Cryptanalysis of public-key cryptosystems that use subcodes of algebraic geometry codes

Alain Couvreur, Irene Márquez-Corbella and Ruud Pellikaan

Abstract We give a polynomial time attack on the McEliece public key cryptosystem based on subcodes of algebraic geometry (AG) codes. The proposed attack reposes on the distinguishability of such codes from random codes using the Schur product. Wieschebrink treated the genus zero case a few years ago but his approach cannot be extent straightforwardly to other genera. We address this problem by introducing and using a new notion, which we call the *t–closure* of a code.

Key words: Algebraic geometry codes, code-based cryptography, Schur products of codes, distinguishers.

1 Introduction

After the original proposal of code based encryption scheme due to McEliece [**?**] which was based on binary Goppa codes, several alternative proposals aimed at reducing the key size by using codes with a higher correction capacity. Among many others, generalised Reed–Solomon (GRS) codes are proposed in 1986 by Niederreiter [**?**] but are subject to a key-recovery polynomial time attack discovered by Sidelnikov and Shestakov [**?**] in 1992. To avoid this attack, Berger and Loidreau [**?**]

Alain Courvreur
INRIA, SACLAY & LIX, CNRS UMR 7161, École Polytechnique 91128 Palaiseau Cedex, e-mail: alain.couvreur@lix.polytechnique.fr

Irene Márquez-Corbella
INRIA, SACLAY & LIX, CNRS UMR 7161, École Polytechnique 91128 Palaiseau Cedex, e-mail: irene.marquez-corbella@inria.fr

Ruud Pellikaan
Departement of Mathematics and Computing Science, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven e-mail: g.r.pellikaan@tue.nl

proposed to replace GRS codes by some random subcodes of small codimension. This proposal has been broken by Wieschebrink [**?**] using Schur products of codes.

Another proposal was to use algebraic geometry (AG) codes, concatenated AG codes or their subfield subcodes [**?**]. The case of AG codes of genus 1 and 2 has been broken by Faure and Minder [**?**]. Then, Marquez et. al. proved that the structure of a curve can be recovered from the very knowledge of an AG code [**?**, **?**] without leading to an efficient attack. Finally a polynomial time attack of the scheme based on AG codes has been obtained by the authors in [**?**]. This attack consists in using the particular behaviour of AG codes with respect to the Schur product to compute a filtration of the public key by AG subcodes, which leads to the design of a polynomial time decoding algorithm allowing encrypted message recovery.

The genus zero case and Berger Loidreau's proposal raises a natural question **what about using subcodes of AG codes?** In this article we propose an attack of this scheme. Compared to the genus zero case, Wieschebrink's attack cannot extend straightforwardly and we need to introduce and use a new notion which we call the *t–closure* of a code. By this manner, we prove subcodes of AG codes to be non secure when the subcode has a small codimension. It is worth noting that choosing a subcode of high codimension instead of the code itself represents a huge loss in terms of error correction capacity and hence is in general a bad choice. For this reason, an attack on the small codimension codes is of interest.

Finally, it hardly needs to be recalled that this result does not imply the end of code-based cryptography since Goppa codes, alternant codes and more generally subfield subcodes of AG codes still resist to any known efficient attack. Their resistance to the presented attack is discussed at the end of the article.

Due to space reasons, many proofs are omitted in this extended abstract.

2 Notation and prerequisites

2.1 Curves and algebraic geometry codes

The interested reader is referred to [**?**, **?**] for further details on the notions introduced in the present subsection. In this article, $\mathscr{X}$ denotes a smooth projective geometrically connected curve of genus g over a finite field $\mathbb{F}_q$. We denote by $P = (P_1, \ldots, P_n)$ an n-tuple of mutually distinct $\mathbb{F}_q$-rational points of $\mathscr{X}$, by D_P the divisor $D_P = P_1 + \cdots + P_n$ and by E an $\mathbb{F}_q$-divisor of degree $m \in \mathbb{Z}$ and support disjoint from that of D_P.

The function field of $\mathscr{X}$ is denoted by $\mathbb{F}_q(\mathscr{X})$. Given an $\mathbb{F}_q$-divisor E on $\mathscr{X}$, the corresponding Riemann-Roch space is denoted by $L(E)$. The *algebraic geometry (AG) code* $\mathscr{C}_L(\mathscr{X}, P, E)$ of length n over $\mathbb{F}_q$ is the image of the evaluation map

$$\mathrm{ev}_P : \begin{cases} L(E) & \longrightarrow & \mathbb{F}_q^n \\ f & \longmapsto & (f(P_1), \ldots, f(P_n)) \end{cases}$$

If $2g-2 < m < n$, then by Riemann-Roch Theorem, $\mathscr{C}_L(\mathscr{X},P,E)$ has dimension $m+1-g$ and minimum distance at least $n-m$.

When the curve is the projective line $\mathbb{P}^1$, the corresponding codes are the so-called *generalised Reed–Solomon* (GRS) codes defined as:

$$\mathrm{GRS}_k(\mathbf{a},\mathbf{b}) := \{(b_1 f(a_1),\ldots,b_n f(a_n)) \mid f \in \mathbb{F}_q[x]_{<k}\}.$$

where $\mathbf{a},\mathbf{b}$ are two n–tuples in $\mathbb{F}_q^n$ such that the entries of $\mathbf{a}$ are pairwise distinct and those of $\mathbf{b}$ are all nonzero and $k < n$.

Remark 1. See [**?**, Example 3.3] for a description of GRS codes as AG codes.

2.2 Schur product

Given two elements $\mathbf{a}$ and $\mathbf{b}$ in $\mathbb{F}_q^n$, the *Schur product* is the component wise multiplication: $\mathbf{a}*\mathbf{b} = (a_1 b_1,\ldots,a_n b_n)$. Let $\mathbf{a} \in \mathbb{F}_q^n$, we set $\mathbf{a}^0 := (1,\ldots,1)$ and by induction we define $\mathbf{a}^{j+1} := \mathbf{a}*\mathbf{a}^j$ for any positive integer j. If all entries of $\mathbf{b}$ are nonzero, we define $\mathbf{b}^{-1} := (b_1^{-1},\ldots,b_n^{-1})$ and thus, $\mathbf{b}^{-j} = \left(\mathbf{b}^j\right)^{-1}$ for any positive integer j.

For two codes $A,B \subseteq \mathbb{F}_q^n$, the code $A*B$ is defined by

$$A*B := \mathrm{Span}_{\mathbb{F}_q}\left\{\mathbf{a}*\mathbf{b} \mid \mathbf{a} \in A \text{ and } \mathbf{b} \in B\right\}.$$

For $B = A$, then $A*A$ is denoted as $A^{(2)}$ and, we define $A^{(t)}$ by induction for any positive integer t.

2.2.1 Application to Decoding, error correcting pairs and arrays

The notion of *error-correcting pair* (ECP) for a linear code was introduced by Pellikaan [**?**, **?**] and independently by Kötter [**?**]. Broadly speaking, given a positive integer t, a t–ECP for a linear code $\mathscr{C} \subseteq \mathbb{F}_q^n$ is a pair of linear codes (A,B) in $\mathbb{F}_q^n$ satisfying $A*B \subseteq \mathscr{C}^\perp$ together with several inequalities relating t and the dimensions and (dual) minimum distances of A,B,C. This data provides a decoding algorithm correcting up to t errors in $O(n^3)$ operations in $\mathbb{F}_q$. ECP's provide a unifying point of view for several classical bounded distance decoding for algebraic and AG codes. See [**?**] for further details.

For an AG code, there always exists a t–ECP with $t = \lfloor \frac{d^*-1-g}{2} \rfloor$, where d^* denotes the *Goppa designed distance* (see [**?**, Definition 2.2.4]). Thus, ECP's allow to correct up to half the designed distance minus $g/2$. Filling this gap and correct up to half the designed distance is possible thanks to more elaborate algorithms based on the so-called *error correcting arrays*. See [**?**, **?**] for further details.

2.2.2 Distinguisher and Cryptanalysis

Another and more recent application of the Schur product concerns cryptanalysis of code-based public key cryptosystems. In this context, the Schur product is a very powerful operation which can help to distinguish some algebraic codes such as AG codes from random ones. The point is that evaluation codes do not behave like random codes with respect to the Schur product: the square of an AG code is very small compared to that of a random code of the same dimension. Thanks to this observation, Wieschebrink [**?**] gave an efficient attack of Berger Loidreau's proposal [**?**] based on subcodes of GRS codes.

Recent attacks consist in pushing this argument forward and take advantage to this distinguisher in order to compute a filtration of the public code by a family of very particular subcodes. This filtration method yields an alternative attack on GRS codes [**?**]. Next it leads to a key recovery attack on wild Goppa codes over quadratic extensions in [**?**]. Finally in the case of AG codes, this approach lead to an attack [**?**] which consists in the computation of an ECP for the public code without retrieving the structure of the curve, the points and the divisor.

3 The attack

Our public key is a non structured generator matrix $\mathbf{G}$ of a subcode C of $\mathscr{C}_L(\mathscr{X},P,E)^{\perp}$ of dimension l, together with the error correcting capacity t. The goal of our attack is to recover the code $\mathscr{C}_L(\mathscr{X},P,E)^{\perp}$ from the knowledge of C and then use the attack of [**?**] which provides a t–ECP and hence a decoding algorithm for $\mathscr{C}_L(\mathscr{X},P,E)$, which yields a fortiori a decoding algorithm for C.

The genus zero case (i.e. the case of GRS codes) proposed in [**?**] was broken by Wieschebrink [**?**] as follows:

- C is the public key contained in some secret $\mathrm{GRS}_k(\mathbf{a},\mathbf{b})$.
- Compute $C^{(2)}$ which is, with a high probability, equal to $\mathrm{GRS}_k(\mathbf{a},\mathbf{b})^{(2)}$, which is itself equal to $\mathrm{GRS}_{2k-1}(\mathbf{a},\mathbf{b}^2)$.
- Apply Sidelnikov Shestakov attack [**?**] to recover $\mathbf{a}$ and $\mathbf{b}^2$, then find $\mathbf{b}$.

Compared to Wieschebrink's approach, our difficulty is that the attack [**?**] is not a key-recovery attack but a blind construction of a decoding algorithm. For this reason, even if $C^{(2)}$ provides probably the code $\mathscr{C}_L(\mathscr{X},P,E)^{(2)}$, it is insufficient for our purpose: we need to find $\mathscr{C}_L(\mathscr{X},P,E)$. This is the reason why we introduce the notion of t–closures.

3.1 The t-closure operation

Definition 1 (t–closure). Let $C \subset \mathbb{F}_q^n$ be a code and $t \geqslant 2$ be an integer. The *t-closure* of C is defined by

$$\overline{C}^t = \left\{ \mathbf{a} \in \mathbb{F}_q^n \mid \mathbf{a} * C^{(t-1)} \subseteq C^{(t)} \right\}.$$

The code C is said to be *t-closed* if $\overline{C}^t = C$.

Proposition 1. *Let $C \in \mathbb{F}_q^n$, then for all $t \geqslant 2$,*

$$\overline{C}^t = \left(C^{(t-1)} * \left(C^{(t)} \right)^{\perp} \right)^{\perp}.$$

Proposition 2. *Let E be a divisor satisfying $\deg(E) \geqslant 2g+1$. Then:*

(i) $\mathscr{C}_L(\mathscr{X}, P, E)^{(t)} = \mathscr{C}_L(\mathscr{X}, P, tE)$.
(ii) $\overline{\mathscr{C}_L(\mathscr{X}, P, E)}^t = \mathscr{C}_L(\mathscr{X}, P, E)$ *if* $\deg(E) \leqslant \frac{n-2}{t}$.

Proof. (i) is proved in [**?**] and is a consequence of [**?**]. For (ii), Proposition 1 shows that

$$\overline{\mathscr{C}_L(\mathscr{X}, P, E)}^t = \left(\mathscr{C}_L(\mathscr{X}, P, E)^{(t-1)} * \left(\mathscr{C}_L(\mathscr{X}, P, E)^{(t)} \right)^{\perp} \right)^{\perp}. \tag{1}$$

Moreover, $\mathscr{C}_L(\mathscr{X}, P, tE)^{\perp} = \mathscr{C}_L(\mathscr{X}, P, (tE)^{\perp})$ where $(tE)^{\perp} = D_P - tE + K$ for some canonical divisor K on $\mathscr{X}$. Thus, $\deg\left((tE)^{\perp}\right) = n - \deg(tE) + 2g - 2$. Since, by assumption, $\deg(E) \leqslant \frac{n-2}{t}$ we have $\deg\left((tE)^{\perp}\right) \geqslant 2g$. Moreover, since $\deg E \geqslant 2g+1$, then, thanks to (i), Equation (1) yields

$$\mathscr{C}_L(\mathscr{X}, P, (t-1)E) * \mathscr{C}_L(\mathscr{X}, P, tE)^{\perp} = \mathscr{C}_L(\mathscr{X}, P, D_P - E + K) = \mathscr{C}_L(\mathscr{X}, P, E)^{\perp}.$$

□

Corollary 1. *Let E be a divisor and $2g+1 \leqslant \deg(E) \leqslant \frac{n-2}{2}$. Then $\overline{\mathscr{C}_L(\mathscr{X}, P, E)}^2 = \mathscr{C}_L(\mathscr{X}, P, E)$.*

Conjecture 1. If $2g+1 \leqslant \deg(E) \leqslant \frac{n-1}{2}$, let C be subcode of $\mathscr{C}_L(\mathscr{X}, P, E)$ of dimension l such that and $2k+1-g \leqslant \binom{l+1}{2}$, where $k = \deg(E)+1-g$ is the dimension of $\mathscr{C}_L(\mathscr{X}, P, E)$, then the probability that $C^{(2)}$ is different from $\mathscr{C}_L(\mathscr{X}, P, 2E)$ tends to 0 when k tends to infinity.

We give a proof along the lines of [**?**, Remark 5] for the special case of subcodes of GRS codes. Our experimental results are in good agreement with this conjecture (see Table 1). The following corollary is central to our attack.

Corollary 2. *If $2g+1 \leqslant \deg(E) \leqslant \frac{n-2}{2}$ and $2k+1-g \leqslant \binom{l+1}{2}$ for $k = \deg(E)+1-g$, then the equality $\overline{C}^2 = \mathscr{C}_L(\mathscr{X}, P, E)$ holds for random l-dimensional subcodes C of $\mathscr{C}_L(\mathscr{X}, P, E)$ with a probability tending to 0 when k tends to infinity.*

3.2 Principle of the attack

The public key consists in $C \subseteq \mathscr{C}_L(\mathscr{X},P,E)^\perp$ and $t = \left\lfloor \frac{d^*-g-1}{2} \right\rfloor$. Set $l := \dim C$. First, let us assume moreover that

$$2g+1 \leqslant \deg(E) \leqslant \tfrac{n-1}{2},\ k = \deg(E)+1-g \ \text{ and } \ 2k-1+g \leqslant \tbinom{l+1}{2}.$$

Step 1. With a high probability, we may assume that $C^{(2)} = \mathscr{C}_L(\mathscr{X},P,2E)$ and hence $\overline{C}^2 = \mathscr{C}_L(\mathscr{X},P,E)$ by Corollary 2. Thus, compute $\overline{C}^2$ by solving a linear system or by applying Proposition 1.
Step 2. Apply the polynomial time attack presented in [**?**] to obtain an ECP, denoted by (A,B), for $\mathscr{C}_L(\mathscr{X},P,E)$. Which yields a decoding algorithm for C.

Estimated complexity: The computation of a closure costs $O(n^4)$ operations in $\mathbb{F}_q$ and the rest of the attack is in $O((\log(t+g))n^4)$ (see [**?**] for further details).

In case $\deg(E) > \frac{n-1}{2}$, then the attack can be applied to several shortenings of C whose 2–closures are computed separately and are then summed up to provide $\mathscr{C}_L(\mathscr{X},P,E)$. This method is described and applied in [**?**, **?**].

This attack has been implemented with MAGMA. To this end L random subcodes of dimension l from Hermitian codes of parameters $[n,k]_q$ were created. It turned out that for all created subcodes a t-ECP could be reconstructed. `Time` represents the average time of the attack obtained with an Intel ® CoreTM 2 Duo 2.8 GHz. The work factor **w** of an ISD attack is given. These work factors have been computed thanks to Christiane Peter's Software [**?**].

q	n	k	t	Time	key size	**w**	l	L
7^2	343	193	54	80 s	83 ko	2^{30}	50	1000
					137 ko	2^{43}	100	1000
					163 ko	2^{62}	150	1000

q	n	k	t	Time	key size	**w**	l	L
9^2	729	521	19	30 min	216 ko	2^{32}	50	500
					670 ko	2^{121}	200	500
					835 ko	2^{178}	400	500

Table 1 Running times of the attack over Hermitian codes

3.3 Which codes are subject to this attack?

Basically, the subcode $C \subseteq \mathscr{C}_L(\mathscr{X},P,E)$ should satisfy:

(i) $\binom{\dim C+1}{2} \geqslant \dim \mathscr{C}_L(\mathscr{X},P,2E)$;
(ii) $2g+1 \leqslant \deg E \leqslant \frac{n-2}{2}$;

The left-hand inequality of (ii) is in general satisfied. On the other hand, as explained above, the right-hand inequality of (ii) can be relaxed by using a shortening trick. Constraint (i) is more central since a subcode which does not satisfies it will probably behave like a random code and it can be checked that a random code is in

general 2–closed. Thus, computing the 2–closure of such a subcode will not provide any significant result. On the other hand, for an AG code of dimension k, subcodes which do not satisfy (i) have dimension smaller than $\sqrt{2k}$ and choosing such very small subcodes and decode them as subcodes of $\mathscr{C}_L(\mathscr{X},P,E)$ would represent a big loss of efficiency. In addition, if these codes are too small they can be subject to generic attacks like information set decoding.

3.3.1 Subfield subcodes still resist

Another class of subcodes which resist to this attack are the subcodes C such that $\overline{C}^2 \subsetneq \mathscr{C}_L(\mathscr{X},P,E)$. It is rather difficult to classify such subcodes but there is a very identifiable family: the subfield subcodes. Let $\mathbb{F}$ be a proper subfield of $\mathbb{F}_q$ (here we assume q to be non prime) and let $C := \mathscr{C}_L(\mathscr{X},P,E) \cap \mathbb{F}^n$ (and then apply a base field extension if one wants to have an $\mathbb{F}_q$–subcode). The point is that $C^2 \subseteq (\mathscr{C}_L(\mathscr{X},P,E)^{(2)}) \cap \mathbb{F}_q^n$ and the 2-closure of C will in general differ from $\mathscr{C}_L(\mathscr{X},P,E)$. For this reason, subfield subcodes resist to this kind of attacks. Notice that even in genus zero: subfield subcodes of GRS codes still resist to filtration attacks unless for the cases presented in [**?**].