A Constructive Count of Rotation Symmetric Functions

Pantelimon Stănică Department of Mathematics Auburn University Montgomery, Montgomery, AL 36124-4023, USA E-mail: pstanica@mail.aum.edu Subhamoy Maitra Applied Statistics Unit, Indian Statistical Institute 203 B. T. Road, Calcutta 700 108, INDIA E-mail: subho@isical.ac.in

Abstract

In this paper we present a constructive detection of minimal monomials in the algebraic normal form of rotation symmetric Boolean functions (immune to circular translation of indices). This helps in constructing rotation symmetric Boolean functions by respecting the rules we present here.

Keywords: Cryptography, Rotation Symmetric Boolean Functions, Algebraic Normal Form, Combinatorial Problems.

1 Introduction

In [2], Pieprzyk and Qu studied some functions, which they called *rotation symmetric* (*RotS*), as components in the rounds of a hashing algorithm. The study of *RotS* functions was continued in [1, 4]. When efficient evaluation of the function is important, for instance in the implementation of MD4, MD5 or HAVAL, the *RotS* property is desirable, since one can reuse evaluations from previous iterations. We can simply view such a hashing algorithm as a sequence of iterations where each iteration takes some input $X = (X_k, \ldots, X_0)$ and a message block M and produces the output $Y = (Y_k, \ldots, Y_0)$ using the rule $Y = M + F(X_{k-1}, \ldots, X_0) + RotS(X_k, s)$. Note that M, X_i, Y_i are blocks of N-bits, and $RotS(X_k, s)$ is the circular rotation of the block X_k by s positions to the left, and F is another cryptographic primitive. It is important to study the component $RotS(X_k, s)$ of such a hashing algorithm (for more information see [2], for a way to reuse previous evaluations).

As it is the case with every cryptographic property, one is interested to count the objects satisfying that property. This motivates us to look at Boolean functions satisfying various criteria and try to select functions necessary for cryptographic design. We need to know how big the pool of choices is and how to generate functions in that pool.

Let V_n be the vector space of dimension n over the two element field \mathbf{Z}_2 (= V_1). A Boolean function on n variables may be viewed as a mapping from V_n into V_1 . We interpret a Boolean function $f(x_1, \ldots, x_n)$ as the output column of its *truth table*, i.e., a binary string of length 2^n , $f = [f(0, 0, \ldots, 0), f(1, 0, \ldots, 0), f(0, 1, \ldots, 0), \ldots, f(1, 1, \ldots, 1)].$

Throughout the paper, by $a \le i \le b$, we mean that a, b, i are integers and i takes the values $a, a + 1, \ldots, b - 1, b$. If $x_i \in \{0, 1\}$ for $1 \le i \le n$, and $0 \le k \le n - 1$, we define

$$\rho_n^k(x_i) = x_{i+k} \quad \text{if } i+k \le n, \\ = x_{i+k-n} \quad \text{if } i+k > n.$$

Let $(x_1, x_2, \ldots, x_n) \in V_n$. We can extend the definition of ρ_n^k on tuples and monomials as follows:

$$\rho_n^k(x_1, x_2, \dots, x_n) = (\rho_n^k(x_1), \rho_n^k(x_2), \dots, \rho_n^k(x_n)) \text{ and } \rho_n^k(x_{i_1}x_{i_2}\cdots) = \rho_n^k(x_{i_1})\rho_n^k(x_{i_2})\cdots$$

Definition 1. A Boolean function f is rotation symmetric (RotS) if for each input $(x_1, \ldots, x_n) \in V_n$, $f(\rho_n^k(x_1, \ldots, x_n)) = f(x_1, \ldots, x_n)$ for any $0 \le k \le n - 1$.

Given a binary string $x = (x_1, \ldots, x_n)$, we define the *weight* of x, denoted by $wt(x_1, \ldots, x_n)$, as the number of 1's in x. Further, + is the addition operator over GF(2). An n-variable Boolean function $f(x_1, \ldots, x_n)$ can be seen as a multivariate polynomial over GF(2), that is,

$$f(x_1, \dots, x_n) = a_0 + \sum_{i=1}^n a_i x_i + \sum_{1 \le i < j \le n} a_{ij} x_i x_j + \dots + a_{12\dots n} x_1 x_2 \dots x_n,$$

where the coefficients $a_0, a_i, a_{ij}, \ldots, a_{12...n} \in \{0, 1\}$. 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. A Boolean function is said to be *homogeneous* if its ANF contains terms of the same degree only.

Let us denote

$$G_n(x_1, \dots, x_n) = \{\rho_n^k(x_1, \dots, x_n), \text{ for } 0 \le k \le n-1\},\$$

that is, the orbit of (x_1, \ldots, x_n) under the action of ρ_n^k , $0 \le k \le n-1$. It is clear that $G_n(x_1, \ldots, x_n)$ generates a partition in the set V_n . Let g_n be the number of such partitions. We found in [4] that the number of *RotS* functions is exactly

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

 ϕ being Euler's *phi*-function. It turns out that the sequence g_n counts also the number of *n*-bead necklaces with 2 colors when turning over is not allowed, or output sequences from a simple *n*-stage cycling shift register, or binary irreducible polynomials whose degree divides *n* (see [3]). The proof needs Burnside's lemma (see [4] for a more detailed discussion).

Further the following results have been proved in [4] regarding RotS functions of some specific degree. Consider *n*-variable RotS Boolean functions. The number of

- (i) degree w homogeneous functions is $2^{g_{n,w}} 1$,
- (ii) degree w functions is $(2^{g_{n,w}}-1)2^{\sum_{i=0}^{w-1}g_{n,i}}$ and
- (iii) functions with degree at most w is $2^{\sum_{i=0}^{w} g_{n,i}}$,

where $g_{n,w}$ is defined as follows [4]: consider $G_n(x_1, \ldots, x_n)$, where $wt(x_1, \ldots, x_n)$ is exactly w, and define $g_{n,w}$ as the number of partitions over the n bit binary strings of weight w (total number $\binom{n}{w}$), determined by G_n . Further, denote by $h_{n,w}$ the number of distinct sets $G_n(x_1, \ldots, x_n)$, where $wt(x_1, \ldots, x_n) = w$ and $|G_n(x_1, \ldots, x_n)| = n$, that is, the number of long cycles of weight w. It is easy to see that $h_{n,w} \leq g_{n,w}$. Write k|'m, if k $(1 < k \leq m)$ is a proper divisor of m. The following result was obtained in [4]:

(*i*)
$$g_{n,w} = \frac{1}{n} \binom{n}{w}$$
, if $gcd(n, w) = 1$. Also, $g_{n,0} = g_{n,n} = 1$.
(*ii*) $g_{n,w} = \frac{1}{n} \left(\binom{n}{w} - \sum_{k|'gcd(n,w)} \frac{n}{k} \cdot h_{\frac{n}{k},\frac{w}{k}} \right) + \sum_{k|'gcd(n,w)} h_{\frac{n}{k},\frac{w}{k}}$, if $w < n$.

However, the combinatorial results of [4] renders a nonconstructive count as opposed to the results of this paper. By using a different method, we find necessary and sufficient conditions for minimal monomials to generate cycles in homogeneous RotS functions. This in turn helps in the enumeration of RotS functions of certain degree.

2 The Results

We start with the definition of the *short algebraic normal form* (SANF) of a *RotS* function. By abuse of notation we use G_n further on the monomials defining

$$G_n(x_{i_1}x_{i_2}\dots x_{i_d}) = \{\rho_n^k(x_{i_1}x_{i_2}\dots x_{i_d}), \text{ for } 0 \le k \le n-1\}.$$

We write a *RotS* function $f(x_1, \ldots, x_n)$ in the form

$$a_0 + a_1 x_1 + \sum a_{1j} x_1 x_j + \ldots + a_{12\ldots n} x_1 x_2 \ldots x_n,$$

where the coefficients $a_0, a_1, a_{1j}, \ldots, a_{12\dots n} \in \{0, 1\}$, and the existence of a representative term $x_1x_{i_2}\ldots x_{i_d}$ implies the existence of all the terms from $G_n(x_1x_{i_2}\ldots x_{i_d})$ in the ANF. This representation of f is called the *short algebraic normal form* (SANF) of f. Note that the number of

terms in each summation (Σ) corresponding to same degree terms depends on the number of short and long cycles.

A cycle is called *long* if the minimum N satisfying $\rho_n^{N+1}(x_1, \ldots, x_{j_d}) = x_1, \ldots, x_{j_d}$ is n-1, i.e., $|G_n(x_1x_{i_2}\ldots x_{i_d})| = n$. A cycle is called *short* if the minimum N satisfying $\rho_n^{N+1}(x_1, \ldots, x_{j_d}) = x_1, \ldots, x_{j_d}$ is strictly smaller than n-1, i.e., $|G_n(x_1x_{i_2}\ldots x_{i_d})| = N < n$. These cycles are completely determined by their *minimal monomial*, i.e., the lexicographically first term $x_{j_1}x_{j_2}\cdots x_{j_d}$ (it is clear that j_1 must be 1).

Assume throughout that $d \ge 1$. First note that a degree d homogeneous RotS function is a sum of degree d RotS long cycles,

$$\sum_{k=0}^{n-1} \rho_n^k(x_{j_1} x_{j_2} \cdots x_{j_d}), j_1 < \cdots < j_d$$

or degree d RotS short cycles

$$\sum_{k=0}^{N-1} \rho_n^k(x_{j_1} x_{j_2} \cdots x_{j_d}), j_1 < \cdots < j_d.$$

Therefore, there is an equivalence between the cycles of RotS functions and their minimal monomials. Using this observation we obtain our first result.

Theorem 2. The number of homogeneous RotS functions of degree $d \ge 1$ equals

$$2^{m(d)} - 1,$$

where m(d) is the number of minimal monomials of degree d.

Proof. Let m_i , i = 1, ..., m(d) be the minimal monomials. It is obvious that any *RotS* function is a sum of cycles determined by these monomials. Since the constant 0 function is not counted, we get the result.

Corollary 3. The number of RotS functions of degree $d \ge 1$ (not necessarily homogeneous) is

$$(2^{m(d)} - 1) \cdot 2^{\sum_{i=1}^{d-1} m(i)}$$

Proof. In a degree $d \operatorname{RotS}$ functions we must have at least a degree d homogeneous RotS cycle. Using the previous theorem and the fact that RotS cycles of lower degree may or may not appear, the number of these being $2^{\sum_{i=1}^{d-1} m(i)}$, we get the count.

The number of degree d monomials is obviously $\binom{n}{d}$. How many of these monomials can occur as minimal terms? We shall give a constructive answer to this question.

We treat the case of degree 2 and 3 separately, to clear up some of the issues in the general degree d case. We take first the degree 2 *RotS* functions, with cycles of the form $f_2(x_1, \dots, x_n) = x_1x_{l+1} + x_2x_{l+2} + \dots$.

Theorem 4. The number of degree 2 homogeneous RotS functions is $2^{\lfloor \frac{n}{2} \rfloor} - 1$.

Proof. It suffices to prove that the number of minimal monomials is $\left\lfloor \frac{n}{2} \right\rfloor$.

First, take *n* to be even. If $l = \frac{n}{2}$, then f_2 is a short cycle. If $l \leq \frac{n}{2} - 1$, it is easily seen that x_1x_{l+1} is minimal. If $l \geq \frac{n}{2}$, then $\rho_n^{n-l}(x_1x_{l+1}) = x_{n-l+1}x_1$, which is less than x_1x_{l+1} , so x_1x_{l+1} is not minimal. Therefore, the number of *RotS* cycles in this case is $\frac{n}{2}$.

Now, if n is odd, the same analysis renders the number of RotS cycles to be $\frac{n-1}{2}$.

Now we take the case of degree 3 RotS functions, with long cycles of the form

$$f_3(x_1, \cdots, x_n) = \sum_{k=0}^{n-1} \rho_n^k(x_1 x_{1+r} x_{1+r+s}),$$

and short cycles of the form

$$f_3(x_1,\cdots,x_n) = \sum_{k=0}^{N-1} \rho_n^k(x_1x_{1+r}x_{1+r+s}),$$

with $\rho_n^N(x_1x_{1+r}x_{1+r+s}) = x_1x_{1+r}x_{1+r+s}$ (N < n minimum with this property).

Theorem 5. The number of degree 3 RotS long cycles equals the number of pairs $1 \le r, s \le n-1$ satisfying either of the following conditions:

(i)
$$s > r$$
 and $s + 2r < n$.
(ii) $s = r < \frac{n}{3}$.

Moreover, there is only one short cycle if and only if $n \equiv 0 \pmod{3}$, generated by the minimal monomial $x_1x_{1+\frac{n}{3}}x_{1+\frac{2n}{3}}$.

Proof. We want to find all minimal monomials of degree 3, which generate long cycles. Take an arbitrary monomial with indices $\{1, r + 1, r + s + 1\}$, which we assume to be minimal. By applying ρ_n^i , we get monomials with indices $\{i + 1, r + i + 1, r + s + i + 1\}$. It follows that for any i, with $1 \le i \le n - 1$, the term $\rho_n^i(x_1x_{r+1}x_{r+s+1})$ follows $x_1x_{r+1}x_{r+s+1}$ in lexicographical order. Therefore, we have $\{i, r+i, r+s+i\} > \{0, r, r+s\}$ (modulo n) (we assume the indices in increasing lexicographical order), which will hold, except, if either $r+i \equiv 0 \pmod{n}$ or $r+s+i \equiv 0 \pmod{n}$. Since $i \le n - 1$, we obtain that either i = n - r or i = n - r - s. Case 1. If i = n - r, then $\{0, s, n - r\} > \{0, r, r + s\}$ (obviously, n - r > s). Thus, we obtain that a necessary condition for our monomial to be minimal is to have s > r, or s = r and $r < \frac{n}{3}$ (not sufficient, yet).

Case 2. If i = n - r - s, then we need $\{0, n - r - s, n - s\} > \{0, r, r + s\}$. We get that, either n - r - s > r, or, n - r - s = r and n - s > r + s, that is, another necessary condition for our monomial to be minimal is to have 2r < n - s, or 2r = n - s and $r > \frac{n}{3}$.

A similar analysis (which will be done fully in our general theorem), renders one short cycle if and only if $n \equiv 0 \pmod{3}$, generated by $x_1 x_{1+\frac{n}{3}} x_{1+\frac{2n}{3}}$.

Corollary 6. The number of degree 3 RotS cycles is

$$n \cdot \left\lfloor \frac{n-1}{3} \right\rfloor - \frac{3 \lfloor \frac{n-1}{3} \rfloor \left(\lfloor \frac{n-1}{3} \rfloor + 1 \right)}{2}$$

plus one short cycle if and only if $n \equiv 0 \pmod{3}$.

Proof. The number of pairs in case (*ii*) of Theorem 5 is $\lfloor \frac{n-1}{3} \rfloor$. In case (*i*), we need $r \leq \lfloor \frac{n-1}{3} \rfloor$ and r < s < n - 2r. Thus, the number of pairs in case (*i*) is

$$\sum_{r=1}^{\lfloor \frac{n-1}{3} \rfloor} (n-3r-1) = (n-1) \left\lfloor \frac{n-1}{3} \right\rfloor - \frac{3 \lfloor \frac{n-1}{3} \rfloor (\lfloor \frac{n-1}{3} \rfloor + 1)}{2}.$$

Hence the result.

Now, we treat the general case. Let a degree d (homogeneous) RotS long cycle be given by

$$f(x_1, \cdots, x_n) = \sum_{j=0}^{n-1} \rho_n^j (x_1 \, x_{1+i_1} \cdots x_{1+i_1+\dots+i_{d-1}}).$$

We shall find all monomials $x_1 x_{1+i_1} \cdots x_{1+i_1+\dots+i_{d-1}}$, that are minimal, thus counting the degree $d \operatorname{RotS}$ cycles and giving in the same time a way to list them. For arbitrary j, ρ_n acts on a minimal monomial in the following way

$$\rho_n^j(x_1 \, x_{1+i_1} \cdots x_{1+i_1+\dots+i_{d-1}}) = x_{1+j} \, x_{1+i_1+j} \cdots x_{1+i_1+\dots+i_{d-1}+j}.$$

For $1 \le k \le d-1$, take $j = n - i_1 - \cdots - i_k$. Since $x_1 x_{1+i_1} \cdots x_{1+i_1+\cdots+i_{d-1}}$ is minimal, it follows that, using the lexicographical order,

$$\{0, i_{k+1}, i_{k+1} + i_{k+2}, \dots, i_{k+1} + \dots + i_{d-1}, n - i_1 - \dots - i_k, n - i_2 - \dots - i_k, \dots, n - i_k\} > \{0, i_1, i_1 + i_2, \dots, i_1 + \dots + i_{d-1}\}.$$
(2)

We observe immediately that $i_1 \leq i_{k+1}$, for any k. We distinguish two cases. Case 1. If k = d - 1, then we need

$$\{0, n - i_1 - \dots - i_{d-1}, \dots, n - i_{d-1}\} > \{0, i_1, i_1 + i_2, \dots, i_1 + \dots + i_{d-1}\}.$$

It implies that the indices satisfy either:

$$C: n > 2i_1 + i_2 + \dots + i_{d-1}, \text{ or}$$

$$C_1: n = 2i_1 + i_2 + \dots + i_{d-1} \text{ and } n > i_1 + 2i_2 + i_3 + \dots + i_{d-1} \iff i_1 > i_2), \text{ or}$$

$$C_2: n = 2i_1 + i_2 + \dots + i_{d-1} \text{ and } i_1 = i_2 > i_3, \text{ or}$$

$$\vdots$$

$$C_{d-2}$$
: $n = 2i_1 + i_2 + \dots + i_{d-1}$ and $i_1 = i_2 = i_3 = \dots = i_{d-2} > i_{d-1}$.

But the conditions $i_1 = \cdots = i_s > i_{s+1}$, occurring in C_s $(1 \le s \le d-2)$, contradict the first observation that $i_1 \le i_{s+1}$, $s \ge 1$. Therefore, the only condition in this case that the indices must satisfy is

$$2i_1 + i_2 + \dots + i_{d-1} < n.$$

Case 2. If $1 \le k \le d-2$, then the inequality (2) implies that the indices satisfy either of the following conditions:

$$\begin{split} P_1^h: & i_1 < i_{k+1}, \text{ or } \\ P_2^h: & i_1 = i_{k+1}, i_2 < i_{k+2}, \text{ or } \\ P_3^k: & i_1 = i_{k+1}, i_2 = i_{k+2} \text{ and } i_3 < i_{k+3}, \text{ or } \\ \vdots \\ P_{d-k-1}^k: & i_1 = i_{k+1}, i_2 = i_{k+2}, \dots, i_{d-k-2} = i_{d-2} \text{ and } i_{d-k-1} < i_{d-1}, \text{ or } \\ P_{d-k-1}^k: & i_1 = i_{k+1}, \dots, i_{d-k-1} = i_{d-1} \text{ and } n - \sum_{a=1}^k i_a > \sum_{b=1}^{d-k} i_b, \text{ or } \\ P_{d-k+1}^k: & i_1 = i_{k+1}, \dots, i_{d-k-1} = i_{d-1}, n = \sum_{a=1}^k i_a + \sum_{b=1}^{d-k} i_b \text{ and } n - \sum_{a=2}^k i_a > \sum_{b=1}^{d-k+1} i_b, \text{ or, } \\ & \text{in general,} \\ P_{d-k+l}^k: & Q_{d-k+l}^k \text{ and } \sum_{s=1}^{d-k+1} i_s + \sum_{t=l+1}^k i_t < n, \text{ for } 0 \le l \le k-1, \\ & \text{where } Q_{d-k+l}^k \text{ is the condition } i_1 = i_{k+1} \text{ and } \dots i_{d-1-k} = i_{d-1} \text{ and } n - \sum_{a=s}^k i_a = \sum_{b=1}^{d-k+s-1} i_b, \text{ for all } \\ & 1 \le s \le l, \text{ that is, } Q_{d-k+l}^k \text{ is the condition } n = \sum_{a=1}^k i_a + \sum_{b=1}^{d-k} i_b \text{ and } i_j = i_{k+j}, 1 \le j \le d-1-k \text{ and } \\ & i_s = i_{d-k+s}, 1 \le s \le l-1. \end{split}$$

To have a short cycle, we need $j = n - i_1 - \dots - i_k$, with $\rho_n^j(x_1x_{1+i_1} \cdots x_{1+i_1+\dots+i_{d-1}}) = x_1x_{1+i_1} \cdots x_{1+i_1+\dots+i_{d-1}}$. We see that if k = d - 1, then $i_1 = i_2 = \dots = i_{d-1} = \frac{n}{d}$, and the

minimal monomial for this unique short cycle is $x_1 x_{1+\frac{n}{d}} \cdots x_{1+(d-1)\frac{n}{d}}$. If k < d-1, then a minimal monomial for a short cycle needs to satisfy

$$\{0, i_{k+1}, i_{k+1} + i_{k+2}, \dots, i_{k+1} + \dots + i_{d-1}, n - i_1 - \dots - i_k, n - i_2 - \dots - i_k, \dots, n - i_k\} = \{0, i_1, i_1 + i_2, \dots, i_1 + \dots + i_{d-1}\},$$
(3)

which implies

$$S_k: \quad i_{k+1} = i_1, i_{k+2} = i_2, \dots, i_{d-1} = i_{d-1-k};$$
$$i_1 = i_{d-k+1}, i_2 = i_{d-k+2}, \dots, i_{k-1} = i_{d-1};$$
$$n = i_1 + i_2 + \dots + i_{d-1} + i_k.$$

Assuming $n - i_1 - \cdots - i_k$ is the order of ρ_n on $x_1 x_{1+i_1} \cdots$, we need to impose also the conditions that for any $j = n - i_1 - \cdots - i_l$ (l > k),

$$\rho_n^j(x_1 \, x_{1+i_1} \cdots \, x_{1+i_1+\dots+i_{d-1}}) > x_1 \, x_{1+i_1} \cdots \, x_{1+i_1+\dots+i_{d-1}}.$$

Similar to the case of long cycles, we obtain that the indices must satisfy in addition to S_k , one of the following conditions

$$P_1^l, P_2^l, \cdots \quad (l > k).$$

Putting together the previous results we get our general theorem $(\lor, \land$ are the logical *or*, respectively, *and*).

Theorem 7. The number of degree d RotS long cycles is equal to the number of sequences $1 \le i_1, i_2, \dots, i_{d-1} \le n-1$ satisfying

$$\bigwedge_{k=1}^{d-2} \left(\bigvee_{s=1}^{d-1} P_s^k\right) \bigwedge C,$$

Moreover, the number of degree d RotS short cycles is equal to the number of sequences $1 \leq i_1, i_2, \dots, i_{d-1} \leq n-1$ satisfying

$$\bigvee_{k=1}^{d-2} \left(S_k \wedge \bigwedge_{l=k+1}^{d-1} (\bigvee_{s=1}^{d-1} P_s^l) \right)$$

plus one more, if $n \equiv 0 \pmod{d}$.

We regard the previous result as a summarizing or listing theorem. The count is certainly not as simple as the one of $g_{n,w}$ presented in the introduction, but it has the advantage that one can construct *RotS* Boolean functions by respecting the rules of Theorem 7. Certainly, it is possible to get the exact count in some particular cases (which we have done in Theorem 4 and Corollary 6 for n = 2, 3), but it seems elusive to get, for general n, the count $g_{n,w}$ using Theorem 7. Acknowledgment: The authors like to acknowledge the anonymous reviewers for their professional comments that improved both the editorial and technical quality of this paper. We also thank the editor for his patience.

References

- T.W. Cusick and P. Stănică. Fast Evaluation, Weights and Nonlinearity of Rotation-Symmetric Functions. *Discrete Mathematics*, pp. 289-301, vol. 258, no. 1-3, 2002.
- [2] J. Pieprzyk and C.X. Qu. Fast Hashing and Rotation-Symmetric Functions. Journal of Universal Computer Science, pp. 20-31, vol. 5, no. 1, 1999.
- [3] N. J. A. Sloane. On single-deletion-correcting codes. Codes and Designs Ray-Chaudhuri Festschrift, pp. 273-292, (Eds. K. T. Arasu and Á. Seress), 2002.
- [4] P. Stănică and S. Maitra. Rotation Symmetric Boolean Functions Count and Cryptographic Properties. Proceedings of R. C. Bose Centenary Symposium on Discrete Mathematics and Applications, December 2002, Indian Statistical Institute, Calcutta. Full version available at http://www.isical.ac.in/~subho.