Constructing good covering codes for applications in Steganography

Jürgen Bierbrauer
Department of Mathematical Sciences
Michigan Technological University
HOUGHTON (MI) 49931 (USA)
Jessica Fridrich
Department of Electrical and Computer Engineering
Binghamton University
BINGHAMTON (NY) 13902-6000

October 20, 2006

Abstract

Application of covering codes to data embedding improves embedding efficiency and security of steganographic schemes. In this paper, we describe several familes of covering codes constructed using the blockwise direct sum of factorizations. We show that non-linear constructions offer better performance compared to simple linear covering codes currently used by steganographers. Implementation details are given for a selected code family.

1 Introduction

Steganography is the art of stealth communication. Its purpose is to make communication undetectable. The steganography problem is also known as the prisoners' dilemma formulated by Simmons [28]. Alice and Bob are imprisoned and want to hatch an escape plan. They are allowed to communicate via a channel monitored by a warden. If the warden finds out that they are communicating secretly, he throws them into solitary confinement. Thus, the prisoners need to design a method to exchange messages without raising the warden's suspicion.

The prisoners hide their messages in innocuous-looking cover objects by slightly modifying them (obtaining stego objects). The embedding process is usually driven by a stego key, which is a secret shared between Alice and Bob. It is typically used to select a subset of the cover object and the order in which the cover object elements are visited during embedding.

The most important property of any steganographic communication is statistical undetectability. In other words, the warden should not be able to distinguish between cover and stego objects. Formal description of this requirement in information-theoretic terms was given by Cachin [4]. If the communication channel that Alice and Bob use is distortion-free, we speak about the passive warden scenario.

Digital multimedia files, such as digital images, audio, or video, are conveniently used in steganography today because they are digitized forms of physical quantities, such as photon counts or voltages, and thus contain certain small level of noise. Because of the presence of this indeterministic component, steganographers hope that part of this component can be replaced with pseudo-random (e.g., encrypted) message bits, thus obtaining a secure steganographic method.

Intuitively, the fewer changes the embedding process incurs, the smaller the chance that the embedding modifications will be detectable. We acknowledge, though, that the number of changes is not the only important factor influencing the security of the steganographic scheme. The choice of the cover object and the character of modifications play an equally important role. For example, it is known that embedding in spatial domain of a decompressed JPEG image can be easily detectable even when only one embedding change is carried out [15]. Furthermore, the impact of embedding realized by flipping LSBs of pixels (Least Significant Bit) is relatively easy to detect even at very low embedding rates [22]. Nevertheless, it is true that for two steganographic schemes with the same embedding mechanism, the one that introduces fewer embedding changes will be more secure.

Steganographers use the concept of embedding efficiency to quantify how effectively a given algorithm embeds data. The embedding efficiency is defined [32] as the average number of random message bits embedded using one embedding change. There is evidence that schemes with low embedding efficiency offer worse security than schemes with higher efficiency. For example, the popular JPEG steganography program OutGuess [26] embeds messages in DCT coefficients (Discrete Cosine Transform) in two passes. In the first pass, it embeds with efficiency 2 by matching the LSBs of DCT coefficients with message bits. In the second pass, more changes are imposed on the previously non-visited DCT coefficients. While this has the benefit of preserving the global DCT histogram, the embedding efficiency decreases significantly. On the other hand, the Model based Steganography (MBS) [27] without deblocking preserves even more statistics than OutGuess and does so at a higher embedding efficiency. Steganalysis of both schemes [14] indicates that MBS is significantly harder to detect than OutGuess.

The importance of high embedding efficiency for steganography and the relevance of covering codes to this problem were recognized for the first time by Crandall [7], who showed that linear codes can markedly improve the embedding efficiency. He called this type of embedding "matrix embedding", which was made popular in the stego community by Westfeld in his F5 algorithm [32].

Crandall refers to an unpublished article by Bierbrauer [2] that provides deeper insight into this problem from the point of view of a coding theorist. The connection between linear covering codes and steganography has also appeared in the paper by Galand and Kabatiansky [17] who addressed both the passive and active warden scenarios.

In this paper, we describe and extend the original Bierbrauer's work. We believe that the steganographic community will benefit from this work as it formulates the problem of embedding efficiency in coding-theoretic language and makes a connection with a large body of work in coding. Moreover, we point out the importance of certain families of codes to steganography and show that non-linear codes have better performance than known linear constructions, e.g., matrix embedding.

In Section 2, the connection between covering functions and steganography is formally established. Coding-theoretic bounds and constructions are subject of the following two Sections 3 and 4. In Section 5, we study some good families of non-linear codes and in Section 6 we give specific examples of the best known covering functions. The embedding efficiency of steganographic schemes that use these covering functions is compared and contrasted to theoretical bounds in Section 7. To enable practical implementation of steganographic schemes that use the proposed constructions, in Section 8 we describe the details of the non-linear Nordstrom-Robinson code and some covering functions related to it. The paper is concluded in

Section 9.

2 The link to coding theory

For concreteness, we assume that the cover object used for communication is a grayscale digital image whose pixels are integeres between 0 and 255. We assign a bit to each pixel value (the LSB of the grayscale value). We will further assume that the embedding mechanism is flipping the LSB, while stressing that other embedding operations or bit assignments are certainly possible. We also assume that the sender can use all pixels for embedding, i.e., the embedding is not constrained to any selection channel [13]. Possible directions one can take for application of covering codes to non-trivial selection channels (wet paper codes) were briefly discussed in [13].

Let us assume that the embedding proceeds by blocks. The cover image is divided into disjoint segments of N pixels. Let $x = (x_1, x_2, \ldots, x_N)$ be the bitstring formed by their least significant bits. Here we view the entries, the bits, as elements of the field $\mathbb{F}_2 = \{0, 1\}$. Formally we can write $x \in \mathbb{F}_2^N$. Assume the secret message has been encoded as a bitstring. We scan this bitstring and divide it into segments of length n, for some number $n < N$. What we want to construct is a suitable function f, which maps bitstrings of length N to bitstrings of length n, formally

$$f : \mathbb{F}_2^N \longrightarrow \mathbb{F}_2^n,$$

which allows us to extract n bits of the secret message. This means that for given $x \in \mathbb{F}_2^N$ (the LSBs of the corresponding segment of the cover image) and $y \in \mathbb{F}_2^n$ (a segment of the secret message) we want to replace x by x' such that $f(x') = y$. An important question is the relation between x and x'. If x and x' differ in 3 of their N coordinates, then that means that 3 of our segment of N pixels need to be changed. Our goal is to keep that number of changes to a minimum – maximize the embedding efficiency. The number of coordinates where the entries of two strings x, x' differ is a basic notion of coding theory. It is the **Hamming distance** $d(x, x')$. If we want to control the worst case, then we fix an upper bound ρ on the embedding distortion $d(x, x')$. This leads to the following notion:

Definition 1. *A* **covering function** $COV(\rho, N, n)$ *is a mapping*

$$f : \mathbb{F}_2^N \longrightarrow \mathbb{F}_2^n$$

which satisfies the following: for every $x \in \mathbb{F}_2^N$, $y \in \mathbb{F}_2^n$ there is some $x' \in \mathbb{F}_2^N$ such that $d(x, x') \leq \rho$ and $f(x') = y$.

The basic question is then: for which ρ, N, n do $COV(\rho, N, n)$ exist? Call N the **length,** n the **redundancy** and ρ the **covering radius.** The following design problems arise:

- We want n/N, the **relative redundancy** to be large (large embedding capacity).
- We want ρ/N, the **relative covering radius,** to be small to have good embedding efficiency.
- Finally, there should be an effective algorithm that calculates x'.

We now translate the above into the terminology of coding theory. In coding theory, a **code** is defined simply as a subset of the space of all tuples of a certain length N over some alphabet, where N is the **length** of the code. We speak of a binary code if the alphabet has two elements. Historically, coding theory developed in the context of information transmission over noisy channels. Typically in these applications the most important parameter is the **minimum distance** d : any two different elements of the code should be at Hamming distance $\geq d$, in other words: if two elements of the code (**codewords**) are different, then they are very different. In our context, the basic parameter is the covering radius:

Definition 2. *Let $\mathcal{C} \subset \mathbb{F}_2^N$. The* **covering radius** *of the code $\mathcal{C}$ is the smallest number ρ such that any N-tuple is at Hamming distance $\leq \rho$ from some codeword.*

Informally, one speaks of error-correcting codes if the minimum distance is the important parameter, of covering codes if one is more interested in the covering radius. While the minimum distance concerns only the distances between codewords (in a way it ignores the ambient space $\mathbb{F}_2^N$), the covering radius is defined in terms of the embedding of the code in its ambient space.

Definition 1 demands that the inverse image $f^{-1}(y)$ be a covering code of radius ρ for every choice of $y \in \mathbb{F}_2^n$. It follows that $\mathbb{F}_2^N$ is the disjoint union of 2^n such covering codes. Clearly, this is an equivalent description of a covering function:

Theorem 1. *The following are equivalent:*

- *A covering function $COV(\rho, N, n)$.*
- *A partition of $\mathbb{F}_2^N$ into 2^n covering codes of covering radius ρ.*

The notion of covering functions was introduced in [2, 1] in a slightly more general form, using arbitrary alphabets. Definition 1 is the binary case. Covering codes are classical objects in coding theory. A recent book on the subject is *Covering codes* by Cohen, Honkala, Litsyn and Lobstein [6]. By Theorem 1 a covering function is equivalent with a partition of the ambient space into covering codes. These partitions have been studied by the coding community. In Etzion-Greenberg [11], they appear under the name *covering by coverings.*

There is also a graph-theoretic link. In fact, the graph-theoretic problem is more general. It applies to any graph G. The problem we are interested in arises as the special case when G is a Hamming graph (the vertices are the bitstrings of length N, two bitstrings form an edge if their distance is 1). A ρ-**dominating** set D of graph G is defined as a set of vertices, such that each vertex of G has distance $\leq \rho$ from a vertex of D. The ρ-**domatic number** of graph G is the smallest number of subsets when the vertices are partitioned into ρ-dominating sets. This notion seems to go back to Zelinka [33] and Carnielli [5]. More information on the subject is in Östergård [24].

Here is an example also discussed in [32]: start from the matrix

$$H = \begin{pmatrix} 1 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 \end{pmatrix}$$

whose entries are elements of $\mathbb{F}_2$. Consider the linear mapping $f : \mathbb{F}_2^7 \longrightarrow \mathbb{F}_2^3$ defined by $f(x_1, x_2, x_3, x_4, x_5, x_6, x_7) = (y_1, y_2, y_3)$, where

$$y_1 = x_1 + x_4 + x_5 + x_7,\ y_2 = x_2 + x_4 + x_6 + x_7,\ y_3 = x_3 + x_5 + x_6 + x_7.$$

This function can be described in terms of matrix H. In fact, y_i is the dot product of x and the i-th row of H. We claim that f is a $COV(1, 7, 3)$.

For example, $f(0011010) = 100$. Assume $y = 111$. We claim that it is possible to replace $x = 0011010$ by x' such that $f(x') = 111$ and $d(x, x') = 1$. In fact, we claim more: the coordinate where x has to be changed is uniquely determined. In our case, this is coordinate number 6, so $x' = 0011000$. Here is the general embedding rule: form $f(x) + y$ (in the example this is 011). Find the column of H which has these entries (in our example, this is the sixth column). This marks the coordinate where x needs to be changed to embed payload y. This procedure indicates how H and f were constructed and how this can be generalized: the columns of H are simply all nonzero 3-tuples in some order.

In general, we start from our choice of n and write a matrix H whose columns consist of all nonzero n-tuples. Then H has $N = 2^n - 1$ columns. The covering function $f : \mathbb{F}_2^N \longrightarrow \mathbb{F}_2^n$ is defined by way of the dot products with the rows of H, just as in the example $n = 3$. Then f is a covering function of radius 1.

Theorem 2. *For every n there is a $COV(1, 2^n - 1, n)$.*

These covering functions are well-known not only in coding theory but also in the steganographic community [32]. They are equivalent to the binary **Hamming codes.** By definition, f is **linear** (over $\mathbb{F}_2$). Every linear covering function can of course be described in terms of an (n, N)-matrix H. Obviously the radius will be $\leq \rho$ if and only if every vector from $\mathbb{F}_2^n$ can be written as a linear combination of at most ρ columns of H. As 0-columns and repeated columns are not helpful for this purpose, we may as well assume that the columns of H are distinct and nonzero. The code $f^{-1}(0)$ is a linear covering code of radius ρ. Vice versa, it is also clear that the existence of such a covering code can be used to construct H and f. The matrix H is known as a **check matrix** of the code.

Theorem 3. *The following are equivalent:*

- *A linear $COV(\rho, N, n)$*
- *A binary linear code of length N and dimension $N - n$ of covering radius ρ.*
- *A collection of N nonzero bitstrings of length n with the property that every element of $\mathbb{F}_2^n$ can be written as a sum of at most ρ bitstrings from the collection.*

The description of covering functions in terms of covering codes was first given by Crandall [7] who references [2]. The textbook [1] contains a description, which is more general in that it considers arbitrary alphabets. Here we concentrate on the binary case as it is by far the most interesting. Galand and Kabatiansky [17] treat the case of the description corresponding to linear covering codes. In Section 4, we are going to see that non-linear constructions can in fact be very powerful.

3 Coding-theoretic bounds and linear constructions

Having established a coding theoretic description gives us a strategic advantage as coding theory is a highly developed area with numerous links to other mathematical disciplines. The deepest and most important of those links is the construction, due to Goppa and Manin in the early 1980s, of linear codes from algebraic curves (see [18]). Unfortunately this has not been a source of interesting covering codes yet.

In this section, we first establish some useful bounds and then give examples of linear covering functions that can be used in steganography directly or as ingredients in more advanced constructions described in Section 5. We start with some simple recursive constructions.

Proposition 1. *If $COV(\rho_i, N_i, n_i)$ exist for $i = 1, 2, \ldots,$ then $COV(\sum \rho_i, \sum N_i, \sum n_i)$ exists.*

The existence of $COV(\rho, N, n)$ implies the existence of

$$COV(\rho + 1, N, n),\ COV(\rho, N + 1, n),\ COV(\rho, N, n - 1)$$

and of $COV(c \cdot \rho, c \cdot N, c \cdot n)$ for every natural number c.

This hardly needs a proof. For the first property simply write your bitstring of length $\sum N_i$ as a concatenation of strings of length N_i each and apply the $COV(\rho_i, N_i, n_i)$ to the corresponding segment. The rest is equally obvious.

In order to obtain a bound, observe that the existence of a $COV(\rho, N, n)$ implies the existence of a covering code with at most 2^{N-n} codewords. Each codeword determines $\sum_{i=0}^{\rho} \binom{N}{i}$ vectors at Hamming distance at most ρ. Adding up all those numbers must give us at least 2^N, the number of all vectors in our space:

Theorem 4 (sphere covering bound). *If $COV(\rho, N, n)$ exists, then*

$$\sum_{i=0}^{\rho} \binom{N}{i} \geq 2^n.$$

As a trivial example, Theorem 4 tells us that $COV(1, N, n)$ can exist only if $N \geq 2^n - 1$. This shows that the Hamming codes give optimal parameters. Recall that we wish to maximize the embedding efficiency – minimize N when ρ, n are given. A less trivial example is $COV(3, N, 11)$, where Theorem 4 tells us

$$1 + N + \binom{N}{2} + \binom{N}{3} \geq 2^{11} = 2048.$$

For $N = 23$ we have equality. The corresponding code exists. It is the single most famous code, the binary Golay code. We see that it defines a $COV(3, 23, 11)$. The cases when Theorem 4 is satisfied with equality correspond to **perfect codes.** We conclude that the Hamming codes and the binary Golay code are perfect. Unfortunately, the binary Golay code is the only non-trivial binary perfect code with covering radius > 1, so the situation will never be quite as nice again.

Linear covering functions can be considered from a geometric point of view as well. This is related to the third description in Theorem 3: the nonzero bitstrings of length n can be seen as points of a geometry. This geometry is the $(n-1)$-dimensional projective geometry $PG(n-1, 2)$ defined over $\mathbb{F}_2$. In this language, a linear $COV(2, N, n)$ is equivalent with a set K of N points in $PG(n-1, 2)$ having the property that each point of $PG(n-1, 2)$ is on a line containing 2 points of K. An extremely innocent example is obtained when $n = 3$. The corresponding geometry is known as the **Fano plane** $PG(2, 2)$, see Figure 1.

The marked quadrangle has the required property that each of the remaining 3 points is on some line through two of the quadrangle points. This defines a $COV(2, 4, 3)$. The linear $COV(2, N, n)$ for $n \leq 7$ and minimal N have been completely classified in Davydov-Marcugini-Pambianco [10]. A family of linear covering functions of radius 2, which often yields the best known values, was constructed in Gabidulin-Davydov-Tombak [16]:

$$COV(2, 5 \cdot 2^{a-1} - 1, 2a + 1) \text{ for } a \geq 1. \tag{1}$$

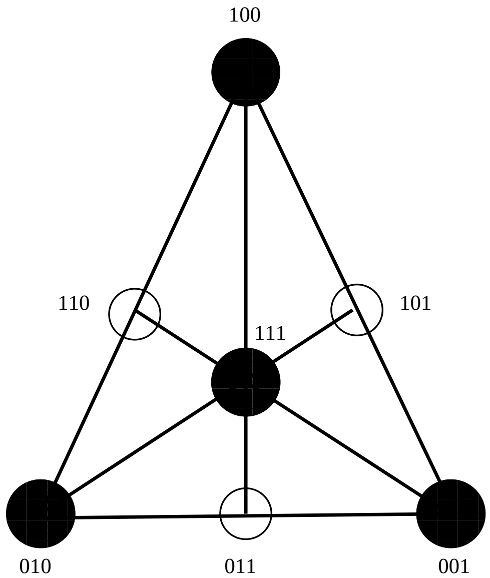

Figure 1: Fano plane and $COV(2, 4, 3)$

The smallest member of the family is the $COV(2, 4, 3)$, which we just constructed in the Fano plane. The next parameters are

$$COV(2, 9, 5),\ COV(2, 19, 7),\ COV(2, 39, 9),\ COV(2, 79, 11).$$

The Hamming code H_m ($m \geq 3$) is known to have a subcode B_m of codimension m (a **primitive BCH-code**), which has covering radius 3. The corresponding parameters as a covering function are therefore

$$COV(3, 2^m - 1, 2m) \text{ for } m \geq 3. \tag{2}$$

The embedding efficiency of steganographic schemes based on the binary Golay code and the code families (1) and (2) is discussed in Section 7.

4 The blockwise direct sum

Many of the best known covering codes and covering functions make use of non-linear codes. Virtually all known interesting constructions make use of a certain recursive procedure, the **blockwise direct sum** BDS. In order to apply it, we need in fact a refinement of the concept of a partition of the ambient space into subcodes:

Definition 3. *Let $\mathcal{D} \subset \mathcal{C} \subset \mathbb{F}_2^N$. We say that $\mathcal{C}/\mathcal{D}$ is a* **factorization** *if $\mathcal{C}$ can be written as the disjoint union of cosets of $\mathcal{D}$ and $\mathbb{F}_2^N$ is a disjoint union of cosets of $\mathcal{C}$.*

Here a coset of $\mathcal{D}$ is a set of the form $\mathcal{D} + x$, in other words a translate. The number of participating cosets of $\mathcal{D}$ in $\mathcal{C}$ is of course $|\mathcal{C}|/|\mathcal{D}|$, the **index** of $\mathcal{D}$ in $\mathcal{C}$. In all cases that we consider, the index will have the form 2^n. We define n to be the **dimension** of $\mathcal{C}/\mathcal{D}$ (or **codimension** of $\mathcal{D}$ in $\mathcal{C}$) in these cases. The **redundancy** k of $\mathcal{C}/\mathcal{D}$ is defined as the redundancy of $\mathcal{C}$, its codimension in ambient space. We will write U for the ambient space. Observe that whenever two **linear** codes form a chain, e.g., $\mathcal{D} \subset \mathcal{C}$, then they form a factorization. The **length** is the dimension of ambient space.

As an example, consider the factorization U/H_m, where H_m is the Hamming code. The length is $2^m - 1$, we have $dim(U/H_m) = m$ and the redundancy is 0. In general, factorizations of redundancy 0 (where the larger of the chain of codes is ambient space) are precisely covering functions. We need a notion which applies to factorizations and generalizes the covering radius (see Honkala [20]):

Definition 4. *Let $\mathcal{C}/\mathcal{D}$ be a factorization. For every x in the ambient space let $m(x)$ be the minimum of the distances from x to one of the cosets and $M(x)$ the maximum. The* **norm** *$\nu = \nu(\mathcal{C}/\mathcal{D})$ is the maximum, taken over all x, of $m(x) + M(x)$.*

In order to get a feeling for this notion, consider the case when $\mathcal{C} = U$. Then each $x \in U$ is contained in one of the cosets defining the factorization. It follows $m(x) = 0$. The norm is therefore the maximum of the $M(x)$. As all cosets of $\mathcal{D}$ have the same structure, in particular the same covering radius, it follows that the norm simply equals the covering radius ρ of $\mathcal{D}$. To sum this up: a factorization $U/\mathcal{D}$ is nothing but a $COV(\rho, N, n)$, where N is the length, $n = dim(U/\mathcal{D})$ and $\rho = \nu(U/\mathcal{D})$ is the norm.

The BDS is a simple and effective construction which uses as input two factorizations of equal dimension and outputs a factorization of larger length. This is relevant to our problem as we can control the covering radius of the output factorization in terms of the norms of the input.

Definition 5. *Let $\mathcal{C}_1/\mathcal{D}_1$ and $\mathcal{C}_2/\mathcal{D}_2$ be factorizations of lengths N_i and equal dimension n. Number the participating cosets $\mathcal{D}_1(i)$ and $\mathcal{D}_2(i)$ for $i = 1, \dots, 2^n$. The blockwise direct sum $(\mathcal{C}_1/\mathcal{D}_1) \vee (\mathcal{C}_2/\mathcal{D}_2)$ is defined by*

$$\mathcal{C} = \cup_{i=1}^{2^n} \mathcal{D}_1(i) \times \mathcal{D}_2(i).$$

The length of the BDS is the sum $N_1 + N_2$ of the lengths. Its dimension is obvious. It is also clear that $(\mathcal{C}_1 \times \mathcal{C}_2)/\mathcal{C}$ is a factorization. The BDS is well-known in the theory of error-correcting codes. It can be used to construct codes with large minimum distance. For the theory of covering codes, it seems to be indispensible. We just convinced ourselves that it works well on the level of factorizations. The main point is that we can control the covering radius:

Theorem 5. *Let $\mathcal{C}$ be the blockwise direct sum of two factorizations with identical dimension n, lengths N_i, norms ν_i and redundancies k_i, as in Definition 5. Then $\mathcal{C}$ has length $N_1 + N_2$ and redundancy $k_1 + k_2 + n$. The covering radius of $\mathcal{C}$ satisfies*

$$\rho(\mathcal{C}) \leq \lfloor (\nu_1 + \nu_2)/2 \rfloor.$$

Proof. The number of elements of $\mathcal{C}$ is obvious. Let (x, y) in the ambient space. Choose j, k such that $d(x, \mathcal{D}_1(j))$ and $d(y, \mathcal{D}_2(k))$ are minimal. It follows from the definition of the norm that the sum of the distances from (x, y) to $\mathcal{D}_1(j) \times \mathcal{D}_2(j)$ and to $\mathcal{D}_1(k) \times \mathcal{D}_2(k)$ is at most $\nu_1 + \nu_2$. One of the two distances must be $\leq (\nu_1 + \nu_2)/2$. $\square$

Let us express the concepts of a factorization and of the BDS in terms of covering functions. The factorization in the terminology of Definition 3 is equivalently described by a mapping $f = (f_l, f_r) : \mathbb{F}_2^N \longrightarrow \mathbb{F}_2^{k+n}$ where $f_l(x) \in \mathbb{F}_2^k,\ f_r(x) \in \mathbb{F}_2^n,\ f^{-1}(0, 0) = \mathcal{D}$, each $f^{-1}(a, b)$ is a coset of $\mathcal{D}$ and $\mathcal{C}$ is the union of the $f^{-1}(0, b)$.

If $f_1 = (f_{1,l}, f_{1,r})$ and $f_2 = (f_{2,l}, f_{2,r})$ describe the factorizations $\mathcal{C}_1/\mathcal{D}_1$ and $\mathcal{C}_2/\mathcal{D}_2$ in Definition 5, then the BDS is defined by

$$(f_1 \vee f_2)(x, y) = (f_{1,l}(x), f_{2,l}(y), f_{1,r}(x) + f_{2,r}(y)) \in \mathbb{F}_2^{k_1 + k_2 + n}.$$

All we need in order to put the BDS to work are good factorizations to use as inputs. It turns out that a famous family of non-linear codes, the Preparata codes, are extremely valuable ingredients for this machinery.

5 Some families of good factorizations

We know that each $COV(\rho, N, n)$ is nothing but a factorization of redundancy 0, length N and dimension n. It can therefore itself be used as ingredient in the BDS. A factorization we know from Section 3 is H_m/B_m, of length $2^m - 1$, dimension m and redundancy m. The norm is clearly ≤ 4 as H_m has covering radius 1 and B_m has covering radius 3.

An important non-linear factorization is furnished by a famous family of non-linear codes, the Preparata codes. Codes with their parameters were first constructed by Preparata [25]. The version we need is the following:

Theorem 6. *For every even $m \geq 4$ there is a subcode $P_m \subset H_m$ such that H_m/P_m is a factorization of dimension $m-1$ and norm 3. More precisely we have that the covering radius of P_m is 3 and that every vector $x \in U \notin H_m$ has distance at most 2 from P_m.*

The factorization of Theorem 6 is the most important non-linear ingredient in constructions of covering codes and covering functions [19]. The smallest member P_4 of the Preparata family was constructed in 1967 [23]. We denote its extension $\overline{P}_4$ by $\mathcal{NR}$ (for the notion of an extension see the paragraph preceding Table 5 below where some elementary facts are explained). This is the famous Nordstrom-Robinson code. It is also the smallest member of the **Kerdock codes,** a family of non-linear codes closely related to the Preparata codes. The Nordstrom-Robinson code has an unusually large group of automorphisms (of order $8! = 40,320$) and is optimal in many respects. It can be found inside the binary Golay code. No linear codes with similar properties as $\mathcal{NR}$ can exist. There are numerous links from the Preparata and Kerdock codes to other mathematical areas, such as finite geometries and group theory. It had been observed early on that the Preparata and Kerdock codes behave like pairs of binary linear codes related by duality, which sounds strange as they are not linear. An explanation for this phenomenon was given in [19]. There are families of linear codes defined over the alphabet $Z_4 = \mathbb{Z}/4\mathbb{Z}$, the integers mod 4, which map to the Preparata and Kerdock codes under the Gray map γ. The Gray map takes $0 \in Z_4$ to 00, the zero-divisor 2 to the pair 11 and the units $1, 3 \in Z_4$ to the pairs of weight 1. It is the only non-linear element in the construction. The original observation that the Preparata and Kerdock codes behave almost as if they were dual linear codes is explained by the fact that their preimages under γ

are in fact dual Z_4-linear codes. The same feature explains why H_m/P_m is a factorization. An explicit proof is in Wan's book *Quaternary codes* [31].

In order to understand the factorizations given in the following table, we recall some elementary facts and constructions. The **sum zero code** consists of the bitstrings of even weight. It has codimension 1 in ambient space, being the dual of the repetition code. We denote it by A. In algebra it is also known as the augmentation ideal. If C is a code of length N then its extension $\overline{C}$ has length $N+1$. It has the same number of codewords as C and is defined such that it is contained in the sum zero code A of length $N+1$. If C/D is a factorization, then $\overline{C}/\overline{D}$ is a factorization as well. Let ν be the norm of C/D. The norm of $\overline{C}/\overline{D}$ is then the even number among $\{\nu+1, \nu+2\}$. We arrive at the following list of factorizations.

Some factorizations				
factorization	length	dim	red	norm
U/H_m	2^m-1	m	0	1
$U/\overline{H}_m$	2^m	$m+1$	0	2
$A/\overline{H}_m$	2^m	m	1	2
H_m/B_m	2^m-1	m	m	4
U/GDT_m	$5\cdot 2^{m-1}-1$	$2m+1$	0	2
H_m/P_m	2^m-1	$m-1$	m	3
U/P_m	2^m-1	$2m-1$	0	3
$\overline{H}_m/\overline{P}_m$	2^m	$m-1$	$m+1$	4
$U/\overline{P}_m$	2^m	$2m$	0	4
$A/\overline{P}_m$	2^m	$2m-1$	1	4

Here m has to be even and ≥ 4 whenever P_m is involved. Recall that H_m are the Hamming codes, B_m is the BCH-code introduced in Section 3, and GDT_m is the code from [16] mentioned in the same section.

In the next section, we give examples of the best known covering functions. The embedding efficiency of steganographic schemes that use these covering functions is discussed in Section 7.

6 The best known covering functions

The best known covering functions are obtained by application of the BDS to the factorizations given in the table at the end of the preceding section. The examples are from Etzion-Greenberg [11] and from Struik's dissertation [30]. Observe that the BDS can be constructed whenever we have factorizations of equal dimension.

Application to H_m/P_m and $A/\overline{H}_{m-1}$ (both of dimension $m-1$) yields, with $m=2a$,

$$COV(2, 6\cdot 4^{a-1}-1, 4a),\ a\geq 2. \tag{3}$$

The first members of this family are

$$COV(2,23,8),\ COV(2,95,12),\ COV(2,383,16),\ COV(2,1535,20).$$

The pair U/GDT_m and H_{2m+2}/P_{2m+2} yields

$$COV(2, 4^{m+1}+5\cdot 2^{m-1}-2, 4m+3) \text{ for } m\geq 1. \tag{4}$$

The first members of this family are

$$COV(2,19,7),\ COV(2,72,11),\ COV(2,274,15),\ COV(2,1062,19).$$

As both H_m/P_m and $\overline{H}_m/\overline{P}_m$ have dimension $m-1$, we can form the BDS. It has length 2^m-1+2^m, redundancy $m+(m-1)+(m+1)=3m$ and covering radius 3. Let $m=2a$. This yields

$$COV(3, 2\cdot 4^a-1, 6a),\ a\geq 2. \tag{5}$$

The smallest examples are

$$COV(3,31,12),\ COV(3,127,18) \text{ and } COV(3,511,24).$$

These BDS can also be used as ingredients. In fact,

$$(H_m/P_m)\vee(\overline{H}_m/\overline{P}_m)\subset H_m\times(H_m\times\mathbb{F}_2),$$

and this pair forms a factorization of dimension m and norm $3+2=5$. This gives us the following two additional factorizations to complement Table 5 where $m=2a\geq 4$:

factorization	length	dim	red	norm
above	2^{m+1}	m	$2m$	5
extension	$2^{m+1}+1$	m	$2m+1$	6

Using as second ingredient $A/\overline{H}_m$ and forming the BDS we obtain, with $m = 2a$, the family

$$COV(3, 3 \cdot 4^a - 1, 6a + 1) \text{ for } a \geq 2, \tag{6}$$

whose smallest members are

$$COV(3, 47, 13),\ COV(3, 191, 19),\ COV(3, 767, 25).$$

Forming the BDS of both table entries instead yields, with $m = 2a$,

$$COV(5, 4^{a+1} - 1, 10a + 1) \text{ for } a \geq 2, \tag{7}$$

with the following smallest members

$$COV(5, 63, 21) \text{ and } COV(5, 255, 31).$$

The BDS of H_m/P_m and H_{m-1}/B_{m-1} yields a covering function of length $2^m - 1 + 2^{m-1} - 1$, covering radius 3 and redundancy $3m - 1$. Letting $m = 2(a + 1)$ this becomes

$$COV(3, 6 \cdot 4^a - 2, 6a + 4) \text{ for } a \geq 1, \tag{8}$$

with the smallest members

$$COV(3, 22, 10),\ COV(3, 94, 16),\ COV(3, 382, 22).$$

In the following table we collected what seems to be the best known bounds on N for $COV(\rho, N, n)$ in the range $\rho \leq 5,\ n \leq 25$. If the entry consists of one number, then this is the minimum N. If the entry has the form $(, N)$, then N is an upper bound and we do not know of a reasonable lower bound. The entry is in boldface if the construction uses non-linear codes.

$n\backslash\rho$	1	2	3	4	5
2	3	2			
3	7	4	3		
4	15	5	5	4	
5	31	9	6	6	5
6	63	$(12,13)$	7	7	7
7	127	$(16,19)$	11	8	8
8	255	**23**	$(13,14)$	9	9
9	511	$(,39)$	$(16,18)$	13	10
10	1023	$(,51)$	$(20,22)$	$(14,16)$	11
11		$(,\mathbf{72})$	23	$(17,19)$	15
12		$(,\mathbf{95})$	$(30,\mathbf{31})$	$(19,23)$	$(16,18)$
13		$(,159)$	$(,\mathbf{47})$	$(,25)$	$(,19)$
14		$(,215)$	$(,63)$	$(,29)$	$(,23)$
15		$(,\mathbf{274})$	$(,71)$	$(,36)$	$(,27)$
16		$(,\mathbf{383})$	$(,\mathbf{94})$	$(,\mathbf{46})$	$(,31)$
17		$(,639)$	$(,126)$	$(,62)$	$(,35)$
18		$(,863)$	$(,\mathbf{127})$	$(,\mathbf{74})$	$(,41)$
19		$(,\mathbf{1062})$	$(,\mathbf{191})$	$(,82)$	$(,\mathbf{46})$
20		$(,\mathbf{1535})$	$(,254)$	$(,90)$	$(,\mathbf{54})$
21		$(,2559)$	$(,308)$	$(,122)$	$(,\mathbf{63})$
22		$(,3455)$	$(,\mathbf{382})$	$(,\mathbf{144})$	$(,\mathbf{82})$
23		$(,\mathbf{4167})$	$(,\mathbf{510})$	$(,\mathbf{158})$	$(,94)$
24		$(,\mathbf{6143})$	$(,\mathbf{511})$	$(,190)$	$(,104)$
25		$(,\mathbf{10,239})$	$(,767)$	$(,238)$	$(,120)$

As a basis for the best known linear constructions we used the extensive tables from [6]. Some more recent values are from Kaikkonen-Rosendahl [21], Davydov [8] and Etzion-Mounits [12]. For $\rho = 2$ and small n see also [10]. A complete census of the linear covering functions with $\rho = 3, n \leq 6$ and $\rho = 4, n \leq 8$ is in [9].

Finally, we use the direct sum construction of Proposition 1. As an example, the direct sum of the binary Golay code $COV(3, 23, 11)$ and the non-linear $COV(2, 23, 8)$ yields

$$COV(5, 46, 19). \tag{9}$$

Two other entries in the table

$$COV(4,74,18), COV(5,82,22), \tag{10}$$

were obtained as direct sum of the computer-generated $COV(2,51,10)$ by Kaikkonen-Rosendahl [21] with $COV(2,23,8)$ ((3) with $a=2$) and $COV(3,31,12)$ ((5) with $a=2$).

The smallest open problem is the existence of a $COV(2,12,6)$. If it existed there would have to exist a covering code of length 12 and radius 2 with $M \leq 64$ codewords. The current lower bound on M is 62, the smallest number of codewords for which such a covering code is known to exist is $M = 78$.

As the (extended) Preparata codes work pretty well in the construction of covering codes and covering functions, it might be expected that other families of Z_4-linear binary codes like the Goethals codes would produce good results as well.

7 Performance

Having derived some families of good covering codes, we now study the embedding efficiency that these codes offer to steganographers. As explained in the introduction, an important concept in steganography is the embedding efficiency, which is defined as the ratio between the number of embedded bits and the number of embedding changes. Using the notation $COV(\rho, N, n)$ for the covering function, we will call the ratio $\alpha = n/N$ the **relative capacity** and $e = n/\rho$ the **embedding efficiency** (in bits per embedding change).

In Figure 2, we show the embedding efficiency as a function of $1/\alpha$ for the binary Hamming code, the binary Golay code, and the families (1)–(8). The upper bound on e for a fixed α can be obtained from the sphere-covering bound (see, e.g., [13])

$$\epsilon \leq \frac{\alpha}{H^{-1}(\alpha)}, \tag{11}$$

where $H^{-1}(\alpha)$ is the inverse of the binary entropy function $H(x) = -x\log_2 x - (1-x)\log_2(1-x)$ on the interval $[0, 1/2]$. Observe that the recursive constructions of Proposition 1 imply that each $COV(\rho, N, n)$ gives us constructions not only of its asymptotic parameters $(N/n, n/\rho)$ but of an infinite set of such parameter pairs that is dense in the region to the right and down of $(N/n, n/\rho)$. For example, This observation is important for practical applications in steganography – the sender should choose the covering code with relative capacity α slightly above the relative message length that he wants to communicate to the recipient.

Members of the family (5) and (7) lead to the highest embedding efficiency, providing significant improvement over the simple binary Hamming codes.

The last issue that needs to be addressed for practical applications is the implementation complexity. This is the topic of the next section where we discuss implementation details for the family of codes (3), as an example.

8 Implementation

We now give the details for the factorizations and covering functions needed to implement a steganographic scheme that uses the covering codes (3) for $m = 4$.

Start with a description of the Nordstrom-Robinson code $\mathcal{NR} = \overline{P}_4$. This is a binary code of length 16, with 2^8 codewords, minimum distance 6 and covering radius 4. It is the smallest member of the Preparata family and the single most famous non-linear code. Among its exceptional properties is the presence of a huge group of symmetries, of order 8! For an introduction see [3]. As mentioned earlier, the charm of the Preparata codes is that they are essentially linear over Z_4. This means that there is an underlying code, which is linear over Z_4, and the binary code in question is the image of this Z_4-linear code under the Gray map γ, where

$$\gamma : \ 0 \mapsto 00, \ 1 \mapsto 10, \ 3 \mapsto 01, \ 2 \mapsto 11.$$

In order to construct the Nordstrom-Robinson code start from the binary matrix

$$\left(\begin{array}{c|c} 1000 & 0111 \\ 0100 & 1011 \\ 0010 & 1101 \\ 0001 & 1110 \end{array}\right), \tag{12}$$

which generates the extended Hamming code $[8, 4, 4]_2$ (length 8, dimension 4, minimum distance 4). As this code is equal to its orthogonal with respect to the ordinary dot product (it is self-dual) it is tempting to lift this matrix

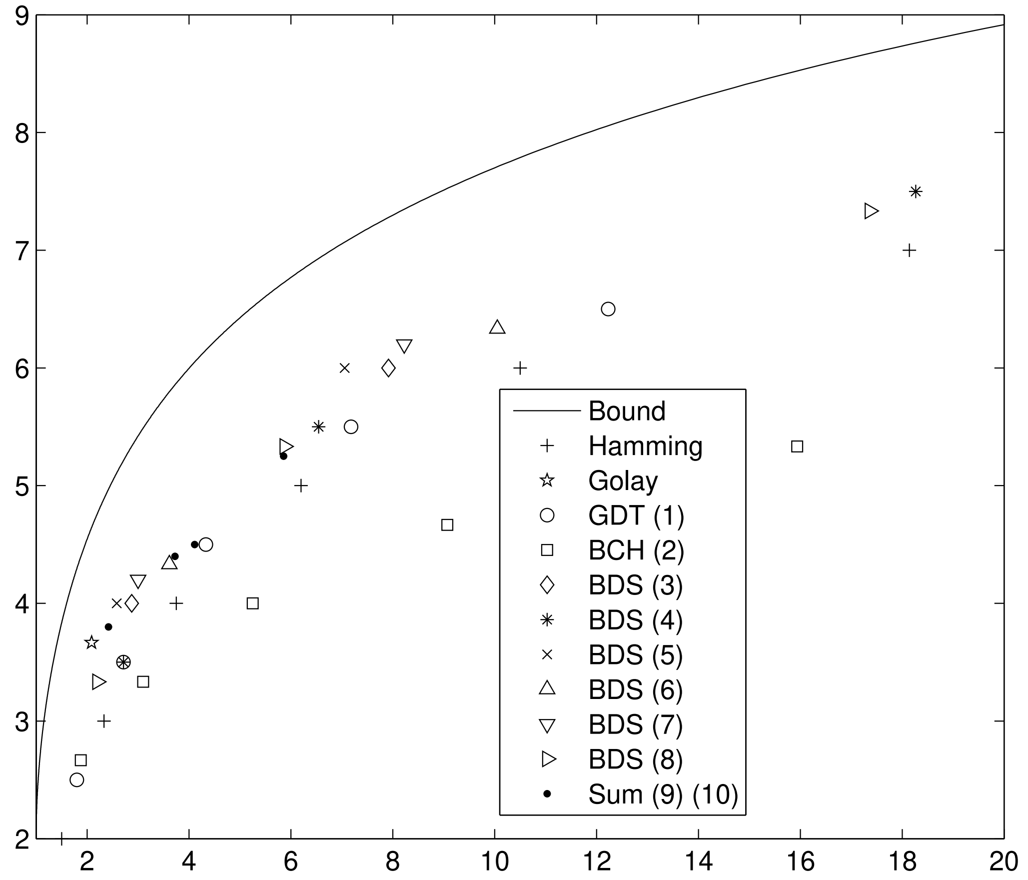

Figure 2: Embedding efficiency n/ρ as a function of $1/\alpha$ for various covering functions $COV(\rho, N, n)$.

to a matrix with entries in Z_4. Observe that the factor ring of Z_4 mod $\{0,2\}$ is the binary field $\mathbb{F}_2$. Lifting means that each entry $0 \in \mathbb{F}_2$ should become 0 or 2 in Z_4 and each $1 \in \mathbb{F}_2$ should be replaced by 1 or 3 in Z_4. We want the lift to have the property that it is self-dual over Z_4. After a moment's thought this leads to the matrix

$$G = (I|P) = \left(\begin{array}{c|c} 1000 & 2333 \\ 0100 & 1231 \\ 0010 & 1123 \\ 0001 & 1312 \end{array} \right).$$

The self-dual Z_4-linear code generated by the rows of this matrix is known as the **octacode** $\mathcal{N}$. The Nordstrom-Robinson code is its image under the Gray map: $\mathcal{NR} = \overline{P}_4 = \gamma(\mathcal{N})$. The length and number of codewords are as promised and it is not hard to check that the minimum distance and covering radius are as claimed. The codewords of $\mathcal{N}$ are $x(a,b,c,d) = (l|r)$ where $l = (a,b,c,d),\ a,b,c,d \in Z_4$. and

$$r = s(l) = (2a+b+c+d, -a+2b+c-d, -a-b+2c+d, -a+b-c+2d).$$

The proof of the following lemma can be left as an easy exercise, using the self-duality of $\mathcal{N}$.

Lemma 1. *Let $x \in \mathcal{N}$ and $\nu_i(x)$ for $i \in Z_4$ the frequency of i as an entry of x. Then the $\nu_i(x)$ have the same parity.*

In order to see that $U/\mathcal{NR}$ is a factorization, observe that $\mathcal{N}$ is **systematic:** in the projection on the left half of parameters each quaternary quadruple occurs precisely once. It follows that the binary code $\mathcal{NR}$ is systematic as well: in the projection on the left half of parameters each binary 8-tuple occurs precisely once. Systematic codes can always be embedded in factorizations. In fact, the $(0,y)$, where $y \in Z_4^4$ are representatives of pairwise disjoint cosets of $\mathcal{N}$. The union of those cosets is the ambient space $U = Z_4^8$. It follows that the same is true for $\mathcal{NR}$. The codewords of $\mathcal{NR}$ are $(\gamma(l), \gamma(s(l)))$, where $l \in Z_4^4$. Write $(x,y) \in \mathbb{F}_2^{8+8}$ as

$$(x,y) = (\gamma(l), y) = (\gamma(l), \gamma(s(l))) + (0, y + \gamma(s(l))).$$

This decomposition enables us to formulate the following proposition.

Proposition 2. *A $COV(4,16,8)$ corresponding to the factorization $\mathbb{F}_2^{16}/\mathcal{NR}$ is given by*

$$f(x,y) = y + \gamma(s(\gamma^{-1}(x))).$$

Here $x, y \in \mathbb{F}_2^8$.

As an example, consider $(x,y) = (00100111, 10011001)$. Then $\gamma^{-1}(x) = 0132$, $s(0132) = 2332$, $\gamma(2332) = 11010111$ and

$$f(x,y) = 10011001 + 11010111 = 01001110.$$

The observation that systematic codes can be embedded in factorizations is from Stinson [29], where a characterization of resilient functions is given in terms of factorizations.

The factorization $U/\mathcal{NR}$ itself is not an interesting covering function. It yields good results when used as an ingredient in the BDS. We mentioned and used a factorization $\overline{H}_4/\mathcal{NR}$ of length 16, dimension 3 and redundancy 5. Here $\overline{H}_4$ is the extended Hamming code, a linear $[16,11,4]$-code. At first we have to see that $\mathcal{NR}$ is in fact contained in the Hamming code. Let

$$M = \left(\begin{array}{c|c} 11000000 & 00111111 \\ 00110000 & 11001111 \\ 00001100 & 11110011 \\ 00000011 & 11111100 \\ 01010101 & 01010101 \end{array} \right).$$

This is a check matrix of the extended Hamming code $\overline{H}_4 = [16,11,4]_2$ (equivalently: no 3 columns of M add to the 0-column). It is orthogonal to all codewords of $\mathcal{NR}$ (for the first 4 rows of M this is obvious, for the last row use Lemma 1). It follows $\mathcal{NR} \subset \overline{H}_4$. Let

$$T = \langle \gamma(2200), \gamma(2020), \gamma(2002) \rangle.$$

Then $(0,T) \subset \overline{H}_4$ as those vectors are orthogonal to the rows of M. The first half of coordinates shows that the cosets $\mathcal{NR} + (0,t),\ t \in T$ are pairwise disjoint. This defines a factorization $\overline{H}_4/\mathcal{NR}$.

Let us calculate the covering function $f = (f_1, f_2) : \mathbb{F}_2^{16} \longrightarrow \mathbb{F}_2^8$ which describes the factorization $\overline{H}_4/\mathcal{NR}$. Let $r_i, i = 1, \ldots, 5$ be the rows and $s_j, j = 1, \ldots, 16$ the columns of M. Let $z = (x|y) \in \mathbb{F}_2^{8+8}$ and denote by e_i the elementary vectors. The first section simply is the syndrome:

$$f_1(z) = (z \cdot r_1, z \cdot r_2, z \cdot r_3, z \cdot r_4, z \cdot r_5) = (\sigma | z \cdot r_5) \in \mathbb{F}_2^{4+1}.$$

Let the $\mathbb{F}_2$-linear mapping $\beta : T \longrightarrow \langle e_6, e_7, e_8 \rangle$ be defined by

$$\beta(\gamma(2200)) = e_6,\ \beta(\gamma(2020)) = e_7,\ \beta(\gamma(2002)) = e_8.$$

In order to calculate f_2, proceed as follows:

- If σ has odd weight, then $f_1(z) = s_i$ is a column of M. Let $z' = z + e_i, \gamma^{-1}(z') = (l, r)$. Then $\gamma(r - s(l)) = (0, t)$ for $t \in T$. Let $f_2(z) = \beta(t)$.
- If σ has even weight, then $f_1(z) = s_1 + s_i$ for a uniquely determined column s_i of M. Let $z' = z + e_1 + e_i$ and continue as in the preceding case.

As an example, consider $z = 11001001|00011001$. The syndrome is the sum of columns number $1, 2, 5, 8, 12, 13, 16$ of M :

$$f_1(z) = 10111.$$

As $\sigma = 1011$ has weight 3, we have that $f_1(z) = s_{11}$ is a column of M. It follows $z' = 11001001|00111001 \in \overline{H}_4$. Then

$$l = 2013,\ r = 0213,\ s(l) = 0033, r - s(l) = 0220 \in \gamma^{-1}(0, T)$$

as promised. Applying γ and β yields $f_2(z) = e_6 + e_7$:

$$f(z) = (f_1(z)|f_2(z)) = 10111|110.$$

This can be adapted to obtain the covering function g describing the factorization H_4/P_4 of the shortened codes: Let $z \in \mathbb{F}_2^{15}$. Add a parity check bit in the beginning to obtain $z' \in \mathbb{F}_2^{16}$ of even weight. Then

$$g(z) = (g_1(z)|g_2(z)) = (z' \cdot r_2, z' \cdot r_3, z' \cdot r_4, z' \cdot r_5|f_2(z')) \in \mathbb{F}_2^{4+3}.$$

As an example, let $z = 11000001|0001100$. Then $z' = 11000001|00011001$. Application of f as before yields $f(z') = 10010|011$. Now $g(z)$ is obtained by removing the first bit:

$$g(z) = 0010|011.$$

This function $g = (g_1, g_2)$ is one of the ingredients in the construction of $COV(2, 23, 8)$ (family (3) in the beginning of Section 6). The second ingredient corresponds to the factorization $A/\overline{H}_3$ of length 8 and dimension

3. Here $\overline{H}_3$ is the extended Hamming code $[8,4,4]_2$ again. Using its generator matrix as given in the beginning of the present section, we can write the corresponding covering function as $h(y)=(h_1(y),h_2(y))$, where $y\in\mathbb{F}_2^8$ and

$$h_1(y)=y_1+\cdots+y_8,\ h_2(y)=(y_1+y_6+y_7+y_8,$$

$$y_2+y_5+y_7+y_8,\ y_3+y_5+y_6+y_8).$$

This leads to the construction of a $COV(2,23,8)$ as $g\vee h$, concretely

$$(g\vee h)(x,y)=(g_1(x),h_1(y),g_2(x)+h_2(y))\in\mathbb{F}_2^8,$$

where, of course, $x\in\mathbb{F}_2^{15},\ y\in\mathbb{F}_2^8$.

So far, we constructed $COV(2,23,8)$, the extraction function f that allows us to extract 8 secret message bits from a block of 23 pixels. To complete the description of the steganographic scheme, we need to explain the embedding mechanism. So, let $(x,y)\in\mathbb{F}_2^{15+8}$ be given such that $f(x,y)=(g\vee h)(x,y)=(a,b,c)\in\mathbb{F}_2^{4+1+3}$ and let $(A,B,C)\in\mathbb{F}_2^{4+1+3}$ be the section of the secret message that we wish to embed. We need to describe how to change (x,y) in at most 2 coordinates such that the resulting bitstring is mapped to (A,B,C). Naturally, we use matrix M above and its submatrix M' obtained by omitting the first row of M. The linear mapping h is based on matrix

$$N=\begin{pmatrix}11111111\\10000111\\01001011\\00101101\end{pmatrix}.$$

Observe that the columns of M' are all quadruples, starting with the 0 quadruple and the columns of N are all quadruples that start with 1. We can describe the embedding procedure. Number the columns of M' from 0 to 15, those of N from 1 to 8.

- Assume $b+B=1, A=a$. Let $(1,c+C)$ be column number j of N. Change bit number j of y, leave x unchanged.

- Assume $b+B=1, A\neq a$. Choose j as above and choose $i\leq 15$ such that $a+A$ is column i of M'. Change the i-th bit of x and the j-th bit of y.

- Let $B = b, A = a$. Change y in two coordinates j_1 and j_2 such that the corresponding columns of N add to $(0, c + C)$.
- Let $B = b, A \neq a$. Consider the 8 pairs of columns of M' that add to $A + a$. This corresponds to 7 vectors $x'_1, \dots, x'_7$ at distance 2 from x and one vector x'_8 at distance 1 (the corresponding column of M' being $A = a$). The values $g_2(x'_i)$ are all different. Choose i such that $g_2(x'_i) + g_2(x) = C + c$.

9 Conclusion

In this paper, we show that certain families of non-linear codes can achieve markedly better performance (higher embedding efficiency) for applications in steganography than simple linear codes currently in use. We construct the codes using the blockwise direct sum of code factorizations. For practitioners, we provide a detailed description of one selected family of covering functions.

The smallest open problem in constructing good families of coverings is the existence of $COV(2, 12, 6)$, as remarked in Section 6. A more general problem is to use the known families of good Z_4-linear codes for the construction of covering codes and covering functions. An even more ambitious aim is to bring algebraic-geometric codes into play. Finally, the theory of covering functions should not be restricted to the binary case.

References

[1] J. Bierbrauer: *Introduction to Coding Theory,* Chapman and Hall, CRC Press 2005.

[2] J. Bierbrauer: *Crandall's problem,* unpublished, available from http://www.ws.binghamton.edu/fridrich/covcodes.pdf 1998.

[3] J. Bierbrauer: *Nordstrom-Robinson code and A_7-geometry, Finite Fields and Their Applications,* to appear.

[4] C. Cachin: *An information-theoretic model for steganography*, in D. Aucsmith (ed.): *Information Hiding. 2nd International Workshop*, LNCS vol. 1525, Springer-Verlag Berlin Heidelberg (1998), 306-318.

[5] W. A. Carnielli: *On covering and coloring problems for rook domains, Discrete Mathematics* ***57*** (1985), 9–16.

[6] G. Cohen, I. Honkala, S. Litsyn, A. Lobstein: *Covering Codes,* North Holland, Amsterdam 1997. ISBN 0-444-82511-8.

[7] R. Crandall: *Some notes on steganography.* Posted on steganography mailing list `http://os.inf.tu-dresden.de/~westfeld/crandall.pdf` (1998).

[8] A. A. Davydov: *New constructions of covering codes, Designs, Codes and Cryptography* ***22*** (2001), 305–316.

[9] A. A. Davydov, G. Faina, S. Marcugini, F. Pambianco: *Locally optimal (nonshortening) linear covering codes and minimal saturating sets in projective spaces, IEEE Transactions on Information Theory* ***51*** (2005), 4378–4387.

[10] A. A. Davydov, S. Marcugini, F. Pambianco: *Minimal* 1*-saturating sets and complete caps in binary projective spaces, Journal of Combinatorial Theory A* ***113*** (2006), 647–663 .

[11] T. Etzion and G. Greenberg: *Constructions for perfect mixed codes and other covering codes, IEEE Transactions on Information Theory* ***39*** (1993), 209–214.

[12] T. Etzion and B. Mounits: *Quasi-perfect codes with small distance, IEEE Transactions on Information Theory* ***51*** (2005), 3938–3946.

[13] J. Fridrich, M. Goljan, and D. Soukal: *Efficient Wet Paper Codes*, in M. Barni, J. Herrera-Joancomarti, S. Katzenbeisser, and F. Perez-Gonzales (eds.): *Information Hiding. 7th International Workshop*, LNCS vol. 3727, Springer-Verlag Berlin Heidelberg (2005), 204-218.

[14] J. Fridrich: *Feature-Based Steganalysis for JPEG Images and its Implications for Future Design of Steganographic Schemes*, in J. Fridrich (ed.): *Information Hiding. 6th International Workshop*, LNCS vol. 3200, Springer-Verlag Berlin Heidelberg (2004), 67–81.

[15] J. Fridrich, M. Goljan, and R. Du: *Steganalysis Based on JPEG Compatibility*, in Tescher et al. (eds.): *Proc. SPIE Multimedia Systems and Applications IV*, SPIE vol. 4518, Denver, CO August (2001), 275–280.

[16] E. M. Gabidulin, A. A. Davydov and I. M. Tombak: *Codes with covering radius* 2 *and other new covering codes, IEEE Transactions on Information Theory* ***37*** (1991), 219-224.

[17] F. Galand and G. Kabatiansky: *Information hiding by coverings, Proceedings of the IEEE Information Theory Workshop 2004,* 151–154.

[18] V. D. Goppa: *Codes on algebraic curves, Soviet Math. Doklady* ***24*** (1981), 170–172.

[19] A. R. Hammons, Jr, P. V. Kumar, A. R. Calderbank, N. J. A. Sloane, P. Solé: *The Z_4-linearity of Kerdock, Preparata, Goethals and related codes, IEEE Transactions on Information Theory* ***40*** (1994), 301–319.

[20] I. S. Honkala: *On (k, t)-subnormal covering codes, IEEE Transactions on Information Theory* ***37*** (1991), 1203–1206.

[21] M. K. Kaikkonen and P. Rosendahl: *New covering codes from an ADS-like construction, IEEE Transactions on Information Theory* ***49*** (2003), 1809–1812.

[22] A. Ker: *A General Framework for Structural Analysis of LSB Replacement*, in M. Barni, J. Herrera-Joancomarti, S. Katzenbeisser, and F. Perez-Gonzales (eds.): *Information Hiding. 7th International Workshop*, LNCS vol. 3727, Springer-Verlag Berlin Heidelberg (2005), 296–311.

[23] A. W. Nordstrom and J. P. Robinson: *An optimum nonlinear code, Information and Control* ***11*** (1967), 613–616.

[24] P. Östergård: *A coloring problem in Hamming spaces, European Journal of Combinatorics* ***18*** (1997), 303–309.

[25] F. P. Preparata: *A class of optimum nonlinear double-error-correcting codes, Information and Control* ***13*** (1968), 378–400.

[26] N. Provos: *Defending Against Statistical Steganalysis*, 10th USENIX Security Symposium, Washington, DC, 2001.

[27] P. Sallee: *Model Based Steganography*, in T. Kalker, I. J. Cox, and Yong Man Ro (eds.): *International Workshop on Digital Watermarking*, LNCS vol. 2939, Springer-Verlag Berlin Heidelberg (2004), 154–167.

[28] G. J. Simmons: *The Prisoners' Problem and the Subliminal Channel*, in D. Chaum (ed.): *Advances in Cryptology: Proceedings of Crypto 83*, Plenum Press (1984), 51–67.

[29] D. R. Stinson: *Resilient functions and large sets of orthogonal arrays*, *Congressus Numerantium* **92**, (1993), 105–110.

[30] R. Struik: *Covering Codes,* Ph.D. dissertation, Eindhoven 1994.

[31] Z. X. Wan: *Quaternary codes,* World Scientific 1997.

[32] A. Westfeld: *High Capacity Despite Better Steganalysis (F5–A Steganographic Algorithm)*, in I. S. Moskowitz (ed.): *Information Hiding. 4th International Workshop*, LNCS vol. 2137, Springer-Verlag, Berlin Heidelberg (2001), 289–302.

[33] B. Zelinka: *On k-domatic numbers of graphs, Czechoslovak Math. Journal* ***33*** (1983), 309–311.