Low Rank Parity Check codes and their application to cryptography

Philippe Gaborit[1] Gaetan Murat[1] Olivier Ruatta[1] Gilles Zémor[2]

[1]Université de Limoges, [2]Université de Bordeaux

WCC 2013

Summary

Coding and cryptography

Cryptography needs different difficult problems

- factorization
- discrete log
- SVP for lattices
- syndrome decoding problem

For code-based cryptography, the security of cryptosystems is usually related to the problem of syndrome decoding.

Syndrome decoding problem

Syndrome decoding

For a given syndrome s, find x of small Hamming weight such that $Hx^t = y$ *with H a random matrix.*

Problem studied for many years with a well known complexity.
Caracteristics :

- NP-hard
- usually fast
- A priori resisting to quantum computer

Best known attacks : Information Set Decoding and variations : FS '09, MMT '11, BJMT '12
Complexity of attacks seem converging to a certain stabilization.

Codes and cryptography

Finding alternative to RSA and NT based system : major issue

- **McEliece cryptosystem** : usually lead to very large public keys (a few hundred thousand) - too large for general utilization
- **lattices :** very close to codes but with a different metric : NTRU '95 : double circulant structure : first non number theory based system with small keys, the LWE, Ring-LWE, ...
- **codes** : different systems proposed in the '00's : structure + structure $\rightarrow$ too much structure for attack (30.000b)
- **recently 2012** : MDPC codes for crypto, NTRU like system , with random small weight double-circulant codes 4800b.
- **rank metric** ??

Rank metric codes

The rank metric is defined in finite extensions.

- $\mathbb{F}_q$ a finite field with q a power of a prime.
- $\mathbb{F}_{q^m}$ an extension of degree m of $\mathbb{F}_q$.
- $B = (b_1, ..., b_m)$ a basis of F_{q^m} over F_q.

$\mathbb{F}_{q^m}$ can be seen as a vector space on $\mathbb{F}_q$.

- $\mathcal{C}$ a linear code over $\mathbb{F}_{q^m}$ of dimension k and length n.
- G a $k \times n$ generator matrix of the code $\mathcal{C}$.
- H a $n \times (n-k)$ parity check matrix of $\mathcal{C}$, $GH = 0$.

Rank metric

Words of the code $\mathcal{C}$ are n-uplets with coordinates in $\mathbb{F}_{q^m}$.

$$v = (v_1, \ldots, v_n)$$

with $v_i \in \mathbb{F}_{q^m}$.
Any coordinate $v_i = \sum_{j=1}^{m} v_{ij} b_j$ with $v_{ij} \in \mathbb{F}_q$.

Métrique rang

v has rank r iff the rank of $V = (v_{ij})_{ij}$ is r.

Rank syndrome decoding problem (RSD)

Syndrome decoding

Let H be a $((n-k)\times n)$ matrix over F_{q^m} with $k \leq n$, $s \in F_{q^m}^k$ and r an integer. The problem is to fin d x such that $\mathrm{rank}(x) = r$ and $Hx^t = s$

- induces short public keys
- not proven NP-hard (does not mean it is not !)

Support analogy

The **support of a word in Hamming metric** $x(x_1, x2, \cdots, x_n)$ is the set of positions $x_i \neq 0$

- how to recover a small Hamming word associated to a given syndrome ?

1) find the support of the word (guess ! !)
2) solve a system to recover the coordinates values

• **Support of a word in rank metric**

The support of a word $x(x_1, x2, \cdots, x_n)$ of rank r is a space E of dim r such that $\forall x_i, x_i \in E$.

- how to recover a word associated to a given syndrome ?

1) find the support (guess it !)

2) solve a system from the syndrome equations to recover the $x_i \in E$

remark : for Hamming Newton binomial, for rank distance Gaussian binomial ! $\rightarrow$ complexity grows faster.

$\Rightarrow$ rank metric induces smaller parameters for a given complexity

Best known attacks

A.Ourivski et T.Johannson '02 :

- basis enumeration : $\leq (k+r)^3 q^{(r-1)(m-r)+2}$ (improvement on the polynomial part of Chabaud-Stern '96)
- coordinate enumeration : $\leq (k+r)^3 r^3 q^{(r-1)(k+1)}$

More recently (2012) Gaborit, Schrek, Ruatta :

- $(m(n-k))^3 q^{\frac{(k+1)(r-1)m}{n}}$, generalization of ISD for rank metrix
- $O(r^3 k^3 q^{r\lceil \frac{(r+1)(k+1)-(n+1)}{r} \rceil})$ with algebraic attacks

Rank metric and cryptography

Classical setting for code based crypto : the MacELiece scheme

• Gabidulin codes are the analogous of Reed-Solomon codes
→ possible to design a system based on Gabidulin codes : GPT cryptosystem '91
problem : as Reed-Solomon codes : the Gabidulin codes are difficult to hide

→ attacks (OJ '02,..) and new constructions (FL '05) and new attacks (Overbeke '06,..)
probably possible to eventually find a resistant construction but doubt on structural attacks

Decoding in rank metric

- Gabidulin $[n, k]$ codes over F_{q^n} decode up to $r = (n - k)/2$
- simple construction possible to decode random errors up to GVR, but slow decoding and difficult to hide

Are there alternatives ?

What does exist in Hamming distance ?

- Reed-Solomon codes and derivatives (BCH, Goppa, ...)
- LDPC codes : dual matrix with low weight

LRPC codes

LDPC : dual with low weight (ie : small support)
$\rightarrow$ equivalent for rank metric : dual with small rank support

Definition

A Low Rank Parity Check (LRPC) code of rank d, length n and dimension k over F_{q^m} is a code such that the code has for parity check matrix, a $(n-k) \times n$ matrix $H(h_{ij})$ such that the sub-vector space of F_{q^m} generated by its coefficients h_{ij} has dimension at most d. We call this dimension the weight of H.

In other terms : all coefficients h_{ij} of H belong to the same 'low' vector space $F < F_1, F_2, \cdots, F_d >$ of F_{q^m} of dimension d.

Decoding LRPC codes

Idea : as usual recover the support and then deduce the coordinates values.

Let $e(e_1, ..., e_n)$ be an error vector of weight r, ie : $\forall e_i : e_i \in E$, and dim(E)=r. Suppose $H.e^t = s = (s_1, ..., s_{n-k})^t$.

$$e_i \in E < E_1, ..., E_r >, h_{ij} \in F < F_1, F_2, \cdots, F_d >$$

$$\Rightarrow s_k \in < E_1 F_1, .., E_r F_d >$$

$\Rightarrow$ if $n - k$ is large enough, it is possible to recover the product space $< E_1 F_1, .., E_r F_d >$

Decoding LRPC codes

Syndrome $s(s_1, .., s_{n-k})$ **:** $S =< s_1, .., s_{n-k} >\subset< E_1 F_1, .., E_r F_d >$

Suppose $S =< E.F > \Rightarrow$ possible to recover E.

Let $S_i = F_i^{-1}.S$, since

$$S =< E.F >=< F_i E_1, F_i E_2, .., F_i E_r, ... >\rightarrow E \subset S_i$$

$$\mathbf{E = S_1 \cap S_2 \cap \cdots \cap S_d}$$

General decoding of LRPC codes

Let $y = xG + e$

1. **Syndrome space computation**
   Compute the syndrome vector $H.y^t = s(s_1, \cdots, s_{n-k})$ and the syndrome space $S =< s_1, \cdots, s_{n-k} >$.
2. **Recovering the support E of the error**
   $S_i = F_i^{-1}S,\ E = S_1 \cap S_2 \cap \cdots \cap S_d,$
3. **Recovering the error vector e** Write $e_i (1 \leq i \leq n)$ in the error support as $e_i = \sum_{i=1}^{n} e_{ij} E_j$, solve the system $H.e^t = s$.
4. **Recovering the message x**
   Recover x from the system $xG = y - e$.

Decoding of LRPC

- **Conditions of success**

- $S =< F.E >\Rightarrow$ rd $\leq$ n-k.
- possibility that $dim(S) \neq n-k \Rightarrow$ probabilistic decoding with error failure in $q^{-(n-k-rd)}$
- if $d = 2$ can decode up to $(n-k)/2$ errors.

- **Complexity of decoding** : very fast symbolic matrix inversion $O(m(n-k)^2)$

- **Comparison with Gabidulin codes :** probabilistic, decoding failure, but as fast.

Application to cryptography

- a new family of decodable codes with a low structure, no use of isometry

- McEliece setting :

Public key : G LRPC code : $[n, k]$ of weight d which can decode up to errors of weight r

Public key : $G' = MG$

Secret key : M

- **Encryption**

c= mG' +e , e of rank r

- **Decryption**

Decode $H.c^t$ in e, then recover m.

- Smaller size of key : double circulant LRPC codes : H=(I A), A circulant matrix

Application to cryptography

- **Attacks on the system**

- message attack : decode a word of weight r for a $[n, k]$ random code

- structural attack : recover the LRPC structure
$\rightarrow$ a $[n, n-k]$ LRPC matrix of weight d contains a word with $\frac{n}{d}$ first zero positions. Searching for a word of weight d in a $[n - \frac{n}{d}, n - k - \frac{n}{d}]$ code.

- **Attack on the double circulant structure**

as for lattices or codes (with Hamming distance) no specific more efficient attack exists exponentially better than decoding random codes.

Parameters

n	k	m	q	d	r	failure	public key	security
74	37	41	2	4	4	-22	1517	80
94	47	47	2	5	5	-23	2397	120
68	34	23	2^4	4	4	-80	3128	100

Conclusion

- LRPC : new family of rank codes with an efficient probabilistic decoding algorithm
- Application to cryptography in the spirit of NTRU and MDPC
- Very small size of keys, comparable to RSA
- More studies need to be done but very good potentiality