De Bruijn Sequences and Complexity of Symmetric Functions

Christelle Rovetta and Marc Mouffron

EADS Defence and Security
Cyber Security Customer Solutions Centre, France

Abstract. A multivalued function is a function from a set E_q^n to a set E_m, where E_k is a set which contains k elements. These functions are used in cryptography: cipher design, hash function design and in theoretical computer science. In this paper, we study the representation of these functions with Multivalued Decision Diagrams (MDD). This representation can be used both to measure complexity and to implement efficiently functions in hardware.
We are especially interested in symmetric functions whose MDDs are shown to have much lower size than classical functions MDDs. One major result is to determine exactly their MDD's maximum size. Notably, we highlight the links between De Bruijn sequences and the most complex symmetric functions and exhibit new functions for $q = 2$ and any m. Enumeration of these functions are supplied, they are numerous enough to allow many applications.

Key words: Symmetric functions, functions over finite sets, hardware implementation, MDD, De Bruijn sequences.

1 Introduction

Today cryptography is spreading everywhere in a lot of devices and peculiarly small, mobile, low energy and low cost pieces of equipment such that Bluetooth earpieces, RFID tags, sensors. To design and implement cryptographic algorithms on these devices, there is an eager need of small footprint Boolean or multivalued functions achieving a good trade-off in term of complexity, implementation and cryptographic properties.

Some works already deal with complexity issues of the partially symmetric Boolean functions [5] and the symmetric multivalued functions [3,6]. They show that decision diagrams [2] are well suited to benefit symmetries and we enhance this further on with excellent results for symmetric multivalued functions. Decision Diagrams (either binary, BDD, or multivalued, MDD) are able both to provide a measure of the complexity of these functions and to achieve an efficient implementation. This is why our analysis makes huge use of this tool.

An asset of symmetric functions is that their MDD have bounds [3,6] of small order, but no result expresses exactly the maximum size. No result neither provides functions achieving this maximum. We investigate this in order to establish the exact maximum value of these MDD, with different levels of reductions [2].

In this paper, we study the structure of symmetric functions MDD and prove their maximum size. We exhibit the functions reaching this maximum value (the "hard" symmetric functions) and give an external characterization of such functions linked to De Bruijn sequences. We then derive some counting results on these "hard" symmetric functions. In this short note some results are shortened and some proofs are omitted, they will be provided in the final paper.

2 Definitions and Notations

2.1 Multivalued Functions

Let us consider $E_k = \{0, 1, \ldots, k-1\}$ the set of k elements, and n, q, m positive integers. Let C_n^k be the choose k among n coefficient. Card(E) stands for the cardinal of set E.

Definition 1. *Given any positive integers n, q, m, we call multivalued function any function from E_q^n to E_m.*

The set of multivalued functions is denoted by $M_n(q, m)$. It contains m^{q^n} functions. Any function $f \in M_n(q, m)$ is characterized by a vector $f_v \in E_m^{q^n}$ called its *value vector*, consisting in the evaluations of the function at every possible input:

$$f_v = (f(0, \ldots, 0), f(0, \ldots, 0, 1), f(0, \ldots, 0, 2), \ldots, f(q-1, \ldots, q-1)) \quad . \tag{1}$$

2.2 Symmetric Multivalued Functions

Definition 2. *A multivalued function $f : E_q^n \longrightarrow E_m$, is symmetric, if f is invariant under any permutation of its input's variables:*
$\forall \sigma \in S_n,\ f(x_1, \ldots, x_n) = f(x_{\sigma(1)}, \ldots, x_{\sigma(n)})$.

The set of these functions is denoted by $SM_n(q, m)$.

Definition 3. *We call symmetry class of $x = (x_1, \ldots, x_n) \in E_q^n$, the set $P_{n,q}(x)$ defined by*

$$P_{n,q}(x) := \left\{ y \in E_q^n \ ; \exists \text{ a permutation } \sigma \text{ such that } y = (y_1, \ldots, y_n) = (x_{\sigma(1)}, \ldots, x_{\sigma(n)}) \right\} \quad (2)$$

We set as representative of the class of symmetry, the smallest element in the lexicographical order.

According to the lemma of Andrew [1], we deduce that there are C_{n+q-1}^{q-1} symmetry classes in E_q^n. We call j^{th} symmetry class of E_q^n the class of symmetry whose representative is classified j^{th} among all representatives according to the lexicographical order.

A symmetric multivalued function can be represented by a vector with values in E_m, whose length equals C_{n+q-1}^{q-1}. The components of the vector are the evaluations of the function for each representative of the symmetry classes. The j^{th} component of the vector corresponds to the evaluation of the function for the j^{th} symmetry class. We call this vector a *simplified vector* of the function:

$$f_{sv} = (f(0, \ldots, 0), \ldots, f(0, \ldots, 0, q-1), f(0, \ldots, 0, 1, 1), \ldots, f(q-1, \ldots, q-1)) \ . \quad (3)$$

There are $m^{C_{n+q-1}^{q-1}}$ possible simplified vectors, i.e. as many as symmetric multivalued functions.

2.3 Multivalued Decision Diagrams

A Multivalued Decision Diagram (MDD) is a generalization of a Binary Decision Diagram (BDD). In his famous article " Graph-Based Algorithms for Boolean Function Manipulation " [2], Bryant defined a reduction procedure *reduce* which reduces a BDD in a single and optimal way. This procedure applies two rules: the fusion rule and the suppression rule. Bryant's procedures can easily be generalized to deal with the MDD.

Definition 4. *The fusion rule says that two nodes are fusioned if their subgraphs are isomorphic. The suppression rule says that a node is deleted if it has only one child node.*

Definition 5. *A Reduced Ordered Multivalued Decision Diagram (ROMDD) is a MDD reduced by the full reduction procedure (i.e. fusion and suppression rules).*

Definition 6. *A Quasi Reduced Ordered Multivalued Decision Diagram (QROMDD) is a MDD reduced by using only the fusion rule.*

2.4 Complexity

The number of nodes of a MDD is called the *size* of the MDD. We call *height* of a node its distance to the top. The set of nodes having a same height $k \in \{0, 1, \ldots, n\}$ is called *level* k of the MDD.

Definition 7. *Let $f \in M_n(q, m)$, we define:*

- *its complexity, noted $c(f)$, the size of its QROMDD.*
- *its reduced complexity, noted $c_R(f)$, the size of its ROMDD.*

3 Symmetric Functions Representation by Simplified MDD

We put forward an optimized representation of symmetric functions by MDD called a *simplified MDD* which is a QROMDD. The idea is to associate a symmetry class to each node of the MDD.

The *simplified MDD* of a symmetric function $f \in SM_n(q, m)$ is defined as follows.
Let $n_{k,j}$ be the j^{th} node from the left of level $k \in \{0, \ldots, n\}$ and $j \in \{0, \ldots, k+1\}$, then $n_{k,j}$ represents the j^{th} symmetry class of the set E_q^k.

- If $k = n$ then $n_{k,j}$ is terminal. Its value is equal to $f(x_j)$ where $x \in E_q^n$ is the representative of the j^{th} class of symmetry of the set E_q^n.
- Else, $n_{k,j}$ is not terminal and it has q distinct sons. Each son represents a distinct symmetry class of the set E_q^{k+1}.

A simplified MDD has a height equal to $n+1$ and has C_{k+q-1}^{q-1} nodes on each of its level k.
Thus, the simplified MDD has C_{q+n}^{q} nodes, which enables us to deduce an upper bound for the complexity of any symmetric multivalued function.

Lemma 1. *Let $f \in SM_n(q,m)$, then $c(f) \leq C_{n+q}^{q}$.*

4 Maximum Complexity of Symmetric Functions

A symmetric multivalued function $f \in SM_n(q,m)$ is called *symmetric (multivalued) hard* if its complexity is the largest among the functions of its set, i.e. if

$$c(f) = \max\{c(g); g \in SM_n(q,m)\} \quad . \tag{4}$$

We denote by $HSM_n(q,m)$ the set of the hard symmetric multivalued functions and by $SC_n(q,m)$ the complexity of the functions in $HSM_n(q,m)$, i.e.

$$SC_n(q,m) := \max\{c(f), f \in SM_n(q,m)\} \tag{5}$$

and

$$HSM_n(q,m) := \{f \in SM_n(q,m); c(f) = SC_n(q,m)\} \quad . \tag{6}$$

For all integer $k \in \{0,1,\ldots,n\}$, we define $SR_{n,q,m}(k) = \min\left(C_{k+q-1}^{q-1}, m^{C_{n-k+q-1}^{q-1}}\right)$.

We call *symmetric inflection level* $h(n,q,m)$, the integer $h \leq n$ such that if $n \geq 2$

$$\begin{cases} C_{h+q-2}^{q-1} < m^{C_{n-h+q-2}^{q-1}} \\ C_{h+q-1}^{q-1} \geq m^{C_{n-h+q-1}^{q-1}} \end{cases} \quad \text{i.e.} \quad \begin{cases} SR_{n,q,m}(h-1) = C_{h+q-2}^{q-1} \\ SR_{n,q,m}(h) \quad\quad = m^{C_{n-h+q-1}^{q-1}} \end{cases}$$

We take by convention $h(0,q,m) = 0$ and $h(1,q,m) = 1$.

For all integer $n \geq 1, q \geq 2, m \geq 2$, we call *pair of symmetric inflection* associated with the parameters, the pair $(SR_{n,q,m}(h(n,q,m)), SR_{n,q,m}(h(n,q,m)-1))$.
In the multivalued case, we notice that:

- C_{k+q-1}^{q-1} is the number of nodes of a simplified MDD on the level k,
- $m^{C_{n-k+q-1}^{q-1}}$ is the number of symmetric functions with $n-k$ variables.

Theorem 1. $\forall n \geq 1,\ SC_n(q,m) = \sum_{k=0}^{n} SR_{n,q,m}(k) = C_{q+h(n,q,m)-1}^{q} + \sum_{k=q+h(n,q,m)-1}^{n} m^{C_{n-k+q-1}^{q-1}}$.

5 Simplified Value Vector in the Case $q = 2$ and any m

5.1 General Results for any n

For the particular case $q = 2$ and any m, the simplified value vector can be read directly on the last level of the simplified MDD.

Theorem 2. *Let (l,k) be the inflection symmetric pair associated to the parameters $(n,2,m)$. When n takes all the values between $a + m^a - 2$ and $a + m^{a+1} - 1$, i.e. $n = a + m^a + b - 2$, for all $b \in \{0,\ldots,(m-1)m^a\}$, then*

(i) $h(n,2,m) = m^a + b + 1 = n - a + 1$,
(ii) l *is equal to* m^a,
(iii) k *is equal to* $h(n,2,m)$.

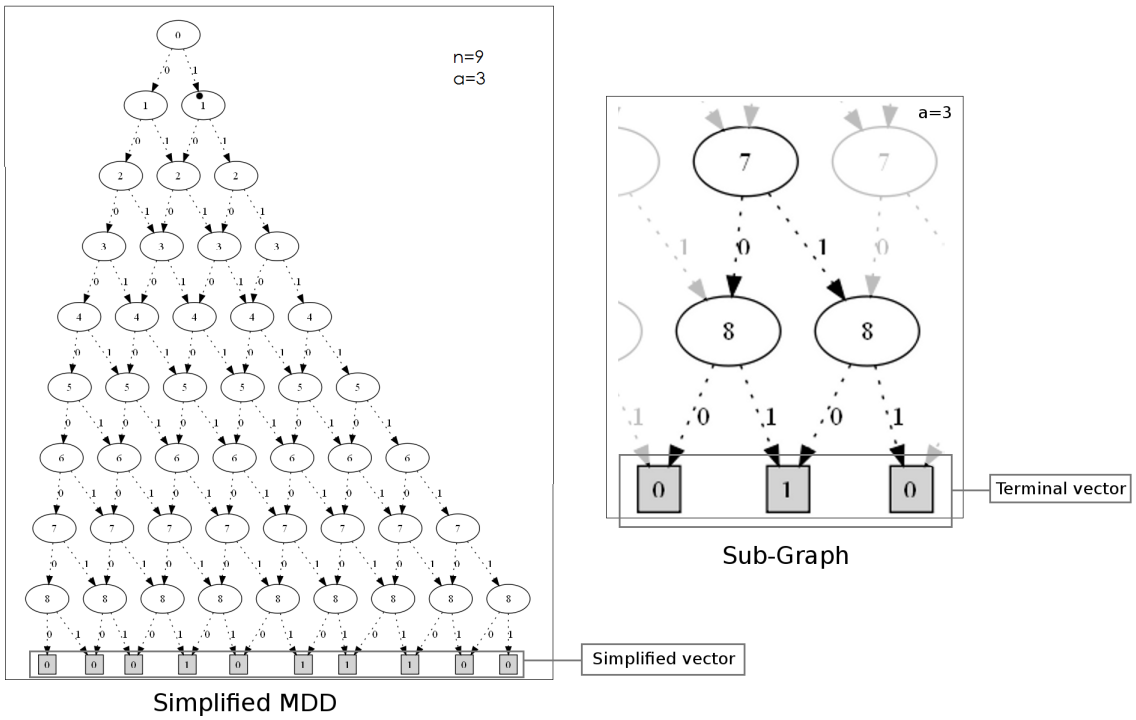


Fig. 1. Simplified MDD and sub-graph of $f \in SM_9(2,2)$. In this case, $a = 3$

Let G_f be the simplified MDD associated to the symmetric function $f \in SM_n(q,m)$. We call *sub-graph* sG_f of G_f any sub-graph of height equals to $n+1-h(n,2,m) = a$ and whose root is of level $h(n,2,m)$ in G_f. We call *terminal vector* into a sub-graph sG_f, the values of the terminal nodes read from left to right. The terminal vector of a subgraph has a length equal to a.

By definition, a simplified MDD has $C^1_{h(n,2,m)+1} = h(n,2,m)+1$ sub-graphs sG_f in the case $q = 2$.

Lemma 2. *A function is hard if and only if the number of its sub-graphs sG_f being not isomorphic is maximum.*

Theorem 3. *Let $n = a + m^a + b - 2$ where $a \geq 0$, $b \in \{0, \ldots, (m-1)m^a\}$. Then a function $f \in SM_n(2,m)$ is hard if and only if in its simplified vector, it appears m^a subsequences of a consecutive different letters.*

Theorem 4. *Let's $f \in HSM_n(2,m), n \geq 1, a \geq 1, b \geq 0$, such that*
$n = a + m^a - 2 + b$ *then* $C_R(f) = m + \sum_{k=0}^{n-1} min(k+1, m^{n+1-k} - m)$

5.2 De Bruijn Sequences and Terminal Vectors

According to Theorem 3, for $n = m^a + a - 2$, $h(n,2,m) + 1 = m^a$, i.e. the simplified MDD of a hard symmetric function has the same number of nodes at the level $h(n,2,m)$ as its QROMDD. So we deduce the following corollary.

Corollary 1. *Let n and a be two positive integers such that $n = a + m^a - 2$ and let $f \in SM_n(2,m)$. Then f is hard if and only if in its simplified vector, it appears exactly m^a distinct subsequences with length equal to a.*

This property on the simplified vector is pertaining to De Bruijn sequences [7]. Let A be an alphabet of m letters. We call a De Bruijn sequence $B(m,a)$ a cyclic sequence such that each subsequence with length equal to a appears exactly once.
A simplified MDD of a symmetric function with $n = m^a + a - 2$ variables contains $m^a + a - 1$ terminal nodes. The following theorem establishes the link between the De Bruijn sequences and the simplified vectors of the hard symmetric functions.

Theorem 5. *Let $n \geq 1$ and $a \geq 0$ be integers such that $n = a + m^a - 2$. Then the simplified value vector of $f \in HSM_n(2, m)$ is a rotation of a De Bruijn sequence $B(m, a)$ at the end of which one the $a - 1$ first letters of the sequence are concatenated.*

Corollary 2. *Let $n \geq 1$ and $a \geq 0$ be integers such that $n = a + m^a - 2$. Then the number of hard symmetric functions of parameters $(n, 2, m)$ is equal to $m!^{m^{a-1}}$.*

Proposition 1. *Let $n \geq 1$ be an integer such that there exists $a \geq 0$ and $b \geq 0$ with $n = a + m^a - 2 + b$. Then the number of hard symmetric functions of parameters $(n, 2, m)$ is greater or equal to: $m!^{m^{a-1}} \times m^b$.*

5.3 Algebraic Degree of Hard Symmetric Boolean Functions

The following theorem establishes the link between the periodicity of the simplified vector of a symmetric Boolean function and its algebraic degree [4].

Theorem 6. *[4] Let $f \in SM_n(2, 2)$, then its simplified value vector, $v_s = (v_s(0), \ldots, v_s(n))$ is periodic with period 2^t, $2^t < n$, if and only if $\deg(f) \leq 2^t - 1$.*

For $n = a + 2^a - 2$, the simplified value vector v_s of a hard symmetric function is periodic with period 2^a, and by properties of De Bruijn sequence this vector cannot be periodic with period 2^k, where $k > a$. Thus, according to the theorem 6 and its contraposition, we obtain the following result.

Theorem 7. *The hard symmetric Boolean functions with $n = a + 2^a - 2$ variables have degree belonging to interval $[2^{a-1}, \ldots, 2^a - 1]$.*

6 Conclusion

We were interested primarily in the multivalued symmetric functions from a set E_q^n to a set E_m and their representations by QROMDD. We initially set out an efficient way to represent these functions by a MDD, then we gave a formula for the exact value of the complexity of the hard symmetric functions. For $q = 2$ (and any m) we could generalize the results concerning the simplified vectors of these hard functions. We highlighted the links between De Bruijn sequences and these simplified vectors from the hard symmetric functions when $n = a + m^a - 2$. We thus could count these hard functions in this particular case. For the general value of n, we stated another link with De Bruijn sequences but we were only able to give a lower bound of the number of functions.
These hard symmetric functions can be a good compromise for a use in cryptography; being symmetric they have a low number of nodes but being hard they also appear among the most robust. We have also shown that their algebraic degree takes interesting values. The further characterization of more detailed cryptographic properties of these functions will be of great interest. The generalization of our results to higher values of q can offer good new functions.

References

1. Andrews G.E.: The theory of partitions. Encyclopedia of Mathematics and Its Applications **2** Addison-Wesley Publishing Company Reading, Mass. (1976).
2. Bryant R.E.: Graph-Based Algorithms for Boolean Function Manipulation. IEEE Transactions on Computers C35 **8** 677–691 (1986) .
3. Butler Jon T., Herscovici David S., Sasao Tsutomu, and Barton Robert J.: "Average and Worst Case Number of Nodes in Decision Diagrams of Symmetric Multiple-Valued Functions", IEEE Transaction on Computers, Vol. 46, N° 4, (April 1997).
4. Canteaut A., Videau M.: Symmetric Boolean Functions. IEEE Transactions on Information Theory **51** N° 8 2791–2811 (2005).
5. Heinrich-Litan L. and Molitor P.: "Least Upper Bounds for the Size of OBDDs Using Symmetry Properties", IEEE Transaction on Computers, Vol. 49, N° 4, 271–281, (April 2000).
6. Mouffron M.: "Balanced Alternating and Symmetric Functions over Finite Sets", Workshop on Boolean Functions Cryptography and Applications, (BFCA08), Copenhagen, Denmark, 27–44, (2008).
7. Rosenfeld V.R.: Enumerating De Bruijn Sequences. MATCH Commun. Math. Comput. Chem. **45** 71-83 (2002).

A APPENDIX: Algebraic Degree of Super Hard and Hard Symmetric Boolean Functions

n	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
1	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	2	4	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	4	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4		8	6	4	2	4	8	8	0	0	0	0	0	0	0	0	0	0	0	0	0
5			4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6				8	6	12	20	16	0	0	0	0	0	0	0	0	0	0	0	0	0
7					4	8	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8						32	30	20	20	20	16	12	8	4	8	32	32	0	0	0	0
9							64	46	28	14	4	0	0	0	0	0	0	0	0	0	0
10								104	64	32	14	0	0	0	0	0	0	0	0	0	0
11									140	92	60	40	28	20	40	136	128	0	0	0	0
12										136	46	24	12	4	8	32	32	0	0	0	0
13											148	72	32	16	40	192	192	0	0	0	0
14												116	48	20	48	200	192	0	0	0	0
15													96	32	64	32	0	0	0	0	0
16														128	336	448	444	372	324	328	296
17															608	720	772	736	644	516	340
18																1408	1540	1560	1264	916	666
19																	3560	3378	2856	2308	1664
20																		6688	5856	4804	3584
21																			10368	7806	5464
22																				16896	12340
23																					24324
TOTAL	8	14	14	12	12	56	126	194	252	294	288	264	224	224	1152	3200	6892	12734	21312	33574	48678

Table 1. Algebraic degree of Super Hard Symmetric Functions for $n = 3, \ldots, 23$

n	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	4	4	4	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3		4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4			0	2	4	2	0	0	0	0	0	0	0	0	0	0	0	0	0
5				12	10	4	0	0	0	0	0	0	0	0	0	0	0	0	0
6					24	22	16	16	16	0	0	0	0	0	0	0	0	0	0
7						20	8	0	0	0	0	0	0	0	0	0	0	0	0
8							16	0	4	12	16	12	12	8	4	0	0	0	0
9								0	12	28	30	24	8	4	0	0	0	0	0
10									48	66	64	44	20	10	4	0	0	0	0
11										104	100	100	92	84	84	64	64	64	0
12											192	136	88	42	8	0	0	0	0
13												328	236	148	88	64	64	64	0
14													396	248	164	128	128	128	0
15														384	144	64	0	0	0
16															416	128	0	16	44
17																256	0	64	184
18																	0	208	532
19																		800	1128
20																			2056
TOTAL	6	8	4	18	38	48	40	16	80	210	402	644	852	928	912	704	256	1344	3944

Table 2. Algebraic degree of Hard Symmetric Functions for $n = 2, \ldots, 20$

B APPENDIX: Complexities of Symmetric Functions from E_2^n

n	$max\ c_R(f)$	Number of super hard functions	$max\ c(f)$	Number of hard functions
1	3	2	3	2
2	5	2	5	6
3	7	8	8	8
4	10	14	12	4
5	14	14	16	18
6	19	12	21	38
7	25	12	27	48
8	31	56	34	40
9	38	126	42	16
10	46	194	50	80
11	55	252	59	210
12	65	294	69	402
13	76	288	80	644
14	88	264	92	852
15	101	224	105	928
16	115	224	119	912
17	129	1152	134	704
18	144	3200	150	256
19	160	6892	166	1344
20	177	12734	183	3944
21	195	21312	201	9276
22	214	33574	220	To be completed
23	234	48678	240	To be completed
24	255	65040	261	To be completed
25	277	81348	283	To be completed
26	300	94376	306	To be completed
27	324	103944	330	To be completed
28	349	107744	355	To be completed
29	375	99744	381	To be completed
30	402	95232	408	To be completed
31	430	81408	436	To be completed
32	459	61440	465	To be completed
33	489	61440	495	To be completed

Table 3. Complexity of Symmetric Functions from E_2^n