Bounds for the Multicovering Radii of Reed-Muller Codes with Applications to Stream Ciphers

I. Honkala (honkala@utu.fi)
Dept. of Mathematics, University of Turku, 20014 Turku, Finland

A. Klapper (klapper@cs.uky.edu)*
Dept. of Computer Science, 763H Anderson Hall, University of Kentucky, Lexington, KY, 40506-0046.

Abstract. The *multicovering radii* of a code are recent generalizations of the covering radius of a code. For positive m, the m-covering radius of C is the least radius t such that every m-tuple of vectors is contained in at least one ball of radius t centered at some codeword. In this paper upper bounds are found for the multicovering radii of first order Reed-Muller codes. These bounds generalize the well-known Norse bounds for the classical covering radii of first order Reed-Muller codes. They are exact in some cases. These bounds are then used to prove the existence of secure families of keystreams against a general class of cryptanalytic attacks. This solves the open question that gave rise to the study of multicovering radii of codes.

Keywords: Error correcting code, stream cipher, covering radius, Reed-Muller code.

1. Introduction

In this paper we derive upper bounds on the multicovering radii of first order Reed-Muller codes. In some cases the bounds are shown to be exact. We then use these bounds to strengthen an earlier result concerning the existence of families of stream ciphers that are asymptotically secure against a certain very general class of cryptanalytic attack.

Multicovering radii, which were introduced recently [7], are defined as follows. We let $F_2 = \{0, 1\}$ denote the field with two elements. We also denote the Hamming distance between two vectors $\mathbf{u}$ and $\mathbf{v}$ by $\mathrm{dist}(\mathbf{u}, \mathbf{v})$ and the complement of a vector $\mathbf{v}$ by $\bar{\mathbf{v}}$. Vectors are indicated by boldface lower case letters, and the components of the vector $\mathbf{v}$ are $v_1, v_2, \cdots$.

DEFINITION 1.1. *Let C be a binary code of length n. Let m be a positive integer. The m-*covering radius *of C, denoted by $t_m(C)$, is the*

* Project sponsored by the National Science Foundation under grant number NCR-9706078.

smallest integer t such that every m-tuple of vectors in F_2^n is contained in a ball of radius t around at least one codeword in C. That is, $\forall \mathbf{v}^1, \cdots, \mathbf{v}^m \in F_2^n : \exists \mathbf{c} \in C : \forall i = 1, \cdots, m : \mathrm{dist}(\mathbf{c}, \mathbf{v}^i) \leq t.$

This is a natural generalization of the classical notion of covering radius, which is exactly the case when $m = 1$ [2, 3, 5]. When $m > 1$, it is known that there is no code of length m, dimension n, and m-covering radius t if

$$t < \left\lceil \frac{n + \lfloor \log_2(m) \rfloor - 1}{2} \right\rceil .$$

This bound is not in general tight, but no general tight bound is known.

The notion of multi-covering radii first arose in an investigation of the existence of stream ciphers secure against a large class of attacks. A *stream cipher* is a cryptosystem in which the cipher text is the bitwise exclusive-or of the message with a binary pseudorandom *key sequence.* The key sequence is known to the sender and receiver. A known plaintext attack on a stream cipher reveals precisely those bits of the key sequence that correspond to the known plaintext bits. Thus the security of a stream cipher depends on the inability of an adversary to infer the key sequence from a subsequence (generally with some constraints on the resources of the adversary).

By a *sequence generator* we mean a finite automaton with output. Many of the specific attacks on stream ciphers that have been considered in the literature have the general form: input a prefix of a sequence; find a sequence generator (usually in a specified class of generators) whose output agrees with the prefix. A model for this general type of attack was recently considered by the second author [8]. Attacks can be deterministic – meaning the output from the generator found must agree with the original sequence exactly – or probabilistic – meaning the output from the generator found must agree with the original sequence in significantly more than half the bits. A class of sequences can be secure against all attacks *infinitely often* – meaning that for each attack, there are infinitely many sequences in the class that resist the attack – or *almost everywhere* – meaning that for each attack all but finitely many sequences in the family resist the attack. It was shown previously that there are families of sequences that are secure against all deterministic attacks almost everywhere and there are families that are secure against all probabilistic attacks infinitely often, but the existence of families that are secure against all probabilistic attacks almost everywhere was left open. It was apparent from this earlier study that an affirmative answer could be proved if tight enough bounds on the multicovering radii of Reed-Muller codes could be found.

In fact, it was this question that led to the invention of the notion of multicovering radii.

In Section 2 we find upper bounds on the multicovering radii of strength two codes. First order Reed-Muller codes are strength two codes, so these bounds apply to them. In Section 3 we show that in some cases these bounds are tight. In Section 4 we improve the bounds from Section 2 for higher strength codes. In Section 5 we use the bounds from Section 2 to prove the existence of families of efficiently generated sequences that are secure in the strong sense discussed above.

2. Norse Bound for Multi-Covering Radius

Let m be a positive integer. Let $\mathbf{v} \in F_2^r$ and $b \in F_2$, and let $\mathbf{c}^{\mathbf{v},b}$ be the vector indexed by F_2^r whose $\mathbf{u}$th coordinate is

$$\mathbf{c}_{\mathbf{u}}^{\mathbf{v},b} = b + \sum_{i=1}^{m} u_i v_i = b + \mathbf{v} \cdot \mathbf{u}.$$

Then $RM(1,r)$, the rth first order Reed-Muller code, consists of all vectors $\mathbf{c}^{\mathbf{v},b}, \mathbf{v} \in F_2^r, b \in F_2$. Let $n = 2^r$, so $RM(1,r)$ is a $[n, r+1, n/2]$ code [9]. $RM(1,r)$ is an example of a *strength 2 code*. That is, every pair of coordinates in $RM(1,r)$ takes on any fixed pair of values for exactly one quarter of the codewords. Since $\mathbf{c}^{\mathbf{v},0}$ and $\mathbf{c}^{\mathbf{v},1}$ are complementary, $RM(1,r)$ is closed under complement as well.

It has long been known [6] that the ordinary covering radius t of $RM(1,r)$ is bounded by

$$t_1(RM(1,r)) \leq \left\lfloor \frac{2^r - 2^{r/2}}{2} \right\rfloor.$$

This bound is known as the *Norse Bound* and it is this that we generalize. It is further known that equality holds if r is even [10].

THEOREM 2.1. *Let C be any strength 2 code of length n. Suppose C is closed under complement. Then*

$$t_m(C) \leq \left\lfloor \frac{n + \sqrt{mn/2}}{2} \right\rfloor.$$

Proof. Let $\mathbf{x} \in F_2^n$. Then

$$\sum_{\mathbf{c} \in C} (2\mathrm{dist}(\mathbf{x}, \mathbf{c}) - n)^2 \ = \ \sum_{\mathbf{c} \in C} \left(\sum_{j=1}^{n} (-1)^{c_j + x_j} \right)^2$$

$$
\begin{aligned}
&= \sum_{j,k=1}^{n} (-1)^{x_j+x_k} \sum_{\mathbf{c}\in C} (-1)^{c_j+c_k} \\
&= \sum_{j=1}^{n} (-1)^{x_j+x_j} |C| \\
&= n|C|.
\end{aligned}
$$

It follows that for any $\mathbf{x}^1, \cdots, \mathbf{x}^m \in F_2^n$,

$$\sum_{\mathbf{c}\in C} \sum_{i=1}^{m} (2\text{dist}(\mathbf{x}^i, \mathbf{c}) - n)^2 = mn|C|.$$

Therefore there is at least one $\mathbf{c} \in C$ such that

$$\sum_{i=1}^{m} (2d_i - n)^2 \leq mn,$$

where $d_i = \text{dist}(\mathbf{x}^i, \mathbf{c})$. Without loss of generality, assume that

$$|2d_1 - n| \geq |2d_2 - n| \geq \cdots \geq |2d_i - n|.$$

Note that all the values $|2\text{dist}(\mathbf{x}, \mathbf{c}) - n|$ remain the same if we replace $\mathbf{c}$ by its complement. Consequently, by replacing $\mathbf{c}$ with its complement (which is also a codeword) if necessary, we may assume that $2d_1 - n \leq 0$. We can then estimate for every $i \geq 2$

$$
\begin{aligned}
(2d_i - n)^2 &\leq \frac{1}{2}\left((2d_1 - n)^2 + (2d_i - n)^2\right) \\
&\leq \frac{1}{2} \sum_{j=1}^{m} (2d_j - n)^2 \\
&\leq \frac{1}{2} mn.
\end{aligned}
$$

Hence $2d_i - n \leq \sqrt{mn/2}$, and therefore for all $i \geq 2$

$$d_i \leq \frac{1}{2}\left(n + \sqrt{mn/2}\right).$$

Because $d_1 \leq n/2$, this also holds for $i = 1$, proving our claim.

□

COROLLARY 2.2. *For any $r > 0$ we have*

$$t_m(RM(1,r)) \leq 2^{r-1} + \left\lfloor \sqrt{m 2^{r-3}} \right\rfloor.$$

Proof. The first order Reed-Muller codes are known to be strength 2 codes.

□

3. Exact Values

In this section we show that the bound in Corollary 2.2 is tight if m is an odd power of 2 and r is even; and if m is an even power of 2 and r is odd; and for $m = 2$ and $r \in \{1, 3, 5\}$. These results lead to general lower bounds, although a significant gap between the lower and upper bounds remains.

THEOREM 3.1. *Assume that C is a binary self-complementary code and has strength two. If C has covering radius $\frac{1}{2}(n - \sqrt{n})$ then $t_2(C) = \frac{1}{2}(n + \sqrt{n})$.*

Proof. In the proof of Theorem 2.1 we saw that

$$\sum_{\mathbf{c} \in C} (2\mathrm{dist}(\mathbf{x}, \mathbf{c}) - n)^2 = n|C|. \tag{1}$$

If the covering radius of C equals $\frac{1}{2}(n - \sqrt{n})$, we know that there exists a vector $\mathbf{x} \in F_2^n$ such that

$$\mathrm{dist}(\mathbf{x}, \mathbf{c}) \geq \frac{1}{2}(n - \sqrt{n}) \tag{2}$$

for all $\mathbf{c} \in C$. Since C is self-complementary, the complement $\overline{\mathbf{c}}$ of every codeword $\mathbf{c} \in C$ also belongs to C, and so by (2),

$$\begin{aligned} \mathrm{dist}(\mathbf{x}, c) &= n - \mathrm{dist}(\mathbf{x}, \overline{\mathbf{c}}) \\ &\leq n - \frac{1}{2}(n - \sqrt{n}) \\ &= \frac{1}{2}(n + \sqrt{n}). \end{aligned}$$

In other words,

$$|2\mathrm{dist}(\mathbf{x}, \mathbf{c}) - n| \leq \sqrt{n}. \tag{3}$$

Together with (1) this implies that in fact equality holds for all $\mathbf{c} \in C$ in (3), i.e.,

$$\mathrm{dist}(\mathbf{x}, \mathbf{c}) \in \left\{ \frac{1}{2}(n - \sqrt{n}), \frac{1}{2}(n + \sqrt{n}) \right\}.$$

Let us now take $\mathbf{x}^1 = \mathbf{x}$ and $\mathbf{x}^2 = \overline{\mathbf{x}}$, the complement of $\mathbf{x}$. Then for all $\mathbf{c} \in C$ we have $\mathrm{dist}(\mathbf{c}, \mathbf{x}^1) = \frac{1}{2}(n + \sqrt{n})$ or $\mathrm{dist}(\mathbf{c}, \mathbf{x}^2) = \frac{1}{2}(n + \sqrt{n})$, proving that $t_2(C) \geq \frac{1}{2}(n + \sqrt{n})$.

□

Since the first-order Reed-Muller codes of even order satisfy the second Norse bound with equality, we obtain the following corollary,

COROLLARY 3.2. *For all* s,

$$t_2(RM(1,2s)) = 2^{2s-1} + 2^{s-1}.$$

Now consider larger values of m.

THEOREM 3.3. *If* C *is self-complementary, then* $t_m(C) = n$ *for all* $m \geq |C|$.

Proof. This is trivial, because thanks to the condition $m \geq |C|$ we can choose $\mathbf{x}^1, \cdots, \mathbf{x}^m$ to include all the codewords of C.

□

In particular, when C is a first-order Reed-Muller code this means that $t_m(RM(1,r)) = n$ whenever $m \geq 2^{r+1}$.

THEOREM 3.4. *Assume that* $m = 2^{2t+1}$. *Then for all* $s \geq t \geq 0$,

$$t_m(RM(1,2s)) = \frac{n + \sqrt{mn/2}}{2} = 2^{2s-1} + 2^{s+t-1}.$$

Proof. We keep m fixed and prove the result by induction on s. In fact we prove the stronger result that there also exists a *self-complementary* set T of vectors $\mathbf{y}^1, \cdots, \mathbf{y}^m$ such that

$$\max_{1 \leq i \leq m} \operatorname{dist}(\mathbf{y}^i, \mathbf{c}) \geq t_m(RM(1,2s))$$

for every $\mathbf{c} \in RM(1,2s)$.

When $s = t$ we have $m = 2n = |RM(1,2s)|$. By Theorem 3.4,

$$\begin{aligned} t_m(RM(1,2s)) &= n && (4)\\ &= \frac{n + \sqrt{mn/2}}{2}. && (5) \end{aligned}$$

Furthermore, we can take $T = RM(1,2s)$.

Assume then that the formula is known to be correct for $s-1$ and that furthermore we know there is a self-complementary set of vectors $\mathbf{y}^1, \cdots, \mathbf{y}^m$ such that

$$\max_{1 \leq i \leq m} \operatorname{dist}(\mathbf{y}_i, \mathbf{c}) \geq t_m(RM(1, 2(s-1))) \tag{6}$$

for every $\mathbf{c} \in RM(1, 2(s-1))$. We show that the same is true for s.

Using the recursive property

$$RM(1,r) = \{(\mathbf{u}, \mathbf{u} + \mathbf{v}) \mid \mathbf{u} \in RM(1, r-1), \mathbf{v} \in RM(0, r-1)\}$$

twice, we see that all the codewords of $RM(1,2s)$ are of one of the types

$$(\mathbf{u},\mathbf{u},\mathbf{u},\mathbf{u}),(\mathbf{u},\overline{\mathbf{u}},\mathbf{u},\overline{\mathbf{u}}),(\mathbf{u},\mathbf{u},\overline{\mathbf{u}},\overline{\mathbf{u}}),(\mathbf{u},\overline{\mathbf{u}},\overline{\mathbf{u}},\mathbf{u}) \tag{7}$$

where $\mathbf{u}\in RM(1,2(s-1))$. We now take

$$\mathbf{x}^i=(\mathbf{y}^i,\mathbf{y}^i,\mathbf{y}^i,\overline{\mathbf{y}^i})$$

for $i=1,\cdots,m$. The vectors $\mathbf{y}^1,\cdots,\mathbf{y}^m$ form a self-complementary set, so the same is true of the vectors $\mathbf{x}^1,\cdots,\mathbf{x}^m$. Consider the distances $\mathrm{dist}(\mathbf{x}^i,\mathbf{c})$ for $\mathbf{c}\in RM(1,2s)$. If $\mathbf{c}$ is of the first type in (7), then

$$\begin{aligned}\mathrm{dist}(\mathbf{x}^i,\mathbf{c}) &= \mathrm{dist}((\mathbf{y}^i,\mathbf{y}^i,\mathbf{y}^i,\overline{\mathbf{y}^i}),(\mathbf{u},\mathbf{u},\mathbf{u},\mathbf{u}))\\ &= 3\mathrm{dist}(\mathbf{y}^i,\mathbf{u})+(n/4-\mathrm{dist}(\mathbf{y}^i,\mathbf{u}))\\ &= 2\mathrm{dist}(\mathbf{y}^i,\mathbf{u})+n/4,\end{aligned}$$

where $n=2^{2s}$. By (4),

$$\max_{1\le i\le m}\mathrm{dist}(\mathbf{x}^i,\mathbf{c})\ge 2t_m(RM(1,2(s-1)))+n/4.$$

The same is true also if $\mathbf{c}$ is of the type $(\mathbf{u},\overline{\mathbf{u}},\mathbf{u},\overline{\mathbf{u}})$ or of the type $(\mathbf{u},\mathbf{u},\overline{\mathbf{u}},\overline{\mathbf{u}})$. Assume finally that $\mathbf{c}=(\mathbf{u},\overline{\mathbf{u}},\overline{\mathbf{u}},\mathbf{u})$. Because $RM(1,2(s-1))$ is self-complementary, we know that $\overline{\mathbf{u}}\in RM(1,2(s-1))$. Applying (4) to $\overline{\mathbf{u}}$ we get

$$\begin{aligned}\min_{1\le i\le m}\mathrm{dist}(\mathbf{y}^i,\mathbf{u}) &= n/4-\max_{1\le i\le m}\mathrm{dist}(\mathbf{y}^i,\overline{\mathbf{u}})\\ &\le n/4-t_m(RM(1,2(s-1))),\end{aligned}$$

which together with

$$\begin{aligned}\mathrm{dist}(\mathbf{x}^i,\mathbf{c}) &= \mathrm{dist}((\mathbf{y}^i,\mathbf{y}^i,\mathbf{y}^i,\overline{\mathbf{y}^i}),(\mathbf{u},\overline{\mathbf{u}},\overline{\mathbf{u}},\mathbf{u}))\\ &= \mathrm{dist}(\mathbf{y}^i,\mathbf{u})+3(n/4-\mathrm{dist}(\mathbf{y}^i,\mathbf{u}))\\ &= 3n/4-2\mathrm{dist}(\mathbf{y}^i,\mathbf{u})\end{aligned}$$

implies that also in this case

$$\begin{aligned}\max_{1\le i\le m}\mathrm{dist}(\mathbf{x}^i,\mathbf{c}) &\ge 3n/4-2(n/4-t_m(RM(1,2(s-1)))\\ &= 2t_m(RM(1,2(s-1)))+n/4.\end{aligned}$$

Consequently,

$$\begin{aligned}t_m(RM(1,2s)) &\ge 2\frac{n/4+\sqrt{m(n/4)/2}}{2}+\frac{n}{4}\\ &= \frac{n+\sqrt{mn/2}}{2},\end{aligned}$$

and the claim follows from Theorem 2.1.

□

In exactly the same way we obtain the following result for the codes $RM(1, 2s+1)$.

THEOREM 3.5. *Assume that* $m = 2^{2t}$ *with* $t > 0$. *Then for all* $s \geq t-1$

$$t_m(RM(1, 2s+1)) = \frac{n + \sqrt{mn/2}}{2} = 2^{2s} + 2^{s+t-1}.$$

Finally, we consider odd values of r in the case $m = 2$. For $r = 1$, $RM(1, r)$ is the set of all length 2 vectors. This code is known to have 2-covering radius equal to 1 (Proposition 3, [7]).

In general, $RM(r-2, r)$ is the extended hamming code $\hat{\mathcal{H}}_r$. In particular, $RM(1, 3) = \hat{\mathcal{H}}_3$. The unextended Hamming codes with $r \geq 3$ are known to have 2-covering radius equal to 2^{r-1} (Proposition 5, [7]). Furthermore, extending a code by adding an overall parity check increases the m-covering radius by 1 (Corollary 1, [7]). Therefore

$$\begin{aligned} t_2(\hat{\mathcal{H}}_r) &= t_2(RM(r-2, r)) \\ &= 2^{r-1} + 1. \end{aligned}$$

In particular, $t_2(RM(1, 3)) = 5$.

THEOREM 3.6.

$$t_2(RM(1, 5)) = 18.$$

Proof. By Theorem 2.1, $t_2(RM(1, 5)) \leq 18$. We show that also $t_2(RM(1, 5)) \geq 18$.

We know that

$$RM(1, 5) = \{(\mathbf{c}, \mathbf{c}), (\mathbf{c}, \overline{\mathbf{c}}) \mid \mathbf{c} \in RM(1, 4)\}. \tag{8}$$

The code $RM(1, 4)$ has covering radius 6 and 2-covering radius 10. Let $\mathbf{y}^1$ be any point at distance 6 from $RM(1, 4)$ and $\mathbf{y}^2$ its complement. Without loss of generality, $\mathbf{y}^1$ has weight 6 and $\mathbf{y}^2$ has weight 10. By the proof of Theorem 3.1, we know that $\mathrm{dist}(\mathbf{y}^1, \mathbf{c}), \mathrm{dist}(\mathbf{y}^2, \mathbf{c}) \in \{6, 10\}$ for all $\mathbf{c} \in RM(1, 4)$. Now take

$$\mathbf{x}^1 = (\mathbf{y}^1, \mathbf{0}), \qquad \mathbf{x}^2 = (\mathbf{y}^2, \mathbf{1}),$$

where $\mathbf{0}$ and $\mathbf{1}$ are the all-zero and all-one words of length 16. The codewords $(\mathbf{0}, \mathbf{0})$, $(\mathbf{1}, \mathbf{0}) \in RM(1, 4)$ are at least distance $6 + 16$ from $\mathbf{x}^2$; and the codewords $(\mathbf{0}, \mathbf{1})$, $(\mathbf{1}, \mathbf{1}) \in RM(1, 4)$ are at least distance $6 + 16$ from $\mathbf{x}^1$. It therefore suffices to consider the other codewords

$(\mathbf{c},\mathbf{c}),(\mathbf{c},\overline{\mathbf{c}}) \in RM(1,4)$, where $c \neq \mathbf{0},\mathbf{1}$, i.e., c and $\overline{\mathbf{c}}$ each has weight 8. However,

$$\operatorname{dist}((\mathbf{c},\mathbf{c}),(\mathbf{y}^1,\mathbf{0})) = \operatorname{dist}(\mathbf{c},\mathbf{y}^1) + 8$$

and

$$\operatorname{dist}((\mathbf{c},\overline{\mathbf{c}}),(\mathbf{y}^2,\mathbf{1})) = \operatorname{dist}(\mathbf{c},\mathbf{y}^2) + 8,$$

so that one of these distances always equals 18 and the other 14.

□

In exactly the same way we can prove the following more general result. For $s = 1$ and $s = 2$ the lower and upper bounds coincide, but already for $s = 3$ we only get $68 \leq t_2(RM(1,7)) \leq 69$.

THEOREM 3.7.

$$2^{2s} + 2^{s-1} \leq t_2(RM(1,2s+1)) \leq 2^{2s} + \sqrt{2}\cdot 2^{s-1}.$$

The previous results and the monotonicity of $t_m(C)$ in m can be used to give lower and upper bounds for $t_m(RM(1,r))$ in the general case. Together Theorems 2.1, 3.4, and 3.5 give the following immediate corollary.

THEOREM 3.8. *For all $m \geq 2$ and $r \geq \lfloor \log_2(m) \rfloor - 2$,*

$$\begin{aligned} 2^{r-1} + \sqrt{m2^{r-5}} &= \frac{n + \sqrt{mn/8}}{2} \\ &\leq t_m(RM(1,r)) \\ &\leq \frac{n + \sqrt{mn/2}}{2} \\ &= 2^{r-1} + \sqrt{m2^{r-3}}. \end{aligned}$$

4. Higher Strength Codes

In this section we consider a strength t code C. This means that if we fix any t-tuple of indices in the codewords, then every binary t-tuple occurs the same number of times in these positions as we let the codewords vary. As it turns out, only the case when the strength is even leads to useful bounds.

THEOREM 4.1. *Let C be any strength $2s$ code of length n. Then*

$$t_m(C) \leq \frac{n}{2} + \left(\frac{(2s)!m}{2^{s+1}s!}\right)^{1/2s}\sqrt{n}.$$

Proof. Let $\mathbf{x} \in F_2^n$. Let

$$\Gamma_{\mathbf{x}} = \sum_{\mathbf{c} \in C} (2\mathrm{dist}(\mathbf{x}, \mathbf{c}) - n)^{2s}.$$

Then

$$\begin{aligned} \Gamma_{\mathbf{x}} &= \sum_{\mathbf{c} \in C} \left(\sum_{i=1}^{n} (-1)^{c_i + x_i} \right)^{2s} \\ &= \sum_{i_1, \cdots, i_{2s}} (-1)^{\sum_{j=1}^{2s} x_{i_j}} \sum_{\mathbf{c} \in C} (-1)^{\sum_{j=1}^{2s} c_{i_j}}. \end{aligned}$$

For any vector $\mathbf{i} = (i_1, \cdots, i_{2s})$ with $1 \leq i_j \leq n$ and index k, let $\tau_k(\mathbf{i})$ be the number j such that $i_j = k$. Then

$$\Gamma_{\mathbf{x}} = \sum_{\mathbf{i}} (-1)^{\sum_{k=1}^{n} \tau_k(\mathbf{i}) x_k} \sum_{\mathbf{c} \in C} (-1)^{\sum_{k=1}^{n} \tau_k(\mathbf{i}) c_k}.$$

For a given $\mathbf{i}$, if any $\tau_k(\mathbf{i})$ is odd, then the inner sum is zero (since C has strength $2s$). Otherwise the inner sum equals $|C|$. We want to bound the number of vectors $\mathbf{i}$ such that $\tau_k(\mathbf{i})$ is even for all k. All such $\mathbf{i}$ can be obtained in the following way: partition the set $\{1, 2, \cdots, 2s\}$ of indices into a union of s 2-element subsets and for each subset assign one of the values $1, 2, \cdots, n$. Since the partitioning can be done in exactly $(2s)!/(2^s s!)$ ways, there are at most

$$\frac{(2s)!}{2^s s!} n^s$$

such vectors $\mathbf{i}$. Consequently,

$$\Gamma_{\mathbf{x}} \leq |C| \frac{(2s)!}{2^s s!} n^s.$$

Now suppose we are given any m vectors $\mathbf{x}^1, \cdots, \mathbf{x}^m$. Then

$$\sum_{j=1}^{m} \Gamma_{\mathbf{x}^j} \leq m|C| \frac{(2s)!}{2^s s!} n^s$$

It follows that for some $\mathbf{c} \in C$,

$$\sum_{j=1}^{m} (2\mathrm{dist}(\mathbf{x}^j, \mathbf{c}) - n)^{2s} \leq m \frac{(2s)!}{2^s s!} n^s.$$

The theorem follows using the same argument as in the proof of Theorem 2.1.

□

Using Stirling's approximation, we see that for large s,

$$\left(\frac{(2s)!}{2^s s!}\right)^{1/2s} \sim \sqrt{s}.$$

Of course, trivially,

$$\frac{(2s)!}{2^s s!} \leq s^s,$$

which proves the following corollary

COROLLARY 4.2. *Let C be any strength $2s$ code of length n. Then*

$$t_m(C) \leq \frac{n}{2} + \sqrt[2s]{m/2}\sqrt{sn}.$$

5. Security of Stream Ciphers

We recall some of the definitions used previously to study the existence of secure stream ciphers [8].

DEFINITION 5.1.
A keystream generator *(or simply generator) is a 4-tuple* (S, F, g, s_0) *such that*

1. *S is a finite set (the* states*);*
2. *$F : S \to S$ is a function (the* state change function*);*
3. *$g : S \to \{0,1\}$ is a function (the* output function*); and*
4. *s_0 is an element of S (the* initial state*).*

A keystream generator outputs an infinite eventually periodic binary sequence by iterating the output and state change operations:

$$g(s_0), g(F(s_0)), g(F(F(s_0))), \cdots.$$

The *length* of a generator is $\lceil \log(|S|) \rceil$, the number of bits required to represent the states. We often use generators whose state space S is a set of n bit vectors $\mathbf{x} = (x_0, \cdots, x_{n-1})$ for some n, whence the length is n. We further generally use generators whose output functions are of the form $g(\mathbf{x}) = x_0$. In this case the generator is completely determined by F and s_0, and we often abuse the notation by identifying the generator with F. Any generator can be replaced by one of this form by possibly increasing the length by one.

The state change and output functions of our generators are described as circuits using binary gates. Such circuits can be encoded as binary strings [1]. The *size* of a generator F is the minimum number of gates in a circuit that computes the function F. This corresponds to evaluation time in a software implementation. The *depth* is the depth of the minimum depth circuit that computes F. This corresponds to evaluation time in a hardware implementation.

A *family of (keystream) generators,* $\mathcal{F}$, is an infinite collection of keystream generators. If B is an infinite eventually periodic binary sequence with eventual period period(B) and $0 < r \leq \text{period}(B)$, then a generator (F, s) with output sequence B' is said to *r-approximate* B if for any k,

$$|\{i, k \leq i \leq k + \text{period}(B) - 1 : b_i = b'_i\}| \geq r.$$

If $0 < r(p) \leq p$ is any function, then the $(\mathcal{F}, r)$-span of B is the least integer n such that B can be $r(\text{period}(B))$-approximated by a generator in $\mathcal{F}_n$ (or ∞ if there is no such n). The $(\mathcal{F}, r)$-span of B is denoted by $\lambda_{\mathcal{F},r}(B)$.

Let $\delta(n)$ be the maximum over all length n generators F in $\mathcal{F}$ of the depth of F. We say $\mathcal{F}$ is

1. *fast* if $\delta(n) \in \mathcal{O}(\log(n))$.
2. *short* if whenever $F \in \mathcal{F}$ generates sequence B, then $\lambda_{\mathcal{F},m}(B)$ is $\mathcal{O}(\log(\text{period}(B)))$.

DEFINITION 5.2. *Let T be an algorithm, let $\mathcal{F}$ be a family of generators, and let $0 < r(p) \leq p$. We say that T is an r-*effective $\mathcal{F}$-synthesizing algorithm *if*

1. *it runs in polynomial time;*
2. *when given the input $b_0, \cdots, b_{k-1}$, T outputs the encoding of a generator $(F, s) \in \mathcal{F}$ such that the first k output bits of F with initial state s are $b_0, \cdots, b_{k-1}$; and*
3. *there is a polynomial $g(n)$ such that if B is any eventually periodic sequence and $n = \lambda_{\mathcal{F},r}(B)$, then on input $b_0, \cdots, b_{k-1}$ with $k \geq g(n)$, T outputs $F \in \mathcal{F}$ of length n that r-approximates B.*

THEOREM 5.3. *Let $h(n)$ be subexponential (in the sense that $h \in \mathcal{O}(a^n)$ for every $a > 1$), let $\epsilon > 1/2$, and let*

$$r(p) = \frac{p}{2} + p^{\epsilon}.$$

There exists a collection $\mathcal{B}$ of sequences such that

1. *$\mathcal{B}$ can be generated by a family $\mathcal{F}$ of fast short generators; and*

2. *for every r-effective register synthesizing algorithm T generating a family $\mathcal{F}'$ of registers, and for all but finitely many sequences $B \in \mathcal{B}$*

$$\lambda_{\mathcal{F}',r}(B) \geq h(\log(\text{period}(B))).$$

Proof. Let $T_1, T_2, \cdots$ be an enumeration of the r-effective register synthesizing algorithms. We construct the sequence $\mathcal{B}$ in stages. At the mth stage we construct B^m which simultaneously has large Hamming distance from every sequence generated by $T_1, \cdots, T_m$ with input a large enough prefix of B^m. We do so using Reed-Muller codes and our bounds on their multi-covering radii.

Recall that a *linear feedback shift register* (LFSR) of length k is a keystream generator with state set F_2^k, state change function of the form

$$F(x_1, \cdots, x_k) = (x_2, \cdots, x_k, f(x_1, \cdots, x_k))$$

for some linear function f, and with output function

$$g(x_1, \cdots, x_k) = x_1.$$

The function f can be computed in depth $\log(k)$. For every k there are LFSRs whose output sequence has period $2^k - 1$. Such sequences are known as *m-sequences.*

The output from the generators we construct consists of an m-sequence of period $2^k - 1$ followed by a $RM(n, 1)$ codeword $\mathbf{c}$ for some k and n. The first step is to see that it is possible to construct a fast short generator that outputs such a sequence. The generator consists of two parts: an LFSR that generates the m-sequence, and a generator that outputs $\mathbf{c}$. The overall generator can be made to output the m-sequence, then switch to the generator of $\mathbf{c}$. This is accomplished by detecting the last state of the LFSR with an AND of k bits. This takes depth $\log(k)$. When $\mathbf{c}$ has been output, the generator switches back to the LFSR similarly.

Let $\mathbf{c} = \mathbf{c}^{\mathbf{v},b} \in RM(1,n)$, $\mathbf{v} \in F_2^n$, $b \in F_2$. We can construct a generator such that one period is $\mathbf{c}$ by modifying a maximum period LFSR. We first modify a LFSR of period $2^n - 1$ so it enters the all zero state after the state $(1, 0, \cdots, 0)$ and enters the state $(0, \cdots, 0, 1)$ after the all zero state. This requires at most depth $\log(n)$ and one extra bit of state. Now $\mathbf{c}$ can be generated by the output function

$$g(x_1, \cdots, x_n) = \sum_{i=1}^{n} v_i x_i + b.$$

This function can be computed by a circuit of depth $\mathcal{O}(\log(n))$.

The combined sequence, one period of which is the m-sequence of period 2^k-1 followed by $\mathbf{c}$, has period 2^n+2^k-1 and is generated by a generator of length $n+k+3$ (two extra bits are used for output and for switching between the two modes of operation) and depth

$$\mathcal{O}(\log(n)+\log(k)) = \mathcal{O}(\log(n+k+3)).$$

Furthermore,

$$\begin{aligned} n+k+3 &\leq 3\max(n,k) \\ &\leq 3\log(2^n+2^k-1), \end{aligned}$$

so this is a fast short generator.

We want sequences that are far from given sequences, but in the Hamming metric, if $\mathbf{c}$ is close to $\mathbf{b}$, then the complement $\overline{\mathbf{c}}$ of $\mathbf{c}$ is far from $\mathbf{b}$: $\mathrm{dist}(\overline{\mathbf{c}},\mathbf{b}) = 2^n - \mathrm{dist}(\mathbf{c},\mathbf{b})$ if the length of the code is 2^n. The first order Reed-Muller code is closed under complementation, so by Theorem 2.1, there is a first order Reed-Muller codeword whose distance from every element of any given set of m sequences of length 2^n is at least

$$\frac{2^n}{2} - \frac{\sqrt{2mn}}{4}.$$

For each r-synthesis algorithm T_i, let $\mathcal{F}^i$ be the family of generators that is output by T_i. We assume that T^i is successful when given

$$g_i(\lambda_{\mathcal{F}^i,r}(B))$$

bits of any sequence B, with g_i a polynomial. At the mth stage of the diagonalization we want to find a fast generator F_m, as described above, with output B^m so that $\lambda_{\mathcal{F}^i,r}(B^m)$ is large for every $i=1,\cdots,m$.

Let $g_i(x) < x^\ell$ for $i=1,\cdots,m$. Let

$$a(k) = \max\left(\frac{k}{\epsilon}, \frac{\log(m)}{2\epsilon-1}\right).$$

Let $k' = a(k)+k$ be large enough that

$$h(k') < (2^k-1)^{1/\ell}$$

and

$$p = 2^k + 2^{a(k)} - 1,$$

is larger than the period of any sequence B^j, $j=1,\cdots,m-1$. Let $n=a(k)$. Thus the generator constructed above first generates an m-sequence of period

$$\begin{aligned} 2^k - 1 &> h(a(k)+k)^\ell \\ &> g_i(h(a(k)+k)) \\ &\geq g_i(h(\log(2^k + 2^{a(k)} - 1))) \\ &= g_i(h(\log(p))). \end{aligned}$$

It follows that if a sequence B of period p satisfies

$$\lambda_{\mathcal{F}^i,r}(B) < h(\log(p)),$$

$i = 1, \cdots, m$, and agrees with the given m-sequence on its first $2^k - 1$ bits, then the sequence generated by the output of T^i, $i = 1, \cdots, m$, must agree with B on at least $r(p)$ bits.

We choose a first order Reed-Muller codeword $\mathbf{c} \in RM(n,1)$ so that whatever sequence T^i outputs given the $2^k - 1$ bits of the initial m-sequence, $i = 1, \cdots, m$, the last 2^n bits disagree with the codeword on at least

$$\frac{2^n}{2} - \frac{(2m2^n)^{1/2}}{4}$$

bits. Thus the output of T^i is correct on at most

$$2^k - 1 + \frac{2^n}{2} + \frac{(2m2^n)^{1/2}}{4} < r(2^n + 2^k - 1)$$

bits by the choice of n and k. Let B^m be the sequence one of whose periods is the m-sequence followed by the codeword $\mathbf{c}$. Then

$$\begin{aligned} \lambda_{\mathcal{F}^i,r}(B^m) &> h\left(\log\left(\frac{2^n + 2^k - 1}{2}\right)\right) \\ &> h(\log(\mathrm{period}(B^m))) \end{aligned}$$

for $i = 1, \cdots, m$.

□

6. Conclusions

We have given bounds on the multi-covering radii of first order Reed-Muller codes and have shown that these bounds are tight in some cases. It is apparent from our lower bounds, however, that even if the bounds are tightened our cryptographic results will not be improved. This would require reducing the $2^{n/2}$ term in the bounds. This may be possible using bounds on the multi-covering radii of the dth order

Reed-Muller codes, $d > 1$. Thus we leave finding such bounds as an interesting open problem.

References

1. J. Balcázar, J. Díaz, and J. Gabarró, *Structural Complexity I*, Springer-Verlag, Berlin (1988).
2. G.D. Cohen, M.G. Karpovsky, H.F. Mattson, Jr., and J.R. Schatz, Covering radius – survey and recent results, IEEE Trans. Info. Theory, Vol. IT-31 (1985) pp. 328-343.
3. G. Cohen, I. Honkala, S. Litsyn and A. Lobstein, *Covering Codes*, North-Holland, Amsterdam (1997).
4. G. Cohen and S. Litsyn, On the covering radius of Reed-Muller codes, Discrete Mathematics, Vol. 106-107 (1992) pp. 147-155.
5. G.D. Cohen, S.N. Litsyn, A.C. Lobstein, and H.F. Mattson, Jr., *Covering Radius 1985-1994,* Dept. Informatique, Ecole Nationale Supérieure des Télécommunications, Technical Report 94 D 025 (1994).
6. T. Helleseth, T. Kløve, and J. Mykkeltveit, On the covering radius of binary codes, IEEE Trans. Info. Theory, Vol. IT-24 (1978) pp. 627-628.
7. A. Klapper, The Multicovering radii of codes, IEEE Trans. Info. Theory, Vol. 43 (1997) pp. 1372-1377.
8. A. Klapper, On the existence of secure keystream generators, to appear, J. Cryptology.
9. F.J. MacWilliams and N.J.A. Sloane, *The Theory of Error Correcting Codes,* Amsterdam: North-Holland (1977).
10. O. Rothaus, On 'bent' functions, J. Combin. Thy., Se. A, vol. 20 (1976) pp. 300-305.

Footnotes

Affiliation of author 1:
Dept. of Mathematics, University of Turku, 20014 Turku, Finland

Affiliation of author 2:
Dept. of Computer Science, 763H Anderson Hall, University of Kentucky, Lexington, KY, 40506-0046.

* Project sponsored by the National Science Foundation under grant number NCR-9706078.

Mailing Address:

Andrew Klapper
Department of Computer Science
763H Anderson Hall
University of Kentucky
Lexington, KY 40506-0046
USA
e-mail:klapper@cs.uky.edu
phone: (606) 269-7218
fax: (606) 323-1971