Recent attacks on McEliece schemes based on Goppa codes

J.-P. Tillich

June 9, 2014

1. The McEliece cryptosystem

- 1978 McEliece cryptosystem based on Goppa codes.

- Secret Key : A generator matrix $\boldsymbol{G}$ of an $[n,k]_q$ code $\mathcal{C}$ having an efficient t–correcting algorithm;

- Public Key : $\boldsymbol{G}' := S\boldsymbol{G}P$, where $S \in \mathrm{GL}(k, \mathbb{F}_q)$ and P is an $n \times n$ permutation matrix;

- Encryption : $m \in \mathbb{F}_q^k \longmapsto y \stackrel{\text{def}}{=} m\boldsymbol{G}' + e$ with $|e| = t$.

- Decryption : $y \longmapsto yP^{-1} = mS\boldsymbol{G} + eP^{-1} \longmapsto mS \longmapsto m$.

Advantages/drawbacks

Advantages

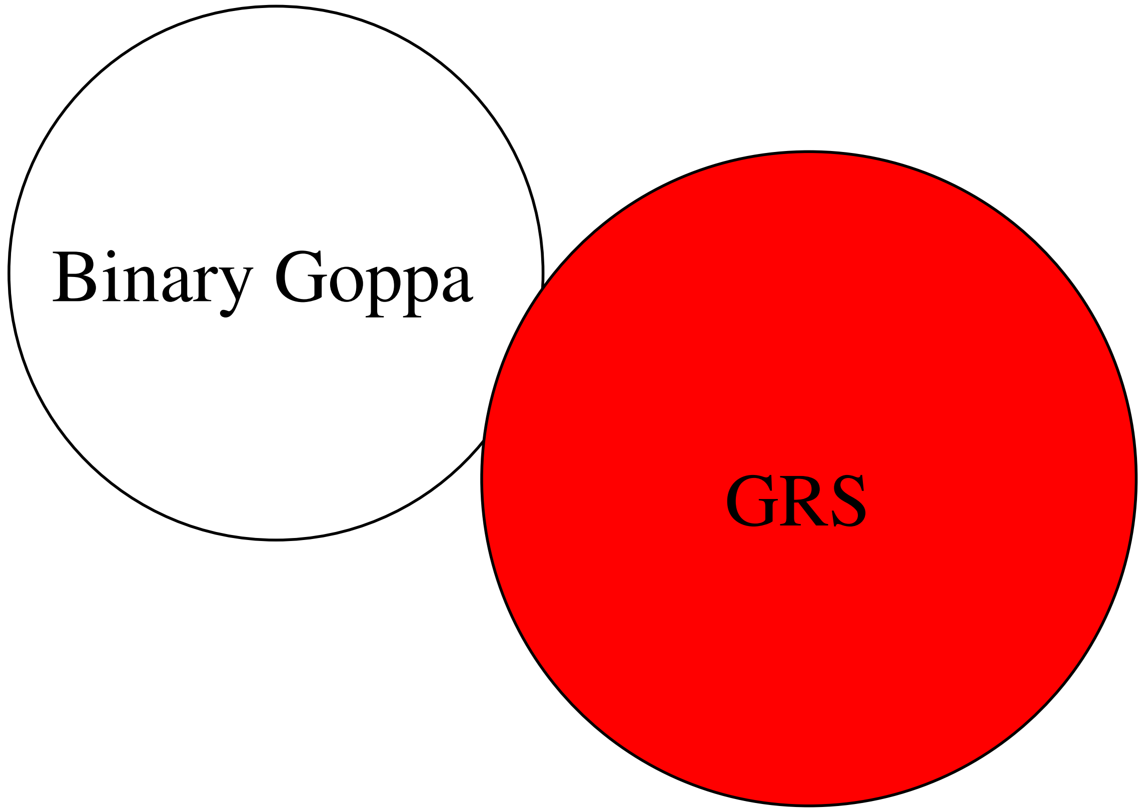
- Post Quantum;
- Efficient encryption and decryption (compared to RSA, El Gamal): the original McEliece has encryption ≈ 5 times faster than RSA 1024, decryption ≈ 150 times faster than RSA 1024.

Drawbacks

- Huge size of the keys: the original proposal (McEliece 1978) has a 67ko key (more than 500 times RSA 1024 for a similar security).

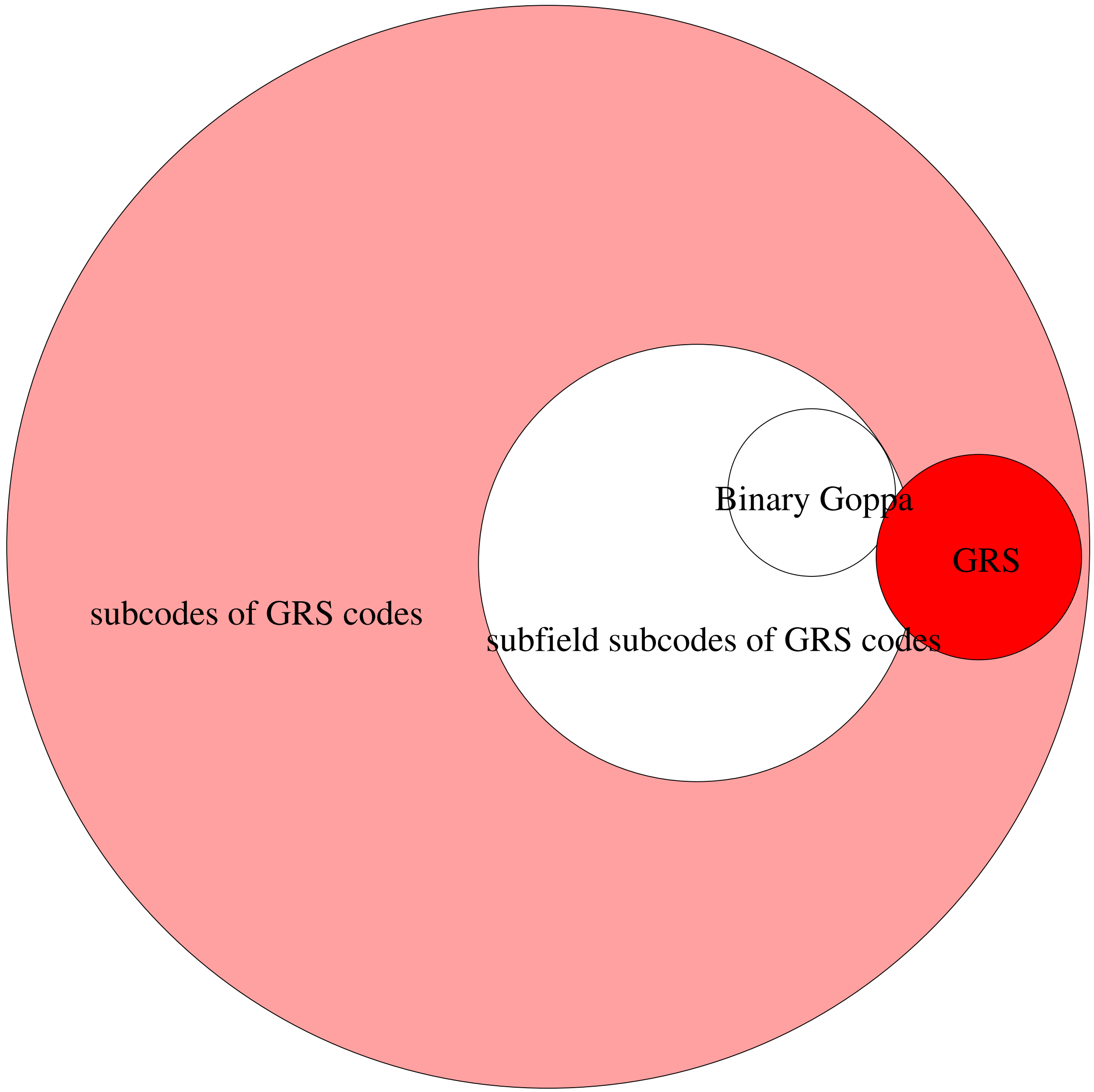
Variants based on generalized Reed-Solomon codes

- 1986 Niederreiter variant based on GRS codes.
- 1992 Sidelnikov-Shestakov attack.
- 2006 Wieschebrink, reparation of the Niederreiter scheme by adding random columns to the generator matrix.
- 2011 Baldi-Bianchi-Chiaraluce-Rosenthal-Schipani, reparation of the Niederreiter scheme by changing the permutation matrix Π into $\Pi + R$ where R is of rank one.
- 2011, Bogdnanov-Lee, homomorphic public-key encryption scheme based on Reed-Solomon codes.
- 2013, Couvreur-Gaborit-Gauthier-Otmani-Tillich, attack on all these variants based on square code considerations.
- 2013 Couvreur-Gaborit-Gauthier-Otmani-Tillich, filtration attack on GRS codes.


Binary Goppa
GRS

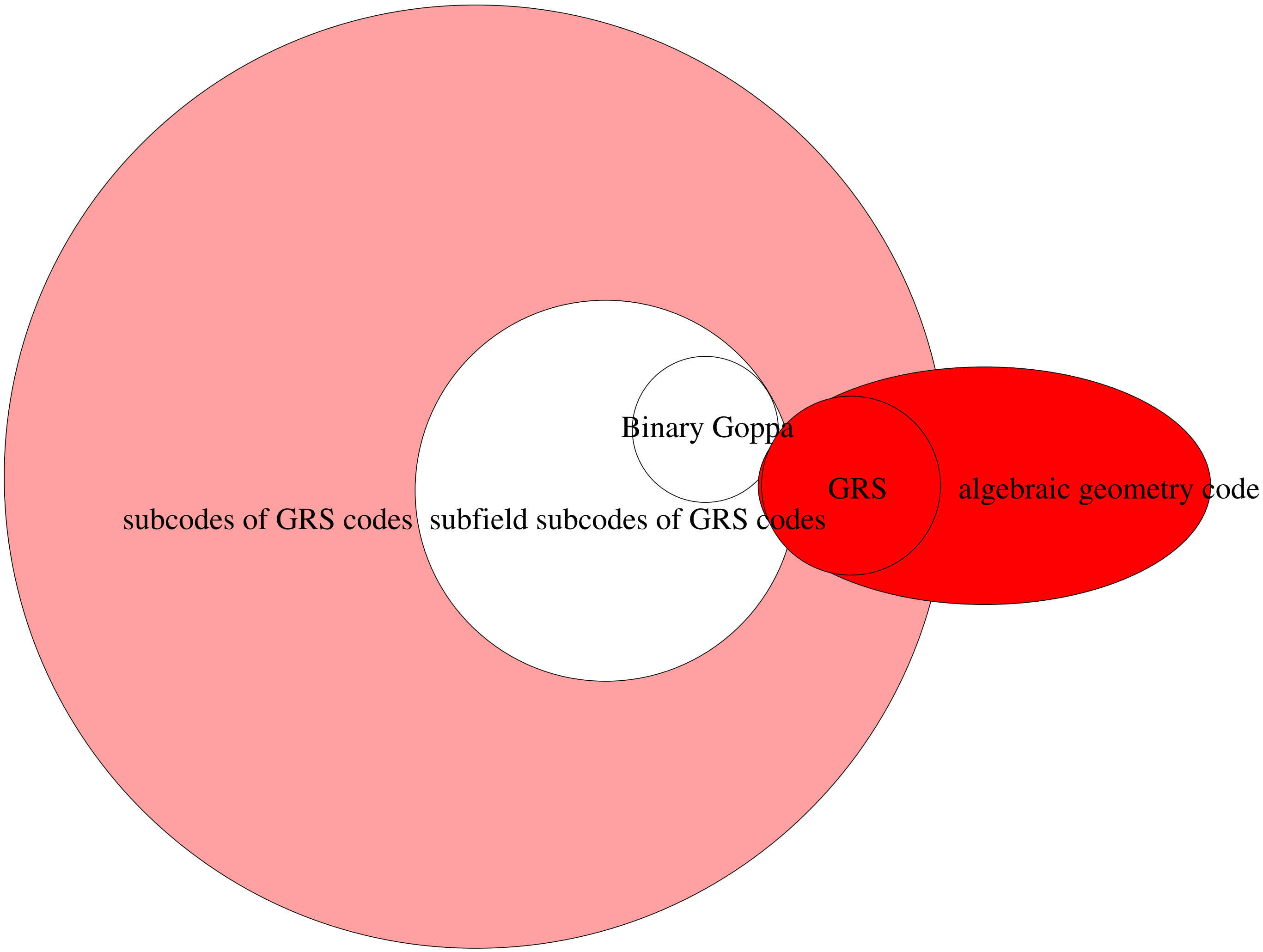
Variants based on subcodes of generalized Reed-Solomon codes.

- 2005 Berger-Loidreau : subcodes of generalized Reed-Solomon codes.
- 2010 Wieschebrink : attack by square code considerations.


Binary Goppa
GRS
subcodes of GRS codes
subfield subcodes of GRS codes

Variants based on algebraic geometric codes

- 1996 : proposed by Janwa-Moreno.
- 2008 : Attacked by Faure-Minder for hyperelliptic curves of genus $\leqslant 2$.
- 2014 : Attacked in general by recovering an error-correcting pair from square code and filtration considerations by Couvreur-Màrquez Corbella-Pellikaan.


Binary Goppa
GRS
algebraic geometry code
subcodes of GRS codes
subfield subcodes of GRS codes

Variants based on Reed-Muller codes.

- 1994 Suggested by Sidelnikov.
- 2007 Attack by Minder-Shokrollahi in sub-exponential time by recovering the structure from minimal codewords.
- 2013 Chizhov-Borodin refinement of the attack by square code considerations.

Alternant/Goppa codes with symmetry

- 2005 Gaborit : quasi-cyclic subcodes of BCH codes.
- 2007 Otmani-Tillich-Dallot : attack.
- 2009 Berger-Cayrel-Gaborit-Otmani : quasi-cyclic alternant codes.
- 2009 Misoczki-Barreto : quasi-dyadic Goppa codes.
- 2010 Faugère-Otmani-Perret-Tillich/Gauthier-Leander : almost all 2009 schemes were broken with an algebraic attack (possible because of the reduction of the number of unknowns).

Other variants

- 199. a zillion propositions with LDPC codes.
- 2000 Monico-Rosenthal-Shokrollahi : attack.
- 2007: Baldi-Chiaraluce"repairing" the LDPC schemes by taking sums of permutation matrices.
- 2007 Otmani-Tillich-Dallot : attack.
- 2008 Baldi-Bodrato-Chiaraluce : a new version.
- 2012 Misoczki-Tillich-Barreto-Sendrier : MDPC codes.
- 2012 Löndahl-Johansson : convolutional codes.
- 2013 Landais-Tillich : attack.

2. Algebraic attacks through square codes

Generalized Reed-Solomon codes

Definition 1. [Generalized Reed-Solomon code] *Let k and n be integers such that $1 \leqslant k < n \leqslant q$ where q is a power of a prime number. The generalized Reed-Solomon code* $\mathbf{GRS}_k(\boldsymbol{x}, \boldsymbol{y})$ *of dimension k is associated to a pair $(\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{F}_q^n \times \mathbb{F}_q^n$ where $\boldsymbol{x}$ is an n-tuple of distinct elements of $\mathbb{F}_q$ and the entries y_i are arbitrary nonzero elements in $\mathbb{F}_q$.* $\mathbf{GRS}_k(\boldsymbol{x}, \boldsymbol{y})$ *is defined as:*

$$\mathbf{GRS}_k(\boldsymbol{x}, \boldsymbol{y}) \stackrel{def}{=} \Big\{(y_1 p(x_1), \dots, y_n p(x_n)) : p \in \mathbb{F}_q[X], \deg p < k\Big\}.$$

$\boldsymbol{x}$ is the support and $\boldsymbol{y}$ the multiplier.

[Sidelnikov-Shestakov1992]: recover from an arbitrary generator matrix of a GRS code $\mathcal{C}$, a tuple $(\boldsymbol{x}, \boldsymbol{y})$ such that $\mathcal{C} = \mathbf{GRS}(\boldsymbol{x}, \boldsymbol{y})$ (all what is needed to decode $\mathcal{C}$ efficiently).

The square code

Definition 2. [Componentwise product] *Given two vectors* $\boldsymbol{a} = (a_1, \ldots, a_n)$ *and* $\boldsymbol{b} = (b_1, \ldots, b_n) \in \mathbb{F}_q^n$, *we denote by* $\boldsymbol{a} \star \boldsymbol{b}$ *the componentwise product*

$$\boldsymbol{a} \star \boldsymbol{b} \stackrel{\text{def}}{=} (a_1 b_1, \ldots, a_n b_n)$$

Definition 3. [Product of codes & square code] *The* star product code *denoted by* $\mathcal{A} \star \mathcal{B}$ *of* $\mathcal{A}$ *and* $\mathcal{B}$ *is the vector space spanned by all products* $\boldsymbol{a} \star \boldsymbol{b}$ *where* $\boldsymbol{a}$ *and* $\boldsymbol{b}$ *range over* $\mathcal{A}$ *and* $\mathcal{B}$ *respectively. When* $\mathcal{B} = \mathcal{A}$, $\mathcal{A} \star \mathcal{A}$ *is called the* square code *of* $\mathcal{A}$ *and is rather denoted by* $\mathcal{A}^2$.

Dimension of the square code

$\mathcal{A}$ and $\mathcal{B}$ codes with respective bases $(\boldsymbol{a}_i)$ and $(\boldsymbol{b}_j)$.

1. $\dim(\mathcal{A} \star \mathcal{B}) \leqslant \dim(\mathcal{A}) \dim(\mathcal{B})$ (generated by the $\boldsymbol{a}_i \star \boldsymbol{b}_j$'s)

2. $\dim(\mathcal{A}^2) \leqslant \binom{\dim(\mathcal{A}) + 1}{2}$ (generated by the $\boldsymbol{a}_i \star \boldsymbol{a}_j$'s with $i \leqslant j$)

What is wrong with generalized Reed-Solomon codes ?

When $\mathcal{C}$ is a random code of length n, with high probability

$$\dim(\mathcal{C}^2) = \min\left\{\binom{\dim(\mathcal{C})+1}{2}, n\right\}$$

When $\mathcal{C}$ is a generalized Reed-Solomon code

$$\dim(\mathcal{C}^2) = \min\left\{2\dim(\mathcal{C}) - 1, n\right\}$$

The explanation

$$\boldsymbol{c} = (y_1 p(x_1), \ldots, y_n p(x_n)),\ \boldsymbol{c}' = (y_1 q(x_1), \ldots, y_n q(x_n)) \in \mathbf{GRS}_k(\boldsymbol{x}, \boldsymbol{y})$$

where p and q are two polynomials of degree at most $k-1$.

$$\boldsymbol{c} \star \boldsymbol{c}' = \left(y_1^2 p(x_1) q(x_2), \ldots, y_n^2 p(x_n) q(x_n)\right) = \left(y_1^2 r(x_1), \ldots, y_n^2 r(x_n)\right)$$

where r is a polynomial of degree $\leqslant 2k-2$.

$$\Longrightarrow \boldsymbol{c} \star \boldsymbol{c}' \in \mathbf{GRS}_{2k-1}(\boldsymbol{x}, \boldsymbol{y}^2)$$

3. Couvreur-Otmani-Tillich : filtration attack

1st polynomial-time attack on McEliece based on certain Goppa codes.

A filtration for GRS codes

A new attack on McEliece based on GRS codes.

known : $C_0 = \mathbf{GRS}_k(\boldsymbol{x}, \boldsymbol{y})$

unknown : $\boldsymbol{x}$,$\boldsymbol{y}$.

$$C_0 = \mathbf{GRS}_k(\boldsymbol{x}, \boldsymbol{y}) \supseteq C_1 = \mathbf{GRS}_{k-1}(\boldsymbol{x}, \boldsymbol{y}) \supseteq \cdots \supseteq C_{k-1} = \mathbf{GRS}_1(\boldsymbol{x}, \boldsymbol{y})$$

The point:

- $C_{k-1} = \{\alpha \boldsymbol{y}, \alpha \in \mathbb{F}_q\}$
- $\boldsymbol{y}$ known $\Rightarrow$ $\boldsymbol{x}$ by solving a linear system.

Square code considerations and the filtration

Assumption : We know $C_0 = \textbf{GRS}_k(\boldsymbol{x}, \boldsymbol{y})$.

Bold assumption : we also know $C_1 = \textbf{GRS}_{k-1}(\boldsymbol{x}, \boldsymbol{y})$

Proposition 1. $C_2 = \textbf{GRS}_{k-2}(\boldsymbol{x}, \boldsymbol{y})$ *is the set of* $\boldsymbol{c}$ *satisfying*

$$\begin{cases} \boldsymbol{c} \in \textbf{GRS}_{k-1}(\boldsymbol{x}, \boldsymbol{y}) \\ \boldsymbol{c} \star \textbf{GRS}_k(\boldsymbol{x}, \boldsymbol{y}) \subseteq \textbf{GRS}_{k-1}(\boldsymbol{x}, \boldsymbol{y})^{\star 2} \end{cases}$$

Viewing codewords as polynomials

Consider $\boldsymbol{c} \in \mathbf{GRS}_{k-1}(\boldsymbol{x}, \boldsymbol{y})$, then there exists a polynomial $p(X)$ in $\mathbb{F}_q[X]$ of degree $\leqslant k-2$ such that

$$c_i = y_i p(x_i)$$

$$\begin{array}{rcl} \boldsymbol{c} \star \mathbf{GRS}_k(\boldsymbol{x}, \boldsymbol{y}) & \subseteq & \mathbf{GRS}_{k-1}(\boldsymbol{x}, \boldsymbol{y})^{\star 2} \\ & \Downarrow & \\ (y_i p(x_i) y_i \underbrace{q(x_i)}_{\deg \leqslant k-1})_i & \in & \underbrace{\mathbf{GRS}_{k-1}(\boldsymbol{x}, \boldsymbol{y})^{\star 2}}_{\deg \leqslant 2k-4} \quad \text{for all } q \text{ of } \deg < k \\ & \Downarrow & \\ \deg p & \leqslant & k-3 \end{array}$$

Polynomial point of view

$$C_0 = \textbf{GRS}_k(\boldsymbol{x}, \boldsymbol{y}) \supseteq C_1 = \textbf{GRS}_{k-1}(\boldsymbol{x}, \boldsymbol{y}) \supseteq \cdots \supseteq C_{k-1} = \textbf{GRS}_1(\boldsymbol{x}, \boldsymbol{y})$$

corresponds to

$$\mathbb{F}_q[z]_{<k} \;\supseteq\; \mathbb{F}_q[z]_{<k-1} \;\supseteq\; \cdots \;\supseteq\; \mathbb{F}_q[z]_{<1}$$

Elementary linear algebra

Computing a basis of the $\boldsymbol{c}$ satisfying

$$\begin{cases} \boldsymbol{c} \in \mathbf{GRS}_{k-1}(\boldsymbol{x}, \boldsymbol{y}) \\ \boldsymbol{c} \star \mathbf{GRS}_k(\boldsymbol{x}, \boldsymbol{y}) \subseteq \mathbf{GRS}_{k-1}(\boldsymbol{x}, \boldsymbol{y})^{\star 2} \end{cases}$$

can be done by elementary linear algebra : solving a linear system.

A better filtration

$\mathbf{GRS}_{k-1}(\boldsymbol{x}, \boldsymbol{y})$ unknown, consider instead the filtration corr. to

$$\mathbb{F}_q[z]_{<k} \supseteq z\mathbb{F}_q[z]_{<k-1} \supseteq \cdots \supseteq z^{\ell}\mathbb{F}_q[z]_{<k-\ell} \supseteq \cdots$$

The first two terms are known.

- The first $\mathcal{C} = \mathbf{GRS}_k(\boldsymbol{x}, \boldsymbol{y})$
- The second: its shortening in the first position (w.l.o.g. we may assume $x_1 = 0$).

$$\begin{pmatrix} 1 & * & \dots & * \\ 0 & a'_{11} & \dots & a'_{1,n-1} \\ \vdots & \vdots & & \vdots \\ 0 & a'_{k-1,1} & \dots & a'_{k-1,n-1} \end{pmatrix}$$

What about alternant/Goppa codes ?

Definition 1. *Let $\boldsymbol{x} \in \mathbb{F}_{q^m}^n, \boldsymbol{y} \in \mathbb{F}_{q^m}^n$ be as in the definition of GRS codes. The alternant code $\textbf{Alt}_r(\boldsymbol{x}, \boldsymbol{y})$ is defined by*

$$\textbf{Alt}_r(\boldsymbol{x}, \boldsymbol{y}) \stackrel{def}{=} \textbf{GRS}_r(\boldsymbol{x}, \boldsymbol{y})^{\perp} \cap \mathbb{F}_q^n$$

Proposition 1.

$$\begin{aligned} \dim \textbf{Alt}_r(\boldsymbol{x}, \boldsymbol{y}) &\geqslant n - mr \\ d_{min}\textbf{Alt}_r(\boldsymbol{x}, \boldsymbol{y}) &\geqslant r + 1 \end{aligned}$$

Goppa codes

Definition 2. *Let* $\boldsymbol{x} \in \mathbb{F}_{q^m}^n$ *be a support and* $\Gamma \in \mathbb{F}_{q^m}[z]$ *such that* $\forall i, \Gamma(x_i) \neq 0$, *then the Goppa code* $\mathbf{Gop}(\boldsymbol{x}, \Gamma)$ *is defined by*

$$\mathbf{Gop}(\boldsymbol{x}, \Gamma) = \mathbf{Alt}_{\deg \Gamma}(\boldsymbol{x}, \boldsymbol{y}),$$

with $y_i = \frac{1}{\Gamma(x_i)}$.

Proposition 2. *Its parameters are given by*

$$\begin{aligned} \dim \mathbf{Gop}(\boldsymbol{x}, \Gamma) &\geqslant n - m \deg \Gamma \\ d_{min} \mathbf{Gop}(\boldsymbol{x}, \Gamma) &\geqslant \deg \Gamma + 1 \end{aligned}$$

Wild Goppa codes

Theorem 1. [Sugyiama et al. 1978] *Let* $\boldsymbol{x} \in \mathbb{F}_{q^m}^n$ *and* $\gamma \in \mathbb{F}_{q^m}[z]$ *squarefree, then*

$$\mathbf{Gop}(\boldsymbol{x}, \gamma^{q-1}) = \mathbf{Gop}(\boldsymbol{x}, \gamma^q)$$

Such a code is called a wild Goppa code. Parameters :

$$\begin{aligned} \dim \mathbf{Gop}(\boldsymbol{x}, \gamma^{q-1}) &\geqslant n - m(q-1)\deg\gamma \\ \mathrm{d_{min}}\mathbf{Gop}(\boldsymbol{x}, \gamma^{q-1}) &\geqslant q\deg\gamma + 1. \end{aligned}$$

$\approx$ twice the error correction capacity in the binary case!

Distinguishing alternant codes from random codes

We have

$$\begin{aligned}\textsf{\textbf{Alt}}_r(\boldsymbol{x},\boldsymbol{y}) &= \textsf{\textbf{GRS}}_r(\boldsymbol{x},\boldsymbol{y})^{\perp}\cap\mathbb{F}_q^n\\ &= \textsf{\textbf{GRS}}_{n-r}(\boldsymbol{x},\boldsymbol{y}')\cap\mathbb{F}_q^n\end{aligned}$$

and

$$\dim \textsf{\textbf{Alt}}_r(\boldsymbol{x},\boldsymbol{y}) \geqslant n - mr.$$

Fact 1. *To distinguish we need*

$$2(n-r) < n \quad\Longrightarrow\quad r > n/2,$$

however

$$m > 1 \quad\Longrightarrow n - mr < 0.$$

Distinguisher on the dual code

- 2011 Faugère-Gauthier-Otmani-Perret-Tillich : it is possible to distinguish alternant codes of high rate from random codes.
- 2012 Márquez Corbella-Pellikaan : equivalent description of the distinguisher in terms of the square of the dual of the alternant code.

Wild + $m = 2$

Theorem 2. [Couvreur , Otmani, Tillich 2013] *If* $m = 2$ *and* $\gamma \in \mathbb{F}_{q^2}[z]$ *an irreducible polynomial of degree* r

1. $\mathbf{Gop}(\boldsymbol{x}, \gamma^{q-1}) = \mathbf{Gop}(\boldsymbol{x}, \gamma^{q+1})$;

2. $\dim \mathbf{Gop}(\boldsymbol{x}, \gamma^{q}) \geqslant n - \underbrace{m}_{=2} r(q-1) + r(r-2)$

Distinguishing wild Goppa codes for $m = 2$

Theorem 3. [Couvreur, Otmani, Tillich 2014] *The square of the shortening of such a wild Goppa in a positions has an abnormal dimension when $a \in \{a^-, \ldots, a^+\}$ and*

$$\begin{aligned} a^- &= n - 2r(q+1) - 1 \\ a^+ &= \max\left\{ a \geqslant 0 \;\middle|\; \begin{array}{c} 3(n-a) - 4r(q+1) - 2 \leqslant \\ \min\left\{ n - a, \binom{n-a-2r(q-1)+r(r-2)}{2} \right\} \end{array} \right\} \end{aligned}$$

Figures

Table 1: Largest value of q for which we can distinguish $\mathbf{Gop}(\boldsymbol{x}, \gamma^{q-1})$ with γ irreducible of degree r.

r	2	3	4	5
q	9	19	37	64

Couvreur-Otmani-Tillich 2014 : filtration attack

Public key $\mathcal{C}$ is a wild Goppa code **Gop**$(\boldsymbol{x}, \gamma^{q-1})$, with $m = 2$.

Fact 2. *W.l.o.g. we may assume*

$$x_0 = 0 \quad \textit{et} \quad x_1 = 1.$$

Filtration attack, Step 1

By using the same technique as for GRS codes, we compute the filtration

$$C_0 = \mathcal{C} \subseteq C_1 \subseteq \cdots \subseteq C_{q+1}$$

associated to

$$\mathbb{F}_{q^2}[z]_{<s} \supseteq z\mathbb{F}_{q^2}[z]_{<s-1} \supseteq \cdots \supseteq z^{q+1}\mathbb{F}_{q^2}[z]_{<s-(q+1)}$$

where $s = n - r(q+1)$.

$$C_0 \star C_t \subseteq C_{\lfloor t/2 \rfloor} \star C_{\lceil t/2 \rceil}$$

Step 2

Lemma 1.

$$\boldsymbol{x}^{\star(-(q+1))} \star \mathcal{C}_{q+1} \subseteq \mathcal{C}.$$

Sketch of proof :

Let $\boldsymbol{c} \in \mathcal{C}_{q+1}$ and $p_{\boldsymbol{c}}$ be the corresponding polynomial $p_{\boldsymbol{c}}$ is of the form

$$p_{\boldsymbol{c}}(z) = z^{q+1} f(z), \qquad \deg q_{\boldsymbol{c}} \leqslant s - (q+1).$$

For all $x \in \mathbb{F}_{q^2}$, $x^{q+1} \in \mathbb{F}_q$ (this is $\mathrm{N}_{\mathbb{F}_{q^2}/Fq}(x)$).

If $x_i^{q+1} q(x_i) \in \mathbb{F}_q$ for all i, then $q(x_i) \in \mathbb{F}_q$ and therefore to q corresponds the codeword $\boldsymbol{x}^{\star-(q+1)} \star \boldsymbol{c} \in \mathcal{C}$

Sketch of the whole attack

- **Step 1.** Compute

$$\mathcal{C} = \mathcal{C}_0 \supseteq \mathcal{C}_1 \supseteq \mathcal{C}_2 \supseteq \cdots \supseteq \mathcal{C}_{q+1}$$

- **Step 2.** From $\mathcal{C}_{q+1}$, one can compute $\boldsymbol{x}^{\star(q+1)} = (x_0^{q+1}, x_1^{q+1}, \ldots, x_{n-1}^{q+1})$. (It uses the norm over $\mathbb{F}_{q^2}$.)

  Reapplying Step 1 and 2, one can also compute: $(\boldsymbol{x} - \mathbf{1})^{\star(q+1)} = ((x_0 - 1)^{q+1}, (x_1 - 1)^{q+1}, \ldots, (x_{n-1} - 1)^{q+1})$

  Step 3. Deduce from $\boldsymbol{x}^{\star(q+1)}$ and $(\boldsymbol{x} - \mathbf{1})^{\star(q+1)}$ the support $\boldsymbol{x}$ up to Galois action.

- **Step 4.** A bit more technique to deduce $\boldsymbol{x}$ and the Goppa Polynomial γ.

Complexity and running time

Complexity : $O(n^4\sqrt{n} + n^4(q^2 - n))$ (recall that $n \leqslant q^2$).

Table 2: Running times with an Intel® Xeon 2.27GHz

$[q, n, k, r]$	[29,781, 516,5] h	[29, 791, 575, 4] h	[29,794,529,5] h
Average time	16min	19.5min	15.5min

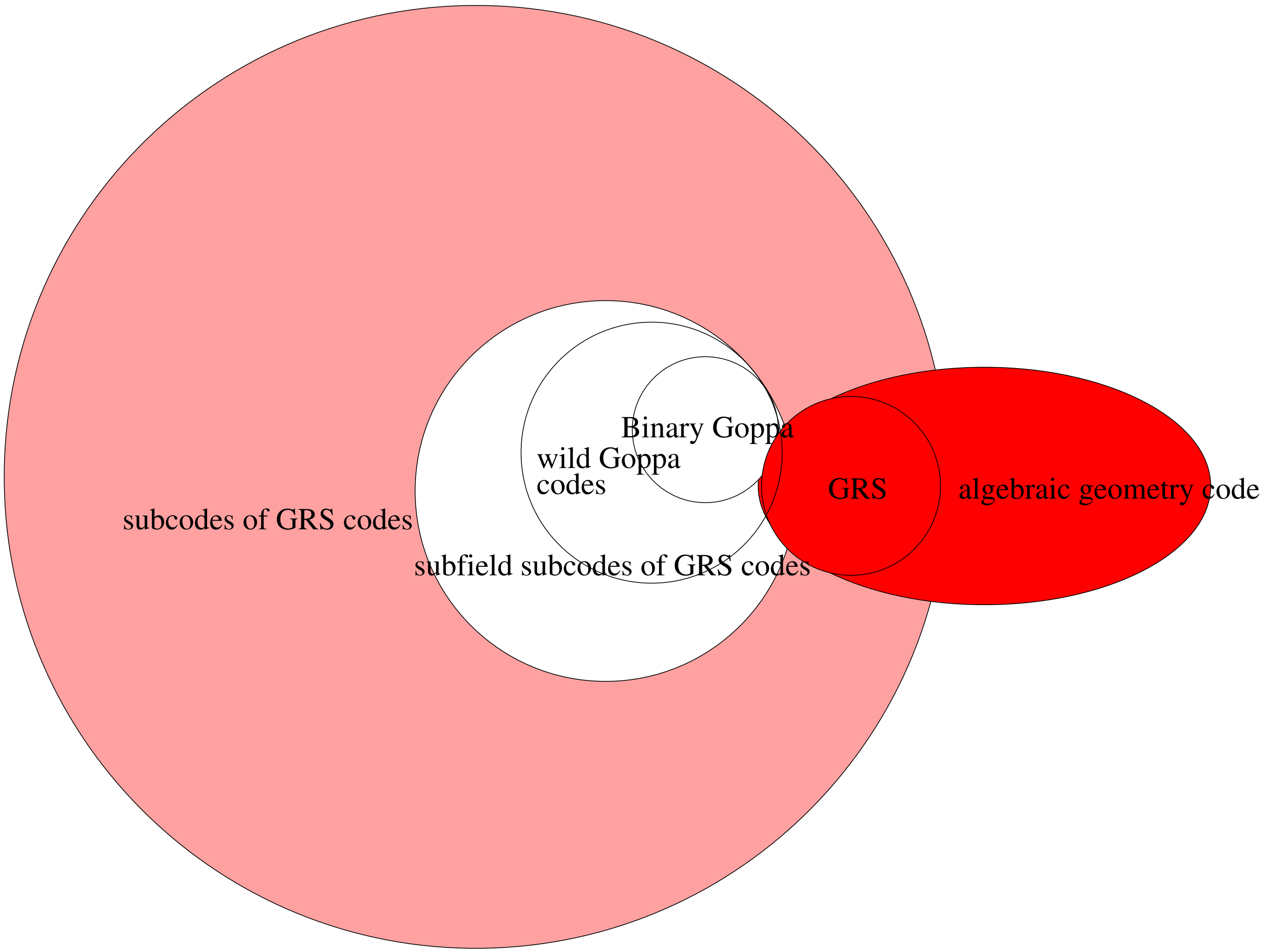
(q, n, k, r)	[31, 795, 563, 4] h	[31,813, 581,4] h	[31, 851, 619, 4] h
Average time	31.5min	31.5min	27.2min

(q, n, k, r)	[32,841,601,4] h	[31, 900, 228, 14]
Average time	49.5min	24min

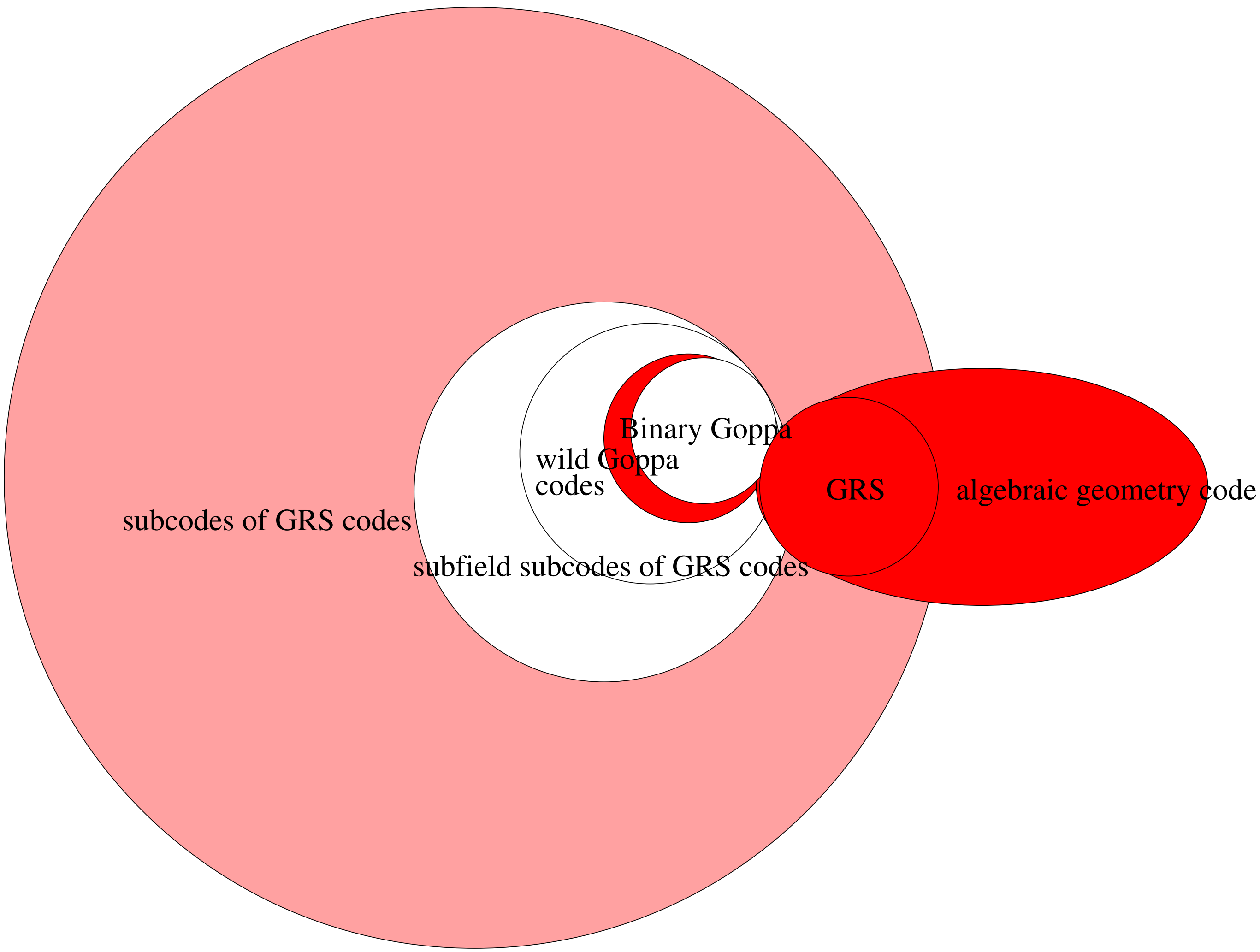
Proposed parameters (Bernstein, Lange, Peters 2010)

Never proposed parameters (More than 2^{130} possible choices for γ and security > 125 bits with respect to ISD)

The old picture



The new picture



Conclusion

- Goppa codes are not necessarily immune to square code attacks.
- Distinguisher $\Rightarrow$ attack.
- Question : are other distingushable codes breakable? For instance high rate Goppa codes (distinguisher on the dual).
- Polynomial time attacks on Reed-Muller codes ?
- Polynomial time attacks on subcodes of algebraic geometry codes?
- other families of codes (MDPC,. . .)?