i, l) = \pm 8$ *iff* $wd(f_j, l) = \mp 8$ *for* $1 \leq i \neq j \leq 2$*. (5). Let the number of places where* $wd(f_i, l) = \pm 16$ *is* x *and the number of places where* $wd(f_i, l) = \pm 16$ *is* y *for* $1 \leq i \neq j \leq 2$*. Then the number of places where* $wd(f, l) = \pm 16$ *is* $4x + y = 64$*.*

Proof : Item 1 follows from the definition of 1-resilient functions.

Now we prove Item 2. If $wd(f_1, l) = 16$, for $ndg(l) = 2$, then $wd(f_2, l) = -16$, since $wd(f, ll) = 0$ as f is $(7, 2, 4, 56)$ and $ndg(ll) = 2$. In that case $wd(f, ll^c) = 32$ and hence $nl(f) = 48$ which is a contradiction.

Item 3 follows similarly as otherwise nonlinearity of f will decrease.

Item 4 follows as $wd(f, l)$ must be divisible by $2^{2+2} = 16$ [16].

In item 5, the relation $4x + y = 64$ follows from the Parseval's relation on square of Walsh distances. ■

Note that it is not clear whether a $(7, 2, 4, 56)$ function f (if exists) can always be seen as concatenation of two $(6, 2, 3, 24)$ functions f_1, f_2. However, we have the following result. The proof is similar to the proof of Proposition 3.2.

Proposition 3.3 *Let f be a $(7, 2, 4, 56)$ function made by concatenation of two $(6, 2, 3, 24)$ functions f_1, f_2. Then we have the following results. (1) $wd(f_i, l) = 0$, for $ndg(l) \leq 2$, $1 \leq i \leq 2$.*
(2) $wd(f_i, l) = \pm 16$ iff $wd(f_j, l) = 0$ for $ndg(l) > 2$, $1 \leq i \neq j \leq 2$.

For a Boolean function f, we define $NZ(f) = \{\overline{\omega} \mid W_f(\overline{\omega}) \neq 0\}$, where W_f is the Walsh transform of f. If we can find two $(6, 2, 3, 24)$ functions f_1, f_2 such that $NZ(f_1) \cap NZ(f_2) = \emptyset$, then $f = (1 \oplus X_7)f_1 \oplus X_7 f_2$ will be a $(7, 2, 4, 56)$ function. Hence, if we can generate the database of all $(6, 2, 3, 24)$ functions, then we can check for concatenation of any two of them.

However, preparation of the database containing $(6, 2, 3, 24)$ and $(6, 1, 4, 24)$ functions in turn needs concatenation of 5 variable Boolean functions. For this we concentrate on the Berlekamp-Welch paper [1]. All Boolean functions on 5 variables are divided into 48 equivalence classes (see [1]), where the functions f and g are equivalent iff there exist an invertible 5×5 binary matrix M, two binary vectors a and b and a binary scalar c, such that $g(x) = f(Mx + a) + bx + c$. Note that f and g have the same algebraic degree and nonlinearity.

Proposition 3.4 *Let f be a $(7, 2, 4, 56)$ function made by concatenation of four functions h_1, h_2, h_3, h_4 on 5 variables. Then each of the h_i's are balanced and $nl(h_i) \geq 8$ for $i = 1, 2, 3, 4$.*

Now our algorithm is as follows. We denote by $\mathcal{O}(f)$ the linear transformation $f(Mx+a)+bx+c$. We take two representative Boolean functions h_1, h_2 and try all the concatenation of h_1 and $\mathcal{O}(h_2)$ (for all possible different values of M, a, b, c, M nonsingular) to generate the 6-variable functions. Then we check for $(6, 2, 3, 24)$ and $(6, 1, 4, 24)$ functions.

However, there are 21 equivalence classes with nonlinearity greater than or equal to 8. It seems computationally infeasible to generate all the 5-variable functions of those equivalence classes by necessary linear transformation and concatenate any two of them to generate 6 variable functions and then checking for $(6, 2, 3, 24)$ and $(6, 1, 4, 24)$ functions. It is clear that we can discard some of the combinations considering the algebraic degree and weight distributions of the five variable functions. Still then the search space explodes.

At this point we take a calculated risk. There are only four equivalence classes of 5-variable functions with nonlinearity 12. We concentrate on those only, leaving the other 17 equivalence classes with nonlinearity 8 and 10.

Thus we have the following strategy. Consider the following four functions [1, Table 1] presented in terms of their algebraic normal forms (ANF) representing the four equivalence classes. The weight distributions are also presented here. This means, that for the function h_1, and for $l \in L(5)$, there are 12 such l's so that $d(h_1, l) = 12, 20$, there are 16 such l's so that $d(h_1, l) = 14, 18$, and there are 4 such l's so that $d(h_1, l) = 16$.

ANF	12, 20	14, 18	16
$h_1 = X_2X_3X_4X_5 \oplus X_1X_2X_3 \oplus X_2X_4 \oplus X_3X_5$	12	16	4
$h_2 = X_1X_2X_3 \oplus X_1X_4 \oplus X_2X_5$	16	0	16
$h_3 = X_1X_2X_3 \oplus X_1X_4X_5 \oplus X_2X_3 \oplus X_2X_4 \oplus X_3X_5$	16	0	16
$h_4 = X_1X_2 \oplus X_3X_4$	16	0	16

Table 1. Representative functions of 5 variables with nonlinearity 12.

Now the algorithm for generating the database of $(6,2,3,24)$ and $(6,1,4,24)$ functions are as follows. Once again note that this database is not exhaustive as we consider the five variable functions with nonlinearity 12 only.

1. For $i = 1,2,3,4$ concatenate h_i and $\mathcal{O}(h_j)$, $j = i$ *to* 4 for all possible linear transformations of h_j (all possible options for M, a, b, c, M nonsingular). Thus their are total 10 pair of cases with h_i, h_j to be checked using all possible linear transformation for h_j.

2. Check whether the function is a $(6,1,4,24)$ function. In this case, store the function f in database if $wd(f,l) = 0$ or ± 8, for $l \in L(6)$, $ndg(l) = 2$ (see Proposition 3.2). Otherwise reject it.

3. If the function is a $(6,2,3,24)$ function, then store it in the database.

4. Reject all other functions.

Moreover, it should be considered that the search space can be further reduced keeping in mind the following constraints when we consider the algorithm.

1. The function $f = h_1\ \mathcal{O}(h_i)$ (concatenation of $h_1, \mathcal{O}(h_i)$), for $i = 2,3,4$, will not generate any $(6,1,4,24)$ or $(6,2,3,24)$ function as the algebraic degree of f becomes 5. This completely discards 3 out of 10 cases and reduces the search space to 70% of the original.

2. The function $f = h_4\ \mathcal{O}(h_4)$ will not generate any $(6,1,4,24)$ function as the algebraic degree of f becomes at most 3. Here we only need to check for $(6,2,3,24)$ functions.

We started generating the functions and storing the $(6,1,4,24)$ and $(6,2,3,24)$ functions in two separate databases. At the same time we have started concatenating any two of these six variable functions inside each of the databases separately. Note that we did not wait for generating all the functions in the two databases. We started constructing 7-variable functions as soon as the 6-variable functions were generated. We estimated that the program will run for 30 days using a 500 MHz Pentium on Linux platform. Fortunately, in between half an hour, we found one $(7,2,4,56)$ function, which is generated from the concatenation of two $(6,2,3,24)$ functions. We terminated the program then. However, getting such a function in a very short time gives the idea that there are a lot of $(7,2,4,56)$ functions available in the search space we have concentrated on. The truth table of the function is as follows. Note that $h_1, h_2, h_3.h_4$ are all 5-variable functions.

$$h_1 = 01100110001111000101101010010110, h_2 = 10011001011010011010010111000011$$
$$h_3 = 00001011011101101011010111001000, h_4 = 11011100011000101010000100011111$$

Here h_1h_2, h_3h_4 *are both* $(6,2,3,24)$ *functions. The function* $h_1h_2h_3h_4$ *is a* $(7,2,4,56)$ *function. Note that the function* $h_1h_3h_2h_4$ *is also a* $(7,2,4,56)$ *function. This can be seen as a concatenation of two* $(6,1,4,24)$ *functions* h_1h_3, h_2h_4.

From the above discussion we get the following theorem.

Theorem 3.1 *It is possible to construct a* $(7,2,4,56)$ *function.*

This completely solves the maximum nonlinearity issue of resilient functions on 7 variables. The following table shows the maximum nonlinearity corresponding to each order of resiliency and note that it is possible to construct such functions. Also it is very clear from [20, 3, 14] that all these functions possess the maximum possible algebraic degree $(6 - m)$ where m is the order of resiliency.

order of resiliency	1	2	3	4	5
maximum achievable nonlinearity	56	56	48	32	0

Table 2. Nonlinearity results for 7-variable Boolean functions.

In the next section we discuss about the recursive constructions which will generate interesting resilient functions on higher number of variables.

3.1 Recursive Construction

Here we use two general construction techniques for generating resilient functions on higher number of variables from functions on lower number of variables.

Construction 1. [18, 2, 11, 16] *Let f be an $(n, m, n-m-1, x)$ function. Let $F \in \Omega_{n+1}$ be defined as $F(Y, \overline{X}) = (1 \oplus Y)f(\overline{X}) \oplus Y(a \oplus f(\overline{X} \oplus \overline{\alpha}))$. Now, (1) either α is an all zero vector and $a = 1$ (2) or $\overline{\alpha}$ is an all one vector and $a = m \bmod 2$. Then F is an $(n+1, m+1, n-m-1, 2x)$ function.*

Next we present a modification of Tarannikov's construction [20]. In Tarannikov's construction two functions are required as inputs and the functions must satisfy certain properties. However, this is a disadvantage in certain situations. Here we modify the construction so that it requires only one function as input and also the resulting construction becomes somewhat easier to understand. In the modified construction, the input function must be of certain form for the construction to work. It is easier in general to get functions in this form than the property required in the original Tarannikov's construction [20].

We say that an $(n, m, -, -)$ function f is in the *desired* form if it is of the form $f_1 f_2$, where f_1, f_2 are $(n-1, m, -, -)$ functions. Here note that the $(7, 2, 4, 56)$ function we have found can be seen as concatenation of two $(6, 2, 4, 24)$ functions. Hence this $(7, 2, 4, 56)$ function is in *desired* form.

Construct(f) { /* Here f is an $(n, m, -, x)$ function in the *desired* form $f_1 f_2$, where f_1, f_2 are both $(n-1, m, -, -)$ functions. */
1. $F = f f^c f^c f$.
2. $g = f_1 f_1^c$ and $h = f_2 f_2^c$.
3. $G = g h h^c g^c$.
4. $F_1 = FG$.
5. Return the function F_1.
}

Note that in the language of [20], the function G above is said to depend quasilinearly on the pair of variables (X_{n+2}, X_{n+1}).

Theorem 3.2 *The function F_1 in Construct(f) is an $(n+3, m+2, -, 2^{n+1}+4x)$ function in the* **desired** *form.*

Proof : In Step 1, the function F is clearly an $(n+2, m+2, -, 4x)$ function.
Claim : In Step 2, the function G is an $(n+2, m+2, -, 4x)$ function.
Proof of Claim : Clearly both g and h are $(n, m+1, -, -)$ functions. The function gh is of the form $f_1 f_1^c f_2 f_2^c$. If we interchange the variables X_{n+1} and X_n for the function gh, we get a function in the form $f_1 f_2 f_1^c f_2^c$ which is actually $f f^c$. Hence we have $nl(gh) = 2nl(f) = 2x$. Let $\Lambda \in L(n+2)$. We can write Λ in one of the forms $llll, ll^c ll^c, lll^c l^c, ll^c l^c l$. We compute
1. $wd(G, llll) = wd(ghh^c g^c, llll) = wd(hh^c, ll) + wd(gg^c, ll) = 0 + 0 = 0$.
2. $wd(G, ll^c ll^c) = wd(ghh^c g^c, ll^c ll^c) = 2wd(gh, ll^c)$.
3. $wd(G, lll^c l^c) = wd(ghh^c g^c, lll^c l^c) = 2wd(gh, ll)$.

4. $wd(G, ll^cl^cl) = wd(ghh^cg^c, ll^cl^cl) = 0$.
If Λ is nondegenerate on at most $(m+2)$ variables, then l is nondegenerate on at most $(m+1)$ variables. Hence $wd(g,l) = wd(h,l) = wd(h,l^c) = 0$ and so $wd(G,\Lambda) = 0$. Thus G is $(m+2)$-resilient. Further, by the above calculation we have $nl(G) = 2nl(gh) = 4nl(f) = 4x$. This completes the proof of the claim.

Since F and G are both $(n+2, m+2, -, 4x)$ functions, the function F_1 is clearly an $(n+3, m+2, -, y)$ function in the *desired* form. Thus it is sufficient to show that $y = 2^{n+1} + 4x$. This is proved by showing that $NZ(F) \cap NZ(G) = \emptyset$. Let Λ be in $L(n+2)$, such that $wd(F,\Lambda) \neq 0$. Then clearly Λ is of the form ll^cl^cl. Thus it is enough to show that for any such Λ, $wd(G,\Lambda) = 0$. But this is what has been shown in item 4 above. ■

Starting with a function f in the *desired* form $Construct(f)$ is repeatedly used in the manner *while (true)* $\{f = Construct(f)\}$. Thus we summarize the construction as follows.

Construction 2. *Let f be an $(n, m, n-m-1, x = 2^{n-1} - 2^{m+1})$ function in* **desired** *form and $f = g_1g_2$, where $g_1, g_2 \in \Omega_{n-1}$. Let $F = g_1g_2g_1^cg_2^cg_1^cg_2^cg_1g_2g_1g_1^cg_2g_2^cg_2^cg_2g_1^cg_1$. Then F is an $(n+3, m+2, n-m, 4x + 2^{n+1} = 2^{n+2} - 2^{m+3})$ function in* **desired** *form.*

In [16], an $(n, m, n-m-1, 2^{n-1} - 2^{m+1})$ function is called a *saturated maximum degree* function and its spectrum is three valued. For such a function we must necessarily have $m > \lfloor \frac{n}{2} \rfloor - 2$. From this a notion of a sequence of Boolean functions, each of which is a saturated maximum degree function with maximum possible nonlinearity was proposed in [16].

Definition 3.1 *For $i \geq 0$ we define SS(i) as follows. An SS(0) is a sequence $f_{0,0}, f_{0,1}, \ldots$, where $f_{0,0}$ is a $(3,0,2,2)$ function and $f_{0,j}$ is a $(3+j, j, 2, 2^{j+1})$ function for $j > 0$. For $i > 0$, an SS(i) is a sequence $f_{i,0}, f_{i,1}, \ldots$, where $f_{i,0}$ is a $(3+2i, i, 2+i, 2^{2+2i} - 2^{1+i})$ function. Also for $j > 0$, $f_{i,j}$ is a $(3+2i+j, i+j, 2+i, 2^{2+2i+j} - 2^{1+i+j})$ function.*

Note that all functions in an SS(t) have the same degree $2+t$. It is also important to see that given a function f of some SS(t), Construction 1 generates all the consecutive functions of SS(t). However, given a function f of some SS(t), Construction 2 generates one function each of SS$(t+p)$ for all $p \geq 1$. That is Construction 2 generates functions in different saturated sequences.

Construction of SS(0) and SS(1) are already known [16]. The initial functions for an SS(i) for $i > 1$ were not known earlier. Here the $(7,2,4,56)$ function is the initial function of SS(2). Thus, from this function, using the Construction 1, one can generate all the functions of SS(2). Note that this sequence was earlier known from the 2nd function onwards, where the second function of SS(2) is an $(8,3,4,112)$ function.

Starting from $(7,2,4,56)$ function (which is in *desired* form), Construction 2 generates one function each of SS$(2+p)$ (which is again in *desired* form) for all $p \geq 1$. These are basically $(p+1)$th functions of SS$(2+p)$. These functions were not known earlier. For $p = 1$, we get the $(10,4,5,480)$ function, which was posed as an open problem in [16]. Construction 2, when used recursively, generates an infinite sequence of Boolean functions on $n = 7+3p$ variables $(p \geq 0)$ with order of resiliency $m = 2+2p$, algebraic degree $4+p$ and nonlinearity $2^{n-1} - 2^{m+1}$, which were not known earlier. These functions are $(7,2,4,56), (10,4,5,480), (13,6,6,3968), (16,8,7,32256), \ldots$ and so on. Note that all these functions with moderate number of input variables have immediate use in stream cipher systems as combining Boolean functions.

Tarannikov [20] has presented construction of n-variable, m-resilient functions with nonlinearity $2^{n-1} - 2^{m+1}$ for $\frac{2n-7}{3} \leq m \leq n-2$. Our 7-variable function can be taken as an initial function when Construction 2 will be used recursively. This gives that for $n = 7+3p$, $m = 2+2p$, we can construct resilient functions with nonlinearity $2^{n-1} - 2^{m+1}$ and algebraic degree $n-m-1$. This improves Tarannikov's bound on m as for $n = 7+3p$, it is possible to construct n-variable, m-resilient functions with nonlinearity $2^{n-1} - 2^{m+1}$ for $\frac{2n-7}{3} - 1 \leq m \leq n-2$.

Now we mention the issue regarding the 9-variable resilient functions. We concentrate on the initial function of SS(3), the $(9,3,5,240)$ function. The existence of this function is not yet known. We are currently searching this function in the following manner using computer program.
1. Construct an $(8,3,4,112)$ function $f_1 = hh^c$, where h is a $(7,2,4,56)$ function. That is we apply the Construction 1 here.
2. Construct an $(8,3,4,112)$ function f_2 using Construction 2 on a function g which is a $(5,1,3,12)$ function.
3. Construct $F = f_1 f_2$.
4. If $NZ(f_1) \cap NZ(f_2) = \emptyset$, then F will be a $(9,3,5,240)$ function.

In this situation we can update the table of interesting Boolean functions on small number of variables than what presented in [16]. The * marked entries are the functions which we construct here.

n	
7	(7, 2, 4, 56)*
8	(8, 1, −, 116)
9	(9, 1, −, 244), (9, 2, 6, 240), (9, 3, 5, 240)
10	(10, 1, −, 492), (10, 1, −, 488), (10, 2, −, 488), (10, 4, 5, 480)*

Table 3. Construction of these functions were posed as open question in [16].

Also it is important to note that using the weight divisibility results of resilient functions involving the algebraic degree [3, 14], it can be shown that the $(8,1,-,116),(10,1,-,492),(10,2,-,488)$ functions, if at all exist, must be $(8,1,6,116),(10,1,8,492),(10,2,7,488)$ functions.

4 Correlation Immune Functions on 5 and 6 variables

In this section we will particularly consider the $[5,2,3,12],[6,3,3,24]$ and $[6,1,5,26]$ functions. These functions provide the best possible trade-off among the parameters order of correlation immunity, nonlinearity and algebraic degree.

First we will show the construction of $[5,2,3,12],[6,3,3,24]$ functions. Our technique is completely new which is based on spectral analysis of such functions. The $[5,2,3,12],[6,3,3,24]$ functions have also been considered earlier in [20]. However, there only one example of $[6,3,3,24]$ function has been given and a $[5,2,3,12]$ function has been derived from the 6-variable one. We will provide a systematic construction technique of $[5,2,3,12],[6,3,3,24]$ functions.

Next we provide a construction to show that the $[6,1,5,26]$ function exists. It was the last important question that was unanswered for 6-variable correlation immune functions. This result also shows that the upper bound on nonlinearity of correlation immune functions [16] is tight in this case.

4.1 $[5,2,3,12],[6,3,3,24]$ Functions

Here we first provide a construction method for $[5,2,3,12]$ functions. The necessary conditions for the existence of a $[5,2,3,12]$ functions are best viewed in terms of the Walsh spectra of the function. Clearly, the function F must have at least 15 zeros in its Walsh spectra since $W_F(\overline{\omega}) = 0$ for $wt(\overline{\omega}) = 1,2$. Note that $\overline{\omega} \in \{0,1\}^5$. Furthermore, the maximum absolute value in the Walsh spectra must be equal to 8. There are just 2 equivalence classes in [1] satisfying these criteria and with algebraic degree 3. These are h_2, h_3 in Table 1. Each of these 2 classes have the same distribution of Walsh coefficients up to complementation, that is the spectra will always be given either as $W_{F_1}(\overline{\omega})$ or $W_{F_2}(\overline{\omega})$, which is shown in the table below. Note that the spectra given by $W_{F_1}(\overline{\omega})$ is the inversion of $W_{F_2}(\overline{\omega})$ which corresponds to the complementation of the Boolean function's truth table.

$W_{F_1}(\overline{\omega})$	$\#\{W_{F_1}(\overline{\omega})\}$	$W_{F_2}(\overline{\omega})$	$\#\{W_{F_2}(\overline{\omega})\}$
0	16	0	16
8	10	8	6
-8	6	-8	10

Table 4. Walsh transform values.

We consider the case where the Walsh spectra contains 16 zero values, 10 values of $+8$ and six values of -8 as in W_{F_1}. We fix 15 zero values in the spectra at the places where $wt(\overline{\omega}) = 1, 2$. Also we fix $W_F(\overline{0}) = -8$. Now there are 16 places left and we can place another zero at any one of these places in 16 ways. Now we have 15 more places left and we choose 5 places and put -8, the remaining 10 places are filled with $+8$. Similarly we have to check for W_{F_2}. Thus the total number of cases we check is N_{per}, will be $N_{per} = 2 \cdot 16 \cdot \binom{15}{5}$. Thus, instead of doing an exhaustive search for $[5, 2, 3, 12]$ functions we simply create all permitted permutations, N_{per} in number, and check if the Walsh spectra is a valid spectra of a Boolean function. If the spectra is valid, then we get the Boolean function back using inverse Walsh transform and the function must be correlation immune of order 2. We obtain in total 384 distinct $[5, 2, 3, 12]$ functions in this way.

Next we provide the following result.

Proposition 4.1 *Let F be an $[n, m, d, x]$ function where m is even. Then FF^r (concatenation of the truth table of the function F and its reverse) will be an $[n + 1, m + 1, d, 2x]$ function.*

Proof : It is very clear that FF^r is an $[n+1, m, d, 2x]$ function. That it is also correlation immune of order $m + 1$ follows from the result that for any linear function l which is nondegenerate on odd number of variables, $l^r = l^c$. Here $m + 1$ is odd. ■

Hence given a $[5, 2, 3, 12]$ function F, we can get a $[6, 3, 3, 24]$ function FF^r.

4.2 $[6, 1, 5, 26]$ Function

In this section we provide a construction of $[6, 1, 5, 26]$ function. First we need the following result.

Proposition 4.2 *Any $[6, 1, -, 26]$ function must be a $[6, 1, 5, 26]$ function.*

Proof : The maximum nonlinearity of a six variable function is 28 (bent function). Consider an 1-ci function F on 6 variables. A correlation immune function F can not be bent. Hence, $nl(F) < 28$. It is known [14] that for an n variable, m-ci, degree d function f, the weights of $f \oplus l$, ($l \in L(n)$), are always divisible by $2^{m+\lfloor \frac{n-m-1}{d} \rfloor}$. Here $n = 6, m = 1$ and hence $nl(F) \leq 28 - 2^{1+\lfloor \frac{4}{d} \rfloor}$. Note that, if F has nonlinearity 26, then d must be 5. If $d < 5$, then nonlinearity of F will decrease further, and it can be at most 24. Thus, F must be of degree 5. ■

Lemma 4.1 *Let if possible F be a $[6, 1, 5, 26]$ function. Then it is possible to write $F = (1 \oplus X_6)f_1 \oplus X_6 f_2$, where f_1 and f_2 are 5-variable functions each having nonlinearity 11 and degree 5.*

Proof : The degree of F is 5. Without loss of generality we consider $X_5 \ldots X_1$ is a degree 5 term in the ANF of F. This is because we can permute the input variables to do this. We put $f_1(X_5, \ldots, X_1) = F(X_6 = 0, X_5, \ldots, X_1)$ and $f_2(X_5, \ldots, X_1) = F(X_6 = 1, X_5, \ldots, X_1)$. Thus both

f_1, f_2 are of degree 5 and hence of odd weight and so $nl(f_1), nl(f_2) \leq 11$. It can be proved that if any of $nl(f_1)$ or $nl(f_2)$ is < 11, then $nl(F) < 26$. ■

The importance of this result is that, given a $[6,1,5,26]$ function F, we can always permute the input variables so that the term $X_5 \dots X_1$ of degree 5 stays in F and then we can consider it as concatenation of two 5-variable functions with nonlinearity 11. Hence if we concatenate all pairs of such 5-variable functions, either we will find a $[6,1,5,26]$ function, or we can conclude that such a function does not exist. A closer inspection of the Berlekamp's paper leaves just 4 out of 49 equivalence classes with nonlinearity 11 whose functions may be concatenated to possibly obtain a $[6,1,5,26]$ function. The following table provides the algebraic normal form of the representative functions with their weight distribution. This means, that for the function h_1, and for $l \in L(5)$, there are 6 such l's so that $d(h_1, l) = 11, 21$, there are 10 such l's so that $d(h_1, l) = 13, 19$, and there are 16 such l's so that $d(h_1, l) = 15, 17$. Note that $A = X_1X_2X_3X_4X_5$.

ANF	11, 21	13, 19	15,17
$h_1 = A \oplus X_1X_2 \oplus X_3X_4$	6	10	16
$h_2 = A \oplus X_1X_2X_3 \oplus X_1X_4 \oplus X_2X_5$	6	10	16
$h_3 = A \oplus X_1X_2X_3 \oplus X_1X_4X_5 \oplus X_3X_5 \oplus X_2X_4 \oplus X_2X_3$	6	10	16
$h_4 = A \oplus X_1X_2X_3 \oplus X_1X_4X_5 \oplus X_4X_5 \oplus X_3X_5 \oplus X_2X_4 \oplus X_2X_3$	4	16	12

Table 5. Representative functions of 5 variables with nonlinearity 11.

Thus, concatenating each representative of the i-th class, $i = 1, \dots, 4$, with all nonsingular affine transformations applied to a representant of the j-th class, $j = i, \dots, 4$ and finally checking the possibility of obtaining a correlation immune function with nonlinearity 26 will answer the question if there exists a $[6,1,5,26]$ function. *We ran this computer program and found a* $[6,1,5,26]$ *function. The function is the concatenation of* h_1 *and* h_2 *where* $h_1 =$ 01001001110100001010100101000101 *and* $h_2 =$ 10000010101001110011100110000001 *are both of nonlinearity 11.*

Theorem 4.1 *It is possible to construct a* $[6,1,5,26]$ *function.*

References

[1] E. R. Berlekamp and L. R. Welch. Weight distributions of the cosets of the $(32,6)$ Reed-Muller code. *IEEE Transactions on Information Theory*, IT-18(1):203–207, January 1972.

[2] P. Camion, C. Carlet, P. Charpin, and N. Sendrier. On correlation immune functions. In *Advances in Cryptology - CRYPTO'91*, pages 86–100. Springer-Verlag, 1992.

[3] C. Carlet. On the coset weight divisibility and nonlinearity of resilient and correlation immune functions. *Preprint*, 2000.

[4] S. Chee, S. Lee, D. Lee, and S. H. Sung. On the correlation immune functions and their nonlinearity. In *Advances in Cryptology, Asiacrypt 96*, number 1163 in Lecture Notes in Computer Science, pages 232–243. Springer-Verlag, 1996.

[5] C. Ding, G. Xiao, and W. Shan. *The Stability Theory of Stream Ciphers.* Number 561 in Lecture Notes in Computer Science. Springer-Verlag, 1991.

[6] E. Filiol and C. Fontaine. Highly nonlinear balanced Boolean functions with a good correlation-immunity. In *Advances in Cryptology - EUROCRYPT'98*. Springer-Verlag, 1998.

[7] X. Guo-Zhen and J. Massey. A spectral characterization of correlation immune combining functions. *IEEE Transactions on Information Theory*, 34(3):569–571, May 1988.

[8] T. Johansson and F. Jonsson. Fast correlation attacks based on turbo code techniques. In *Advances in Cryptology - CRYPTO'99*, number 1666 in Lecture Notes in Computer Science, pages 181–197. Springer-Verlag, August 1999.

[9] T. Johansson and F. Jonsson. Improved fast correlation attacks on stream ciphers via convolutional codes. In *Advances in Cryptology - EUROCRYPT'99*, number 1592 in Lecture Notes in Computer Science, pages 347–362. Springer-Verlag, May 1999.

[10] T. Johansson and F. Jonsson. Fast correlation attacks through reconstruction of linear polynomials. In *Advances in Cryptology - CRYPTO 2000*, number 1880 in Lecture Notes in Computer Science, pages 300–315. Springer Verlag, 2000.

[11] S. Maitra and P. Sarkar. Highly nonlinear resilient functions optimizing Siegenthaler's inequality. In *Advances in Cryptology - CRYPTO'99*, number 1666 in Lecture Notes in Computer Science, pages 198–215. Springer Verlag, August 1999.

[12] W. Meier and O. Staffelbach. Fast correlation attack on stream ciphers. In *Advances in Cryptology - EUROCRYPT'88*, volume 330, pages 301–314. Springer-Verlag, May 1988.

[13] E. Pasalic and T. Johansson. Further results on the relation between nonlinearity and resiliency of Boolean functions. In *IMA Conference on Cryptography and Coding*, number 1746 in Lecture Notes in Computer Science, pages 35–45. Springer-Verlag, 1999.

[14] P. Sarkar. Spectral domain analysis of correlation immune and resilient boolean functions. *Preprint*, 2000.

[15] P. Sarkar and S. Maitra. Construction of nonlinear Boolean functions with important cryptographic properties. In *Advances in Cryptology - EUROCRYPT 2000*, number 1807 in Lecture Notes in Computer Science, pages 485–506. Springer Verlag, 2000.

[16] P. Sarkar and S. Maitra. Nonlinearity bounds and constructions of resilient boolean functions. In *Advances in Cryptology - CRYPTO 2000*, number 1880 in Lecture Notes in Computer Science, pages 515–532. Springer Verlag, 2000.

[17] J. Seberry, X. M. Zhang, and Y. Zheng. On constructions and nonlinearity of correlation immune Boolean functions. In *Advances in Cryptology - EUROCRYPT'93*, pages 181–199. Springer-Verlag, 1994.

[18] T. Siegenthaler. Correlation-immunity of nonlinear combining functions for cryptographic applications. *IEEE Transactions on Information Theory*, IT-30(5):776–780, September 1984.

[19] T. Siegenthaler. Decrypting a class of stream ciphers using ciphertext only. *IEEE Transactions on Computers*, C-34(1):81–85, January 1985.

[20] Y. V. Tarannikov. On resilient Boolean functions with maximum possible nonlinearity. *Cryptology ePrint Archive, eprint.iacr.org, No. 2000/005*, 2000.

[21] Y. Zheng and X. M. Zhang. Improving upper bound on nonlinearity of high order correlation immune functions. In *SAC 2000*, Lecture Notes in Computer Science (to be published). Springer Verlag, 2000.