*There exist Boolean functions on n (odd) variables having nonlinearity $> 2^{n-1} - 2^{\frac{n-1}{2}}$ if and only if $n > 7$

Selçuk Kavut†, Subhamoy Maitra‡ and Melek D. Yücel†

Abstract

For the first time we find a Boolean function on 9 variables having nonlinearity 241. Such functions are discovered using a suitably modified steepest-descent based iterative heuristic search in the class of rotation symmetric Boolean functions. This shows that there exist Boolean functions on n (odd) variables having nonlinearity $> 2^{n-1} - 2^{\frac{n-1}{2}}$ if and only if $n > 7$.

Keywords: Boolean Functions, Combinatorial Problems, Cryptography, Heuristic Search, Nonlinearity, Rotational Symmetry.

1 Introduction

Boolean functions with very high nonlinearity is one of the most challenging problems in the area of cryptography and combinatorics. The problem is also related to covering radius of first order Reed-Muller code. On even number of variables n, there maximum possible nonlinearity $2^{n-1} - 2^{\frac{n}{2}-1}$ is attained for the well known bent functions [2, 5]. However, for the case when n is odd, the situation is more complicated and very few results are available since 1972 as follows.

1. **Negative results.** In 1972 [1], it has been shown that the maximum nonlinearity of 5-variable Boolean functions is 12 and in 1980 [3] it has been shown that the maximum nonlinearity of 7-variable Boolean functions is 56. Thus for odd $n \leq 7$, the maximum nonlinearity of n-variable functions is $\leq 2^{n-1} - 2^{\frac{n-1}{2}}$.

*This is a working draft only to announce the result. We will come up with a detailed draft soon with more details.

†Department of Electrical and Electronics Engineering, Middle East Technical University – ODTÜ, 06531 Ankara, Turkey. Email: kavut@metu.edu.tr, yucel@eee.metu.edu.tr

‡Applied Statistics Unit, Indian Statistical Institute, 203 B T Road, Kolkata 700 108, India. Email: subho@isical.ac.in

2. **Positive results.** In 1983 [4], it has been shown that one can get Boolean functions on 15 variables having nonlinearity 16276 and using this result one can show that for odd $n \geq 15$, it is possible to get Boolean functions having nonlinearity $2^{n-1} - 2^{\frac{n-1}{2}} + 20 \cdot 2^{\frac{n-15}{2}}$.

The question for $n = 9, 11, 13$ stayed completely open; the maximum nonlinearity known for these cases was $2^{n-1} - 2^{\frac{n-1}{2}}$ and there was no proof or evidence (till this work) whether there are functions having nonlinearity strictly greater than that.

A Boolean function on n variables may be viewed as a mapping from $V_n = \{0,1\}^n$ into $\{0,1\}$. The *truth table* of a Boolean function $f(x_1, \ldots, x_n)$ is a binary string of length 2^n, $f = [f(0,0,\cdots,0),\ f(1,0,\cdots,0),\ f(0,1,\cdots,0),\ \ldots, f(1,1,\cdots,1)]$. The *Hamming weight* of a binary string S is the number of 1's in S denoted by $wt(S)$. An n-variable function f is said to be *balanced* if its truth table contains an equal number of 0's and 1's, i.e., $wt(f) = 2^{n-1}$. Also, the *Hamming distance* between equidimensional binary strings S_1 and S_2 is defined by $d(S_1, S_2) = wt(S_1 \oplus S_2)$, where $\oplus$ denotes the addition over $GF(2)$.

An n-variable Boolean function $f(x_1, \ldots, x_n)$ 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, $f(x_1, \ldots, x_n)$ can be written as

$$a_0 \oplus \bigoplus_{1 \leq i \leq n} a_i x_i \oplus \bigoplus_{1 \leq i < j \leq n} a_{ij} x_i x_j \oplus \ldots \oplus a_{12\ldots n} x_1 x_2 \ldots x_n,$$

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 the degree of f and denoted by $deg(f)$.

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.

Let $x = (x_1, \ldots, x_n)$ and $\omega = (\omega_1, \ldots, \omega_n)$ both belonging to $\{0,1\}^n$ and $x \cdot \omega = x_1\omega_1 \oplus \ldots \oplus x_n\omega_n$. Let $f(x)$ be a Boolean function on n variables. Then the *Walsh transform* of $f(x)$ is a real valued function over $\{0,1\}^n$ which is defined as

$$W_f(\omega) = \sum_{x \in \{0,1\}^n} (-1)^{f(x) \oplus x \cdot \omega}.$$

In terms of Walsh spectrum, the nonlinearity of f is given by

$$nl(f) = 2^{n-1} - \frac{1}{2} \max_{\omega \in \{0,1\}^n} |W_f(\omega)|.$$

1.1 Rotation Symmetric Boolean Functions

Let $x_i \in \{0,1\}$ for $1 \leq i \leq n$. For $1 \leq k \leq n$, we define

$$\begin{aligned} \rho_n^k(x_i) &= x_{i+k}, \quad \text{if } i+k \leq n, \text{ and} \\ &= x_{i+k-n}, \quad \text{if } i+k > n. \end{aligned}$$

Let $(x_1, x_2, \ldots, x_{n-1}, x_n) \in V_n$. We can extend the definition of ρ_n^k to n-tuples as

$$\rho_n^k(x_1, x_2, \ldots, x_n) = (\rho_n^k(x_1), \rho_n^k(x_2), \ldots, \rho_n^k(x_n)).$$

Definition 1 *A Boolean function f is called* Rotation Symmetric *if for each input $(x_1, \ldots, x_n) \in \{0,1\}^n$, $f(\rho_n^k(x_1, \ldots, x_n)) = f(x_1, \ldots, x_n)$ for $1 \leq k \leq n$.*

Consider the set of vectors

$$G_n(x_1, \ldots, x_n) = \{\rho_n^k(x_1, \ldots, x_n), \text{ for } 1 \leq k \leq n\}.$$

Note that $G_n(x_1, \ldots, x_n)$ generates an orbit in the set V_n. Let g_n be the number of such orbits. Using Burnside's lemma, it can be shown that

$$g_n = \frac{1}{n} \sum_{k|n} \phi(k)\, 2^{\frac{n}{k}},$$

ϕ being Euler's *phi*−function. It can be easily checked that $g_n \approx \frac{2^n}{n}$. Since $2^{g_n} << 2^{2^n}$, the number of n-variable RSBFs is much smaller than the total space of Boolean functions. Thus attempting a search in this space is more encouraging than attempting a search in the complete space of Boolean functions.

2 Search Strategy

Our search strategy uses a steepest-descent like iterative algorithm, where each iteration step has the input Boolean function f and the output Boolean function f_{min}. At each iteration step, a cost function is calculated within a pre-defined neighborhood of f and the Boolean function having the smallest cost is chosen as the iteration output f_{min}. In some rare cases, the cost of f_{min} may be larger than or equal to the cost of f. This is the crucial part of the search strategy, which provides the ability to escape from local minima and its distinction from the steepest-descent algorithm. Our steepest-descent based search technique minimizes the cost until a local minimum is attained, then it takes a step in the direction of non-decreasing cost. That is, whenever possible, the cost is minimized; otherwise, a step in the reverse direction is taken. The deterministic step in the reverse direction corresponds to the smallest possible cost increase within the pre-defined neighborhood of the preceding Boolean function, which also makes it possible to escape from the local minima. The following is the truth table of a 9-variable function $f(x_1, \ldots, x_9)$ having nonlinearity $2^{9-1} - 2^{\frac{9-1}{2}} + 1 = 241$.

```
10010111011111110011111111111010000011101111101010101110110010 01
01010101111110001111101011001101110011001010100110100000100000 11
01110110011001101110101111000000111110101000100011100000101100 11
11110100111000001000100110000011110010000100010110010001010111 10
01111111011111000010110000101001111111001100101110100001000000 01
11101010100110001100000010000101111010000001000110001011010111 10
11111110001000011110100100010001100001001000001110000101000111 10
11100001100101010010000100110110100101110001011001110110111010 01
```

Thus it is clear that the function $g(y_1, y_2) \oplus f(x_1, \ldots, x_9)$ is an 11-variable function with nonlinearity $2^{11-1} - 2^{\frac{11-1}{2}} + 2 = 994$ where $g(y_1, y_2)$ is a 2-variable bent function. Similarly $h(y_1, y_2, y_3, y_4) \oplus f(x_1, \ldots, x_9)$ is a 13-variable function with nonlinearity $2^{13-1} - 2^{\frac{13-1}{2}} + 4 = 4036$ where $h(y_1, y_2, y_3, y_4)$ is a 4-variable bent function.

This proves the following result.

Theorem 1 *there exist Boolean functions on n (odd) variables having nonlinearity $> 2^{n-1} - 2^{\frac{n-1}{2}}$ if and only if $n > 7$.*

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] J. F. Dillon. Elementary Hadamard Difference sets. PhD Thesis, University of Maryland, 1974.

[3] J. J. Mykkeltveit. The covering radius of the $(128, 8)$ Reed-Muller code is 56. *IEEE Transactions on Information Theory*, IT-26(3):359–362, 1980.

[4] N. J. Patterson and D. H. Wiedemann. The covering radius of the $(2^{15}, 16)$ Reed-Muller code is at least 16276. *IEEE Transactions on Information Theory*, IT-29(3):354–356, 1983. See correction in IT-36(2):443, 1990.

[5] O. S. Rothaus. On bent functions. *Journal of Combinatorial Theory, Series A*, 20:300–305, 1976.