

# CLASSIFICATION OF CUBIC BOOLEAN FUNCTIONS IN 7 VARIABLES

An Braeken, Svetla Nikova

Department Electrical Engineering, ESAT/COSIC,  
Katholieke Universiteit Leuven, Kasteelpark Arenberg 10,  
B-3001 Heverlee-Leuven, Belgium  
`an.braeken,svetla.nikova@kuleuven.ac.be`

Yuri Borissov

Institute of Mathematics and Informatics,  
Bulgarian Academy of Sciences,  
8 G.Bonchev, 1113 Sofia, Bulgaria  
`yborisov@moi.math.bas.bg`

*Based on a classification of invariants, we show how to efficiently classify Boolean functions in 6 and 7 variables of degree 3. The sizes of the orbits for every class is determined in the same time.*

## INTRODUCTION

A well-known and widely used approach in the study of algebraic objects is the investigation of their sub-objects and quotient objects. Consider the set of Boolean functions on  $\mathbb{F}_2^n$  of degree less than or equal to  $r$  for  $0 \leq r \leq n$ , which can also be seen as the Reed-Muller code  $RM(r, n)$  of order  $r$ . The automorphism group of  $RM(r, n)$  is equal to the general affine group  $AGL(n, 2)$  for all  $1 \leq r \leq n$  [MS91, Theorem 24]. For  $-1 \leq s < r \leq n$ , the quotient space of  $RM(r, n)$  by the sub-code  $RM(s, n)$  is denoted by  $RM(r, n)/RM(s, n)$ . Consequently, two Boolean functions  $f_1, f_2$  of  $RM(r, n)/RM(s, n)$  are said to be equivalent over  $RM(s, n)$  if  $f_1(\bar{x}) = f_2(\bar{x}A \oplus \bar{b}) \bmod RM(s, n)$ . If  $s = 1$ , this means that

$$f_1(\bar{x}) = f_2(\bar{x}A \oplus \bar{a}) \oplus \bar{x} \cdot \bar{B} \oplus b, \quad \forall x \in \mathbb{F}_2^n, \quad (1)$$

where  $A$  is a nonsingular binary  $n \times n$ -matrix,  $b$  is a binary constant, and  $\bar{a}, \bar{B}$  are  $n$ -dimensional binary vectors. If  $s = 0$ , then  $\bar{B}$  is zero and the functions  $f_1$  and  $f_2$  are said to be equivalent over  $RM(0, n)$ . If  $\bar{B}, b$  are 0, the functions are equivalent over  $RM(-1, n)$ . A property is called  $RM(s, n)$  invariant property if it is invariant over  $RM(s, n)$ . For  $s = 1$ , it is called affine equivalence. Note that

on  $RM(r, n)/RM(r - 1, n)$  for  $0 \leq r \leq n$ , the action of  $AGL(n, 2)$  is reduced to the action of the general linear group  $GL(n, 2)$ .

An overview of known and new invariant properties over  $RM(s, n)$  for  $s = 1, 0, -1$  is presented. Based on these invariants, we show how efficiently to classify the Boolean functions of 6 variables and 7 variables of degree 3. The advantage of this classification is that in the same time the sizes of the orbits for each class are computed. This number determines the density of that class and can be used in order to compute the number of Boolean functions which are correlation-immune, resilient, or satisfy a certain order of propagation characteristics following the method presented in [BBNP04].

## Invariant Properties

Equivalence classes provide a powerful tool in both the construction and analysis of Boolean functions for cryptography. In particular, rather than considering the entire space of  $2^{2^n}$  functions a reduced view can be found in the consideration of only one function from each class. We start by giving an overview of the invariant properties over  $RM(s, n)$  for  $s = 1, 0, -1$ . The computational complexity of determining if the Boolean function satisfies the property is an important measure for the efficiency of the property.

### Invariant Properties over $RM(1, n)$

A well-known invariant property is the distribution of the absolute Walsh and autocorrelation values which is denoted as a set of tuples in which the first element represents the absolute value in the spectrum and the second element the number of occurrence. Consequently, also other properties which are derived from the Walsh and autocorrelation spectrum like nonlinearity, dimension of the linear space, GAC indicators, distance to functions with non-zero linear structures [MS89], etc., are invariant properties. Relations between the Walsh and autocorrelation values for two equivalent functions are derived in [Pre93]. The Walsh and autocorrelation spectrum can be determined with complexity  $O(n2^n)$  using Fast Fourier Transform [CT93].

In [LZ96, LVi98], the invariants stabilisator, index and height are introduced. The stabilisator is defined as the order of the group of linear (affine) transformations that keeps the function invariant. The dimension of the largest flat on which  $f$  is affine is called the index. The height of a function is defined as the smallest

dimension of a flat  $V$  for which  $D_V f = 0$ . The properties height and index can be generalized by the number of flats for which the corresponding property is satisfied. Related to the index of a function is the property of weakly normality which was proven to be affine invariant property in [Dub01]. These invariants have very high computational complexity (almost equal to the order of the group of linear (affine) transformations for the stabilisator and the number of flats of certain dimension for the height and index). Consequently, they cannot be used efficiently for determining inequivalence for higher dimensions (i.e.  $n \geq 13..15$ ).

Dillon proved in [Dil74] the invariance of the set of derivatives, *i.e.*, the sets of all  $k$ -dimensional derivatives are invariant for any  $1 \leq k \leq n$ . Consequently, this explains why the height of a Boolean function is an invariant.

In [BL03], the relation between restriction and derivation was noticed. Moreover, it was shown that the set of restrictions with respect to hyperplanes is an invariant. More generally, the set of restrictions with respect to subspaces of dimension  $k$  with  $1 \leq k \leq n$  is invariant.

Fuller and Millan introduced in [FM03] the concept of connection functions, which are functions that differ from the original function with distance one. The set of  $2^n$  connection functions are invariant. Moreover, this concept can be generalized to connection functions of any distance  $1 \leq d \leq 2^n - 1$  (see [MFD04]).

The weight distribution of the minimal words of  $f \oplus g$  for all  $g \in RM(1, n)$  was noticed to be an invariant property in [BM04]. Here a minimal word is a codeword whose support does not contain the support of another codeword. Properties on minimal codewords are derived in [AB98]. The complexity of this invariant can be approximated by  $O(2^{3n})$ . However, this can only be used as distinguisher for Boolean functions of low degree with respect to the number of variables, since the weight of a codeword should be less than  $2^n - \sum_{i=1}^d \binom{n}{i} + 1$  in order to be a minimal word.

The number of bases in the set of vectors with the same absolute value in the Walsh and in the autocorrelation spectrum is shown to be an affine invariant property in [BBNP04]. The proof follows immediately from the fact that between the sets of bases there exists a bijective correspondence. The complexity of the invariant mainly depends on the spectra and is different for each Boolean function.

### Invariant Properties over $RM(0, n)$

The number of flats of dimension  $k$  for  $1 \leq k \leq n$  on which a Boolean function is constant, related to the property of  $k$ -normality, is an  $RM(0, n)$  invariant.

It was noted in [Lea04] that the approach using connection functions, introduced by Fuller and Millan, does not work for bent functions since the Walsh and autocorrelation spectrum of the connection functions are too similar. Therefore Leander presented in [Lea04] a further generalization which can only be used for checking equivalence over  $RM(0, n)$ . Instead of adding a function of low weight a quadratic function of the form  $x_i x_j$  for  $1 \leq i < j \leq n$  was added. This requires  $2^{\frac{(2^n-1)(2^n-2)}{2}}$  functions to be tested.

We now present a new invariant property, which can be used for testing equivalence over  $RM(0, n)$  of bent functions and which is more efficient than the approach of Leander. This set consists of the  $2^{n+1}$  product functions, which are the functions obtained by the product with an affine function. It is easy to prove that if we have two equivalent functions  $f_1, f_2$  with  $f_1(\bar{x}) = f_2(\bar{x}A \oplus \bar{a}) \oplus b$  then their product functions are invariant over  $RM(1, n)$ . This follows from the fact that  $f_1(\bar{x})(\bar{c} \cdot \bar{x} \oplus c_0) = f_2(\bar{x})(\bar{d} \cdot \bar{x} \oplus d_0) \oplus \bar{B}\bar{x} \oplus b'$  if and only if  $\bar{d} = \bar{c}A^{-1}$ ,  $d_0 = \bar{c}A^{-1}\bar{a}^t \oplus c_0$ ,  $\bar{B} = b(\bar{c}A^{-1})$ , and  $b' = b(\bar{c}A^{-1}\bar{a}^t \oplus c_0)$  where  $\bar{c}, \bar{d} \in \mathbb{F}_2^n$  and  $c_0, d_0 \in \mathbb{F}_2$ .

The tuple consisting of the size of the annihilators of a function [MPC04] and its complement is an invariant over  $RM(0, n)$ . Let  $f_1(\bar{x}) = f_2(\bar{x}A \oplus \bar{a}) \oplus b$ , then if  $g$  is annihilator of  $f_1$ ,  $g(\bar{x}A^{-1} \oplus \bar{a}A^{-1})$  is annihilator of  $f_2(\bar{x}) \oplus \bar{b}$ . This is an efficient invariant for functions with small number of variables. But, the complexity of the invariant increases very fast with the dimension.

### Invariant of $RM(-1, n)$

The most efficient invariant is the weight of the function. Balanced functions are invariant, while non-balanced functions split into two classes under  $RM(0, n)$  invariance.

## Equivalence Classes and Sizes of Boolean Functions in 6 variables

**Equivalence Classes** We first show how to derive the classification of  $RM(3, 6)/RM(1, 6)$ . These classes can be obtained from the 6 representatives  $f_i \oplus RM(2, 6)$  for  $1 \leq i \leq 6$  of  $RM(3, 6)/RM(2, 6)$  as given in [Hou96b]. For each representative, we run through all functions consisting only of degree equal

to 2 and distinguish the affine inequivalent cosets of  $RM(1, 6)$  by using as invariants the distribution of the absolute Walsh and autocorrelation values. These indicators suffice to distinguish all 34 affine equivalence classes.

In order to determine the equivalence classes of  $RM(4, 6)/RM(1, 6)$ , we start from the representatives of  $RM(4, 6)/RM(2, 6)$ . These representatives turn out to be the dual of the equivalence classes of  $RM(3, 6)/RM(1, 6)$ . Since the number of representatives is equal for both sets as shown in [Hou96a], we only had to check if the dual representatives are also inequivalent under  $RM(2, 6)$ , which can be done by computing the frequency distribution of the derivatives which belong to  $RM(3, 6)/RM(1, 6)$ . Then, for each representative of  $RM(4, 6)/RM(2, 6)$ , we run through all functions consisting only of terms of degree equal to 2 and distinguish the affine inequivalent cosets of  $RM(1, 6)$ . Here, the invariants of the distribution of the absolute Walsh and autocorrelation values are not sufficient for finding all 2499 different classes. In combination with the invariant of the restrictions with respect to a hyperplane, the problem is solved. Here, for each representative the set of restrictions with respect to all 126 hyperplanes  $c_1x_1 \oplus \dots \oplus c_6x_6 \oplus c_7$  where  $(c_1, \dots, c_7) \in \mathbb{F}_2^7 \setminus \{\bar{0}, \bar{1}\}$  and check to which of the 29 affine equivalence classes of  $RM(4, 5)/RM(1, 5)$  it belongs. Therefore, it suffices to use the invariants of Walsh and autocorrelation spectrum. The classes of  $RM(5, 6)/RM(1, 6)$  and  $RM(6, 6)/RM(1, 6)$  are found in a similar way (with the same invariants but now the restrictions belong to one of the 48 equivalence classes of  $RM(5, 5)/RM(1, 5)$ ). We start again from the representatives of  $RM(5, 6)/RM(2, 6)$  and  $RM(6, 6)/RM(2, 6)$ . In order to obtain the classes of  $RM(5, 6)/RM(2, 6)$  and  $RM(6, 6)/RM(2, 6)$ , we dualize the equivalence classes of  $RM(3, 6)/RM(0, 6)$  and  $RM(3, 6)/RM(-1, 6)$ . Again, after checking the inequivalence over  $RM(2, 6)$  by frequency distribution of the derivatives, they turned out to define the complete set of representatives. The equivalence classes of  $RM(3, 6)/RM(0, 6)$  are found by running through all linear functions. The 120 different classes could then be distinguished by using as invariants the size of the annihilator set together with the number of flats on which a Boolean function is constant. Once the 120 equivalence classes of  $RM(3, 6)/RM(0, 6)$  are obtained, the weight of the function determines if the representative of  $RM(3, 6)/RM(0, 6)$  leads to one or two classes in  $RM(3, 6)/RM(-1, 6)$ .

**Sizes of the Orbits** In order to employ the approach for counting the number of Boolean functions with certain properties as described in [BBNP04], we also

need to know the sizes of these orbits. Let us first describe how to find the sizes of the representatives of  $RM(3, 6)/RM(1, 6)$ . The sizes were computed during the classification phase multiplying the final results by the sizes of the corresponding orbits in  $RM(3, 6)/RM(2, 6)$  given in [Hou96b]. To check these results in the cases of  $f_2$ ,  $f_4$  and  $f_6$  we obtained linear systems for unknown sizes by taking into account the weight distributions of the cosets of  $RM(1, 6)$  and the weight distribution of the corresponding representative of  $RM(3, 6)/RM(2, 6)$  to which these cosets belong. Of course if  $f_1 = 0$  one can use also [MS91, Theorem 1 and Theorem 2, p.436]. The results obtained in these two ways coincide.

Also the sizes of  $RM(3, 6)/RM(0, 6)$  and  $RM(3, 6)/RM(-1, 6)$  were obtained during the classification phase, where we only run through all linear functions and constant functions respectively.

**Remark 1** *Note that Maiorana has already proven in 1991 the existence of 150 357 equivalence classes under  $RM(1, n)$  for Boolean functions with 6 variables by using a computer search based on group theory [Mai91].*

### Equivalence Classes and Sizes of Boolean Functions in 7 variables

We start from the 12 equivalence classes of  $RM(3, 7)/RM(2, 7)$  as given in [Hou96b] and run through all functions containing only terms of degree equal to 2. The fastest invariant in the previous list which leads to the complete classification is the invariant of the restrictions with respect to a hyperplane. For every such function of  $RM(3, 7)/RM(1, 7)$ , we compute all 254 restrictions  $f(\bar{x})$  with respect to the hyperplane  $c_1x_1 \oplus \dots \oplus c_7x_7 \oplus c_8$  where  $(c_1, \dots, c_8) \in \mathbb{F}_2^8 \setminus \{\bar{0}, \bar{1}\}$  and check to which of the 34 affine equivalence classes of  $RM(3, 6)/RM(1, 6)$  they belong. Therefore, it suffices to use the invariants of Walsh and autocorrelation spectrum, together with the determination of the class of  $RM(3, 6)/RM(2, 6)$ . This class can be distinguished by the invariant  $\delta(f) = (\delta_0(f), \delta_1(f), \delta_2(f))$ , where  $\delta_i(f) = |\{\bar{a} \in \mathbb{F}_2^n : D_{\bar{a}}f \text{ is a plateaued function with an amplitude } 2^{n-i}\}|$  for  $0 \leq i \leq 2$ .

The sizes of the orbits were computed in a similar way as for  $n = 6$ . They were computed during the classification phase and multiplied by the sizes of the orbits of  $RM(3, 7)/RM(2, 7)$ , which were derived in [BM04]. A list of all 179 equivalence classes together with the sizes of the orbits of  $RM(3, 7)/RM(1, 7)$  can be found in [BBNP04]. Recently, independently Meng Qing-shu *et. al.* has also derived the equivalence classes of  $RM(3, 7)/RM(1, 7)$  [QsmHgYz05].

## REFERENCES

- [AB98] A. Ashikmin and A. Barg. Minimal vectors in linear codes. *IEEE Transactions on Information Theory*, IT-44(5):2010–2017, 1998.
- [BBNP04] A. Braeken, Y. Borissov, S. Nikova, and B. Preneel. Classification of Boolean functions of 6 variables or less with respect to cryptographic properties. Cryptology ePrint Archive, Report 2004/248, 2004.
- [BL03] E. Brier and P. Langevin. Classification of Boolean cubic forms of nine variables. *IEEE Information Theory Workshop 2003*, pages 179–182, 2003.
- [BM04] Y. Borissov and N. Manev. On minimal codewords in the 3rd order binary Reed-Muller codes. *Serdica*, 30(2-3):303–324, 2004.
- [CT93] James W. Cooley and John W. Tukey. On the origin and publication of the FFT paper. *Current Contents*, (51-52):8–9, 1993.
- [Dil74] J. Dillon. *Elementary Hadamard Difference Sets*. PhD thesis, University of Maryland, 1974.
- [Dub01] Sylvie Dubuc. *Etude des propriétés de dégénérescence et de normalité des fonctions booléennes et construction des fonctions  $q$ -aires parfaitement non-linéaires*. PhD thesis, Université de Caen, 2001.
- [FM03] J. Fuller and W. Millan. Linear redundancy in S-boxes. In *Fast Software Encryption — FSE 2003*, volume 2887 of *Lecture Notes in Computer Science*, pages 74–86. Thomas Johansson, editor, Springer, 2003.
- [Hou96a] X.D. Hou. Covering radius of the Reed-Muller code  $r(1, 7)$  – a simpler proof. *Journal of Theoretical Theory*, 74:337–341, 1996.
- [Hou96b] X.D. Hou.  $gl(m, 2)$  acting on  $r(r, m)/r(r - 1, m)$ . *Discrete Mathematics*, 149:99–122, 1996.
- [Lea04] N.G. Leander. *Normality of Bent Functions, Monomial-and Bimonomial-Bent Functions*. PhD thesis, Ruhr Universität Bochum, 2004.
- [LVi98] P. Langevin, P. Veron, and J.P. Zanotti (ii). Fonction booléennes équilibrés. Technical report, 1998.
- [LZ96] P. Langevin and J.P. Zanotti. Fonction booléennes équilibrés (i). Technical report, 1996.

- [Mai91] J.A. Maiorana. A classification of the cosets of the Reed-Muller code  $r(1,6)$ . *Mathematics of Computation*, 57(195):403–414, 1991.
- [MFD04] W. Millan, J. Fuller, and E. Dawson. New concepts in evolutionary search for Boolean functions in cryptology. *Computational Intelligence*, 20(3):463–474, 2004.
- [MPC04] W. Meier, E. Pasalic, and C. Carlet. Algebraic attacks and decomposition of Boolean functions. In *Advances in Cryptology — EUROCRYPT 2004*, volume 3027 of *Lecture Notes in Computer Science*, pages 474–491. Christian Cachin and Jan Camenisch, editors, Springer, 2004.
- [MS89] W. Meier and O. Staffelbach. Nonlinearity criteria for cryptographic functions. In *Advances in Cryptology — EUROCRYPT 1989*, volume 434 of *Lecture Notes in Computer Science*, pages 549–562. Jean-Jacques Quisquater and Joos Vandewalle, editors, Springer, 1989.
- [MS91] F.J. MacWilliams and N.J.A. Sloane. *The Theory of Error-Correcting Codes*. Elsevier Science Publisher, 1991. ISBN 0-444-85193-3.
- [Pre93] B. Preneel. *Analysis and design of cryptographic hash functions*. PhD thesis, Katholieke Universiteit Leuven, 1993.
- [QsmHgYz05] Meng Qing-shu, Yang min, Zhang Huan-guo, and Liu Yu-zhen. Analysis of affinely equivalent boolean functions, 2005.