Recognizing the structure of permuted reducible codes

R. Overbeck

GK Electronic Commerce,
TU-Darmstadt,
Department of Computer Science,
Cryptography and Computer Algebra Group.
overbeck@cdc.informatik.tu-darmstadt.de

Abstract. In code based cryptography a structured code is used as a trapdoor. In order to prevent an attacker to use the trapdoor, the structure of the code must be hidden. In this paper we analyze the approach to use a permuted reducible code. We show that this strategy has to be used with care and does not allow to reduce the key size for the McEliece PKC or the CFS signature scheme.

Keywords: Public key cryptography, coding theory.

1 Introduction

Public key cryptosystems based on coding theory is an interesting alternative to established public key cryptosystems as RSA or DL based schemes. The former ones have been introduced by R.J. McEliece in 1978 [5] and the initial version remains unbroken for appropriate parameter sets. The drawback of the cryptosystem from [5] is the impractical public key size of about 88KByte for secure instances [1]. To overcome this drawback several variants of McEliece's PKC have been proposed. The employed techniques are highly interesting, as most of them are applicable to all code based cryptosystems and intuitively analogous to the modifications proposed for the basic multivariate schemes [3, 9, 11].

1.1 Our contribution

We analyze the technique of reducible codes from [3] applied to *Hamming distance codes*, i.e. codes which are used to correct errors in a metric whose isometries are permutations. This proposal is very interesting, as it could allow to speed up the CFS signature scheme from [2] as we will see in section 2. In our analysis we consider different approaches like the one from [6] to recognize reducible structures, and try to determine parameter

sets secure against such attacks. The main result of our analysis is that the use of reducible codes does not allow to reduce the key size of the CFS or McEliece scheme.

1.2 Concepts of code based cryptography

Even if R.J. McEliece used binary Goppa codes with irreducible generator polynomials in his original cryptosystem, he led way to a large class of cryptographic systems. Following his ideas, every class of error correcting codes can be used to construct a public key cryptosystem – even if the security status is not known a priori. A pseudo-description of such cryptosystems would be the following:

Definition 1.1. *A McEliece-like code based public key cryptosystem consists of three algorithms:*

*(i) The **key generation** algorithm, which takes in a (set of) security parameter(s) and returns a secret key, which consists of*
   - *a (set of) secret code(s) over a finite field $\mathbb{F}$, which allow(s) to efficiently correct up to t errors according to a certain norm and*
   - *an efficiently invertible transformation, which maps (tuples of) codewords of the secret code(s) to codewords of a public code $\mathcal{G}^{\text{pub}}$.*

   The public key consists of the matrix G^{pub} generating $\mathcal{G}^{\text{pub}}$ and the number r of errors one can correct in $\mathcal{G}^{\text{pub}}$ knowing the secret key.

*(ii) The **encryption** algorithm, which takes in a message $\mathbf{x}$, generates a random vector $\mathbf{e}$ of norm r and returns the ciphertext $\mathbf{c} = \mathbf{x}\mathsf{G}^{\text{pub}} + \mathbf{e}$.*

*(iii) The **decryption** algorithm takes in the ciphertext $\mathbf{c}$, uses secret transformation to recover the error $\mathbf{e}$ and returns the message $\mathbf{x}$.*

The basic security of code based cryptosystems depends on the difficulty of the following two attacks:

(i) **Structural Attack:** Recover the secret transformation and the description of the secret code(s) from $(\mathsf{G}^{\text{pub}}, r)$.
(ii) **Ciphertext-Only Attack:** Recover the original message from the ciphertext and the public key.

The difficulty of the ciphertext-only attack is related to the general decoding problem, while in general the difficulty of structural attacks is not related to any classic coding theoretic problem and mainly depends on the class of codes and the secret transformation used. In this section we present some techniques of how to generate the secret transformation.

We assume, that it is sufficient to know a secret (especially structured) matrix $\mathsf{G} \in \mathbb{F}^{k\times n}$ to correct errors of norm at most t in the secret code. This is true e.g. for Goppa codes and Reed-Muller codes but as well for most other codes, see e.g. [4] and [3]. To hide the structure of the secret code (i.e. G), one can apply one or several transformations. The transformations we consider in this paper are given in Table 1.1. However, there exist further strategies for hiding the secret key.

(i) **Row Scrambler [5]:** Multiply G with a random invertible matrix $\mathsf{S} \in \mathbb{F}^{k\times k}$ from the right. As $\langle \mathsf{G} \rangle = \langle \mathsf{SG} \rangle$, one can use the known error correction algorithm.

(ii) **Column Scrambler / Isometry [5]:** Multiply G with a random invertible matrix $\mathsf{T} \in \mathbb{F}^{n\times n}$ from the left, where T preserves the norm. Obviously one can correct errors of norm up to r in $\langle \mathsf{GT} \rangle$, if G and T are known.

(iii) **Subcode [7]:** Let $0 < l < k$. Multiply G with a random $\mathsf{S} \in \mathbb{F}^{l\times k}$ of full rank from the right. As $\langle \mathsf{SG} \rangle \subseteq \langle \mathsf{G} \rangle$, the known error correction algorithm may be used.

(iv) **Subfield Subcode [5]:** Take the $\mathbb{F}_{\mathrm{SUB}}$-subfield subcode of the secret code for $\mathbb{F}_{\mathrm{SUB}}$ a subfield of $\mathbb{F}$. As before, one can correct errors by the error correcting algorithm for the secret code. However, sometimes one can correct errors of larger norm in the subfield subcode than in the original code, compare [4].

(v) **Reducible Codes [3]:** Choose some $\mathsf{Y} \in \mathbb{F}^{k\times n}$ and take the code generated by

$$\left[\begin{array}{c|c} \mathsf{G} & \mathsf{Y} \\ \hline 0 & \mathsf{G} \end{array}\right].$$

Error correction by the algorithm for the secret code is possible if one corrects errors in sections, beginning from the left. One might extend this strategy by replacing one of the matrices G by a second secret code, compare section 2.

Table 1.1. Strategies for hiding the structure of a code

Note that it is essential to use certain transformations in combination. We would like to remark two further facts: Using a *concatenated code* as in [9], thus for example $\left\langle \left[\begin{array}{c|c} \mathsf{G} & \mathsf{SG} \end{array}\right] \right\rangle$ with an invertible $\mathsf{S} \in \mathbb{F}_{q^m}^{k\times k}$ may be seen as the combination of the reducible code and the subcode modification. One could as well treat the subfield subcode transformation as a subcode transformation for structural attacks, but we prefer treating them separately. Table 1.2 shows a classification of some code based public key cryptosystems and whether resistance against structural attacks may be achieved for appropriate parameter sets (compare [6], [8] and [11]). Examples were chosen among the ones correcting errors of small Hamming norm and using techniques from Table 1.1.

A signature scheme can be built using a keypair of a McEliece-like PKC [2]. If the ratio of decodable syndromes to the total number of

PKC	McEliece	Niederreiter	Modified Niederreiter	Sidelnikov
Class of secret code	GRS [1]	GRS	GRS	Reed-Muller
Row Scrambler	•	•	•	•
Isometry	•	•	•	•
Subcode	-	-	•	- / •
Subfield Subcode	•	-	-	-
Reducible Codes	-	-	-	- / •
Security against structural attacks	√	no	√	no/no

[1] Goppa codes are subfield subcodes of certain GRS codes.

Table 1.2. Classification of code based cryptosystems

syndromes is not too small, one can treat the hash value of a message as a syndrome and try to decrypt it. If one is not able to decode every syndrome, one has to add a random vector to the message before hashing and try decryption several times. The signature is the random vector used and the error corresponding to the syndrome. If a McEliece PKC key pair is used, we call the resulting signature scheme *CFS scheme*.

2 Reducible structures in Hamming distance codes

In this section we want to discuss the use of reducible codes as presented in [3] applied to Hamming distance codes instead of rank distance codes. Let q be (the power of) a prime and let $\mathsf{G}_1 \in \mathbb{F}_q^{k_1 \times n_1}$, $\mathsf{G}_2 \in \mathbb{F}_q^{k_2 \times n_2}$ and $\mathsf{Y} \in \mathbb{F}_q^{k_1 \times n_2}$ generate codes which have a weight distribution indistinguishable from random codes. In the following we assume that an attacker is given a systematic generator matrix G of the code generated by

$$\left[\begin{array}{c|c} \mathsf{G}_1 & \mathsf{Y} \\ \hline 0 & \mathsf{G}_2 \end{array}\right] \mathsf{P} \in \mathbb{F}^{(k_1+k_2)\times(n_1+n_2)}, \tag{1}$$

where $\mathsf{P} \in \mathbb{F}_q^{n \times n}$ with $n = n_1 + n_2$ is a permutation matrix. The code $\langle \mathsf{G} \rangle$ is thus permutation equivalent to a reducible code. We will denote the columns of G belonging to G_1 with $\mathcal{N}_1$ and the other ones with $\mathcal{N}_2$. Note, that the dual matrix of $\mathsf{G}\mathsf{P}^{-1}$ defines a reducible code as well. If we denote the minimum distance of $\langle \mathsf{G}_1 \rangle$ and $\langle \mathsf{G}_2 \rangle$ with d_1 and d_2, G defines an $[n, k = k_1 + k_2, d = \min\{d_1, d_2\}]$ code. The attacker's task is to distinguish between columns of $\mathcal{N}_1$ and $\mathcal{N}_2$. The following example illustrates why it could be tempting to use reducible codes:

Example 1. **(Application to CFS)** One possible secure parameter set for the CFS signature scheme is to use a McEliece key with an $[n = 2^{15}, k = 2^{15} - 11 \cdot 15, d = 23]$ secret Goppa code. Here the fraction of random syndromes, which is decodable for the secret key holder is $\sum_{i=0}^{11} \binom{2^{15}}{i} / 2^{165} \approx \binom{2^{15}}{11} / 2^{165} \approx 2^{-25}$, which is fairly low, thus the signature scheme is much too slow. Replacing the Goppa code with a reducible code where G_1 and G_2 define $[n = 2^{14}, k = 2^{14} - 6 \cdot 14, d = 13]$ Goppa codes, the secret key holder could identify certain syndromes of weight 12, raising the fraction of decodable syndromes to $(\sum_{i=0}^{6} \binom{2^{14}}{6})^2 / 2^{168} \approx \binom{2^{14}}{6}^2 / 2^{168} \approx 2^{-19}$. This would speed up the CFS scheme by the factor $\approx 2^6$. If an attacker could not recover the reducible structure, he would apparently be faced with almost the same problem as in the original CFS scheme. Unfortunately, he can recover the structure in this case as we will see in the following.

2.1 Recovering reducible structures by Gaussian elimination

If (k_1, n_1) and (k_2, n_2) have not been chosen carefully, one can e.g. aim to find many relatively low weight codewords with common support in the dual code: To recognize the reducible structure, the attacker chooses at random k columns of G, say the set $\mathcal{K}$. Now he considers the code $\bar{\mathcal{G}}$ generated by $\left[\mathsf{G}_{.\mathcal{K}} \middle| \mathsf{G}_{.\mathcal{K}^C}\right]$ and calculates its *pseudosystematic* generator matrix $\bar{\mathsf{G}}$: For each $1 \leq i \leq n$, $\bar{\mathsf{G}}_{.\{1,\cdots i\}}$ contains at least $r_i := \mathrm{rank}(\bar{\mathsf{G}}_{.\{1,\cdots i\}})$ columns of weight 1 and $\bar{\mathsf{G}}_{ji} = 0$ for all $j > r_i$. In other words, the attacker tries to compute the systematic generator matrix of $\bar{\mathcal{G}}$ without further column changes, which can be done by Gaussian elimination. If the attacker has chosen $\mathcal{K}$ s.t. it contains more than k_1 columns out of $\mathcal{N}_1$, then there will be k_2 rows in $\bar{\mathsf{G}}$ without influence from $\left[\mathsf{G}_1 \middle| \mathsf{Y}\right]$. It follows that the columns coming from $\mathcal{N}_1 \setminus \mathcal{K}$ will have weight $\approx k_1 - k_1/q$, whereas the columns of $\mathcal{N}_2 \setminus \mathcal{K}$ will have weight $\approx k - k/q$. (Note that columns of weight $w - 1$ in $\bar{\mathsf{G}}_{\{k+1\cdots,n\}}$ identify codewords of weight w in the dual code.) Thus, the probability, that an attacker can recognize the reducible structure by computing the column weights is

$$\mathcal{P} \approx \frac{\sum_{i=k_1}^{k} \binom{i}{n_1}\binom{k-i}{n_2}}{\binom{k}{n}}. \tag{2}$$

In example 1, an attacker could recover the reducible structure in about two Gaussian eliminations. Roughly saying, the number of columns chosen from G_1 will be about $(n_1 \cdot k)/n$ and should be far below k_1 in order to hide the reducible structure efficiently.

Example 2. There are parameter sets, where the above strategy does not succeed, for example if $q = 2, n_1 = n_2 = 1000, k_1 = 900$ and $k_2 = 100$, as $\mathcal{P} \approx 2^{-1000}$. However, in such a case it is easy to correct as much errors as the secret key holder by general decoding since $\langle \mathsf{G}_1 \rangle$ will have minimum distance ≈ 13 (see below and [1]).

2.2 Structural attack by searching low weight codewords

Even if the previously presented attack does not work, an attacker can try to compute many low weight codewords and assume that a recognizable fraction has support $\subseteq \mathcal{N}_2$. The naive way would be to guess an information set, where an short codeword has only one non-zero entry and to identify it by using the Gaussian elimination as above. Finding a single word of weight w in a code by this method has time complexity about $\mathcal{O}(k^2 n) N_w \binom{n-w}{k} / (\binom{n-w}{k-1}\binom{w}{1})$, where N_w denotes the number of pairwise linearly independent weight w codewords, i.e. $\binom{n}{w}(q-1)^{w-1}q^{k-n}$ for random codes. Faster algorithms for binary codes are the ones presented e.g. in [1]. Their time complexity for finding an codeword of weight w may be approximated by $\mathcal{O}(n^3)(q-1)^{\mathcal{O}(1)} 2^{-w \log_2(1-k/n)} \cdot N_w$ (compare [8]).

However, if as in example 2 the code $\langle \left[\mathsf{G}_1 \middle| \mathsf{Y} \right] \rangle$ contains much more codewords of short weight than $\langle \mathsf{G}_2 \rangle$, computing short weighted codewords will hardly allow to distinguish between $\mathcal{N}_1$ and $\mathcal{N}_2$.

Remark 2.1. We consider parameter sets for reducible Hamming distance codes to be secure if $\binom{n_2}{w} q^{k_2 - k_1 + n_1} / \binom{n}{w}$ is very small for all $w \in \{1, \cdots, n_2\}$.

2.3 Application to Reed-Muller codes

Reed-Muller codes were considered for cryptographic use by Sidelnikov [9]. A Reed-Muller code may be described as follows: The Reed-Muller code in m variables of degree r consists of all codewords which can be obtained by evaluating some binomial in $\mathbb{F}_2[x_1, \cdots, x_m]$ of degree at most r at all possible variable assignments, see [4]. Lexicographic ordering of the 2^m possible assignments leads to the following recursive description of the canonical generator matrix $\mathsf{R}(r, m)$, which is reducible:

$$\mathsf{R}(r, m) = \left[\begin{array}{c|c} \mathsf{R}(r, m-1) & \mathsf{R}(r, m-1) \\ \hline 0 & \mathsf{R}(r-1, m-1) \end{array} \right], \tag{3}$$

where $\mathsf{R}(r, m) = \mathsf{R}(m, m)$ for $r > m$ and $\mathsf{R}(0, m)$ is the codeword of length 2^m which is one at all positions. The code $\langle \mathsf{R}(r, m) \rangle$ is a $[2^m, \sum_{i=0}^{r} \binom{m}{i}, d]$ code with $d = 2^{m-r}$ and is a subcode of $\langle \mathsf{R}(r+1, m) \rangle$ [4].

If $\mathsf{P} \in \mathbb{F}_2^{n \times n}$ is a permutation matrix, one might try to recover P from some generator matrix of $\langle \mathsf{R}(r,m)\,\mathsf{P} \rangle$ by the methods presented above. However, this is not true, as the different blocks of G in equation (3) are no random looking codes but iteratively given by equation (3). Nevertheless, the presented approach has two applications: A pseudosystematic generator matrix of a permuted Reed-Muller has only columns of odd weight, which distinguishes it from random cods. Further, minimum weight codewords can be found almost immediately by the pseudosystematic generator matrix of the dual code : A Reed-Muller code has $N_w = 2^r \prod_{i=0}^{m-r-1} \frac{2^{m-i}-1}{2^{m-r-i}-1}$ code words of lowest weight $w = 2^{m-r}$ [4].

The existence of a distinguisher from random codes allows to doubt the security of a cryptosystem built on Reed-Muller codes and indeed, Sidelnikovs cryptosystem was already broken in [6].

2.4 Concatenated codes and similar cases

A quite special situation appears in the case where Y in equation (1) contains many zero lines, as for example for dual codes of concatenated codes. To recognize the structure by a single Gaussian elimination as done in section 2.1, one has to guess either more than k_1 columns out of $\mathcal{N}_1$ or $k_2 + \dim \langle \mathsf{Y} \rangle$ columns out of $\mathcal{N}_2$. Again, the success probability may be determined analogous to equation (2). However, depending on the parameter set, searching for low weight codewords can be faster as most will have support either in $\mathcal{N}_1$ or $\mathcal{N}_2$ for sufficiently small weights.

3 Conclusion

We have presented and analyzed ways to recognize a reducible structure in a code. We conclude that it is possible to recognize a reducible structure with very easy means for reasonable parameter sets. The presented methods apply as well to concatenated codes, as their dual is a reducible code. Parameter sets, where the reducible construction of the code can not be recognized by an attacker exist. However, they seem unsuitable for cryptographic applications.

Because of the strong difference between random codes and Reed-Muller codes, our approach does not reveal the structure immediately, but allows computing low weight codewords in almost polynomial time.

References

1. A. Canteaut and F. Chabaud. A new algorithm for finding minimum-weight words in a linear code: Application to McEliece's cryptosystem and to narrow-sense BCH

codes of length 511. *IEEETIT: IEEE Transactions on Information Theory*, 44, 1998.
2. N. Courtois, M. Finiasz, and N.Sendrier. How to achieve a McEliece-based digital signature scheme. In *Advances in Cryptology - ASIACRYPT 2001*, volume 2248, pages 157–174. Springer-Verlag, 2001.
3. E.M. Gabidulin, A.V. Ourivski, B. Honary, and B. Ammar. Reducible rank codes and their applications to cryptography. *IEEE Transactions on Information Theory*, 49(12):3289–3293, 2003.
4. F.J. MacWilliams and N.J.A. Sloane. *The Theory of Error-Correctiong Codes*. North-Holland Amsterdam, 7 edition, 1992.
5. R.J. McEliece. A public key cryptosystem based on algebraic coding theory. *DSN progress report*, 42-44:114–116, 1978.
6. L. Minder. Breaking the sidelnikov cryptosystem. In *Proc. of Third Workshop on Coding and Systems, University of Zürich*, December 8/9, 2006.
7. H. Niederreiter. Knapsack-type cryptosystems and algebraic coding theory. *Probl. Control and Inform. Theory*, 15:19–34, 1986.
8. N. Sendrier. On the security of the McEliece public-key cryptosystem. In M. Blaum, P.G. Farrell, and H. van Tilborg, editors, *Proceedings of Workshop honoring Prof. Bob McEliece on his 60th birthday*, pages 141–163. Kluwer, 2002.
9. V.M. Sidelnikov. A public-key cryptosystem based on binary Reed-Muller codes. *Discrete Mathematics and Applications*, 4 No. 3, 1994.
10. J. Stern. A method for finding codewords of small weight. *Coding Theory and Applications*, 388:106–133, 1989.
11. Christian Wieschebrink. An attack on a modified niederreiter encryption scheme. In *Public Key Cryptography*, volume 3958 of *LNCS*, pages 14–26, 2006.

A Notations

We give a short repetition of the basic definitions of coding theory. As we limit ourselves to linear codes over finite fields, we make the following definition:

Definition A.2. *An $[n,k]$-code $\mathcal{C}$ over a finite field $\mathbb{F}$ is a k-dimensional subvectorspace of the vector space $\mathbb{F}^n$. We call the code $\mathcal{C}$ an $[n,k,d]$ code if $d = \min_{x,y\in\mathcal{C}} \|x-y\|$ for some norm $\|\cdot\|$. The number of positions of an vector $\mathbf{x} \in \mathbb{F}^n$, which differ from zero is called weight of $\mathbf{x}$ and corresponds to the Hamming norm. The matrix $\mathsf{C} \in \mathbb{F}^{k\times n}$ is a generator matrix for the $[n,k]$ code $\mathcal{C}$ over $\mathbb{F}$, if the rows of C span $\mathcal{C}$ over $\mathbb{F}$. We write $\mathcal{C} = \langle \mathsf{C} \rangle$.*

Any subvectorspace of $\mathcal{C}$ is said to be a subcode of $\mathcal{C}$. If $\mathcal{C}$ is a code over $\mathbb{F}$ and $\mathbb{F}_{\mathrm{SUB}}$ is a subfield of $\mathbb{F}$, then the $\mathbb{F}_{\mathrm{SUB}}$-(subfield) subcode of $\mathcal{C}$ is the code consisting of all words of $\mathcal{C}$ which have only entries in $\mathbb{F}_{\mathrm{SUB}}$. An $\mathbb{F}_{\mathrm{SUB}}$-subfield subcode is a $\mathbb{F}_{\mathrm{SUB}}$-linear code. Since codes are treated as vector spaces, we will often define them by the matrices related to the code:

For ease of notation we will use the following notation throughout the paper: We will identify $\mathbf{x} \in \mathbb{F}^n$ with $(x_1, \cdots, x_n), x_i \in \mathbb{F}$ for $i = 1, \cdots, n$. For any (ordered) subset $\{j_1, \cdots j_m\} = J \subseteq \{1, \cdots n\}$ we denote the vector $(x_{j_1}, \cdots, x_{j_m}) \in \mathbb{F}^m$ with $\mathbf{x}_J$. Similarly, we denote by $\mathsf{M}_{.J}$ the submatrix of a $k \times n$ matrix M consisting of the columns corresponding to the indices of J and $\mathsf{M}_{J'.} = \left(\left(\mathsf{M}^\top\right)_{.J'}\right)^\top$ for any (ordered) subset J' of $\{1, \cdots, k\}$. Block matrices will be given in brackets.