The average radius of codes: survey and new results

Gérard Cohen Carlos Munuera Patrick Solé

***Abstract*—The average radius of a block code is a parameter that occurs naturally in quantization and steganography. We give asymptotic upper and lower bounds on this parameter. In particular we show that for almost all long codes the normalized average radius equals the normalized covering radius. We survey some special graph-theoretic lower bounds.**

***Index Terms*—steganography, covering radius, probability of error**

I. Introduction

The **average radius** of a binary code is the average distance of a vector in ambient space to the code.

$$\tilde{R} = \frac{1}{2^n} \sum_{x \in \mathbb{F}_2^n} d(x, C). \tag{1}$$

It should not confused with the **covering radius**

$$R = \max_{x \in \mathbb{F}_2^n} d(x, C),$$

a parameter that has received a lot of attention in the past [4]. Following [8, Chap. 1, §5], let us denote by α_i the number of cosets of weight i. Then, for a linear $[n, k]$ code the defining formula (1) becomes

$$\tilde{R} = \frac{1}{2^{n-k}} \sum_{i=1}^{R} i\alpha_i. \tag{2}$$

This quantity occurs in steganography as the average number of changes made in the cover to realize an embedding [3], [5], [9]. It also controls the average distortion in vector quantizing when using the code as a codebook [12]. An immediate consequence of the above definitions is that

$$\tilde{R} \leq R,$$

In Section 4, we shall see that this bound is asymptotically tight. However, non asymptotic improvements are possible as in Proposition 2. Consequently, most of this paper will be dedicated to lower bounds on $\tilde{R}$, either direct or by graph theoretic arguments as in Section 5. Since steganography requires coset coding, we shall limit ourselves to *linear* binary codes, and Gray images of $\mathbb{Z}_4$-codes. The material is organized as follows. Section 2 collects the necessary notation and definitions. Noteworthy is a Lemma on the average radius of the Hamming ball. Section 3 contains combinatorial bounds. Asymptotic bounds are in Section 4. Section 5 surveys graph theoretic lower bounds from the nineties. Section 6 compiles some exact values of $\tilde{R}$ computed from known coset weight distribution. Section 7 casts our results into perspective and points out some challenging open problems.

GC and PS are with CNRS/LTCI, Telecom ParisTech, 46 rue Barrault, 75634 Paris Cedex 13, France.CM is with Department of Applied Mathematics, University of Valladolid, Avda Salamanca SN, 47014 Valladolid, Castilla, Spain.

II. Notation and definitions

The notation tends to follow that of [4]. Greek letters normally denote normalized quantities. The parameters of a linear code are length n, dimension $k = n\kappa$, minimum distance $d = n\delta$, covering radius $R = n\rho$, average radius $\tilde{R} = n\tilde{\rho}$. The **packing radius** is $\lfloor \frac{d-1}{2} \rfloor$, and its normalized version $\delta/2$. The parameters of C are collectively denoted by $[n, k, d, R, \tilde{R}]$ and those of its dual $C^\perp$ by $[n, n-k, d^\perp, R^\perp, \tilde{R^\perp}]$. For a set $B \subseteq \mathbb{F}_2^n$ and a vector $x \in \mathbb{F}_2^n$ we denote by $\tilde{d}(x, B)$ the average distance from $b \in B$ to x. Formally

$$\tilde{d}(x, B) = \frac{1}{|B|} \sum_{b \in B} d(x, b).$$

Of special importance is the case of the **mean radius of a Hamming ball**

$$b(n, t) = \tilde{d}(0, B(0, t)),$$

where $B(0, t)$ denotes the Hamming ball of radius t and center 0. The following two results are proved in [9, §5.4].

Lemma 1: If $V(n, t) = |B(0, t)| = \sum_{i=0}^{n} \binom{n}{i}$, then

- $$b(n, t) = n \frac{V(n-1, t-1)}{V(n, t)}.$$

978-1-4577-0595-3/11/$26.00 ©2011 IEEE

- $$\lim_{n\to\infty} \frac{b(n,\tau n)}{n} = \tau,$$

for all $0 < \tau < 0.5$.

Proof: The first point is easily proved by considering the generating function

$$f_n(x) = \sum_{j=0}^{t} \binom{n}{j} x^j$$

and computing $f_n'(1)/f_n(1)$. The second point follows from

$$\lim_{n\to\infty} \frac{b(n,\tau n)}{n} = \lim_{n\to\infty} \frac{\binom{n-1}{\lfloor \tau n \rfloor - 1}}{\binom{n}{\lfloor \tau n \rfloor}} = \lim_{n\to\infty} \frac{\lfloor \tau n \rfloor}{n}.$$

■

By a **Voronoi region** B for the $[n,k]$ code C we shall mean a set of 2^{n-k} coset leaders that constitute a system of distinct representatives for the 2^{n-k} cosets of C into ambient space. This region is, in general, non unique. A Voronoi region is **hereditary** iff it is stable by inclusion for supports. Note that, since the translates of B by elements of C tile the ambient space we get, for all $c \in C$, the property

$$\tilde{d}(c, B + c) = \tilde{R},$$

and, in particular

$$\tilde{d}(0, B) = \tilde{R}.$$

III. Combinatorial bounds

Define the defect to the Hamming bound of an $[n,k]$ code C by the expression

$$HB(n,k,t) = 1 - 2^{k-n} V(n,t).$$

Note that $HB(n,k,t) = 0$ iff C is perfect.

Proposition 1: If the code C has packing radius t then

$$\tilde{R} \geq 2^{k-n} V(n,t) b(n,t) + (t+1) HB(n,k,t),$$

with equality iff the code is perfect or quasi perfect $(R = t+1)$.

Proof: Let

$$S_t = \bigcup_{c\in C} B(c,t).$$

Clearly $|S_t| = 2^k V(n,t)$. Note that if $x \in B(c,t)$, for some $c \in C$ then $d(x,C) = d(x,c)$. On the other hand, if $x \notin S_t$, then $d(x,C) \geq t+1$. From this discussion we see that, on average,

$$\tilde{R} \geq \frac{[|S_t| b(n,t) + |\mathbb{F}_2^n \setminus S_t|(t+1)]}{2^n}.$$

The result follows upon noticing that

$$|\mathbb{F}_2^n \setminus S_t| = 2^n HB(n,k,t).$$

■

An upper bound that sharpens the trivial $\tilde{R} \leq R$ is the following

Proposition 2: If the code C has packing radius t and covering radius R then

$$\tilde{R} \leq 2^{k-n} \left(V(n,R) b(n,R) - \sum_{i=t+1}^{R} i \binom{n-t-i}{i} \right),$$

Proof: As is well-known $\alpha_i = \binom{n}{i}$ for $i \leq t$. On the other hand, since C is not $t+1$-correcting, $\alpha_{t+1} \leq \binom{n}{t+1} - 1$. By using a hereditary Voronoi [4, Thm. 2.414] this bound can be generalized in the case of $t+1 \leq i \leq R$ we have

$$\alpha_i \leq \binom{n}{i} - \binom{n-t-1}{i-t-1}.$$

Indeed a missing coset leader of weight $t+1$ has $\binom{n-t-1}{i-t-1}$ descendants of weight i. The result follows then after substitution into (2). ■

Remark: By using the formula [8, Chap. 1 §5 (24)]

$$1 - P_e = \sum_{i=0}^{n} \alpha_i p^i (1-p)^{n-i}.$$

we see that the estimates on the α_i's in the above proof can be useful to estimate word error rates P_e in the context of transmission on a BSC with transition probability p.

IV. Asymptotics

Let

$$h(x) = -x \log_2(x) - (1-x) \log_2(1-x)$$

denote the binary entropy function and let $h^{-1}(\cdot)$ denote the inverse function over the range $[0, 1/2]$. The following Lemma is the contrapositive of [9, Lemma 5.11 (3)].

Lemma 2: If t is such that $V(n,t) \leq |B|$, then $\tilde{d}(0,B) \geq b(n,t)$.

Proof:

W.l.o.g. we may assume that $|B| = V(n,t)$. Let $a = |B \bigcap B(0,t).|$ We see that then

$$\tilde{d}(0,B) \geq \tilde{d}(0,B(0,t)) + a/|B|.$$

■

Its asymptotic counterpart is as follows.

Proposition 3: Let $0 < \theta < 1$. If $|B| > V(n,n\theta) \approx 2^{nh(\theta)}$, then $\tilde{d}(0,B) \geq n\theta$.

Proof: Combine Lemma (2) with Lemma (1). ■

Let us now use known upper bounds of [7, §2.10] written as $1-\kappa > h(f(\delta))$ to get lower bounds on $\tilde{R} = n\tilde{\rho}$ of the form $\tilde{\rho} \geq f(\delta)$. Thus, we get from the

- Hamming bound: $\tilde{\rho} \geq \delta/2$.
- Elias bound:$\tilde{\rho} \geq 1/2 - (1/2)(1-2\delta)^{1/2} \geq \delta/2$.
- first MRRW bound:

$$\tilde{\rho} \geq h^{-1}(1 - h(1/2 - (\delta(1-\delta))^{1/2}$$

We require a well-known but deep theorem from [4]. We shall say that a property holds for almost all codes (a.a.) iff for large n the proportion of codes of length n with that property is $> 1 - o(1)$.

Theorem 1: Almost all codes of parameters $[n, nr, n\delta, n\rho]$ lie on the Varshamov Gilbert bound for minimum distance and on the sphere covering bound for covering radius, i.e satisfy:

$h^{-1}(1-\kappa) - o(1) \leq \rho \sim \delta \leq h^{-1}(1-\kappa) + o(1)$.

In particular, for those codes:

$$1 - \kappa \geq h(\delta) - o(1) \tag{3}$$

Proof: The assertion for the covering radius comes from [4, Th. 12.3.10]. That for minimum distance from [1, Lemma 1.2]. ■

We are now ready for the main asymptotic result.

Theorem 2: Almost all codes satisfy

$$\delta - o(1) \leq \tilde{\rho} \leq \rho \ \leq \delta + o(1).$$

Proof: We take for B a Voronoi region of C. Then, by (3) for a.a. codes:

$|B| \geq 2^{n(h(\delta)-o(1))}$ and, by Proposition 3 we obtain

$n^{-1}\tilde{d}(0,B) \geq \delta - o(1)$.

But, by the properties of the Voronoi region

$$n^{-1}\tilde{d}(0,B) \sim \tilde{\rho},$$

so that, passing to the limit on n we get $\tilde{\rho} \geq \delta$. ■

Remark: There are infinite families of codes for which $\rho \neq \tilde{\rho}$ for large n. By performing i times a direct product of parity-check matrices, from $C^{\perp}[n, n-k, d, R]$, one gets $C_i^{\perp}[ni, (n-k)i]$; reverting to the primal code, this gives $C_i[ni, ki, d, Ri]$.

For concreteness let $C = BCH_2[15, 7, 5, 3]$ By Proposition 1 $\tilde{R} \approx 2.46$; this gives $C_i[15i, 7i, 5, 3i, 2.46i]$. Indeed $R - \tilde{R} = 0.54i$ goes to infinity, and for the normalized quantities:

$$\rho - \tilde{\rho} > 0.54i/15i = 0.036.$$

V. Graph theoretic approach

The theme of this section is that codes with high symmetry have a high average radius. Let $G(C)$ denote the automorphism group of C. It was proved by graph theoretic methods in [12] that

Theorem 3: If $G(C)$ acts transitively on the coordinates of C, then

$$\tilde{R} \geq \frac{n}{2d^{\perp}}.$$

In particular the conclusion holds for cyclic codes.

Define the coset graph $\Gamma(C)$ as the undirected Cayley graph with vertex set the cosets of C; two cosets being connected iff they differ by a coset of weight one. It is well known that the diameter of this graph is none other than the covering radius of the code [2]. In [12] it is proved that the mean distance $\overline{D}$ is related to the average radius by the formula

$$\tilde{R} = \frac{2^{n-k}-1}{2^{n-k}}\overline{D}.$$

Recall that, following [10], a graph is **orbital regular** if its automorphism group has a subgroup that acts regularly on the set of edges, and, more generally on its orbits on ordered pairs of vertices.

Theorem 4: If $\Gamma(C)$ is orbital regular then

$$R \leq 2\lceil \tilde{R}(n-k-1)\log_e(2)\frac{2^{n-k}}{2^{n-k}-1}\rceil.$$

In particular, the result is true for C cyclic without words of period $< n$.

Proof: The first assertion follows from [11, Cor. 1] and [10, Thm. 2.2] applied to $\Gamma(C)$. The second assertion follows from [11, Thm. 6]. ■

Fig. 1: Exact values of $\tilde{R}$

Code	n	$n-k$	t	$\tilde{R}$	R	Ref.
Hamming	2^m-1	m	1	$1-2^{-m}$	1	[12]
Ext. Hamming	2^m	$m+1$	1	$1.5-2^{-m}$	2	[12]
Doubly trunc. Golay	21	9	2	2.502	3	[2]
Trunc. Golay	22	10	2	2.73	3	[2]
Golay	23	11	3	2.85	3	[2]
$RM(2,5)$	32	16	3	4.33	6	[7, §11.4]
D_{40}	40	20	3	5.34	7	[6]
Ext. Golay	24	12	3	3.35	4	[2]
$BCH(2)$	2^m-1	$2m$	2	$2.5-\frac{2^{m-2}+1}{2^{2m-1}}$	3	Prop. 1
Ext. $BCH(2)$	2^m	$2m+1$	2	$3-(2^m+4)2^{-2m-1}$	3	Prop. 1
Preparata	$2^{2m}-1$	$2m-1$	2	$2+(2^{m-1}+1)2^{-2m+1}$	3	Prop. 1

VI. Exact values of $\tilde{R}$

In this section we extend the table of [12] by using Proposition 1 as well as coset weight distribution from the literature. By t we denote the error correcting capacity. The example of the cyclic Golay shows that we might have $\tilde{R} < t$ for some codes. The terminology "truncated Golay" of [2] means punctured in the sense of [8].

VII. Conclusion and open problems

In the present paper steganography directed our investigation towards a relatively unexplored invariant of a code: the average radius. It is related to, but different from, the covering radius. Both asymptotics and numerical values show that the codes with the lowest average radius for given codimension are codes that are both good packings and good coverings. Indeed, if two codes have the same covering radius the one with the larger packing radius will have the lower average radius. Thus the Preparata codes of the preceding section have a lower average radius than any linear code of the same length, covering radius and codimension.

There are many open problems and directions. For instance, generalizing the bounds of Section V to unrestricted codes is a possible direction. Deriving upper bounds on $\tilde{R}$ as a function of the dual distance that do not come from the immediate $\tilde{R} \leq R$ is a challenge. In general all the problems and techniques that enter the study of the covering radius are worthy of consideration to study the average radius. Considering other alphabets and metrics might lead to interesting problems as well.

References

[1] A. Barg, Complexity isues in Coding theory, in *Handbook of Coding Theory*, vol. I, V.S. Pless, W.C. Huffman eds, North Holland (1998).

[2] Brouwer, A. E., Cohen, A. M.; Neumaier, A. *Distance-regular graphs.* Ergebnisse der Mathematik und ihrer Grenzgebiete (3) , 18. Springer-Verlag, Berlin, 1989.

[3] J. Bierbrauer, J. Fridrich, Constructing good covering codes for applications in steganography, Lecture Notes in Computer Science, Trans. Data Hiding Multimedia Security III (ed. Y. Q. Shi), Springer-Verlag, vol. 4920 (2008) 1–22.

[4] G.Cohen, I.Honkala, S.Litsyn, A.Lobstein, *Covering Codes*, Elsevier, 1997.

[5] F. Galand, G. Kabatiansky, Steganography via Covering Codes, Proceedings of ISIT 2003, Yokohama.

[6] M. Harada, M Ozeki, Extremal self dual codes with the smallest covering radius, Discrete Math 215 (2000) 271–281.

[7] W. Cary Huffman, Vera Pless *Fundamentals of error correcting codes*, Cambridge (2003).

[8] MacWilliams, F. J.; Sloane, N. J. A, *The theory of Error Correcting Codes*, North Holland (1977).

[9] C. Munuera, Steganography from a coding theory point of view, Algebraic Geometric Modelling in Information Theory., World Scientific, to appear.

[10] P. Solé, The edge-forwarding index of orbital regular graphs, Discrete Math. 130 (1994), no. 1-3, 171–176.

[11] P. Solé, Expanding and forwarding. Discrete Appl. Math. 58 (1995), no. 1, 67–78.

[12] P. Solé, J-P. Tillich, On the dual distance and the gap of a binary code. Discrete Math. 192 (1998), no. 1–3, 333–336.