Computational Aspects of Retrieving a Representation of an Algebraic Geometry Code (Extended abstract)

I. Márquez-Corbella[1], E. Martínez-Moro[1], G.R. Pellikaan[2], and D. Ruano[3]

[1] Institute of Mathematics IMUVa
Universidad de Valladolid, Castilla, Spain
imarquez@agt.uva.es,edgar@maf.uva.es
[2] Department of Mathematics and Computing Science
Eindhoven University of Technology, Eindhoven, The Netherlands
g.r.pellikaan@tue.nl
[3] Department of Mathematical Sciences
Aalborg University, Denmark
diego@math.aau.dk

Abstract. Code-based cryptography is an interesting alternative to classic number-theory PKC since it is conjectured to be secure against quantum computer attacks. Many families of codes have been proposed for these cryptosystems such as algebraic geometry codes. In a previous paper [9] we showed that for so called very strong algebraic geometry codes $\mathcal{C} = \mathcal{C}_L(\mathcal{X}, \mathcal{P}, E)$ where $\mathcal{X}$ is an algebraic curve over $\mathbb{F}_q$ and $\mathcal{P} = (P_1, \ldots, P_n)$ is an n-tuple of mutually distinct $\mathbb{F}_q$-rational points of $\mathcal{X}$ and E is a divisor of $\mathcal{X}$ with disjoint support from $\mathcal{P}$ it was shown that an equivalent representation $\mathcal{C} = \mathcal{C}_L(\mathcal{Y}, \mathcal{Q}, F)$ can be found. The n-tuple of points are obtained directly from a generator matrix of C, where the columns are viewed as homogeneous coordinates of these points. The curve $\mathcal{Y}$ is given by $I_2(\mathcal{Y})$, the homogeneous elements of degree 2 of the vanishing ideal $I(\mathcal{Y})$. Furthermore it was shown that $I_2(\mathcal{Y})$ can be computed in an efficient as the kernel of certain linear map. What was not shown was how to get the divisor F and a decoding algorithm in an efficient way. In this talk show some work in progress on the topics needed to be dealt towards an efficient computational approach to this problem.

Keywords: Algebraic geometry codes, McEliece PKC

Introduction

In 1978, McEliece [11] introduced the first public key cryptosystem (PKC) based on the theory of error-correcting codes in particular he proposed to use a classical binary Goppa code. The security of this scheme is based on the hardness of the decoding problem for general linear codes and the hardness of distinguishing a code with the prescribed structure from a random one. Moreover, McEliece scheme an interesting candidate for post-quantum cryptography. An overview of

the state of the art of cryptosystems that are secure against attacks by quantum computers is provided in [3]. Another advantage of this scheme is its fast encryption and decryption functions.

Many attempts to replace Goppa codes with different families of codes have been proven to be insecure as for example using GRS codes such as the original Niederreiter system [12] which was broken by Sidelnikov and Shestakov [13] in 1992.

let $\mathcal{X}$ be an algebraic curve of genus g over the finite field $\mathbb{F}_q$, $\mathcal{P} = (P_1, \ldots, P_n)$ be an n-tuple of mutually distinct $\mathbb{F}_q$-rational points of $\mathcal{X}$ and E a divisor of $\mathcal{X}$ with disjoint support from $\mathcal{P}$ of degree m. We define the *vector space of rational functions associated to* E as the set

$$\mathcal{L}(E) = \{f \in \mathbb{F}_q(\mathcal{X}) \mid f = 0 \text{ or } (f) \geq -E\},$$

and the *linear series* of E as the collection $|E| = \{\ F \mid F \equiv E, F \geq 0\ \}$. Then the following evaluation map

$$\mathrm{ev}_{\mathcal{P}} : \mathcal{L}(E) \longrightarrow \mathbb{F}_q^n$$

is well defined by $\mathrm{ev}_{\mathcal{P}}(f) = (f(P_1), \ldots, f(P_n))$. The *algebraic geometry code* $\mathcal{C}_L(\mathcal{X}, \mathcal{P}, E)$ is the image of $\mathcal{L}(E)$ under the evaluation map $\mathrm{ev}_{\mathcal{P}}$, i.e.

$$\mathcal{C}_L(\mathcal{X}, \mathcal{P}, E) = \{(f(P_1), \ldots, f(P_n) \mid f \in \mathcal{L}(E))\} \subseteq \mathbb{F}_q^n.$$

As consequence of the Riemann-Roch theorem, if $n > m > 2g - 2$ then $\mathcal{C}_L(\mathcal{X}, \mathcal{P}, E)$ has dimension $m + 1 - g$ and minimum distance at least $n - m$.

Recall that GRS codes can be seen as the special class of algebraic geometry codes on the projective line, that is the algebraic curve of genus zero. This result was generalized to curves of genus 1 and 2 by Faure and Minder [5] in 2008. These attacks can be viewed as retrieving the curve, n points on this curve and the divisor E.

Since the initial Niederreiter scheme is completely broken, Berger and Loidreau [2] proposed in 2005 another version which was designed to resist precisely the Sidelnikov-Shestakov attack. The main idea of this variant is to work with subcodes of the original GRS code rather than using the complete GRS code. However Wieschebrink [14] in 2006 presents the first feasible attack to the Berger-Loidreau cryptosystem that allows us to recover the secret key if the chosen subcode is large enough but which was impractical for small subcodes. Furthermore in 2010 Wieschebrink [15] noted that it seems that with high probability the square code of a subcode of a GRS code of parameters $[n, k]$ is itself a GRS code of dimension $2k - 1$.

Therefore we can apply the Sidelnikov-Shestakov attack and thus reconstruct the secret key in polynomial time. Continuing this line of work, in [10], we characterized those subcodes which are weak keys for the Berger-Loidreau cryptosystem. That is, firstly those subcodes which are themselves GRS codes, we have seen that the probability of occurrence of this fact is very small, and secondly those subcodes whose square code is a GRS code of maximal dimension which has high probability of occurrence.

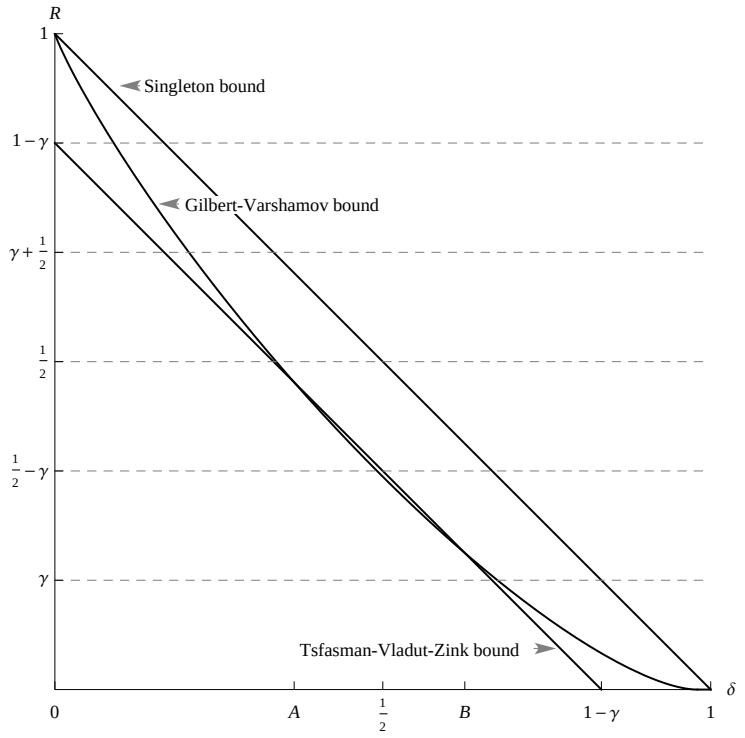

Fig. 1. Bounds on R as a function of the relative minimum distance δ for $q = 49$ and $\gamma = \frac{1}{6}$.

In 1996 Janwa and Moreno [7] proposed to use the collection of AG codes on curves for the McEliece cryptosystem. As we have already explained this system was broken for codes on curves of genus $g \leq 2$ by Faure and Minder [5]. But the security status of this proposal for higher genus was not known.

Definition 1. *A code $\mathcal{C}$ over $\mathbb{F}_q$ is called* very strong algebraic-geometric *(VSAG) if $\mathcal{C}$ is equal to $\mathcal{C}_L(\mathcal{X}, \mathcal{P}, E)$ where the curve $\mathcal{X}$ over $\mathbb{F}_q$ has genus g, $\mathcal{P}$ consists of n points and E has degree m such that*

$$2g + 2 \leq m < \tfrac{1}{2}n \quad or \quad \tfrac{1}{2}n + 2g - 2 < m \leq n - 4.$$

In [9] we proved the following result

Theorem 1. *Let $\mathcal{C}$ be a VSAG code then a VSAG representation can be obtained from its generator matrix. Moreover all VSAG representations of $\mathcal{C}$ are strict isomorphic.*

Theorem 1 implies, **provided we have an efficient procedure for decoding the VSAG representation obtained in the theorem**, that one

should not use VSAG codes for the McEliece PKC system in the range

$$\gamma \leq R \leq \tfrac{1}{2} - \gamma \quad \text{or} \quad \tfrac{1}{2} + \gamma \leq R \leq 1 - \gamma,$$

for $n \to \infty$, since there is an efficient attack by our result. In the same paper, by a shortening argument, we proved that also codes in the range

$$\tfrac{1}{2} - \gamma \leq R \leq 1 - 3\gamma \quad \text{or} \quad 3\gamma \leq R \leq \tfrac{1}{2} + \gamma,$$

for $n \to \infty$, should be excluded. The above mentioned intervals $[\gamma, \frac{1}{2} - \gamma]$, $[\frac{1}{2} + \gamma, 1-\gamma]$, $[\frac{1}{2} - \gamma, 1-3\gamma]$ and $[3\gamma, \frac{1}{2} + \gamma]$ are nonempty if and only if $\gamma \leq \frac{1}{4}$, and the union of these intervals cover the whole interval $[\gamma, 1 - \gamma]$ if and only if $\gamma \leq \frac{1}{6}$.

Work in progress

As it was mention before, a VSAG representation isomorphic to the original code can be built from the public key of the PKC (the scrambled generator matrix of the original code). Indeed, decoding the VSAG representation implies decoding the original code, i.e. breaking the cryptosystem. The purpose of this research is twofold

1. Compute efficiently the VSAG representation, i.e. retrieving the triple given by the curve, a set of points and the divisor defining the functions to be evaluated.
2. Decode the code given by VSAG representation.

Up to now we have made some advances in direction 1. Indeed, if the VSAG representation lies in some of the families of AG codes that are provided with an efficient error correcting procedure this will imply tht the PKC based on the original code would be broken.

Computing the VSAG representation

Let $r = l(E) - 1$ and $\{f_0, \ldots, f_r\}$ be a basis of $\mathcal{L}(E)$. Consider the following map:

$$\varphi_E : \mathcal{X} \longrightarrow \mathbb{P}^r(\mathbb{F}_q)$$

defined by $\varphi_E(P) = (f_0(P), \ldots, f_r(P))$.

If $m > 2g$ then $r = m-g$, so φ_E defines an embedding of the curve $\mathcal{X}$ of degree m in $\mathbb{P}^r$. More precisely, let $\mathcal{Y} = \varphi_E(\mathcal{X})$, $Q_j = \varphi_E(P_j)$ and $\mathcal{Q} = (Q_1, \ldots, Q_n)$. Then $\mathcal{Y}$ is a curve in $\mathbb{P}^{m-g}$ of degree m, φ_E is an isomorphism from $\mathcal{X}$ to $\mathcal{Y}$ and $\varphi_E(E) = \mathcal{Y} \cdot H$ for some hyperplane H of $\mathbb{P}^{m-g}$ that is disjoint from $\mathcal{Q}$. See [6, Theorems 7.33 and 7.40]. Let $F = \varphi_E(E) = \mathcal{Y} \cdot H$. Then $\mathcal{C} = \mathcal{C}_L(\mathcal{Y}, \mathcal{Q}, F)$, that is $(\mathcal{Y}, \mathcal{Q}, F)$ is also a representation of the code $\mathcal{C}$ which is strict isomorphic with $(\mathcal{X}, \mathcal{P}, E)$.

Computing $\mathcal{Y}$. Let $\mathcal{C}$ be a k dimensional subspace of $\mathbb{F}_q^n$ with basis $\{\mathbf{g}_1, \ldots, \mathbf{g}_k\}$. We denote by $S^2(\mathcal{C})$ the second symmetric power of $\mathcal{C}$. If $x_i = \mathbf{g}_i$, then $S^2(\mathcal{C})$ has basis $\{x_i x_j \mid 1 \leq i \leq j \leq n\}$ and dimension $\binom{k+1}{2}$. Furthermore we denote by $\langle \mathcal{C} * \mathcal{C} \rangle$ or $\mathcal{C}^{(2)}$ the square of $\mathcal{C}$, that is the linear subspace in $\mathbb{F}_q^n$ generated by $\{\mathbf{a} * \mathbf{b} | \mathbf{a}, \mathbf{b} \in C\}$. See [4, §4 Definition 6] and [10, 15]. Now we consider the linear map

$$\sigma : S^2(\mathcal{C}) \longrightarrow \mathcal{C}^{(2)},$$

where the element $x_i x_j$ is mapped to $\mathbf{g}_i * \mathbf{g}_j$. The kernel of this map will be denoted by $K^2(\mathcal{C})$.

Proposition 1 (Proposition 15 in [9]). *Let $\mathcal{Q}$ be an n-tuple of points in $\mathbb{P}^r(\mathbb{F}_q)$ not in a hyperplane, $k = r + 1$, $G_{\mathcal{Q}}$ be the $k \times n$ matrix associated to $\mathcal{Q}$ and $\mathcal{C}$ be the subspace of $\mathbb{F}_q^n$ generated by the rows of $G_{\mathcal{Q}}$. Then*

$$I_2(\mathcal{Q}) = \{ \textstyle\sum_{1 \leq i \leq j \leq k} a_{ij} X_i X_j \mid \sum_{1 \leq i \leq j \leq k} a_{ij} x_i x_j \in K^2(\mathcal{C}) \}.$$

Let $\mathcal{Q}$ be an n-tuple of points in $\mathbb{P}^r(\mathbb{F}_q)$ not in a hyperplane. Then $\mathcal{O}(n^2\binom{r}{2})$ is an upper bound on the complexity of the computation of $I_2(\mathcal{Q})$ and a Gröbner basis of this ideal can be computed by straight-forward adaptation of the *Projective version of the classical Buchberger-Möller Algorithm* presented in [1] for the special case where we know that the elements of the reduced Gröbner basis have degree two.

Computing $E = \mathcal{Y} \cdot H$. Let $\mathbf{g}_1, \ldots, \mathbf{g}_k$ be the rows of the chosen generator matrix G of $\mathcal{C}$. By the star product $*$ he vector space $\mathbb{F}_q^n$ is an $\mathbb{F}_q$-algebra. Consider the map of $\mathbb{F}_q$-algebras

$$\varepsilon : \mathbb{F}_q[X_1, \ldots, X_k] \longrightarrow \mathbb{F}_q^n$$

given by $X_i \mapsto \mathbf{g}_i$ for $i = 1, \ldots, k$ and extended by the universal property of $\mathbb{F}_q[X_1, \ldots, X_k]$ as an $\mathbb{F}_q$-algebra.

Let R be the factor ring $R = \mathbb{F}_q[X_1, \ldots, X_k]/I(\mathcal{Y})$. The ideal $I(\mathcal{Y})$ is in the kernel of ε. Hence ε induces a map

$$\varepsilon : R \longrightarrow \mathbb{F}_q^n,$$

that we also denote by ε. Let R_d be the subspace of R given by cosets of homogeneous polynomials of degree d. Then $\varepsilon(R_1) = C$ by construction of ε, and more generally $\varepsilon(R_d) = C^{(d)}$.
Let $f(X)$ be a nonzero linear function in R_1. Then $\varepsilon(f(X)) = \mathbf{g}$ is a nonzero codeword of $\mathcal{C}$ and $\varepsilon(f(X)R_1) = \mathbf{g} * C$.

Let H be the hyperplane given by the linear equation $f(X) = 0$. We may assume without loss of generality after possibly extending the field of constants that $E = \mathcal{Y} \cdot H$ that there is a nonzero function $f \in \mathcal{L}(E)$ such that $(f)_\infty = E$, that means that the divisor of poles of f is equal to E. Let $\mathbf{g} = \mathrm{ev}_{\mathcal{P}}(f) \in \mathcal{C}_L(\mathcal{X}, \mathcal{P}, E) = \mathcal{C}$. Then $\mathbf{g} * C$ is a subspace of $\mathcal{C}^{(2)}$ and the coset $\mathcal{C}^{(2)}/\mathbf{g} * C$ has

dimension $(2m+1-g)-(m+1-g)=m$. Therefore we have an explicitly given $\mathbb{F}_q$-linear map:

$$\mathbb{F}_q[X_1,\ldots,X_k] \longrightarrow \mathcal{C}^{(2)}/\mathbf{g}*C$$

with kernel the ideal $I_2(\mathcal{Y})+(f)$, that is the vanishing ideal of $\mathcal{Y}\cap H$ with multiplicities counted. In this situation there is an efficient (polynomial) algorithm that computes a Gröbner basis of $I_2(\mathcal{Y})+(f)$, see [8].

References

1. J. Abbott, A. Bigatti, M. Kreuzer, and L. Robbiano. Computing ideals of points. *J. Symbolic Comput.*, 30(4):341–356, 2000.
2. T. Berger and P. Loidreau. How to mask the structure of codes for a cryptographic use. *Designs, Codes and Cryptography*, 35:63–79, 2005.
3. D.J. Bernstein. Introduction to post-quantum cryptography. In J. Buchmann D.J. Bernstein and E. Dahmen, editors, *Post-quantum cryptography*, pages 1–14. Springer-Verlag, Berlin, 2009.
4. I. Cascudo, H. Chen, R. Cramer, and X. Xing. Asymptotically good ideal linear secret sharing with strong multiplication overy any fixed finite field. In S. Halevi, editor, *Advances in Cryptology - CRYPTO 2009, Lecture Notes in Computer Science*, volume 5677, pages 466–486, Berlin, 2009. Springer.
5. C. Faure and L. Minder. Cryptanalysis of the McEliece cryptosystem over hyperelliptic codes. In *Proceedings 11th Int. Workshop on Algebraic and Combinatorial Coding Theory, ACCT 2008*, pages 99–107, 2008.
6. J. W. P. Hirschfeld, G. Kochmáros, and F. Torres. *Algebraic curves over a finite field*. Princeton Univ. Press, Princeton, 2008.
7. H. Janwa and O. Moreno. McEliece public crypto system using algebraic-geometric codes. *Designs, Codes and Cryptography*, 8:293–307, 1996.
8. M.G. Marinari, H.M. Möller, and T. Mora. Gröbner basis of ideals defined by functionals with an application to ideals of projective points. *AAECC*, 4(2):103–145, 1993.
9. I. Márquez-Corbella, E. Martínez-Moro, and G.R. Pellikaan. Cryptanalysis of public-key cryptosystems based on algebraic geometry codes. *To appear in Designs, Codes and Cryptography*, pages 20, MFO–Preprint OWP 2012 – 01, http://www.mfo.de/scientific–programme/publications/owp, 2012.
10. I. Márquez-Corbella, E. Martínez-Moro, and R. Pellikaan. The non-gap sequence of a subcode of a generalized Reed-Solomon code. In *To appear in Designs, Codes and Cryptography*, 2012.
11. R. J. McEliece. A public-key cryptosystem based on algebraic coding theory. *DSN Progress Report*, 42–44:114–116, 1978.
12. H. Niederreiter. Knapsack-type crypto systems and algebraic coding theory. *Problems of Control and Information Theory*, 15(2):159–166, 1986.
13. V. M. Sidelnikov and S. O. Shestakov. On the insecurity of cryptosystems based on generalized Reed-Solomon codes. *Discrete Math. Appl.*, 2:439–444, 1992.
14. C. Wieschebrink. An attack on the modified Niederreiter encryption scheme. In *PKC 2006, Lecture Notes in Computer Science*, volume 3958, pages 14–26, Berlin, 2006. Springer.
15. C. Wieschebrink. Cryptanalysis of the Niederreiter public key scheme based on GRS subcodes. In *Post-Quantum Cryptography, Lecture Notes in Computer Science*, volume 6061, pages 61–72, Berlin, 2010. Springer.