Fourteenth International Workshop on Algebraic and Combinatorial Coding Theory
September 7–13, 2014, Svetlogorsk (Kaliningrad region), Russia *pp. 234–239*

On cellular code and their cryptographic applications

Pierre Loidreau Pierre.Loidreau@m4x.org
DGA MI and IRMAR, Université de Rennes 1

Abstract. We present a family of codes generalizing the notion of quasi-cyclic codes. Generator matrices are formed with cells which are square matrices from a given commutative matrix algebra generated by a polynomial. We show that any divisor of the polynomial gives rise to a *projected* code with smaller parameters. By studying how this framework affects quasi-cyclic codes based cryptography, we show that it should be taken into account when designing such cryptosystems.

1 Introduction

Quasi-cyclic codes based cryptography has become a hot subject recently especially since the use of LRPC codes for rank metric and MDPC codes for Hamming metric [3, 4]. The only algebraic structure required is thus quasi-cyclicity, which is employed to reduce drastically the size of the public-key.

In this paper we elaborate a framework generalising the notion of quasi-cyclic codes which is a framework well adapted to the study of the security of quasi-cyclic codes based cryptography. In a first part we generalize the notion of quasi-cyclic codes to what we denote cellular codes because of the form of the generator matrix whose fundamental element is not an element of a finite field, but rather an element of a commutative matrix algebra. In a second part we show that the families of cellular codes are stable via some kind of projection, giving codes with smaller parameters. In a final part we show how these projections can be employed to reduce complexities of attacks on quasi-cyclic codes based cryptography in Hamming metric as well as in rank metric.

2 Cellular codes

We follow the approaches of [1, 2].

2.1 Construction

Let $g(x)$ be a polynomial of degree m with coefficients in a Galois field $K = GF(q)$. Let us define the quotient

$$\mathcal{R}_g \stackrel{def}{=} K[x]/g \cdot K[x] \tag{2.1}$$

The mapping $\psi_g :\ \mathcal{R}_g \to K^m$ defined by

$$a(x) = \sum_{i=0}^{m-1} a_i x^i \ \longmapsto\ \psi_g(a(x)) = (a_0, \ldots, a_{m-1}) \tag{2.2}$$

is a K-linear isomorphism. We consider the mapping $\Phi_g :\ \mathcal{R}_g \to K^{m\times m}$ from $\mathcal{R}_g$ considered as a K-algebra into the ring of $m \times m$ matrices over K defined by

$$a(x) = \sum_{i=0}^{m-1} a_i x^i \ \longmapsto\ \Phi_g(a) = \begin{pmatrix} \psi_g(a(x)) \\ \psi_g(xa(x)) \\ \vdots \\ \psi_g(x^{m-1}a(x)) \end{pmatrix} \tag{2.3}$$

From its construction it is quite clear that Φ_g is a one-to-one morphism of K-algebras. Hence $\Phi_g(\mathcal{R}_g)$ is a commutative subalgebra of $K^{m\times m}$. In the case where $g(x) = x^m - 1$, $\Phi_g(\mathcal{R}_g)$ is exactly the m-dimensional commutative algebra of circulant matrices.

In the rest of the paper whenever there is no ambiguity on the polynomial g we denote by $\psi_g(a(x)) = \mathbf{a}$ and $\Phi_g(a(x)) = \mathbf{A}$. We have

Proposition 1.
For all $a(x),\ b(x) \in \mathcal{R}_g$, $\psi_g(a(x)b(x)) = \mathbf{aB} = \mathbf{bA}$.

Now consider an integer $\ell \geq 1$. The isomorphism ψ_g can be canonically extended to an isomorphism $\psi_g^{(\ell)} :\ \mathcal{R}_g^\ell \to K^{m\ell}$

$$\mathbf{y}(x) = (y_0(x), \ldots, y_{s-1}(x)) \in \mathcal{R}_g^\ell \longmapsto \psi_g^{(\ell)} = (\mathbf{y}_0, \ldots, \mathbf{y}_{s-1}) \in K^{m\ell}$$

Definition 1. *Let $\mathcal{M}$ be a submodule of $\mathcal{R}_g$ of rank s. Then the set $\mathcal{C} = \psi_g^{(\ell)}(\mathcal{M})$ is a K-linear code of length $m \times \ell$. It is called a* g-cellular code *of index ℓ over K associated to the module $\mathcal{M}$.*

The question is now how to obtain generator matrices of such codes. Let $\mathcal{M}$ be a $\mathcal{R}_g$-module of rank s with generator matrix

$$\mathbf{G}_{\mathcal{M}} = \begin{pmatrix} a_{1,1}(x) & \cdots & a_{1,\ell}(x) \\ \vdots & \ddots & \vdots \\ a_{s,1}(x) & \cdots & a_{s,\ell}(x) \end{pmatrix}$$

Proposition 2. *A generator matrix for $\mathcal{C} = \psi_g^{(\ell)}(\mathcal{M})$ is given by :*

$$\mathbf{G} = \begin{pmatrix} \mathbf{A}_{1,1} & \cdots & \mathbf{A}_{1,\ell} \\ \vdots & \ddots & \vdots \\ \mathbf{A}_{s,1} & \cdots & \mathbf{A}_{s,\ell} \end{pmatrix}$$

That is $\psi_g^{(\ell)}(\mathcal{M}) =< \Phi_g(\mathbf{G}_\mathcal{M}) >$. *Therefore the dimension* k *of* $\mathcal{C}$ *satisfies* $k \leq ms$.

¿From this construction it is immediate that quasi-cyclic codes of order m and index ℓ are $(x^m - 1)$-*cellular codes* of index ℓ. Contrarily to quasi-cyclic codes, in general *cellular codes* have a trivial automorphism group.

2.2 Projected codes

We show how to obtain codes with smaller parameters from *cellular codes*. Let g be a polynomial of degree m over a finite field K. Let $\mathcal{M}$ be an $\mathcal{R}_g$-submodule of $\mathcal{R}_g^\ell$ of rank s generated by $\mathbf{G}_\mathcal{M} = (a_{i,j}(x))$ and $\mathcal{C}$ be the associated *g-cellular code*.

Let f be a divisor of g, *i.e.* $g(x) = h(x)f(x)$ in some extension of K, say L (in Hamming metric applications K is generally $GF(2)$, and L is some extension $GF(2^u)$). The K-algebra $\mathcal{R}_g$ can be canonically lifted to an L-algebra, and the application

$$\begin{array}{rccl} \Pi & : & \mathcal{R}_g & \longrightarrow & \mathcal{R}_f = L[x]/f \cdot L[x], \\ & & a(x) & \mapsto & a(x) \bmod f(x) \end{array}$$

is a well defined surjective morphism of L-algebras. As usual we can canonically extend Π into a morphism of vector spaces: $\Pi^{(\ell)} : \mathcal{R}_g^\ell \longrightarrow \mathcal{R}_f^\ell$.

The process to obtain the *f-cellular code* which is the projected code by f of the *g-cellular code* is:

Theorem 1. *Let* $\mathcal{C}$ *be a* g-cellular code *over* K *of index* ℓ, *let* $\mathbf{G}_\mathcal{M} = (a_{i,j}(x))$ *be a generator matrix of the associated* $\mathcal{R}_g$*-module. Let* f *be a divisor of* g *with coefficients in* $L \hookleftarrow K$. *Then the set*

$$\mathcal{C}' = \psi_f^{(\ell)} \circ \Pi^{(\ell)} \circ (\psi_g^{-1})^{(\ell)}(\mathcal{C})$$

is an f-cellular code *over* L *of index* ℓ, *and a generator matrix of the associated* $\mathcal{R}_f$*-module* $\mathcal{M}'$ *is given by* $\mathbf{G}_{\mathcal{M}'} = (a'_{i,j}(x))$, *where* $a'_{i,j}(x) = a_{i,j}(x) \bmod f$.

A direct consequence is:

Corollary 1. *Let* δ *be the degree of the polynomial* f. *Then* $\mathcal{C}'$ *is a* $[n' = \delta\ell, k']$ *code over* L, *where* $k' \leq \delta s$.

2.3 Decoding

Let $g(x)$ be a polynomial of degree m in K. Consider a *g-cellular code* $\mathcal{C}$ over K of index ℓ. Suppose that one receives a vector $\mathbf{y} \in K^{m\ell}$ such that

$$\mathbf{y} = \underbrace{\mathbf{c}}_{\in \mathcal{C}} + \mathbf{e}$$

where $\mathbf{e} = (\mathbf{e}_1, \ldots, \mathbf{e}_\ell)$ is an error vector of weight t in some metric (could be Hamming or rank or any other metric). Let f be a divisor of g of degree δ in some extension field $L \hookleftarrow K$.

Since the mapping $\psi_f^{(\ell)} \circ \Pi^{(\ell)} \circ (\psi_g^{-1})^{(\ell)}$ is L-linear from $K^{m\ell}$ into $L^{\delta\ell}$, we by applying this mapping $\mathbf{y}$ and obtain:

$$\mathbf{y}' = \underbrace{\mathbf{c}'}_{\in \mathcal{C}'} + \mathbf{e}',$$

where $\mathcal{C}'$ has length n and dimension $\leq \delta s$.

We now have to decode in a code with smaller parameter. The effect of $\psi_f^{(\ell)} \circ \Pi^{(\ell)} \circ (\psi_g^{-1})^{(\ell)}$ on the error-vector does not usually preserves the metric. Namely, the projected error-vector is

$$\mathbf{e}' = (\mathbf{e}'_1, \ldots, \mathbf{e}'_\ell),$$

where for all i, $e'_i(x) = e_i(x) \bmod f$.

2.4 Applications to Hamming metric and rank metric decoding of quasi-cyclic codes

Let $\mathcal{C}$ be a quasi-cyclic code of order m and of index ℓ over $K = GF(q)$. Let $f(x)$ be a divisor of $g(x) = x^m - 1$ is some finite field $L = GF(q^u)$.

We address the decoding problem in rank or Hamming metric:

$$\mathbf{y} = \underbrace{\mathbf{c}}_{\in \mathcal{C}} + \mathbf{e}$$

Hamming metric Let the Hamming weight of $\mathbf{e} = (\mathbf{e}_1, \ldots, \mathbf{e}_\ell)$ be equal to t. Let δ be the degree of f, let $wt(\mathbf{e}_i) \stackrel{def}{=} u_i$, and let $e_i(x) = e_i^{(1)}(x) + x^\delta e_i^{(2)}(x)$, $e_i^{(1)}(x)$ and $e_i^{(2)}(x)$ being of respective weight $u_i^{(1)}$ and $u_i^{(2)}$. By projecting on the *f-cellular code* $\mathcal{C}'$, we obtain the following:

$$\mathbf{y}' = \underbrace{\mathbf{c}'}_{\in \mathcal{C}'} + \mathbf{e}',$$

where $wt(\mathbf{e}'_i) = 0$ if there is no error on $\mathbf{e}_i$ and generally $wt(\mathbf{e}'_i) \leq u_i^{(1)} + (wt(f) - 1)u_i^{(2)} = u_i + (wt(f) - 2)u_i^{(2)}$, where $wt(f)$ is the weight of f. Therefore

$$wt(\mathbf{e}') \leq t + (wt(f) - 2) \sum_{i=0}^{\ell} u_i^{(2)}.$$

In the case where $wt(f) = 2$ ($\mathcal{C}'$ is cyclic, negacyclic or constacyclic), the weight of $\mathbf{e}'$ is at most the weight of $\mathbf{e}$.

Therefore, if the weight of $\mathbf{e}'$ is in the range of unique decoding in $\mathcal{C}'$ (less than GV for instance), we recover information on $\mathbf{e}$.

Rank metric Let the rank of $\mathbf{e}$ over some subfield of $K = GF(q)$ be equal to t. By projecting on the *f-cellular code* $\mathcal{C}'$ we obtain:

$$\mathbf{y}' = \underbrace{\mathbf{c}'}_{\in \mathcal{C}'} + \mathbf{e}',$$

Since L is an extension of order u of K, we have

Proposition 3. *$Rk(\mathbf{e}') \leq u \times Rk(\mathbf{e})$.*

If the rank weight of $\mathbf{e}'$ is in the range where it can be recovered for rank metric (less than GV for instance) then we apply general decoding algorithms or algorithms for finding low rank codewords and recover information on $\mathbf{e}'$, that is some elements of the subsbace containing $\mathbf{e}$.

This technique can as well be applied to the search for small weight codewords in a given code.

3 Results for cryptographic applications

If we assume that there is unique decoding for a *random code* up to GV bound we obtain the results that follow.

3.1 MDPC based McEliece cryptosystem

Originally [3] proposed parameters were the code $\mathcal{C}$ has parameters $n = 9600$, $k = 4800$. The Hamming weight of the errors is $t = 84$, and a parity-check matrix has vectors of weight $d = 90$. The work factors with these parameters and using Lee-Brickell algorithm are:

- *Key recovery:* 2^{104} binary operations.
- *Decoding attack:* 2^{104} binary operations.

For rate 1/2 codes we have $d_{gv} \approx 0.11n = 1054 >> 84$. The polynomial $x^{4800} - 1$, has many factors over $GF(2)$. Suppose we wish to decode or find words of weight t. We choose s divisor of 4800, such that $t \approx 0.11n/s$. Here we choose $s = 10$, corresponding to the divisor $x^{480} - 1$. The obtained projected code is still a quasi-cyclic code of length 960 and dimension 480.

The work-factors with the projected parameters and using Lee-Brickell algorithm give:

- *Key recovery:* 2^{103} binary operations.
- *Decoding:* $2^{101.5}$ binary operations.

Though we do not address this issue here, we can take into account that in that case, bits can compensate and the average weight that we have to decode can be much less than 84 or 90.

3.2 LRPC cryptosystem

According to [5] complexities rank metric decoding algorithms for a vector of rank t in a code of length n, dimension k over $GF(q = p^v)$ with base field $GF(p)$ are the following:

$C_1 = t^3k^3p^{\lceil \frac{(t+1)(k+1)-(n+1)}{t} \rceil}$	$C_2 = (n-k)^3v^3 \min(p^{t\lfloor \frac{kv}{n} \rfloor}, p^{(t-1)\lfloor \frac{(k+1)v}{n} \rfloor})$
$C_3 = (k+t)^3t^3p^{(k+1)(t-1)}$	$C_4 = (k+t)^3p^{(t-1)(v-t)+2}$

We study one set of parameters proposed in [4]. In that case, the code $\mathcal{C}$ has parameters $n = 94, k = m = 47,\ v = 47,\ p = 2, r = 5$. The polynomial $g(x) = x^{47} - 1$ can be factorized in $GF(2)$ under the form $g(x) = (x-1)p_1(x)p_2(x)$, where p_1 and p_2 are polynomials of degree 23.

Therefore the decoding can be done in a code $\mathcal{C}'$ with parameters $m = 47,\ n = 46,\ k = 23$. The GV distance for this parameters is $d_{GV} = 13$, ensuring unique decoding. The comparison of complexity results are:

Decod. Complex.	$Log_2(C_1)$	$Log_2(C_2)$	$Log_2(C_3)$	$Log_2(C_4)$
$\mathcal{C}$	129	218	216	187
$\mathcal{C}'$	126	120	117	184

The complexity gain is thus 2^{12} between the two best algorihtms of that list.

References

[1] S. Ling and P. Solé, On the algebraic structure of quasi-cyclic codes I: Finite fields, *IEEE Trans. Inf. Theo.*, 47(7), pages 2751-2759.

[2] K. Lally and P. Fitzpatrick, Algebraic structure of quasi-cyclic codes, *Discr. Appl. Math.*, 111, pages 157-175.

[3] R. Misoczki, J.-P. Tillich, N.Sendrier and P. Barreto, MDPC McEliece: New McEliece variants from moderate density parity-check codes, in *Proc. ISIT 2013*. Successive versions available at http://eprint.iacr.org/2012/409.

[4] P. Gaborit, G. Murat, O. Ruatta and G. Zémor, Low Rank Parity-check codes and their application to cryptography. in *Proc. WCC 2013*.

[5] P. Gaborit, O. Ruatta and J. Schrek, On the complexity of rank syndrome decoding problem, http://arxiv.org/abs/1301.1026.