Error-correcting pairs for a public-key cryptosystem

Ruud Pellikaan and Irene Márquez-Corbella

Discrete Mathematics, Techn. Univ. Eindhoven
P.O. Box 513, 5600 MB Eindhoven, The Netherlands. E-mail: g.r.pellikaan@tue.nl
SECRET Project-Team - INRIA, Paris-Rocquencourt
B.P. 105, 78153 Le Chesnay Cedex France, E-mail: irene.marquez-corbella@inria.fr;

Abstract. Code-based Cryptography (CBC) is a powerful and promising alternative for quantum resistant cryptography. Indeed, together with lattice-based cryptography, multivariate cryptography and hash-based cryptography are the principal available techniques for post-quantum cryptography. Many families of codes have been proposed for these cryptosystems. One of the main requirements is having high performance t-bounded decoding algorithms which in the case of having an error-correcting pair is achieved. In this article the class of codes with a t-ECP is proposed for the McEliece cryptosystem. The hardness of retrieving the t-ECP for a given code is considered. As a first step we give a survey of results about distinguishers of several subclasses. Recent results will be surveyed.

Keywords: Code-based Cryptography, Error-Correcting Pairs, Distinguisher.

1 Introduction

The notion of *public key cryptography* (PKC) was first published in the public domain in 1976 by Diffie and Hellman though Merkle and Hellman had developed some of the key concepts during the same time. In fact Ellis published the same idea already in 1970 and called it *non-secret encryption* but it was not made public until 1997, see [32, p. 279–292]. The advantage with respect to symmetric-key cryptography is that it does not require an initial exchange of secrets between sender and receiver. A *one-way function* is the crucial concept in their papers. After forty years one can still state with [16] that:

"At the heart of any public-key cryptosystem is a one-way function *- a function $y = f(x)$ that is easy to evaluate but for which it is computationally infeasible (***one hopes***) to find the inverse $x = f^{-1}(y)$".*

Well known (**supposedly**) one-way functions are:

1. **Discrete logarithm** for which a group G (written multiplicatively) and an element $a \in G$ are required, then x is an integer and $y = a^x$. The security of the key exchange proposed by Williamson and Diffie-Hellman in 1974 and

1976, respectively depend on the difficulty of finding discrete logarithms in a finite field.

2. **Integer factorization** where $x = (p, q)$ is a pair of distinct prime numbers and $y = pq$ is its product. The security of the cryptosystem proposed by Cocks and Rivest-Shamir-Adleman (RSA) from 1973 and 1978, respectively is based on the hardness of factorizing integers.
3. **Integer knapsack** where an n-tuple of integers $a_1, \ldots, a_n$ is given, then x is an n-tuple of zeros and ones and $y = \sum_{j=1}^{n} x_j a_j$. The Merkle-Hellman public key system from 1978 is based on the difficulty of the integer knapsack problem.
4. **Decoding error-correcting codes** where an n-tuple of vectors $h_1, \ldots, h_n$ in $\mathbb{F}_q^r$ is given, then x is an n-tuple of elements in $\mathbb{F}_q$ and $y = \sum_{j=1}^{n} x_j h_j$. In 1978 McEliece [24] presents the first PKC system based on the difficulty of decoding error-correcting codes.
5. **Elliptic curve discrete logarithm** where G is an elliptic curve group (written additively) over a finite field, $x = P$ is a point on the curve and $y = kP$ is another point on the curve obtained by the multiplication of P with a positive integer k. Elliptic Curve Cryptography (ECC) proposed independently by Koblitz and Miller in 1985 is based on the difficulty of inverting this function in the group of points on an elliptic curve over a finite field.

All known public key cryptosystems depend for their security on an unproven proposition. In the cases (3) and (4) it depends on the assumption that $P \neq NP$. Even if this is indeed the case, it is not shown that most of the cases are difficult to solve, since the theory of NP completeness deals with the worst-case situation. Finally, it may be that in the average the problem is difficult, but that the particular instances used in the system can be broken by a structural attack. That in fact was the fate of the Merkle-Hellman public-key system.

With the discovery of Shor's algorithm anyone with a quantum computer can break in polynomial time all cryptosystems whose security depends on the difficulty of the problems (1), (2) and (5). Post-quantum cryptography gave birth to the next generation of cryptography algorithms, which are designed to run on conventional computers but no attacks by classical or quantum computers are known against them. See [4] for an overview of the state of the art in this area.

It may be the fate of all proposed one-way functions that they will be broken in the future. It may also be the case that some party has already broken some widely used one-way function without revealing it in the public domain for their own benefit. The key difficulty lies in the fact that present day knowledge on lower bounds on the complexity of these functions is still out of reach. This is a sobering and humbling conclusion after so many years of research. So we continue with the practice of proposing and refuting PKC systems.

In 1978 [24] McEliece presented the first PKC system based on the theory of error-correcting codes. Its main advantages are its fast encryption and decryption schemes. However, the main drawback of the original McEliece was its large key size. But this does not mean that code-based cryptography is inher-

ently inefficient. Indeed, there have been impressive achievements in this area (reducing the key size while keeping the same level of security): now [26] there are constructions with compact keys or around 5000 bits for 80 bits of security which is comparable to RSA's public key size. Code-based cryptosystems such as proposed by McEliece [24] and Niederreiter [27] are interesting candidates for post-quantum cryptography. See the survey [5].

2 Code based cryptography

A code C is a linear subspace of $\mathbb{F}_q^n$. The *parameters* of the code are denoted by $[n, k, d]$, where n is its *length*, k its *dimension* and d its *minimum distance*. The problem of *minimum distance decoding* has as input $(G, \mathbf{y})$, where G is a generator matrix of a code C over $\mathbb{F}_q$ of parameters $[n, k, d]$ and $\mathbf{y} \in \mathbb{F}_q^n$ is a received word. The output is a codeword $\mathbf{c} \in C$ of minimal distance to $\mathbf{y}$. One can phrase the problem equivalently in terms of a parity check matrix H of the code. Then the input is $(H, \mathbf{s})$, where $\mathbf{s} \in \mathbb{F}_q^{n-k}$. The output is an $\mathbf{e} \in \mathbb{F}_q^n$ of minimal weight such that $\mathbf{e}H^T = \mathbf{s}$. The relation of the two versions is given by $\mathbf{s} = \mathbf{y}H^T$ the *syndrome* and $\mathbf{e} = \mathbf{y} - \mathbf{c}$ the error vector of the received word $\mathbf{y}$. The *bounded distance decoding problem* depends on a function $t(n, k, d)$. The input is again $(H, \mathbf{s})$ but the output is a word $\mathbf{e}$ (if any) such that $\mathrm{wt}(\mathbf{e}) \leq t(n, k, d)$. Moreover *decoding up to half the minimum distance* is the bounded distance decoding problem such that $t(n, k, d) = \lfloor (d-1)/2 \rfloor$ for all n, k and d. The solution of the decoding problems posed above has two parts [13]. Firstly the *preprocessing* part done at a laboratory or a factory where for an appropriate code C a decoder $\mathcal{A}_C$ is built which is allowed to be time consuming. Secondly the actual operating of the many copies of the decoder for consumers which should work very fast. So we can consider the problem of *minimum distance decoding with preprocessing*. From the error-correction point of view it seems pointless to decode a bad code, but for breaking the McEliece cryptosystem by a general or *generic attack* one must be able to decode efficiently all, or almost all, codes.

The security of code-based cryptosystems is based on the hardness of decoding up to half the minimum distance. The minimum distance decoding problem was shown by Berlekamp-McEliece-Van Tilborg [3] to be NP-hard. However it is not known whether this problem is almost always or in the average difficult. The status of the hardness of decoding up to half the minimum distance is an open problem. McEliece proposed to use binary Goppa codes for his PKC system.

All known minimum distance decoding algorithms for general codes have exponential complexity in the length of the code. The complexity exponent of decoding general binary codes up to half the minimum distance has been lowered from above $1/3$ that is 0.3869 for brute force decoding to below $1/20$ that is 0.04934. See [1, 31]. However there are several classes of codes such as the generalized Reed-Solomon (GRS), BCH, Goppa or algebraic geometry codes which have polynomial decoding algorithms that correct up to a certain bound which is at most half the minimum distance.

In the McEliece PKC system the *public key space* $\mathcal{K}$ is the collection of all generator matrices of a chosen class of codes that have an efficient decoding algorithm that corrects all patterns of t errors, the *plaintext space* is $\mathcal{P} = \mathbb{F}_q^k \times W_{n,q,t}$, where $W_{n,q,t}$ is the collection of all $\mathbf{e} \in \mathbb{F}_q^n$ of weight t, and the *ciphertext space* is $\mathcal{C} = \mathbb{F}_q^n$. The *sample space* is given by $\Omega = \mathcal{P} \times \mathcal{K}$. The *encryption map* $E_G : \mathcal{P} \to \mathcal{C}$ for a given key $G \in \mathcal{K}$ is defined by $E_G(\mathbf{m}, \mathbf{e}) = \mathbf{m}G + \mathbf{e}$. An *adversary* $\mathcal{A}$ is a map from $\mathcal{C} \times \mathcal{K}$ to $\mathcal{P}$. This adversary is successful for $(x, G) \in \Omega$ if $\mathcal{A}(E_G(x), G) = x$.

Let $\mathcal{C}$ be a class of codes such that every code C in $\mathcal{C}$ has an efficient decoding algorithm correcting all patterns of t errors. Let $G \in \mathbb{F}_q^{k \times n}$ be a generator matrix of C. In order to mask the origin of G, take a $k \times k$ invertible matrix S over $\mathbb{F}_q$ and an $n \times n$ permutation or monomial matrix P. Then for the McEliece PKC the matrices G, S and P are kept secret while $G' = SGP$ is made public. Furthermore the (trapdoor) one-way function of this cryptosystem is usually presented as follows:

$$x = (\mathbf{m}, \mathbf{e}) \mapsto y = \mathbf{m}G' + \mathbf{e},$$

where $\mathbf{m} \in \mathbb{F}_q^k$ is the plaintext and $\mathbf{e} \in \mathbb{F}_q^n$ is a random error vector with Hamming weight at most t.

McEliece proposed to use the family of Goppa codes. The problem of bounded distance decoding for the class of codes that have the same parameters as the Goppa codes is difficult in the worst-case. However, it is still an open problem whether decoding up to half the minimum distance is NP-hard which is the security basis of the McEliece cryptosystem.

In 1986 Niederreiter [27] presented a dual version of McEliece cryptosystem which is equivalent in terms of security. Niederreiter's system differs from McEliece's system in the public-key structure (it uses a parity check matrix instead of a generator matrix of the code), in the encryption mechanism (we compute the syndrome of a message by the public key) and in the decryption message. Niederreiter proposed several classes of codes such as alternant codes which contains the Goppa codes as subclass, algebraic geometry codes introduced by Goppa [14] and GRS codes.

Let $H \in \mathbb{F}_q^{(n-k) \times n}$ be a parity check matrix of a code C in $\mathcal{C}$. H is masked by $H' = SHP$, where S is an invertible matrix over $\mathbb{F}_q$ of size $n - k$ and P is an $n \times n$ permutation or monomial matrix. The (trapdoor) one-way function in case of the Niederreiter PKC is presented by

$$x = \mathbf{m} \mapsto y = H'\mathbf{m}^T,$$

where $\mathbf{m} \in \mathbb{F}_q^n$ has weight t.

The security of a *general attack* of the PKC systems of McEliece and Niederrieter is based on two assumptions [5, 15]:

A.1 In the average it is difficult to decode t errors for all codes that have the same parameters as the codes used as key,

A.2 It is difficult to distinguish arbitrary codes from those coming from $\mathcal{K}$.

Concerning the second assumption recent progress is made [8, 10, 11] where it is shown that one can distinguish between high rate Goppa, alternant and random codes.

Assuming the Kerckhoff principle, the attacker knows the class $\mathcal{K}$. A *key recovery* or *structural attack* uses the special structure of codes in the class of $\mathcal{K}$. For instance Sidelnikov-Shestakov gave an adversary that is always successful if one takes for public key space the generator matrices of GRS codes.

It was shown in [9, 17, 28, 29] that the known efficient bounded distance decoding algorithms of GRS, BCH, Goppa and algebraic geometry codes can be described by a basic algorithm using an error-correcting pair. That means that the proposed McEliece cryptosystems that use these classes of codes are in fact using the error-correcting pair as a secret key. Hence the security of these PKC systems is not only based on the inherent intractability of bounded distance decoding but also on the assumption that it is difficult to retrieve efficiently an error-correcting pair.

3 Error-correcting pairs

From now on the dimension of a linear code C will be denoted by $k(C)$ and its minimum distance by $d(C)$. Given two elements $\mathbf{a}$ and $\mathbf{b}$ in $\mathbb{F}_q^n$, the *star multiplication* is defined by coordinatewise multiplication, that is $\mathbf{a}*\mathbf{b} = (a_1b_1, \ldots, a_nb_n)$ while the *standard inner multiplication* is defined by $\mathbf{a} \cdot \mathbf{b} = \sum_{i=1}^{n} a_ib_i$.

Let A, B and C be subspaces of $\mathbb{F}_q^n$. Then $A * B$ is the subspace generated by $\{\mathbf{a} * \mathbf{b} \mid \mathbf{a} \in A \text{ and } \mathbf{b} \in B\}$. And $C^{\perp} = \{\mathbf{x} | \mathbf{x} \cdot \mathbf{c} = 0 \text{ for all } \mathbf{c} \in C\}$ is the *dual* code of C. Furthermore $A \perp B$ if and only if $\mathbf{a} \cdot \mathbf{b} = 0$ for all $\mathbf{a} \in A$ and $\mathbf{b} \in B$.

Consider the following properties:

$E.1$	$(A * B) \perp C$	$E.4$	$d(A) + d(C) > n$
$E.2$	$k(A) >$	$E.5$	$d(A^{\perp}) > 1$
$E.3$	$d(B^{\perp}) > t$	$E.6$	$d(A) + 2t > n$

Let C be a linear code in $\mathbb{F}_q^n$. The pair (A, B) of linear codes over $\mathbb{F}_{q^m}$ of length n is called a *t-error-correcting pair* (ECP) for C if the conditions $E.1$, $E.2$, $E.3$ and $E.4$ hold.

If (A, B) is a pair of codes that satisfies conditions $E.1$, $E.2$, $E.3$, $E.5$ and $E.6$, then $d(C) \geq 2t + 1$ and (A, B) is a t-ECP for C by [29, Corollary 3.4].

The notion of an error-correcting pair for a linear code was introduced in 1988 by Pellikaan [28] and independently by Kötter in [17, 18] in 1992. It is shown that a linear code in $\mathbb{F}_q^n$ with a t-error-correcting pair has a decoding algorithm which corrects up to t errors with complexity $\mathcal{O}(n^3)$.

Generalized Reed-Solomon (GRS) codes are the prime examples of codes that have a t-error-correcting pair. Moreover if C is an $[n, n - 2t, 2t + 1]$ code which has a t-error-correcting pair, then C is a GRS code. This is trivial if $t = 1$, proved for $t = 2$ in [29, Theorem 6.5] and for arbitrary t in [23].

The existence of ECP's for GRS and algebraic geometry codes was shown in [28]. For many cyclic codes Duursma and Kötter in [9, 17, 18] have found ECP's which correct beyond the designed BCH capacity.

The class of GRS codes was mentioned by Niederreiter [27] in his proposal of a code-based PKC. However GRS codes are not suited for a coded-based PKC by the attack of Sidelnikov-Shestakov.

A binary Goppa code with parameters $[1024, 524, 101]$ as proposed by McEliece is no longer secure with nowadays computing power as shown in Peters et al. [31, 30] by improving decoding algorithms for general codes.

The class of subcodes of GRS codes was proposed by Berger-Loidreau [2] for code-based PKC to resist precisely the Sidelnikov-Shestakov attack. But for certain parameter choices this proposal is also not secure as shown by Wieschebrink [33, 34] and Márquez et al. [20].

Algebraic geometry codes were proposed by Niederreiter [27] and Janwa-Moreno [14] for code based PKC systems. This system was broken for low genus at most two [12, 25] and for arbitrary genus by Márquez et al. [7, 19, 21] for certain choices of the parameters.

Subfield subcodes of algebraic geometry codes were proposed by Janwa-Moreno [14] and broken by Couvreur et al. [6] for certain parameters. The class of Goppa codes remains so far unbroken.

4 The ECP one-way function

Let $\mathcal{P}(n, t, q)$ be the collection of pairs (A, B) such that there exist a positive integer m and a pair (A, B) of $\mathbb{F}_{q^m}$-linear codes of length n, that satisfy the conditions E.2, E.3, E.5 and E.6. Let C be the $\mathbb{F}_q$-linear code of length n that is the subfield subcode that has the elements of $A * B$ as parity checks. So

$$C = \mathbb{F}_q^n \cap (A * B)^{\perp}.$$

Then the minimum distance of C is at least $2t + 1$ and (A, B) is a t-ECP for C as was noted before. Let $\mathcal{F}(n, t, q)$ be the collection of $\mathbb{F}_q$-linear codes of length n and minimum distance $d \geq 2t + 1$. Consider the following map

$$\begin{array}{rccc} \varphi_{(n,t,q)} : & \mathcal{P}(n, t, q) & \longrightarrow & \mathcal{F}(n, t, q) \\ & (A, B) & \longmapsto & C. \end{array}$$

The question is whether this map is a one-way function.

If the map $\varphi_{(n,t,q)}$ is indeed difficult to invert, then we will call it the *ECP one-way function* and the code C with parity check matrix W might be used as a public-key in a coding based PKC. Otherwise it would mean that the PKC based on codes that can be decoded by error-correcting pairs is not secure.

Let $\mathcal{K}$ be a collection of generator matrices of codes that have a t-error-correcting pair and that is used for a coded-based PKC system. We address assumption whether we can distinguish arbitrary codes from those coming from $\mathcal{K}$.

5 Acknowledgement

An earlier version of this work was presented for the first time by the second author at the *Code-Based Cryptography Workshop*, May 2012 at the Technical University of Denmark, Lyngby and posted at arXiv [22] and furthermore at the conferences Applications of Computer Algebra 2013 and 2014 at Malaga and Fordham, respectively.

References

1. Becker, A., Joux, A., May, A., Meurer, A.: Decoding random binary linear codes in $2^{n/20}$: How $1 + 1 = 0$ improves information set decoding. In: Advances in cryptology—EUROCRYPT 2012, Lecture Notes in Comput. Sci., vol. 7237, pp. 520–536. Springer, Heidelberg (2012)
2. Berger, T., Loidreau, P.: How to mask the structure of codes for a cryptographic use. Designs, Codes and Cryptography 35, 63–79 (2005)
3. Berlekamp, E., McEliece, R., van Tilborg, H.: On the inherent intractability of certain coding problems. IEEE Transactions on Information 24, 384–386 (1978)
4. Bernstein, D.: Introduction to post-quantum cryptography. In: D.J. Bernstein, J.B., Dahmen, E. (eds.) Post-quantum cryptography, pp. 1–14. Springer-Verlag, Berlin (2009)
5. Biswas, B., Sendrier, N.: McEliece cryptosystem implementation : Theory and practice. In: Post-Quantum Cryptography, Lecture Notes in Computer Science. vol. 5299, pp. 47–62. Springer, Berlin (2008)
6. Couvreur, A., Márquez-Corbella, I., Pellikaan, R.: Cryptanalysis of public-key cryptosystems that use subcodes of algebraic geometry codes. In: 4ICMCTA, Coding Theory and Application, CIM Series in Mathematical Sciences 3, pp. 133–140 (2014)
7. Couvreur, A., Márquez-Corbella, I., Pellikaan, R.: A polynomial time attack against algebraic geometry code based public key cryptosystems. In: IEEE International Symposion on Information Theory ISIT 2014, p. 1446 (2014)
8. Couvreur, A., Gaborit, P., Gauthier-Umaña, V., Otmani, A., Tillich, J.P.: Distinguisher-based attacks on public-key cryptosystems using Reed-Solomon codes. Des. Codes Cryptogr. 73(2), 641–666 (2014)
9. Duursma, I., Kötter, R.: Error-locating pairs for cyclic codes. IEEE Trans. Inform. Theory 40, 1108–1121 (1994)
10. Faugère, J.C., Otmani, A., Perret, L., Tillich, J.P.: Algebraic cryptanalysis of McEliece variants with compact keys. In: Gilbert, H. (ed.) EUROCRYPT 2010, Lecture Notes in Computer Science. vol. 6110, pp. 279–298. Springer, Berlin (2010)
11. Faugère, J.C., Gauthier-Umaña, V., Otmani, A., Perret, L., Tillich, J.P.: A distinguisher for high-rate McEliece cryptosystems. IEEE Trans. Inform. Theory 59(10), 6830–6844 (2013)
12. Faure, C., Minder, L.: Cryptanalysis of the McEliece cryptosystem over hyperelliptic codes. In: Proceedings 11th Int. Workshop on Algebraic and Combinatorial Coding Theory, ACCT 2008. pp. 99–107 (2008)
13. Høholdt, T., Pellikaan, R.: On decoding algebraic-geometric codes. IEEE Transactions on Information 41, 1589–1614 (1995)
14. Janwa, H., Moreno, O.: McEliece public crypto system using algebraic-geometric codes. Designs, Codes and Cryptography 8, 293–307 (1996)

15. Kobara, K., Imai, H.: Semantically secure McEliece public-key cryptosystems - conversions for McEliece PKC. In: PKC 2001, Lecture Notes in Computer Science. vol. 1992, pp. 19–35. Springer, Berlin (2001)
16. Koblitz, N., Menezes, A.: The brave new world of bodacious assumptions in cryptography. Notices Amer. Math.Soc. 57(3), 357–365 (2010)
17. Kötter, R.: A unified description of an error locating procedure for linear codes. In: Proceedings of Algebraic and Combinatorial Coding Theory, pp. 113–117. Voneshta Voda (1992)
18. Kötter, R.: On algebraic decoding of algebraic-geometric and cyclic codes. Ph.D. thesis, Linköping University of Technology, Linköping Studies in Science and Technology, Dissertation no. 419 (1996)
19. Márquez-Corbella, I., Martínez-Moro, E., Pellikaan, R.: On the unique representation of very strong algebraic geometry codes. Designs, Codes and Cryptography pp. 215–230 (2012)
20. Márquez-Corbella, I., Martínez-Moro, E., Pellikaan, R.: The non-gap sequence of a subcode of a generalized Reed-Solomon code. Designs, Codes and Cryptography 66(1-3) (2013)
21. Márquez-Corbella, I., Martínez-Moro, E., Pellikaan, R., Ruano, D.: Computational aspects of retrieving a representation of an algebraic geometry code. J. Symbolic Computation 64, 67–87 (2014)
22. Márquez-Corbella, I., Pellikaan, R.: Error-correcting pairs for a public-key cryptosystem. Preprint arXiv:1205.3647 (2012)
23. Márquez-Corbella, I., Pellikaan, R.: A characterization of MDS codes that have an error-correcting pair. Preprint arXiv:1508.02187 (2015)
24. McEliece, R.: A public-key cryptosystem based on algebraic coding theory. DSN Progress Report 42–44, 114–116 (1978)
25. Minder, L.: Cryptography based on error correcting codes. Ph.D. thesis, 3846 EPFL (2007)
26. Misoczki, R., Tillich, J.P., Sendrier, N., Barreto, P.: MDPC-McEliece: New McEliece variants from moderate density parity-check codes. In: Information Theory Proceedings (ISIT), 2013 IEEE International Symposium on. pp. 2069–2073 (2013)
27. Niederreiter, H.: Knapsack-type crypto systems and algebraic coding theory. Problems of Control and Information Theory 15(2), 159–166 (1986)
28. Pellikaan, R.: On decoding by error location and dependent sets of error positions. Discrete Math. 106–107, 369–381 (1992)
29. Pellikaan, R.: On the existence of error-correcting pairs. Statistical Planning and Inference 51, 229–242 (1996)
30. Peters, C.: Information-set decoding for linear codes over Fq. In: Post-Quantum Cryptography, Lecture Notes in Computer Science. vol. 6061, pp. 81–94. Springer, Berlin (2010)
31. Peters, C.: Curves, codes and cryptography. Ph.D. thesis, Technical University Eindhoven (2011)
32. Singh, S.: The code book: Science of secrecy from ancient Egypt to quantum cryptography. Anchor Books, New York (1999)
33. Wieschebrink, C.: An attack on the modified Niederreiter encryption scheme. In: PKC 2006, Lecture Notes in Computer Science. vol. 3958, pp. 14–26. Springer, Berlin (2006)
34. Wieschebrink, C.: Cryptanalysis of the Niederreiter public key scheme based on GRS subcodes. In: Post-Quantum Cryptography, Lecture Notes in Computer Science. vol. 6061, pp. 61–72. Springer, Berlin (2010)