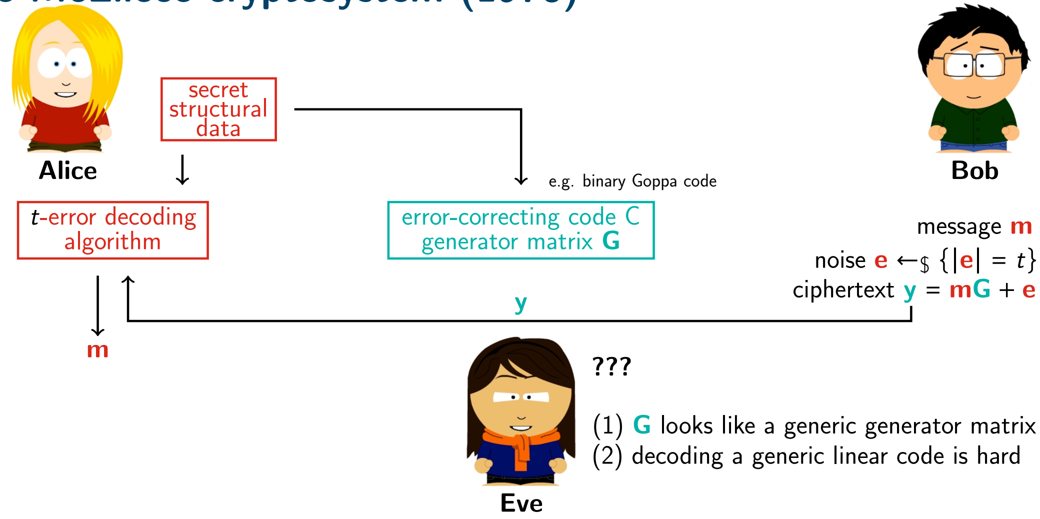
The syzygy distinguisher

Hugues Randriam
ANSSI, Laboratoire de cryptographie
& Télécom Paris, C^2

Eurocrypt 2025, Madrid
2025-05-05

The McEliece cryptosystem (1978)



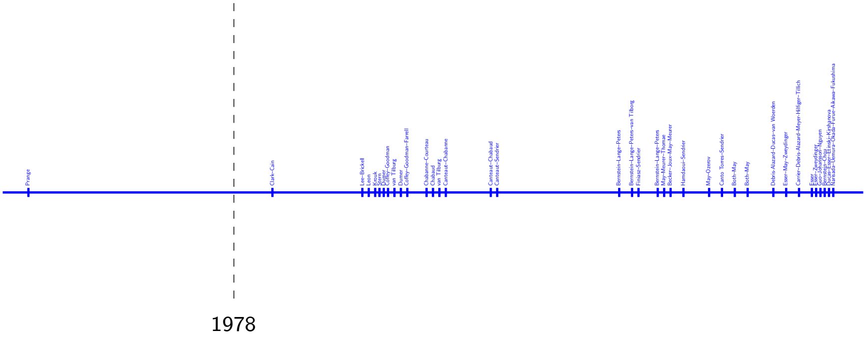
Note:
(1) ad hoc problem, trapdoor similar to those in today's multivariate cryptography
(2) well-studied problem, NP-hard, believed to be quantum-resistant

Stability of McEliece cryptanalysis

Asymptotic complexity for rate R, length $n \to \infty$ codes: $(C + o(1))^{\frac{n}{\log n}}$

Blue: information set decoding — improving C would be a major result!

$C = \frac{1}{(1-R)^{1-R}}$

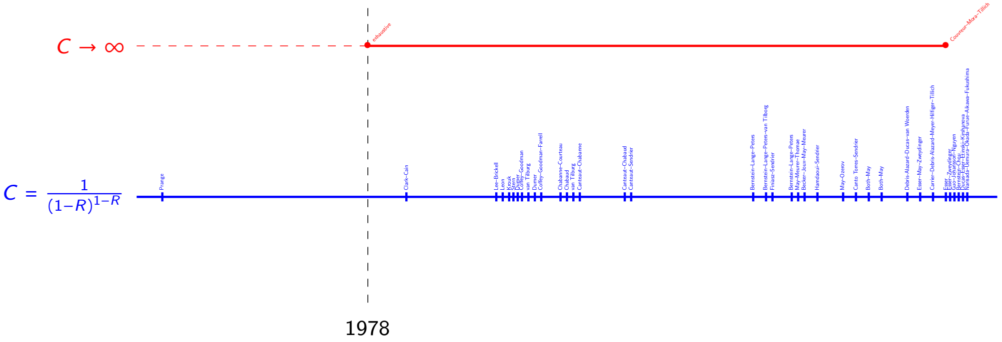


credit: Classic McEliece team

Stability of McEliece cryptanalysis

Asymptotic complexity for rate R, length $n \to \infty$ codes: $(C + o(1))^{\frac{n}{\log n}}$

Blue: information set decoding — improving C would be a major result!
Red: Goppa structure recovery/distinguisher
(unmentioned results only work for extreme regimes or other types or codes, or need additional information)



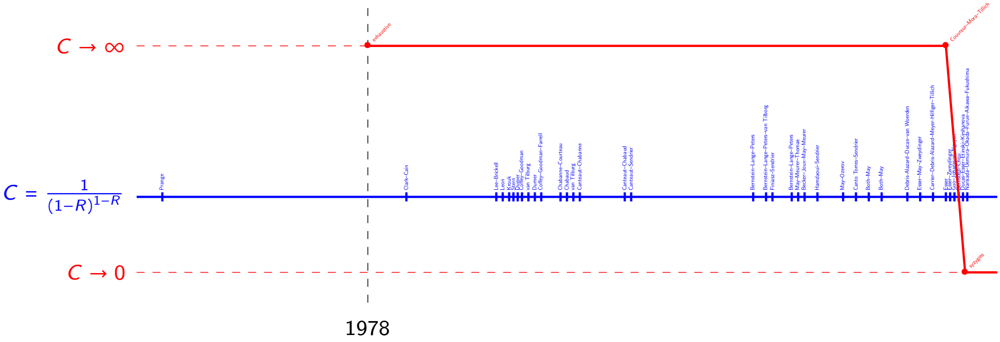
Stability of McEliece cryptanalysis

Asymptotic complexity for rate R, length $n \to \infty$ codes: $(C + o(1))^{\frac{n}{\log n}}$

Blue: information set decoding — improving C would be a major result!

Red: Goppa structure recovery/distinguisher

(unmentioned results only work for extreme regimes or other types or codes, or need additional information)



continuous incremental improvements vs. sudden leaps, potentially devastating

(Dual) Goppa structure

- $\mathbf{x} = (x_1, \ldots, x_n) \in (\mathbb{F}_{2^m})^n$ and $g(X) \in \mathbb{F}_{2^m}[X]$ irreducible of degree t
- construct the generalized Vandermonde matrix

$$\mathbf{H}_{\text{priv}} = \begin{pmatrix} 1/g(x_1) & 1/g(x_2) & \ldots & 1/g(x_n) \\ x_1/g(x_1) & x_2/g(x_2) & \ldots & x_n/g(x_n) \\ \vdots & \vdots & & \vdots \\ x_1^{t-1}/g(x_1) & x_2^{t-1}/g(x_2) & \ldots & x_n^{t-1}/g(x_n) \end{pmatrix} \in (\mathbb{F}_{2^m})^{t \times n}$$

- identify $\mathbb{F}_{2^m} \simeq (\mathbb{F}_2)^m$ (columns), put in reduced row echelon form, get

$$\mathbf{H}_{\text{pub}} \in (\mathbb{F}_2)^{mt \times n}$$

- Goppa structure recovery: find some $(\mathbf{x}, g)$ that give $\mathbf{H}_{\text{pub}}$
- Goppa distinguisher: decide if a given $\mathbf{H}$ comes from some $(\mathbf{x}, g)$

Quadratic relations

- $C \subseteq \mathbb{F}^n$ with basis $\mathbf{c}_1, \ldots, \mathbf{c}_k$
- evaluation map: $\mathrm{ev} : \mathbb{F}[X_1, \ldots, X_k] \to \mathbb{F}^n$, $X_i \mapsto \mathbf{c}_i$
- space of quadratic relations: $I_2(C) = \ker(\mathrm{ev}_2 : \mathbb{F}[X_1, \ldots, X_k]_2 \to \mathbb{F}^n)$
- $\dim(I_2(C)) = \frac{k(k+1)}{2} - \mathrm{rk}(\mathrm{ev}_2) \geq \left(\frac{k(k+1)}{2} - n\right)^+$

Example

- $C = \mathrm{rowspan}_{\mathbb{F}_{2^m}}(\mathbf{H}_{\mathrm{priv}}) = \mathrm{GRS}_t(\mathbf{x}, \mathbf{y})$ where $\mathbf{y} = g(\mathbf{x})^{-1}$
- basis $\mathbf{c}_i = \mathbf{y}\mathbf{x}^{i-1}$ for $1 \leq i \leq t$

then for $a + b = c + d$:

$$\mathbf{c}_a \mathbf{c}_b = \mathbf{c}_c \mathbf{c}_d$$

$$X_a X_b - X_c X_d \in I_2(C)$$

The [FGOPT10] distinguisher

Theorem

There is an explicit lower bound

$$\dim_{\mathbb{F}_2}(I_2(\mathrm{C})) \geq T = T(m, n, t)$$

when $\mathrm{C} = \mathrm{rowspan}_{\mathbb{F}_2}(\mathbf{H}_{\mathrm{pub}})$ *is a dual Goppa code.*

Proof:

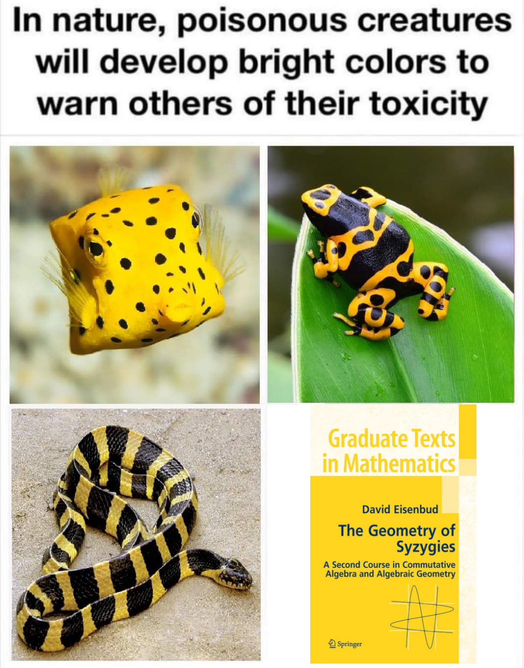
- $\dim_{\mathbb{F}_2}(I_2(\mathrm{C})) = \dim_{\mathbb{F}_{2^m}}(I_2(\mathrm{C}_{\mathbb{F}_{2^m}}))$ because $\mathrm{rk}(\mathrm{ev}_2)$ doesn't depend on the field
- $\mathrm{C}_{\mathbb{F}_{2^m}} = \mathrm{rowspan}_{\mathbb{F}_{2^m}}(\mathbf{H}_{\mathrm{pub}}) = \mathrm{GRS}_t(\mathbf{x}, \mathbf{y}) \oplus \mathrm{GRS}_t(\mathbf{x}^2, \mathbf{y}^2) \oplus \cdots \oplus \mathrm{GRS}_t(\mathbf{x}^{2^{m-1}}, \mathbf{y}^{2^{m-1}})$ (Delsarte)

On the other hand for a random $[n, k]_2$-code (where $k = mt$) w.h.p. [CCMZ15]

$$\dim_{\mathbb{F}_2}(I_2(\mathrm{C})) = \left(\frac{k(k+1)}{2} - n\right)^+$$

→ can distinguish when (very restrictive!)

$$n > \frac{k(k+1)}{2} - T$$



Theorem 2.8. *Let X be a set of 7 points in linearly general position in $\mathbb{P}^3$. There are just two distinct Betti diagrams possible for the homogeneous coordinate ring S_X:*

	0	1	2	3
0	1	–	–	–
1	–	3	–	–
2	–	1	6	3

and

	0	1	2	3
0	1	–	–	–
1	–	3	2	–
2	–	3	6	3

In the first case the points do not lie on any curve of degree 3. In the second case, the ideal J generated by the quadrics containing X is the ideal of the unique curve of degree 3 containing X, which is irreducible.

Figure 1: a distinguisher for $[7, 4]$ GRS codes

Linear resolutions

Start from a basis $Q_1, \ldots, Q_M$ of $I_2(C)$:

- by definition $Q_1, \ldots, Q_M$ satisfy no linear relation with scalar coefficients
- but can have syzygies: linear relations whose coefficients are polynomials
- keep it simple: consider only relations whose coefficients are of degree 1.

Now consider a basis of the space of such relations:

$$\begin{aligned} R_1 &: \quad \ell_{11} Q_1 + \cdots + \ell_{1M} Q_M = 0 \\ &\vdots \\ R_N &: \quad \ell_{N1} Q_1 + \cdots + \ell_{NM} Q_M = 0 \end{aligned}$$

- by definition $R_1, \ldots, R_N$ satisfy no linear relation with scalar coefficients
- but can satisfy linear relations whose coefficients are polynomials
- keep it simple: consider only relations whose coefficients are of degree 1.

Iterate! Relations between relations between relations...

Algebraic geometry view

- $\mathbf{H} \in \mathbb{F}^{k\times n}$, $C = \text{rowspan}_{\mathbb{F}}(\mathbf{H})$
- columns of $\mathbf{H}$ define a set of points $\mathfrak{X} = \{\overline{\mathbf{p}}_1, \dots, \overline{\mathbf{p}}_n\} \subseteq \mathbb{P}^{k-1}(\mathbb{F})$
- homogeneous coordinate ring $S_{\mathfrak{X}}$ is a quotient of $S = \mathbb{F}[X_1, \dots, X_k]$
- $I_2(C)$ = space of quadrics through $\mathfrak{X}$
- the previous slide defines the linear strand of the minimal resolution of $S_{\mathfrak{X}}$
- the dimensions of these syzygy spaces form the first row of its Betti diagram
- all these numbers β_{ij} are code invariants generalizing $\beta_{12} = \dim(I_2(C))$

	0	1	2	3	4	5	6	7	8	9	10	11
0	1	–	–	–	–	–	–	–	–	–	–	–
1	–	55	320	891	1408	1210	320	55	–	–	–	–
2	–	1	11	55	220	650	1672	1870	1221	485	110	11

Figure 2: Betti diagram of the $[23, 12]_2$ Golay code, and its linear strand

Goppa case: the Eagon-Northcott complex

If C is a dual Goppa code, then $I_2(C_{\mathbb{F}_{2^m}})$ contains the 2×2 minors of a matrix of linear forms $\begin{pmatrix} \ell_1 & \ell_2 & \dots & \ell_f \\ \ell'_1 & \ell'_2 & \dots & \ell'_f \end{pmatrix}$:

- these minors are the $\binom{f}{2}$ quadratic forms $Q_{ij} = \ell_i \ell'_j - \ell_j \ell'_i$
- the Q_{ij} admit the $2\binom{f}{3}$ relations
  - $R_{ijk} : \ \ell_i Q_{jk} - \ell_j Q_{ik} + \ell_k Q_{ij} = 0$
  - $R'_{ijk} : \ \ell'_i Q_{jk} - \ell'_j Q_{ik} + \ell'_k Q_{ij} = 0$
- these R_{ijk} and R'_{ijk} admit the $3\binom{f}{4}$ relations
  - $S_{ijkl} : \ \ell_i R_{jkl} - \ell_j R_{ikl} + \ell_k R_{ijl} - \ell_l R_{ijk} = 0$
  - $S'_{ijkl} : \ \ell_i R'_{jkl} - \ell_j R'_{ikl} + \ell_k R'_{ijl} - \ell_l R'_{ijk} + \ell'_i R_{jkl} - \ell'_j R_{ikl} + \ell'_k R_{ijl} - \ell'_l R_{ijk} = 0$
  - $S''_{ijkl} : \ \ell'_i R'_{jkl} - \ell'_j R'_{ikl} + \ell_k R'_{ijl} - \ell_l R'_{ijk} = 0$
- etc., so the length of the linear strand is at least f.

Goppa case: the Eagon-Northcott complex

If C is a dual Goppa code, then $I_2(C_{\mathbb{F}_{2^m}})$ contains the 2×2 minors of a matrix of linear forms $\begin{pmatrix} \ell_1 & \ell_2 & \dots & \ell_f \\ \ell_1' & \ell_2' & \dots & \ell_f' \end{pmatrix}$:

- these minors are the $\binom{f}{2}$ quadratic forms $Q_{ij} = \ell_i \ell_j' - \ell_j \ell_i'$
- the Q_{ij} admit the $2\binom{f}{3}$ relations
  - $R_{ijk} : \ \ell_i Q_{jk} - \ell_j Q_{ik} + \ell_k Q_{ij} = 0$
  - $R_{ijk}' : \ \ell_i' Q_{jk} - \ell_j' Q_{ik} + \ell_k' Q_{ij} = 0$
- these R_{ijk} and R_{ijk}' admit the $3\binom{f}{4}$ relations
  - $S_{ijkl} : \ \ell_i R_{jkl} - \ell_j R_{ikl} + \ell_k R_{ijl} - \ell_l R_{ijk} = 0$
  - $S_{ijkl}' : \ \ell_i R_{jkl}' - \ell_j R_{ikl}' + \ell_k R_{ijl}' - \ell_l R_{ijk}' + \ell_i' R_{jkl} - \ell_j' R_{ikl} + \ell_k' R_{ijl} - \ell_l' R_{ijk} = 0$
  - $S_{ijkl}'' : \ \ell_i' R_{jkl}' - \ell_j' R_{ikl}' + \ell_k R_{ijl}' - \ell_l R_{ijk}' = 0$
- etc., so the length of the linear strand is at least f.

Moreover one can show that this f is unexpectedly close to k.

Random case

- the r-th linear syzygy space is defined iteratively as the left kernel of a Macaulay matrix constructed from the $(r-1)$-th space
- w.h.p. we expect this space is null iff this matrix has # rows < # columns, which happens iff

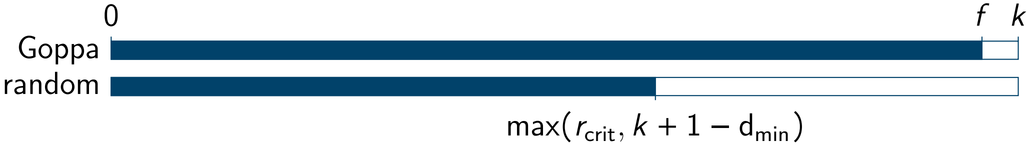
$$r > r_{\text{crit}} = \frac{k(k+1)}{n} \approx kR$$

- Minimal resolution conjecture (warning: false but "true enough") supports this over an infinite field
- another necessary condition, possibly stronger in the finite field case, is

$$r > k + 1 - \mathrm{d}_{\min}$$

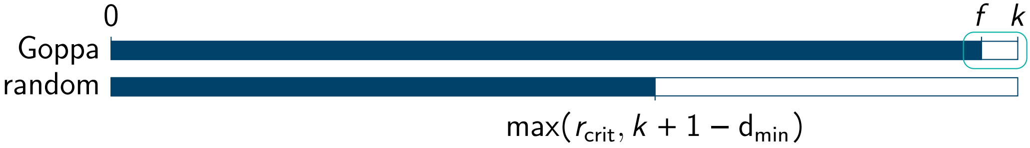
- Heuristic, supported experimentally: no other condition

Shortening



Expected complexity is polynomial in $\binom{k}{r_{\mathrm{crit}}} \approx \binom{k}{Rk}$ thus exponential in k or n.

Shortening



Expected complexity is polynomial in $\binom{k}{r_{crit}} \approx \binom{k}{Rk}$ thus exponential in k or n.

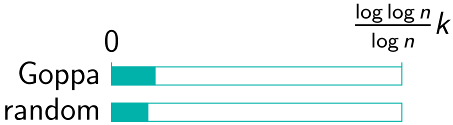
Recall f very close to k, more precisely: $f \approx \left(1 - \frac{\log\log n}{\log n}\right) k$

Shortening changes parameters:

$n \rightsquigarrow n-1, \quad k \rightsquigarrow k-1, \quad f \rightsquigarrow f-1, \quad \mathrm{d}_{min} \rightsquigarrow \geq \mathrm{d}_{min} - 1, \quad R \searrow, \quad r_{crit} \searrow$

- for R small enough, $k + 1 - \mathrm{d}_{min} < r_{crit}$
- we can shorten and still distinguish as long as $f > r_{crit}$
- works up to $k \rightsquigarrow \frac{\log\log n}{\log n} k, \quad R \rightsquigarrow \frac{R}{1-R} \frac{\log\log n}{\log n}$

Shortening



$$\binom{k}{Rk} \rightsquigarrow \binom{\frac{\log\log n}{\log n}k}{\frac{R}{1-R}\left(\frac{\log\log n}{\log n}\right)^2 k} \approx 2^{\frac{R^2}{1-R}\frac{(\log\log n)^3}{(\log n)^2}n}$$ subexponential in $\frac{n}{\log n}$

Recall f very close to k, more precisely: $f \approx \left(1 - \frac{\log\log n}{\log n}\right)k$

Shortening changes parameters:

$n \rightsquigarrow n-1, \quad k \rightsquigarrow k-1, \quad f \rightsquigarrow f-1, \quad \mathrm{d_{min}} \rightsquigarrow \geq \mathrm{d_{min}} - 1, \quad R \searrow, \quad r_{\mathrm{crit}} \searrow$

- for R small enough, $k + 1 - \mathrm{d_{min}} < r_{\mathrm{crit}}$
- we can shorten and still distinguish as long as $f > r_{\mathrm{crit}}$
- works up to $k \rightsquigarrow \frac{\log\log n}{\log n}k, \quad R \rightsquigarrow \frac{R}{1-R}\frac{\log\log n}{\log n}$

Concrete parameters

Best I can deal with in practice is $m = 10$, $n = 1024$, $t = 10$:

- before shortening, dual codes have parameters $[1024, 100]$
- theoretically distinguishable at $r = 10$, but too heavy
- shorten 40 times → shortened codes have parameters $[984, 60]$
- distinguishable at $r = 4$ in practice: $\beta_{3,4} = 30$ for Goppa vs 0 random
- no deviation from the heuristics

`Classic McEliece 348864` has $m = 12$, $n = 3488$, $t = 64$:

- before shortening, dual codes have parameters $[3488, 768]$
- shorten 377 times → shortened codes have parameters $[3111, 391]$
- theoretically distinguishable at $r = 50$, complexity estimate 2^{528} unfeasible

Asymptotic gain $\frac{(\log \log n)^3}{\log n}$ tends to 0 ridiculously slowly!

Conclusion

- Is McEliece broken? — No.
- Will it be broken soon? — I don't know, and I wouldn't bet in any direction.
- Is our understanding of its security stable? — Definitely not!

Two main technical ingredients:

- (fancy♥) higher modules of syzygies, Betti numbers
- (don't underestimate!) f unexpectedly close to k, allows to shorten a lot.

TODO:

- Improve complexity/implementation, theoretically and practically.
- Pursue theoretical study of Betti numbers from coding theory viewpoint.
- This is not a black-box distinguisher, it comes with a lot of structural information → use it (joint with other techniques) for structural recovery?
- Betti numbers are new code invariants. Find other applications?